

US Patent & Trademark Office

Patent Public Search | Text View

United States Patent Application Publication

20250250550

Kind Code

A1

Publication Date

August 07, 2025

Inventor(s)

Novick; Scott J. et al.

ENGINEERED TRANSAMINASE POLYPEPTIDES FOR INDUSTRIAL BIOCATALYSIS

Abstract

The present invention provides engineered transaminase polypeptides useful for the synthesis of chiral amine compounds under industrially relevant conditions. The invention also provides polynucleotides encoding the engineered transaminase polypeptides, host cells capable of expressing the engineered transaminases, and methods of using the engineered transaminases for the production of chiral amine compounds.

Inventors: Novick; Scott J. (Palo Alto, CA), Dellas; Nikki (San Carlos, CA), Alivizo; Oscar (Fremont, CA), Garcia; Ravi David (Los Gatos, CA), Ching; Charlene (San Jose, CA), Entwistle; David (San Carlos, CA)

Applicant: Codexis, Inc. (Redwood City, CA)

Family ID: 64660822

Appl. No.: 19/183477

Filed: April 18, 2025

Related U.S. Application Data

parent US continuation 17742560 20220512 parent-grant-document US 12297462 child US 19183477

parent US division 16615999 20191122 parent-grant-document US 11359183 US division PCT/US2018/034069 20180523 child US 17742560

us-provisional-application US 62519297 20170614

Publication Classification

Int. Cl.: C12N9/10 (20060101)

Background/Summary

CROSS REFERENCE TO RELATED APPLICATIONS [0001] The present application is a continuing application of pending U.S. application Ser. No. 17/742,560, filed May 12, 2022, which is a divisional of U.S. application Ser. No. 16/615,999, filed Nov. 22, 2019, now U.S. Pat. No. 11,359,183, which claims priority under 35 USC § 371 to PCT International Application No. PCT/US2018/034069, filed May 23, 2018, which claims priority to U.S. Prov. Pat. Appln. Ser. No. 62/519,297, filed Jun. 14, 2017, all of which are hereby incorporated by reference in their entireties, for all purposes.

REFERENCE TO SEQUENCE LISTING, TABLE OR COMPUTER PROGRAM

[0002] The official copy of the Sequence Listing is submitted concurrently with the specification as an ST26 formatted .xml file via EFS-Web, with a filename of “CX2-167USP1_ST26.xml”, a creation date of Apr. 18, 2025, and a size of 376,121 bytes. The Sequence Listing file submitted herewith is ST26 formatted version of the the ST25 formatted .txt file with a filename “CX2-167USP1_ST26.txt”, a creation date of Jun. 13, 2017, and a size of 495,524 bytes, and includes no new matter relative to that sequence listing that was filed with the priority U.S. Prov. Pat. Appln. Ser. No. 62/519,297 on Jun. 14, 2017. The Sequence Listing filed via EFS-Web is part of the specification and incorporated in its entirety by reference herein.

FIELD OF THE INVENTION

[0003] The present invention provides engineered transaminase polypeptides useful under industrial process conditions for the production of pharmaceutical and fine chemical amine compounds.

BACKGROUND

[0004] Transaminases (E.C. 2.6.1) catalyze the transfer of an amino group, a pair of electrons, and a proton from an amino donor compound to the keto group of an amino acceptor compound. Transaminase reactions can result in the formation of a chiral amine product compound. As shown in Scheme 1, an amino acceptor compound (B) (which is the keto substrate precursor of a desired chiral amine product (D)) is reacted with an amino donor compound (A) in the presence of a transaminase. The transaminase catalyzes the transfer of the primary amine group of the amino donor compound (A) to the keto group of the amino acceptor compound (B). The transaminase reaction results in a chiral amine product compound (D) (assuming R_{sup.3} is not the same as R_{sup.4}) and a new amino acceptor byproduct (or “carbonyl byproduct”) compound (C) which has a keto group.

##STR00001##

[0005] Chiral amine compounds are frequently used in the pharmaceutical, agrochemical and chemical industries as intermediates or synthons for the preparation of wide range of commercially desired compounds, such as cephalosporine or pyrrolidine derivatives. Typically these industrial applications of chiral amine compounds involve using only one particular stereomeric form of the molecule (e.g., only the (R) or the (S) enantiomer is physiologically active). Transaminases are highly stereoselective and have many potential industrial uses for the synthesis of optically pure chiral amine compounds.

[0006] Examples of the uses of transaminases to make chiral amine compounds include: the enantiomeric enrichment of amino acids (See e.g., Shin et al., Biosci. Biotechnol. Biochem., 65:1782-1788 [2001]; Iwasaki et al., Biotech. Lett., 25:1843-1846 [2003]; Iwasaki et al., Appl.

Microbiol., 69:499-505 [2004]; Yun et al., Appl. Environ. Microbiol., 70:2529-2534 [2004]; and Hwang et al., Enz. Microbiol. Technol., 34:429-426 [2004]); the preparation of intermediates and precursors of pregabalin (e.g., WO 2008/127646); the enzymatic transamination of cycloamine analogs (e.g., WO 2011/017551); the stereospecific synthesis and enantiomeric enrichment of β -amino acids (e.g., WO 2005/005633); the enantiomeric enrichment of amines (See, e.g., U.S. Pat. Nos. 4,950,606; 5,300,437; and 5,169,780); the production of amino acids and derivatives (See e.g., U.S. Pat. Nos. 5,316,943; 4,518,692; 4,826,766; 6,197,558; and 4,600,692); and in the production of the pharmaceutical compounds, sitagliptin, rivastigmine, and vernakalant (See e.g., U.S. Pat. No. 8,293,507; Savile, et al., Sci., 329: 305-9 [2010]; WO2011/159910; and WO2012/024104).

[0007] Wild-type transaminases having the ability to catalyze a reaction of Scheme 1 have been isolated from various microorganisms, including, but not limited to, *Alcaligenes denitrificans*, *Bordetella bronchiseptica*, *Bordetella parapertussis*, *Brucella melitensis*, *Burkholderia malle*, *Burkholderia pseudomallei*, *Chromobacterium violaceum*, *Oceanicola granulosus* HTCC2516, *Oceanobacter* sp. RED65, *Oceanospirillum* sp. MED92, *Pseudomonas putida*, *Ralstonia solanacearum*, *Rhizobium meliloti*, *Rhizobium* sp. (strain NGR234), *Bacillus thuringiensis*, *Klebsiella pneumonia*, *Vibrio fluvialis* (See e.g., Shin et al., Biosci. Biotechnol. Biochem., 65:1782-1788 [2001]), and *Arthrobacter* sp. KNK168 (See e.g., Iwasaki et al., Appl. Microbiol. Biotechnol., 69: 499-505 [2006]; and U.S. Pat. No. 7,169,592). Several of these wild-type transaminase genes and encoded polypeptides have been sequenced (e.g., *Ralstonia solanacearum* [Genbank Acc. No. YP_002257813.1, GI:207739420], *Burkholderia pseudomallei* 1710b [Genbank Acc. No. ABA47738.1, GI:76578263], *Bordetella petrii* [Genbank Acc. No. AM902716.1, GI:163258032], *Vibrio fluvialis* JS17 [Genbank Acc. No. AEA39183.1, GI: 327207066], and *Arthrobacter* sp. KNK168 [GenBank Acc. No. BAK39753.1, GI:336088341]). At least two wild-type transaminases of classes EC 2.6.1.18 and EC 2.6.1-19, have been crystallized and structurally characterized (See e.g., Yonaha et al., Agric. Biol.

[0008] Chem., 47:2257-2265 [1983]).

[0009] Transaminases are known that have (R)-selective or (S)-selective stereoselectivity. For example, the wild-type transaminase from *Arthrobacter* sp. KNK168 is considered (R)-selective and produces primarily (R)-amine compounds from certain substrates (See e.g., Iwasaki et al., Appl.

[0010] Microbiol. Biotechnol., 69: 499-505 [2006]; and U.S. Pat. No. 7,169,592), whereas the wild-type transaminase from *Vibrio fluvialis* JS17 is considered (S)-selective and produces primarily (S)-amine compounds from certain substrates (See e.g., Shin et al., Appl. Microbiol. Biotechnol., 61: 463-471 [2003]).

[0011] Non-naturally occurring transaminases having (R)-selectivity, increased solvent and thermal stability, and other improved properties for the conversion of a wide range of amino acceptor substrates, have been generated by mutagenesis and/or directed evolution of wild-type and other engineered transaminase backbone sequences (See e.g., U.S. Pat. No. 8,293,507 B2; WO2011/005477A1; WO2012/024104; and Savile et al., Sci., 329:305-9 [2010]).

[0012] However, transaminases generally have properties that are undesirable for commercial application in the preparation of chiral amine compounds, such as instability to industrially useful process conditions (e.g., solvent, temperature), poor recognition of, and stereoselectivity for, commercially useful amino acceptor and/or amino donor substrates, and low product yields due to unfavorable reaction equilibrium. Thus, there is a need for engineered transaminases that can be used in industrial processes for preparing chiral amines compounds in an optically active form.

SUMMARY OF THE INVENTION

[0013] The present invention provides engineered polypeptides having transaminase activity, polynucleotides encoding the polypeptides, methods of making the polypeptides, and methods of using the polypeptides for the biocatalytic conversion of amino acceptor substrate compounds (i.e.,

keto group containing compounds) to chiral amine product compounds. The transaminase polypeptides of the present invention have been engineered to have one or more residue differences as compared to a previously engineered transaminase polypeptide (of amino acid sequence SEQ ID NO:2) and associated enhanced solvent and thermal stability relative to previously engineered transaminase polypeptides (See e.g., U.S. Pat. No. 8,293,507 B2; WO2011005477A1; and WO2012024104). The amino acid residue differences are located at residue positions that result in improvement of various enzyme properties, including among others, activity, stereoselectivity, stability, expression, and product tolerance.

[0014] In some embodiments, the engineered transaminase polypeptides of the present invention have been engineered for efficient conversion of the substrate, (R)-5-([1,1'-biphenyl]-4-yl)-2-methyl-4-oxopentanoic acid (referred to herein as “compound (2)”) to its corresponding chiral amine product compound, (2R,4S)-5-([1,1'-biphenyl]-4-yl)-4-amino-2-methylpentanoic acid (referred to herein as “compound (1)”) as shown in Scheme 2.

##STR00002##

[0015] Compound (1) is a precursor to the known NEP (neprilysin) inhibitor compound “sacubitril.” The evolved structural features of the engineered transaminase polypeptides of the present invention, however, also allow for the biocatalytic conversion of a range of ketone substrate compounds of Formula (II) (including compounds other than compound (2)) to their corresponding chiral amine product compounds of Formula (I) (including compounds other than compound (1)) as shown in Scheme 3,

##STR00003##

wherein

[0016] R is C.sub.1-4 alkyl, or C.sub.1-4 alkyl with a terminal carboxylic acid group. The alkyl groups can be branched or unbranched. In some embodiments, the engineered transaminase polypeptide are capable of biocatalytic conversion of compounds of Formula (II) to compounds of Formula (I) having the indicated stereochemical configuration at the stereogenic center marked with an *; in an enantiomeric excess of at least 70% over the opposite enantiomer.

[0017] In some additional embodiments, the evolved structural features of the engineered transaminase polypeptides of the present invention, however, also allow for the biocatalytic conversion of a range of ketone substrate compounds of Formula (III) (including compounds other than compound (2)) to their corresponding chiral amine product compounds of Formula (IV) (including compounds other than compound (1)) as shown in Scheme 4,

##STR00004##

wherein

[0018] R is CO.sub.2H or C.sub.1-4 alkyl, where C.sub.1-4 alkyl is optionally substituted with R.sup.1 and R.sup.2. Where R.sup.1 =H or C.sub.1-4 alkyl, R.sup.2=H or CO.sub.2R.sup.3, R.sup.3=H or C.sub.1-4 alkyl. The alkyl groups R.sup.1, R.sup.2 and R.sup.3 can be linear or branched.

[0019] In some embodiments, the present invention provides engineered transaminases comprising polypeptide sequences having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NOS:4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156, or a functional fragment thereof, wherein said engineered transaminases comprise at least one substitution or substitution set in said polypeptide sequences, and wherein the amino acid positions of said polypeptide sequences are numbered with reference to SEQ ID NO: 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156. In some embodiments, the engineered transaminase comprises a polypeptide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO:4, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 17, 17/259/347, 20, 45/48/78/81/92, 57, 57/319, 85, 86, 157, 158, 161, 203/416, 314, 316, 319,

375/416, 404, 415, 416, 417, 420, 421, 422, and 424, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 4. In some additional embodiments, the engineered transaminase comprises a polypeptide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 18, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 5/17/86/157/259/314/316/422, 17/86/157/203/259/314/316/422, 17/86/157/314/316/347/422, 17/86/157/316/422, 85/86, 86/157/314/316/347/416/422, 86/157/416, 86/203/316/416, 86/203/316/422, 86/316/416, 139/221/310/316, 157/316, 195/316, 207/316, 227/422, 280/381/416/445, 314, 314/316, 316, 416, 417, and 422, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 18. In some further embodiments, the engineered transaminase comprises a polypeptide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 20, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 17/139/314/416/417/422, 17/195/316/414/416/422, 17/314/316/414/416, 17/314/316/414/416/417/422, 20, 31, 54, 57, 83, 86, 86/194,121, 139/195/227/314/316/416, 139/202/207/227/314/316/416/422, 139/207/227/314/414/416/422, 139/227/310/414/416/417, 139/316/414/416/417/422, 146/157, 164/207/213/227/310/316/416, 166, 178/314/316/416, 190/195/207/213/244/314/414/416/417/422, 194/213/310/314/316/422, 195/227/314/316/416, 195/227/314/416/417, 195/227/314/422, 195/227/414/416, 195/227/414/416/422, 195/227/416/417, 195/316/416/417/422, 207/208/220/241/291/314/316, 226, 227, 227/416, 228, 258, 259, 271/310/316/416/422, 282, 310/314/316/414/416/417/422, 314/316/414/416/422, 314/316/416, 314/316/416/422, 316, 319, 320, 322, and 417, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 20. In yet some additional embodiments, the engineered transaminase comprises a polypeptide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 30, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 21, 21/23/53/146/162/395, 21/23/53/146/162/395/432, 21/23/53/162/395/432, 21/23/53/395/432, 21/23/395, 21/53/159/395, 21/53/163/395/432, 23/53/146/159/395/431, 23/53/146/159/395/432, 23/53/146/162/395/432, 23/53/146/163, 23/146/163/432, 23/162, 53/146/159/395, 53/146/162/395, 53/146/162/432, 53/146/163/395, 53/146/395/432, 53/162/163/395/432, 53/162/395, 53/162/395/432, 53/163/395/432, 53/395, 57/86, 57/86/121/146/157, 57/86/121/157/226/259/282/332, 57/86/121/332, 57/86/157/259, 57/86/157/282/332, 57/86/332, 57/121/417/422, 57/146/226/259, 57/146/226/259/282, 57/259, 86, 86/157, 86/226,121, 121/146/157/226/332, 121/208/316/417/422, 121/226/332, 121/316/422, 121/332, 121/416/417/422, 121/417/422, 146, 146/162/395/432, 157, 162/432, 226/282/332, 227, 227/417, 259, 282, 282/332, 316/417/422, 332, 395/432, 416, 416/422, 417, and 422, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 30. In some additional embodiments, the engineered transaminase comprises a polypeptide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 30, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 21, 21/23/53/146/162/395, 21/23/53/146/162/395/432, 21/23/53/162/395/432, 21/23/53/395/432, 21/23/395, 21/53/159/395, 21/53/163/395/432, 23/53/146/159/395/431, 23/53/146/159/395/432, 23/53/146/162/395/432, 23/53/146/163, 23/146/163/432, 23/162, 53/146/159/395, 53/146/162/395, 53/146/162/432, 53/146/163/395, 53/146/395/432, 53/162/163/395/432, 53/162/395, 53/162/395/432, 53/163/395/432, 53/395, 57/86, 57/86/121/146/157, 57/86/121/157/226/259/282/332, 57/86/121/332, 57/86/157/259,

57/86/157/282/332, 57/86/332, 57/121/417/422, 57/146/226/259/282, 57/259, 86, 86/157, 86/226, 121, 121/146/157/226/332, 121/208/316/417/422, 121/226/332, 121/316/422, 121/332, 121/416/417/422, 121/417/422, 146, 146/162/395/432, 157, 162/432, 226/282/332, 227, 227/417, 259, 282, 282/332, 316/417/422, 332, 395/432, 416, 416/422, 417, and 422, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 30. In some further embodiments, the engineered transaminase comprises a polypeptide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 66, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 107, 141, 161, 174, 245, 281, 293, 432, and 453, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 66. In yet some further embodiments, the engineered transaminase comprises a polypeptide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 70, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 2, 4, 5, 5/38/56/346, 13, 17, 17/31, 19, 31, 57, 74, 79, 86, 121, 155, 161, 161/174/432, 163/167, 164, 190/291, 195, 195/408, 203/207/219, 205/419, 207, 207/235/419, 214/228/251, 222, 223, 226/228/230/231/311, 228, 233, 246, 247/271/293, 259, 282, 284, 309, 327, 346, 359, 362, 375, 404, 408, 419, 420, 422, and 443, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 70. In still some additional embodiments, the engineered transaminase comprises a polypeptide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 48, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 21/23/53/163/293, 21/23/53/163/293/395, 21/53/146, 21/53/146/163/293/395/432, 21/53/163/293/395, 21/53/293/432, 23/53/146/163, 23/53/146/163/293/432, 23/53/146/293/432, 23/53/163, 23/53/163/293/395, 23/53/163/293/432, 23/53/163/395, 23/53/293, 23/53/293/395/432, 23/53/293/432, 23/146/293, 53/146/163/293/395/432, 53/146/163/395/432, 53/146/163/432, 53/146/293, 53/163, 53/163/293/395, 53/163/293/395/432, 53/163/293/432, 53/163/432, 53/293, 53/293/395, 53/395, 53/432, 146, and 293, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 48. In some further embodiments, the engineered transaminase comprises a polypeptide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 88, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 157, 163, 163/164, 284, and 332, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 88. In some additional embodiments, the engineered transaminase comprises a polypeptide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 88, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 31, 86, 149, 191/208/211/286, 192/294, 195/207/208/286, 208/294, 260, 284, 286, 294, 313, 395/396, and 424, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 88. In some further embodiments, the engineered transaminase comprises a polypeptide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 110, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 14/31/208/260/286/294/424, 14/31/260/286/294/424, 31, 31/208, 31/208/241/286/294, 31/208/260/286/294, 31/208/260/286/294/313/424, 31/208/260/286/294/424, 31/208/260/286/313/424,

31/208/260/294/313/424, 31/208/260/294/424, 31/208/286/294, 31/208/286/294/313, 31/208/286/294/313/424, 31/208/286/294/424, 31/208/286/313/424, 31/208/286/424, 31/208/294, 31/208/294/313/424, 31/208/294/424, 31/260/286, 31/260/286/294/313, 31/260/286/294/313/424, 31/260/286/313/424, 31/260/286/424, 31/260/294, 31/260/294/313, 31/286/294, 31/286/294/313, 31/286/294/424, 31/294/313/424, 31/294/424, 208/260/286/294/313/424, 208/260/286/294/424, 208/260/286/424, 208/260/294, 208/260/294/313/424, 208/286/294/424, 260/286/294/424, 260/286/424, 286/294/313/424, and 286/313/424, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 110. In some further embodiments, the engineered transaminase comprises a polypeptide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 112, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 16/272, 17, 19/86, 19/86/321, 20, 21, 22, 22/255, 22/255/362/421, 22/294, 22/362, 24/255/362, 31, 41, 42/187/294, 66, 66/190, 66/394, 84, 114, 153, 157, 161, 163, 164, 164/397, 221, 222, 272, 294, 320, 320/435, and 394, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 112. In still some additional embodiments, the engineered transaminase comprises a polypeptide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 120, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 17, 17/21, 17/21/22, 17/22, 17/22/272, 17/22/272/320, 17/31/320/397, 17/84/164, 17/157/164, 17/157/320/397, 17/221/320, 17/272/320, 17/320, 17/397, 19/22/42/320/394, 19/42, 20/320, 25/320, 66/187/320, 66/272, 66/320, 66/394, 164/320/321, 165, 255, 272, 272/320, and 320, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 120. In some further embodiments, the engineered transaminase comprises a polypeptide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 138, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 24, 24/25, 24/110/167, 25, 25/147/167, 27/190/207/389/396, 27/190/207/396, 41/42/66/165, 42, 42/66/165/394, 42/165/272/394, 66/164, 66/164/272, 66/165, 66/394, 163, 164, 164/165/394, 165, 165/272, 165/394, 166, 167, 190, 190/207, 190/207/210, 190/207/210/385/389, 190/207/385, 190/210, 190/385/396, 190/389, 207, 207/210/243/250/385/389, 207/214/396, 207/243, 207/389, 210, 210/385, 214, 214/396, 272/394, 283, 385, 385/389, 389, 389/396, and 394, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 138. In some additional embodiments, the engineered transaminase comprises a polypeptide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 156, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 12/26/30/291, 12/26/73/77/291, 13/45/107/236/277/346/351/410/429, 22/25/110/291/304, 24/147/164/166/167/243/389, 24/166, 24/166/190, 24/166/207, 26/30, 26/30/168/291, 30/77/291, 30/291, 70/77/108/177/196/247/291/309/361/366, 70/77/168/291, 70/177/291/361/366, 70/177/291/372, 70/186, 70/247/291/309/372, 70/291, 70/291/304/361, 74, 74/211, 77/110/291/308, 77/291/361/366, 107/277/346/410/429, 107/346, 110/235/391, 147/164/166/167, 147/164/166/243/272/283/385/389, 147/164/166/272, 163, 166, 166/167/272/389, 166/190, 166/190/207, 166/207, 177/195/196/291/304/309/366, 177/196/247/304/309/361/366, 177/291/309/361, 196/291/304/361, 196/291/309/361, 198/235/391, 207, 247/304/309/366, 253/391, 277, 291, 291/304/309, and 291/309, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 156. In some further embodiments, the engineered transaminase comprises a polypeptide sequence comprising at least

85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, identical to the sequence of at least one engineered transaminase variant set forth in Table 6.1, 7.1, 8.1, 9.1, 10.1, 11.1, 12.1, 13.1, 13.2, 14.1, 15.1, 16.1, 17.1, and/or 18.1. In yet some additional embodiments, the engineered transaminase is a variant engineered polymerase provided in Table 6.1, 7.1, 8.1, 9.1, 10.1, 11.1, 12.1, 13.1, 13.2, 14.1, 15.1, 16.1, 17.1. In some further embodiments, the engineered transaminase comprises a polypeptide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to the sequence of at least one engineered transaminase variant set forth in SEQ ID NOS: 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156. In some additional embodiments, the engineered transaminase comprises a polypeptide sequence comprising SEQ ID NOS: 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156. In some further embodiments, the engineered transaminase comprises a polypeptide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to the sequence of at least one engineered transaminase variant set forth in the even numbered sequences of SEQ ID NOS: 6-172. In yet some additional embodiments, the engineered transaminase comprises a polypeptide sequence forth in the even numbered sequences of SEQ ID NOS: 6-172. In some further embodiments, the engineered transaminase comprises at least one improved property compared to wild-type *V. fluvialis* transaminase. In some additional embodiments, the improved property of the engineered transaminase comprises improved activity on a substrate. In some further embodiments, the substrate comprises compound (1). In yet some additional embodiments, the improved property of the engineered transaminase comprises improved thermostability. In some additional embodiments, the engineered transaminase is purified. The present invention also provides compositions comprising an engineered transaminase provided herein. In some embodiments, the compositions comprise more than one engineered transaminase provided herein.

[0020] The present invention also provides polynucleotide sequences encoding at least one engineered transaminase provided herein. In some embodiments, the polynucleotide sequence encodes at least one engineered transaminase, said polynucleotide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NOS: 3, 17, 19, 29, 47, 65, 69, 87, 109, 111, 119, 137, and/or 155, wherein the polynucleotide sequence of said engineered transaminase comprises at least one substitution at one or more positions. In some further embodiments, the polynucleotide sequence encodes at least one engineered transaminase comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NOS: 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156, or a functional fragment thereof. In yet some additional embodiments, the polynucleotide sequence is operably linked to a control sequence. In still some further embodiments, the polynucleotide sequence is codon optimized.

[0021] The present invention also provides expression vectors comprising at least one polynucleotide sequence encoding an engineered transaminase provided herein. In some embodiments, the expression vector comprises at least one polynucleotide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NOS: 3, 17, 19, 29, 47, 65, 69, 87, 109, 111, 119, 137, and/or 155, wherein the polynucleotide sequence of said engineered transaminase comprises at least one substitution at one or more positions. In some embodiments, the expression vector comprises a polynucleotide sequence encoding at least one engineered transaminase comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NOS: 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156, or a functional fragment thereof.

[0022] The present invention also provides host cells comprising at least one expression vector provided herein. In some embodiments, the host cell comprises at least one polynucleotide sequence provided herein. In some embodiments, the host cell comprises at least one

polynucleotide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NOS: 3, 17, 19, 29, 47, 65, 69, 87, 109, 111, 119, 137, and/or 155, wherein the polynucleotide sequence encoding the engineered transaminase comprises at least one substitution at one or more positions. In some embodiments, the host cell comprises a polynucleotide sequence encoding at least one engineered transaminase comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NOS: 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156, or a functional fragment thereof. In some embodiments, at least one polynucleotide sequence encoding an engineered transaminase is present in at least one expression vector.

[0023] The present invention also provides methods of producing an engineered transaminase in a host cell, comprising culturing the host cell provided herein under suitable conditions, such that at least one engineered transaminase is produced. In some embodiments, the methods further comprise recovering at least one engineered transaminase from the culture and/or host cell. In some additional embodiments, the methods further comprise the step of purifying said at least one engineered transaminase.

[0024] In some embodiments of the present invention, the engineered polypeptide having transaminase activity is immobilized on a solid support, optionally wherein the solid support is selected from a bead or resin comprising polymethacrylate with epoxide functional groups, polymethacrylate with amino epoxide functional groups, styrene/DVB copolymer or polymethacrylate with octadecyl functional groups.

[0025] In some embodiments of the engineered polypeptides having transaminase activity of the present invention, the engineered polypeptide is capable of converting a substrate of compound (2) to a product of compound (1) under suitable reaction conditions. In some embodiments, the engineered polypeptide is capable of converting compound (2) to compound (1) with at least 1.2 fold, 2 fold, 5 fold, 10 fold, 20 fold, 25 fold, 50 fold, 75 fold, 100 fold, 1,000 fold, 10,000 fold, 20,000 fold, 30,000 fold, 40,000 fold, 50,000 fold, 60,000 fold, 70,000 fold, 80,000 fold, 90,000 fold, 95,000 fold, or greater than the activity of a reference sequence (e.g., SEQ ID NO:2, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, or 156), under suitable reaction conditions. In some embodiments, the engineered polypeptide is capable of converting compound (2) to compound (1) with increased activity relative to a reference sequence (e.g., SEQ ID NO:2, 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, or 156), in which the suitable reaction conditions comprise compound (1) at a loading of at least 50 g/L, 0.5 g/L PLP, 1 M isopropylamine, pH 9, and 58° C.

[0026] In some embodiments, the present invention provides a process of preparing compound (1)
##STR00005##

comprising a step of contacting a substrate of compound (2)

##STR00006##

with an engineered polypeptide as disclosed herein in the presence of an amino group donor under suitable reaction conditions.

[0027] In some embodiments, the processes using the engineered polypeptides disclosed herein can be carried out wherein the compound (1) is produced in at least 90%, 97%, 98%, 99% or greater enantiomeric and diastereomeric excess.

[0028] Any of the processes disclosed herein using the engineered polypeptides for the preparation of compound (1) can be carried out under a range of suitable reaction conditions, including but not limited to, ranges of amine donor, pH, temperature, buffer, solvent system, substrate loading, polypeptide loading, cofactor loading, pressure, and reaction time. For example, in some embodiments, the preparation of compounds of compound (1) can be carried out wherein the suitable reaction conditions comprise: (a) substrate loading of about 10 to 100 g/L of substrate compound (e.g., 50 g/L or 75 g/L of compound (2)); (b) of about 0.5 g/L to 5 g/L engineered polypeptide; (c) IPM concentration of about 0.5 to 3 M; (d) PLP cofactor concentration of about

0.1 to 1 g/L; (e) DMSO concentration of about 0% (v/v) to about 20% (v/v); (f) pH of about 8.5 to 11.5; and (g) temperature of about 45° C. to 65° C. In some embodiments, the suitable reaction conditions comprise: (a) about 50 g/L of substrate compound (e.g., compound (2)); (b) about 1 g/L engineered polypeptide; (c) about 1 M isopropylamine (IPM); (d) about 0.5 g/L pyridoxal phosphate (PLP); (e) about pH 9; and (g) about 58° C.

[0029] In some embodiments, the processes using the engineered polypeptides disclosed herein can be carried out wherein the amino group donor is selected from isopropylamine, alanine, 3-aminobutyric acid, or methylbenzylamine. In some embodiments, the amino group donor is isopropylamine.

Description

DESCRIPTION OF THE INVENTION

[0030] For the descriptions provided herein, the use of the singular includes the plural (and vice versa) unless specifically stated otherwise. For instance, the singular forms “a”, “an” and “the” include plural referents unless the context clearly indicates otherwise. Similarly, “comprise,” “comprises,” “comprising” “include,” “includes,” and “including” are interchangeable and not intended to be limiting.

[0031] It is to be further understood that where descriptions of various embodiments use the term “comprising,” those skilled in the art would understand that in some specific instances, an embodiment can be alternatively described using language “consisting essentially of” or “consisting of.”

[0032] Both the foregoing general description, including the drawings, and the following detailed description are exemplary and explanatory only and are not restrictive of this invention. Moreover, the section headings used herein are for organizational purposes only and not to be construed as limiting the subject matter described.

Definitions

[0033] As used herein, the following terms are intended to have the following meanings. In reference to the present invention, the technical and scientific terms used in the descriptions herein will have the meanings commonly understood by one of ordinary skill in the art, unless specifically defined otherwise. Accordingly, the following terms are intended to have the following meanings. In addition, all patents and publications, including all sequences disclosed within such patents and publications, referred to herein are expressly incorporated by reference.

[0034] Unless otherwise indicated, the practice of the present invention involves conventional techniques commonly used in molecular biology, fermentation, microbiology, and related fields, which are known to those of skill in the art. Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described. Indeed, it is intended that the present invention not be limited to the particular methodology, protocols, and reagents described herein, as these may vary, depending upon the context in which they are used. The headings provided herein are not limitations of the various aspects or embodiments of the present invention that can be had by reference to the specification as a whole.

[0035] Accordingly, the terms defined below are more fully defined by reference to the specification as a whole.

[0036] Nonetheless, in order to facilitate understanding of the present invention, a number of terms are defined below. Numeric ranges are inclusive of the numbers defining the range. Thus, every numerical range disclosed herein is intended to encompass every narrower numerical range that

falls within such broader numerical range, as if such narrower numerical ranges were all expressly written herein. It is also intended that every maximum (or minimum) numerical limitation disclosed herein includes every lower (or higher) numerical limitation, as if such lower (or higher) numerical limitations were expressly written herein.

[0037] As used herein, the term “comprising” and its cognates are used in their inclusive sense (i.e., equivalent to the term “including” and its corresponding cognates).

[0038] As used herein and in the appended claims, the singular “a”, “an” and “the” include the plural reference unless the context clearly dictates otherwise. Thus, for example, reference to a “host cell” includes a plurality of such host cells.

[0039] Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation and amino acid sequences are written left to right in amino to carboxy orientation, respectively.

[0040] As used herein, the terms “protein,” “polypeptide,” and “peptide” are used interchangeably herein to denote a polymer of at least two amino acids covalently linked by an amide bond, regardless of length or post-translational modification (e.g., glycosylation, phosphorylation, lipidation, myristilation, ubiquitination, etc.). Included within this definition are D- and L-amino acids, and mixtures of D- and L-amino acids.

[0041] The abbreviations used for the genetically encoded amino acids are conventional and are as follows:

TABLE-US-00001 Three-Letter One-Letter Amino Acid Abbreviation
Alanine Ala A
Arginine Arg R
Asparagine Asn N
Aspartate Asp D
Cysteine Cys C
Glutamate Glu E
Glutamine Gln Q
Glycine Gly G
Histidine His H
Isoleucine Ile I
Leucine Leu L
Lysine Lys K
Methionine Met M
Phenylalanine Phe F
Proline Pro P
Serine Ser S
Threonine Thr T
Tryptophan Trp W
Tyrosine Tyr Y
Valine Val V

[0042] When the three-letter abbreviations are used, unless specifically preceded by an “L” or a “D” or clear from the context in which the abbreviation is used, the amino acid may be in either the L- or D-configuration about α -carbon (C.). For example, whereas “Ala” designates alanine without specifying the configuration about the α -carbon, “D-Ala” and “L-Ala” designate D-alanine and L-alanine, respectively. When the one-letter abbreviations are used, upper case letters designate amino acids in the L-configuration about the α -carbon and lower case letters designate amino acids in the D-configuration about the α -carbon. For example, “A” designates L-alanine and “a” designates D-alanine. When polypeptide sequences are presented as a string of one-letter or three-letter abbreviations (or mixtures thereof), the sequences are presented in the amino (N) to carboxy (C) direction in accordance with common convention.

[0043] As used herein, “hydrophilic amino acid or residue” refers to an amino acid or residue having a side chain exhibiting a hydrophobicity of less than zero according to the normalized consensus hydrophobicity scale of Eisenberg et al., (Eisenberg et al., J. Mol. Biol., 179:125-142 [1984]).

[0044] Genetically encoded hydrophilic amino acids include L-Thr (T), L-Ser (S), L-His (H), L-Glu (E), L-Asn (N), L-Gln (Q), L-Asp (D), L-Lys (K) and L-Arg (R).

[0045] As used herein, “acidic amino acid or residue” refers to a hydrophilic amino acid or residue having a side chain exhibiting a pK value of less than about 6 when the amino acid is included in a peptide or polypeptide. Acidic amino acids typically have negatively charged side chains at physiological pH due to loss of a hydrogen ion. Genetically encoded acidic amino acids include L-Glu (E) and L-Asp (D).

[0046] As used herein, “basic amino acid or residue” refers to a hydrophilic amino acid or residue having a side chain exhibiting a pK value of greater than about 6 when the amino acid is included in a peptide or polypeptide. Basic amino acids typically have positively charged side chains at physiological pH due to association with hydronium ion. Genetically encoded basic amino acids include L-Arg (R) and L-Lys (K).

[0047] As used herein, “polar amino acid or residue” refers to a hydrophilic amino acid or residue

having a side chain that is uncharged at physiological pH, but which has at least one bond in which the pair of electrons shared in common by two atoms is held more closely by one of the atoms. Genetically encoded polar amino acids include L-Asn (N), L-Gln (Q), L-Ser (S) and L-Thr (T). [0048] As used herein, “hydrophobic amino acid or residue” refers to an amino acid or residue having a side chain exhibiting a hydrophobicity of greater than zero according to the normalized consensus hydrophobicity scale of Eisenberg et al., (Eisenberg et al., J. Mol. Biol., 179:125-142 [1984]). Genetically encoded hydrophobic amino acids include L-Pro (P), L-Ile (I), L-Phe (F), L-Val (V), L-Leu (L), L-Trp (W), L-Met (M), L-Ala (A) and L-Tyr (Y).

[0049] As used herein, “aromatic amino acid or residue” refers to a hydrophilic or hydrophobic amino acid or residue having a side chain that includes at least one aromatic or heteroaromatic ring.

[0050] Genetically encoded aromatic amino acids include L-Phe (F), L-Tyr (Y) and L-Trp (W). Although owing to the pKa of its heteroaromatic nitrogen atom L-His (H) it is sometimes classified as a basic residue, or as an aromatic residue as its side chain includes a heteroaromatic ring, herein histidine is classified as a hydrophilic residue or as a “constrained residue” (see below).

[0051] As used herein, “constrained amino acid or residue” refers to an amino acid or residue that has a constrained geometry. Herein, constrained residues include L-Pro (P) and L-His (H). Histidine has a constrained geometry because it has a relatively small imidazole ring. Proline has a constrained geometry because it also has a five membered ring.

[0052] As used herein, “non-polar amino acid or residue” refers to a hydrophobic amino acid or residue having a side chain that is uncharged at physiological pH and which has bonds in which the pair of electrons shared in common by two atoms is generally held equally by each of the two atoms (i.e., the side chain is not polar). Genetically encoded non-polar amino acids include L-Gly (G), L-Leu (L), L-Val (V), L-Ile (I), L-Met (M) and L-Ala (A).

[0053] As used herein, “aliphatic amino acid or residue” refers to a hydrophobic amino acid or residue having an aliphatic hydrocarbon side chain. Genetically encoded aliphatic amino acids include L-Ala (A), L-Val (V), L-Leu (L) and L-Ile (I). It is noted that cysteine (or “L-Cys” or “[C]”) is unusual in that it can form disulfide bridges with other L-Cys (C) amino acids or other sulfanyl- or sulfhydryl-containing amino acids. The “cysteine-like residues” include cysteine and other amino acids that contain sulfhydryl moieties that are available for formation of disulfide bridges. The ability of L-Cys (C) (and other amino acids with —SH containing side chains) to exist in a peptide in either the reduced free —SH or oxidized disulfide-bridged form affects whether L-Cys (C) contributes net hydrophobic or hydrophilic character to a peptide. While L-Cys (C) exhibits a hydrophobicity of 0.29 according to the normalized consensus scale of Eisenberg (Eisenberg et al., 1984, supra), it is to be understood that for purposes of the present invention, L-Cys (C) is categorized into its own unique group.

[0054] As used herein, “small amino acid or residue” refers to an amino acid or residue having a side chain that is composed of a total three or fewer carbon and/or heteroatoms (excluding the α -carbon and hydrogens). The small amino acids or residues may be further categorized as aliphatic, non-polar, polar or acidic small amino acids or residues, in accordance with the above definitions. Genetically-encoded small amino acids include L-Ala (A), L-Val (V), L-Cys (C), L-Asn (N), L-Ser (S), L-Thr (T) and L-Asp (D).

[0055] As used herein, “hydroxyl-containing amino acid or residue” refers to an amino acid containing a hydroxyl (—OH) moiety. Genetically-encoded hydroxyl-containing amino acids include L-Ser (S) L-Thr (T) and L-Tyr (Y).

[0056] As used herein, “amino acid difference” and “residue difference” refer to a difference in the amino acid residue at a position of a polypeptide sequence relative to the amino acid residue at a corresponding position in a reference sequence. The positions of amino acid differences generally are referred to herein as “X_n,” where n refers to the corresponding position in the reference sequence upon which the residue difference is based. For example, a “residue difference at position X₄₀ as compared to SEQ ID NO:2” refers to a difference of the amino acid residue at the

polypeptide position corresponding to position 40 of SEQ ID NO:2. Thus, if the reference polypeptide of SEQ ID NO:2 has a histidine at position 40, then a “residue difference at position X40 as compared to SEQ ID NO:2” refers to an amino acid substitution of any residue other than histidine at the position of the polypeptide corresponding to position 40 of SEQ ID NO:2. In most instances herein, the specific amino acid residue difference at a position is indicated as “XnY” where “Xn” specified the corresponding position as described above, and “Y” is the single letter identifier of the amino acid found in the engineered polypeptide (i.e., the different residue than in the reference polypeptide). In some instances, the present invention also provides specific amino acid differences denoted by the conventional notation “AnB”, where A is the single letter identifier of the residue in the reference sequence, “n” is the number of the residue position in the reference sequence, and B is the single letter identifier of the residue substitution in the sequence of the engineered polypeptide. In some instances, a polypeptide of the present invention can include one or more amino acid residue differences relative to a reference sequence, which is indicated by a list of the specified positions where residue differences are present relative to the reference sequence. In some embodiments, where more than one amino acid can be used in a specific residue position of a polypeptide, the various amino acid residues that can be used are separated by a “/” (e.g., X192A/G). In some embodiments, in which there are variants with multiple substitutions, the substitutions are separated by either a semicolon (;) or a slash (/) (e.g., Y17V;1259T;E347K or Y17V/1259T/E347K).

[0057] The present invention includes engineered polypeptide sequences comprising one or more amino acid differences that include either/or both conservative and non-conservative amino acid substitutions. The amino acid sequences of the specific recombinant carbonic anhydrase polypeptides included in the Sequence Listing of the present invention include an initiating methionine (M) residue (i.e., M represents residue position 1). The skilled artisan, however, understands that this initiating methionine residue can be removed by biological processing machinery, such as in a host cell or in vitro translation system, to generate a mature protein lacking the initiating methionine residue, but otherwise retaining the enzyme's properties. Consequently, the term “amino acid residue difference relative to SEQ ID NO:2 at position Xn” as used herein may refer to position “Xn” or to the corresponding position (e.g., position (X-1)n) in a reference sequence that has been processed so as to lack the starting methionine.

[0058] As used herein, the phrase “conservative amino acid substitutions” refers to the interchangeability of residues having similar side chains, and thus typically involves substitution of the amino acid in the polypeptide with amino acids within the same or similar defined class of amino acids.

[0059] By way of example and not limitation, in some embodiments, an amino acid with an aliphatic side chain is substituted with another aliphatic amino acid (e.g., alanine, valine, leucine, and isoleucine); an amino acid with a hydroxyl side chain is substituted with another amino acid with a hydroxyl side chain (e.g., serine and threonine); an amino acids having aromatic side chains is substituted with another amino acid having an aromatic side chain (e.g., phenylalanine, tyrosine, tryptophan, and histidine); an amino acid with a basic side chain is substituted with another amino acid with a basis side chain (e.g., lysine and arginine); an amino acid with an acidic side chain is substituted with another amino acid with an acidic side chain (e.g., aspartic acid or glutamic acid); and/or a hydrophobic or hydrophilic amino acid is replaced with another hydrophobic or hydrophilic amino acid, respectively. Exemplary conservative substitutions are provided in Table 1.

TABLE-US-00002	
TABLE 1 Exemplary Conservative Amino Acid Substitutions	
Residue	Potential Conservative Substitutions
A, L, V, I	Other aliphatic (A, L, V, I)
G, M	Other non-polar (A, L, V, I, G, M)
D, E	Other acidic (D, E)
K, R	Other basic (K, R)
N, Q, S, T	Other polar (H, Y, W, F)
C, P	Non-polar

[0060] As used herein, the phrase “non-conservative substitution” refers to substitution of an amino acid in the polypeptide with an amino acid with significantly differing side chain properties. Non-

conservative substitutions may use amino acids between, rather than within, the defined groups and affects (a) the structure of the peptide backbone in the area of the substitution (e.g., proline for glycine) (b) the charge or hydrophobicity, or (c) the bulk of the side chain. By way of example and not limitation, an exemplary non-conservative substitution can be an acidic amino acid substituted with a basic or aliphatic amino acid; an aromatic amino acid substituted with a small amino acid; and a hydrophilic amino acid substituted with a hydrophobic amino acid.

[0061] As used herein, “deletion” refers to modification of the polypeptide by removal of one or more amino acids from the reference polypeptide. Deletions can comprise removal of 1 or more amino acids, 2 or more amino acids, 5 or more amino acids, 10 or more amino acids, 15 or more amino acids, or 20 or more amino acids, up to 10% of the total number of amino acids, or up to 20% of the total number of amino acids making up the polypeptide while retaining enzymatic activity and/or retaining the improved properties of an engineered enzyme. Deletions can be directed to the internal portions and/or terminal portions of the polypeptide. In various embodiments, the deletion can comprise a continuous segment or can be discontinuous.

[0062] As used herein, “insertion” refers to modification of the polypeptide by addition of one or more amino acids to the reference polypeptide. In some embodiments, the improved engineered transaminase enzymes comprise insertions of one or more amino acids to the naturally occurring transaminase polypeptide as well as insertions of one or more amino acids to engineered transaminase polypeptides. Insertions can be in the internal portions of the polypeptide, or to the carboxy or amino terminus. Insertions as used herein include fusion proteins as is known in the art. The insertion can be a contiguous segment of amino acids or separated by one or more of the amino acids in the naturally occurring polypeptide.

[0063] The term “amino acid substitution set” or “substitution set” refers to a group of amino acid substitutions in a polypeptide sequence, as compared to a reference sequence. A substitution set can have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or more amino acid substitutions. In some embodiments, a substitution set refers to the set of amino acid substitutions that is present in any of the variant transaminases listed in the Tables provided in the Examples. The term “substitution set” is also used in reference to a group of nucleotide substitutions in a polynucleotide sequence, as compared to a reference sequence.

[0064] As used herein, “fragment” refers to a polypeptide that has an amino-terminal and/or carboxy-terminal deletion, but where the remaining amino acid sequence is identical to the corresponding positions in the sequence. Fragments can typically have about 80%, about 90%, about 95%, about 98%, or about 99% of the full-length transaminase polypeptide, for example the polypeptide of SEQ ID NO:2. In some embodiments, the fragment is “biologically active” (i.e., it exhibits the same enzymatic activity as the full-length sequence).

[0065] As used herein, “isolated polypeptide” refers to a polypeptide that is substantially separated from other contaminants that naturally accompany it (e.g., proteins, lipids, and polynucleotides). The term embraces polypeptides which have been removed or purified from their naturally-occurring environment or expression system (e.g., host cell or in vitro synthesis). The improved transaminase enzymes may be present within a cell, present in the cellular medium, or prepared in various forms, such as lysates or isolated preparations. As such, in some embodiments, the engineered transaminase polypeptides of the present invention can be an isolated polypeptide.

[0066] As used herein, “substantially pure polypeptide” refers to a composition in which the polypeptide species is the predominant species present (i.e., on a molar or weight basis it is more abundant than any other individual macromolecular species in the composition), and is generally a substantially purified composition when the object species comprises at least about 50 percent of the macromolecular species present by mole or % weight. Generally, a substantially pure engineered transaminase polypeptide composition comprises about 60% or more, about 70% or more, about 80% or more, about 90% or more, about 91% or more, about 92% or more, about 93% or more, about 94% or more, about 95% or more, about 96% or more, about 97% or more, about

98% or more, or about 99% of all macromolecular species by mole or % weight present in the composition. Solvent species, small molecules (<500 Daltons), and elemental ion species are not considered macromolecular species.

[0067] In some embodiments, the isolated improved transaminase polypeptide is a substantially pure polypeptide composition.

[0068] As used herein, “substantially pure polynucleotide” refers to a composition in which the polynucleotide species is the predominant species present (i.e., on a molar or weight basis it is more abundant than any other individual macromolecular species in the composition), and is generally a substantially purified composition when the object species comprises at least about 50 percent of the macromolecular species present by mole or % weight. Generally, a substantially pure engineered transaminase polynucleotide composition comprises about 60% or more, about 70% or more, about 80% or more, about 90% or more, about 91% or more, about 92% or more, about 93% or more, about 94% or more, about 95% or more, about 96% or more, about 97% or more, about 98% or more, or about 99% of all macromolecular species by mole or % weight present in the composition. In some embodiments, the isolated improved transaminase polypeptide is a substantially pure polynucleotide composition.

[0069] As used herein, “polynucleotide” and “nucleic acid” refer to two or more nucleosides that are covalently linked together. The polynucleotide may be wholly comprised ribonucleosides (i.e., an RNA), wholly comprised of 2' deoxyribonucleotides (i.e., a DNA) or mixtures of ribo- and 2' deoxyribonucleosides. While the nucleosides will typically be linked together via standard phosphodiester linkages, the polynucleotides may include one or more non-standard linkages. The polynucleotide may be single-stranded or double-stranded, or may include both single-stranded regions and double-stranded regions. Moreover, while a polynucleotide will typically be composed of the naturally occurring encoding nucleobases (i.e., adenine, guanine, uracil, thymine, and cytosine), it may include one or more modified and/or synthetic nucleobases (e.g., inosine, xanthine, hypoxanthine, etc.). Preferably, such modified or synthetic nucleobases will be encoding nucleobases.

[0070] The abbreviations used for the genetically encoding nucleosides are conventional and are as follows: adenosine (A); guanosine (G); cytidine (C); thymidine (T); and uridine (U). Unless specifically delineated, the abbreviated nucleotides may be either ribonucleosides or 2'-deoxyribonucleosides. The nucleosides may be specified as being either ribonucleosides or 2'-deoxyribonucleosides on an individual basis or on an aggregate basis. When nucleic acid sequences are presented as a string of one-letter abbreviations, the sequences are presented in the 5' to 3' direction in accordance with common convention, and the phosphates are not indicated.

[0071] As used herein, “hybridization stringency” relates to hybridization conditions, such as washing conditions, in the hybridization of nucleic acids. Generally, hybridization reactions are performed under conditions of lower stringency, followed by washes of varying but higher stringency. The term “moderately stringent hybridization” refers to conditions that permit target-DNA to bind a complementary nucleic acid that has about 60% identity, preferably about 75% identity, about 85% identity to the target DNA; with greater than about 90% identity to target-polynucleotide. Exemplary moderately stringent conditions are conditions equivalent to hybridization in 50% formamide, 5×Denhart's solution, 5×SSPE, 0.2% SDS at 42° C., followed by washing in 0.2×SSPE, 0.2% SDS, at 42° C. “High stringency hybridization” refers generally to conditions that are about 10° C. or less from the thermal melting temperature T_m as determined under the solution condition for a defined polynucleotide sequence. In some embodiments, a high stringency condition refers to conditions that permit hybridization of only those nucleic acid sequences that form stable hybrids in 0.018M NaCl at 65° C. (i.e., if a hybrid is not stable in 0.018M NaCl at 65° C., it will not be stable under high stringency conditions, as contemplated herein). High stringency conditions can be provided, for example, by hybridization in conditions equivalent to 50% formamide, 5×Denhart's solution, 5×SSPE, 0.2% SDS at 42° C., followed by

washing in 0.1×SSPE, and 0.1% SDS at 65° C. Another high stringency condition is hybridizing in conditions equivalent to hybridizing in 5×SSC containing 0.1% (w:v) SDS at 65° C. and washing in 0.1×SSC containing 0.1% SDS at 65° C. Other high stringency hybridization conditions, as well as moderately stringent conditions, are known to those of skill in the art.

[0072] As used herein, “coding sequence” refers to that portion of a nucleic acid (e.g., a gene) that encodes an amino acid sequence of a protein.

[0073] As used herein, “codon optimized” refers to changes in the codons of the polynucleotide encoding a protein to those preferentially used in a particular organism such that the encoded protein is efficiently expressed in the organism of interest. In some embodiments, the polynucleotides encoding the transaminase enzymes may be codon optimized for optimal production from the host organism selected for expression. Although the genetic code is degenerate in that most amino acids are represented by several codons, called “synonyms” or “synonymous” codons, it is well known that codon usage by particular organisms is nonrandom and biased towards particular codon triplets. This codon usage bias may be higher in reference to a given gene, genes of common function or ancestral origin, highly expressed proteins versus low copy number proteins, and the aggregate protein coding regions of an organism's genome. In some embodiments, the polynucleotides encoding the transaminase enzymes may be codon optimized for optimal production from the host organism selected for expression.

[0074] As used herein, “preferred, optimal, high codon usage bias codons” refers interchangeably to codons that are used at higher frequency in the protein coding regions than other codons that code for the same amino acid. The preferred codons may be determined in relation to codon usage in a single gene, a set of genes of common function or origin, highly expressed genes, the codon frequency in the aggregate protein coding regions of the whole organism, codon frequency in the aggregate protein coding regions of related organisms, or combinations thereof. Codons whose frequency increases with the level of gene expression are typically optimal codons for expression. A variety of methods are known for determining the codon frequency (e.g., codon usage, relative synonymous codon usage) and codon preference in specific organisms, including multivariate analysis, for example, using cluster analysis or correspondence analysis, and the effective number of codons used in a gene (See e.g., GCG CodonPreference, Genetics Computer Group Wisconsin Package; CodonW, John Peden, University of Nottingham; McInerney, *Bioinform.*, 14:372-73 [1998]; Stenico et al., *Nucleic Acids Res.*, 22:437-46 [1994]; and Wright, *Gene* 87:23-29 [1990]). Codon usage tables are available for a growing list of organisms (See e.g., Wada et al., *Nucleic Acids Res.*, 20:2111-2118 [1992]; Nakamura et al., *Nucl. Acids Res.*, 28:292 [2000]; Duret, et al., *supra*; Henaut and Danchin, “*Escherichia coli* and *Salmonella*,” Neidhardt, et al. (eds.), ASM Press, Washington D.C., [1996], p. 2047-2066. The data source for obtaining codon usage may rely on any available nucleotide sequence capable of coding for a protein. These data sets include nucleic acid sequences actually known to encode expressed proteins (e.g., complete protein coding sequences-CDS), expressed sequence tags (ESTS), or predicted coding regions of genomic sequences (See e.g., Uberbacher, *Meth. Enzymol.*, 266:259-281 [1996]; Tiwari et al., *Comput. Appl. Biosci.*, 13:263-270 [1997]).

[0075] As used herein, “control sequence” is defined herein to include all components, which are necessary or advantageous for the expression of a polynucleotide and/or polypeptide of the present invention. Each control sequence may be native or foreign to the polynucleotide of interest. Such control sequences include, but are not limited to, a leader, polyadenylation sequence, propeptide sequence, promoter, signal peptide sequence, and transcription terminator.

[0076] As used herein, “operably linked” is defined herein as a configuration in which a control sequence is appropriately placed (i.e., in a functional relationship) at a position relative to a polynucleotide of interest such that the control sequence directs or regulates the expression of the polynucleotide and/or polypeptide of interest.

[0077] As used herein, “promoter sequence” refers to a nucleic acid sequence that is recognized by

a host cell for expression of a polynucleotide of interest, such as a coding sequence. The control sequence may comprise an appropriate promoter sequence. The promoter sequence contains transcriptional control sequences, which mediate the expression of a polynucleotide of interest. The promoter may be any nucleic acid sequence which shows transcriptional activity in the host cell of choice including mutant, truncated, and hybrid promoters, and may be obtained from genes encoding extracellular or intracellular polypeptides either homologous or heterologous to the host cell.

[0078] As used herein, “naturally occurring” and “wild-type” refers to the form found in nature. For example, a naturally occurring or wild-type polypeptide or polynucleotide sequence is a sequence present in an organism that can be isolated from a source in nature and which has not been intentionally modified by human manipulation.

[0079] As used herein, “non-naturally occurring,” “engineered,” and “recombinant” when used in the present invention with reference to (e.g., a cell, nucleic acid, or polypeptide), refers to a material, or a material corresponding to the natural or native form of the material, that has been modified in a manner that would not otherwise exist in nature. In some embodiments the material is identical to naturally occurring material, but is produced or derived from synthetic materials and/or by manipulation using recombinant techniques. Non-limiting examples include, among others, recombinant cells expressing genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise expressed at a different level.

[0080] As used herein, “percentage of sequence identity,” “percent identity,” and “percent identical” refer to comparisons between polynucleotide sequences or polypeptide sequences, and are determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which either the identical nucleic acid base or amino acid residue occurs in both sequences or a nucleic acid base or amino acid residue is aligned with a gap to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

Determination of optimal alignment and percent sequence identity is performed using the BLAST and BLAST 2.0 algorithms (See e.g., Altschul et al., J. Mol. Biol. 215: 403-410 [1990]; and Altschul et al., Nucl. Acids Res., 25: 3389-3402 [1977]). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information website.

[0081] Briefly, the BLAST analyses involve first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as, the neighborhood word score threshold (Altschul et al., supra). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W , T , and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, $M=5$, $N=-4$, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the

[0082] Numerous other algorithms are available and known in the art that function similarly to BLAST in providing percent identity for two sequences. Optimal alignment of sequences for comparison can be conducted using any suitable method known in the art (e.g., by the local homology algorithm of Smith and Waterman, Adv. Appl. Math., 2:482 [1981]; by the homology alignment algorithm of Needleman and Wunsch, J. Mol. Biol., 48:443 [1970]; by the search for similarity method of Pearson and Lipman, Proc. Natl. Acad. Sci. USA 85:2444 [1988]; and/or by computerized implementations of these algorithms [GAP, BESTFIT, FASTA, and TFASTA in the GCG Wisconsin Software Package]), or by visual inspection, using methods commonly known in the art. Additionally, determination of sequence alignment and percent sequence identity can employ the BESTFIT or GAP programs in the GCG Wisconsin Software package (Accelrys, Madison WI), using the default parameters provided.

[0083] As used herein, “substantial identity” refers to a polynucleotide or polypeptide sequence that has at least 80 percent sequence identity, at least 85 percent identity and 89 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison window of at least 20 residue positions, frequently over a window of at least 30-50 residues, wherein the percentage of sequence identity is calculated by comparing the reference sequence to a sequence that includes deletions or additions which total 20 percent or less of the reference sequence over the window of comparison. In specific embodiments applied to polypeptides, the term “substantial identity” means that two polypeptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap weights, share at least 80 percent sequence identity, preferably at least 89 percent sequence identity, at least 95 percent sequence identity or more (e.g., 99 percent sequence identity). In some preferred embodiments, residue positions that are not identical differ by conservative amino acid substitutions.

[0084] As used herein, “reference sequence” refers to a defined sequence to which another sequence is compared. A reference sequence may be a subset of a larger sequence, for example, a segment of a full-length gene or polypeptide sequence. Generally, a reference sequence is at least 20 nucleotide or amino acid residues in length, at least 25 residues in length, at least 50 residues in length, or the full length of the nucleic acid or polypeptide. Since two polynucleotides or polypeptides may each (1) comprise a sequence (i.e., a portion of the complete sequence) that is similar between the two sequences, and (2) may further comprise a sequence that is divergent between the two sequences, sequence comparisons between two (or more) polynucleotides or polypeptide are typically performed by comparing sequences of the two polynucleotides over a comparison window to identify and compare local regions of sequence similarity. The term “reference sequence” is not intended to be limited to wild-type sequences, and can include engineered or altered sequences. For example, in some embodiments, a “reference sequence” can be a previously engineered or altered amino acid sequence.

[0085] As used herein, “comparison window” refers to a conceptual segment of at least about 20 contiguous nucleotide positions or amino acids residues wherein a sequence may be compared to a reference sequence of at least 20 contiguous nucleotides or amino acids and wherein the portion of the sequence in the comparison window may comprise additions or deletions (i.e., gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The comparison window can be longer than 20 contiguous residues, and includes, optionally 30, 40, 50, 100, or longer windows.

[0086] As used herein, “corresponding to,” “reference to,” and “relative to” when used in the context of the numbering of a given amino acid or polynucleotide sequence refers to the numbering of the residues of a specified reference sequence when the given amino acid or polynucleotide sequence is compared to the reference sequence. In other words, the residue number or residue

position of a given polymer is designated with respect to the reference sequence rather than by the actual numerical position of the residue within the given amino acid or polynucleotide sequence. For example, a given amino acid sequence, such as that of an engineered transaminase, can be aligned to a reference sequence by introducing gaps to optimize residue matches between the two sequences. In these cases, although the gaps are present, the numbering of the residue in the given amino acid or polynucleotide sequence is made with respect to the reference sequence to which it has been aligned. As used herein, a reference to a residue position, such as “X_n” as further described below, is to be construed as referring to “a residue corresponding to”, unless specifically denoted otherwise. Thus, for example, “X₉₄” refers to any amino acid at position 94 in a polypeptide sequence.

[0087] As used herein, when used in reference to a nucleic acid or polypeptide, the term “heterologous” refers to a sequence that is not normally expressed and secreted by an organism (e.g., a wild-type organism). In some embodiments, the term encompasses a sequence that comprises two or more subsequences which are not found in the same relationship to each other as normally found in nature, or is recombinantly engineered so that its level of expression, or physical relationship to other nucleic acids or other molecules in a cell, or structure, is not normally found in nature. For instance, a heterologous nucleic acid is typically recombinantly produced, having two or more sequences from unrelated genes arranged in a manner not found in nature (e.g., a nucleic acid open reading frame (ORF) of the invention operatively linked to a promoter sequence inserted into an expression cassette, such as a vector). In some embodiments, “heterologous polynucleotide” refers to any polynucleotide that is introduced into a host cell by laboratory techniques, and includes polynucleotides that are removed from a host cell, subjected to laboratory manipulation, and then reintroduced into a host cell.

[0088] As used herein, “improved enzyme property” refers to a transaminase that exhibits an improvement in any enzyme property as compared to a reference transaminase. For the engineered transaminase polypeptides described herein, the comparison is generally made to the wild-type transaminase enzyme, although in some embodiments, the reference transaminase can be another improved engineered transaminase. Enzyme properties for which improvement is desirable include, but are not limited to, enzymatic activity (which can be expressed in terms of percent conversion of the substrate at a specified reaction time using a specified amount of transaminase), chemoselectivity, thermal stability, solvent stability, pH activity profile, cofactor requirements, refractoriness to inhibitors (e.g., product inhibition), stereospecificity, and stereoselectivity (including enantioselectivity).

[0089] As used herein, “increased enzymatic activity” refers to an improved property of the engineered transaminase polypeptides, which can be represented by an increase in specific activity (e.g., product produced/time/weight protein) or an increase in percent conversion of the substrate to the product (e.g., percent conversion of starting amount of substrate to product in a specified time period using a specified amount of transaminase) as compared to the reference transaminase enzyme. Exemplary methods to determine enzyme activity are provided in the Examples. Any property relating to enzyme activity may be affected, including the classical enzyme properties of $K_{sub.m}$, $V_{sub.max}$ or $k_{sub.cat}$, changes of which can lead to increased enzymatic activity. Improvements in enzyme activity can be from about 1.5 times the enzymatic activity of the corresponding wild-type transaminase enzyme, to as much as 2 times, 5 times, 10 times, 20 times, 25 times, 50 times, 75 times, 100 times, or more enzymatic activity than the naturally occurring transaminase or another engineered transaminase from which the transaminase polypeptides were derived. In some embodiments, the engineered transaminase enzyme exhibits improved enzymatic activity in the range of 1.5 to 50 times, 1.5 to 100 times greater than that of the parent transaminase enzyme. It is understood by the skilled artisan that the activity of any enzyme is diffusion limited such that the catalytic turnover rate cannot exceed the diffusion rate of the substrate, including any required cofactors. The theoretical maximum of the diffusion limit, or $k_{sub.cat}/K_{sub.m}$, is

generally about 10.sup.8 to 10.sup.9 (M.sup.-1 s.sup.-1). Hence, any improvements in the enzyme activity of the transaminase will have an upper limit related to the diffusion rate of the substrates acted on by the transaminase enzyme. Comparisons of enzyme activities are made using a defined preparation of enzyme, a defined assay under a set condition, and one or more defined substrates, as further described in detail herein. Generally, when lysates are compared, the numbers of cells and the amount of protein assayed are determined as well as use of identical expression systems and identical host cells to minimize variations in amount of enzyme produced by the host cells and present in the lysates.

[0090] As used herein, “increased enzymatic activity” and “increased activity” refer to an improved property of an engineered enzyme, which can be represented by an increase in specific activity (e.g., product produced/time/weight protein) or an increase in percent conversion of the substrate to the product (e.g., percent conversion of starting amount of substrate to product in a specified time period using a specified amount of transaminase) as compared to a reference enzyme as described herein. Any property relating to enzyme activity may be affected, including the classical enzyme properties of $K_{sub.m}$, $V_{sub.max}$ or $k_{sub.cat}$, changes of which can lead to increased enzymatic activity. Comparisons of enzyme activities are made using a defined preparation of enzyme, a defined assay under a set condition, and one or more defined substrates, as further described in detail herein. Generally, when enzymes in cell lysates are compared, the numbers of cells and the amount of protein assayed are determined as well as use of identical expression systems and identical host cells to minimize variations in amount of enzyme produced by the host cells and present in the lysates.

[0091] As used herein, “conversion” refers to the enzymatic transformation of a substrate to the corresponding product.

[0092] As used herein “percent conversion” refers to the percent of the substrate that is converted to the product within a period of time under specified conditions. Thus, for example, the “enzymatic activity” or “activity” of a transaminase polypeptide can be expressed as “percent conversion” of the substrate to the product.

[0093] As used herein, “chemoselectivity” refers to the preferential formation in a chemical or enzymatic reaction of one product over another.

[0094] As used herein, “thermostable” and “thermal stable” are used interchangeably to refer to a polypeptide that is resistant to inactivation when exposed to a set of temperature conditions (e.g., 40-80° C.) for a period of time (e.g., 0.5-24 hrs) compared to the untreated enzyme, thus retaining a certain level of residual activity (e.g., more than 60% to 80%) after exposure to elevated temperatures.

[0095] As used herein, “solvent stable” refers to the ability of a polypeptide to maintain similar activity (e.g., more than 60% to 80%) after exposure to varying concentrations (e.g., 5-99%) of solvent (e.g., isopropyl alcohol, tetrahydrofuran, 2-methyltetrahydrofuran, acetone, toluene, butylacetate, methyl tert-butylether, etc.) for a period of time (e.g., 0.5-24 hrs) compared to the untreated enzyme.

[0096] As used herein, “pH stable” refers to a transaminase polypeptide that maintains similar activity (e.g., more than 60% to 80%) after exposure to high or low pH (e.g., 4.5-6 or 8 to 12) for a period of time (e.g., 0.5-24 hrs) compared to the untreated enzyme.

[0097] As used herein, “thermo- and solvent stable” refers to a transaminase polypeptide that is both thermostable and solvent stable.

[0098] As used herein, “suitable reaction conditions” refer to those conditions in the biocatalytic reaction solution (e.g., ranges of enzyme loading, substrate loading, cofactor loading, temperature, pH, buffers, co-solvents, etc.) under which a transaminase polypeptide of the present invention is capable of transamination. Exemplary “suitable reaction conditions” are provided in the present invention and illustrated by the Examples.

[0099] As used herein, “loading,” such as in “compound loading,” “enzyme loading,” or “cofactor

loading” refers to the concentration or amount of a component in a reaction mixture at the start of the reaction.

[0100] As used herein, “substrate” in the context of a biocatalyst mediated process refers to the compound or molecule acted on by the biocatalyst.

[0101] As used herein “product” in the context of a biocatalyst mediated process refers to the compound or molecule resulting from the action of the biocatalyst.

[0102] As used herein, “equilibration” as used herein refers to the process resulting in a steady state concentration of chemical species in a chemical or enzymatic reaction (e.g., interconversion of two species A and B), including interconversion of stereoisomers, as determined by the forward rate constant and the reverse rate constant of the chemical or enzymatic reaction.

[0103] “Transaminase” or “aminotransferase” are used interchangeably herein to refer to a polypeptide having an enzymatic capability of transferring an amino group (—NH.sub.2), a pair of electrons, and a proton from the primary amine of an amine donor compound to the carbonyl group (C=O) of an amine acceptor compound, thereby converting the amine donor compound into its corresponding carbonyl compound and the carbonyl acceptor compound into its corresponding primary amine compound (See e.g., Scheme 1). Transaminases as used herein include naturally occurring (wild type) transaminase as well as non-naturally occurring engineered polypeptides generated by human manipulation.

[0104] “Amino group donor” or “amino donor” used interchangeably herein to refer to an amino group containing compound which is capable of donating an amino group to an acceptor carbonyl compound (i.e., an amino group acceptor), thereby becoming a carbonyl by-product. Amino group donors have the general structural formula,

##STR00007##

in which each of R.sup.1 , and R.sup.2 , when taken independently, is an alkyl, an alkylaryl group, or aryl group which is unsubstituted or substituted with one or more enzymatically non-inhibiting groups. R.sup.1 can be the same or different from R.sup.2 in structure or chirality. The groups R.sup.1 and R.sup.2 , taken together, may form a ring that is unsubstituted, substituted, or fused to other rings. Typical amino group donors include chiral and achiral amino acids, and chiral and achiral amines.

[0105] “Chiral amine” refers to an amino group containing compound having the general structural formula,

##STR00008##

in which each of R.sup.1 , and R.sup.2 , when taken independently, is an alkyl, an alkylaryl group, or aryl group which is unsubstituted or substituted with one or more groups. R.sup.1 is different from R.sup.2 in structure causing the carbon bearing the amino group (denoted with a \ast) to be stereogenic center. The groups R and R.sup.2 , taken together, may form a ring that is unsubstituted, substituted, or fused to other rings but is otherwise not symmetrical.

[0106] “Carbonyl by-product” refers to the carbonyl compound formed from the amino group donor when the amino group on the amino group donor is transferred to the amino group acceptor in a transamination reaction. The carbonyl by-product has the general structural formula,

##STR00009##

wherein R.sup.1 and R.sup.2 are defined above for the amino group donor.

[0107] “Amino acceptor” and “amine acceptor,” “keto substrate,” are used interchangeably herein to refer to a carbonyl group containing compound that accepts the amino group from an amino group donor in a reaction mediated by a transaminase (See e.g., Scheme 1). In the context of the present invention, the amino acceptor compound for the transaminase can include, among others, compound (2).

[0108] “Cofactor,” as used herein, refers to a non-protein compound that operates in combination with an enzyme in catalyzing a reaction. As used herein, “cofactor” is intended to encompass the vitamin B.sub.6 family compounds PLP, PN, PL, PM, PNP, and PMP, which are sometimes also

referred to as coenzymes.

[0109] “Pyridoxal-phosphate,” “PLP,” “pyridoxal-5'-phosphate,” “PYP,” and “P5P” are used interchangeably herein to refer to the compound that acts as a cofactor in transaminase reactions. In some embodiments, pyridoxal phosphate is defined by the structure 1-(4'-formyl-3'-hydroxy-2'-methyl-5'-pyridyl)methoxyphosphonic acid, CAS number [54-47-7]. Pyridoxal-5'-phosphate can be produced in vivo by phosphorylation and oxidation of pyridoxol (also known as Vitamin B.sub.6). In transamination reactions using transaminase enzymes, the amine group of the amino donor is transferred to the cofactor to produce a keto byproduct, while pyridoxal-5'-phosphate is converted to pyridoxamine phosphate. Pyridoxal-5'-phosphate is regenerated by reaction with a different keto compound (the amino acceptor).

[0110] The transfer of the amine group from pyridoxamine phosphate to the amino acceptor produces an amine and regenerates the cofactor. In some embodiments, the pyridoxal-5'-phosphate can be replaced by other members of the vitamin B.sub.6 family, including pyridoxine (PN), pyridoxal (PL), pyridoxamine (PM), and their phosphorylated counterparts; pyridoxine phosphate (PNP), and pyridoxamine phosphate (PMP).

[0111] “Alkyl” refers to groups of from 1 to 18 carbon atoms, either straight chained or branched, particularly from 1 to 8 carbon atoms, and more particularly 1 to 6 carbon atoms. An alkyl with a specified number of carbon atoms is denoted in parenthesis (e.g., (C.sub.1-C.sub.4) alkyl refers to an alkyl of 1 to 4 carbon atoms).

[0112] “Alkenyl” refers to groups of from 2 to 12 carbon atoms, either straight or branched containing at least one double bond but optionally containing more than one double bond.

[0113] “Alkynyl” refers to groups of from 2 to 12 carbon atoms, either straight or branched containing at least one triple bond but optionally containing more than one triple bond, and optionally containing one or more double bonded moieties.

[0114] “Aryl” refers to an unsaturated aromatic carbocyclic group of from 5 to 14 carbon atoms having a single ring (e.g., phenyl) or multiple condensed rings (e.g., naphthyl or anthryl). For multiple condensed rings, at least one of the rings is aromatic. Representative aryls include phenyl, pyridyl, naphthyl and the like.

[0115] “Arylalkyl” refers to an alkyl substituted with an aryl moiety. Representative arylalkyl groups include benzyl, phenethyl and the like.

[0116] “Arylalkenyl” refers to an alkenyl as defined herein substituted with an aryl group.

[0117] “Arylalkynyl” refers to an alkynyl as defined herein substituted with an aryl group.

[0118] “Heteroaryl” refers to an aromatic heterocyclic group of 5 to 14 ring atoms containing 1 to 4 ring heteroatoms selected from oxygen, nitrogen and sulfur within the ring. Heteroaryl groups can have a single ring (e.g., pyridyl or furyl) or multiple condensed rings (e.g., indolizinyll or benzothienyl). For multiple condensed rings, at least one of the rings is aromatic.

[0119] “Heteroarylalkyl” refers to an alkyl substituted with a heteroaryl moiety as defined herein.

[0120] “Heteroarylalkenyl” refers to an alkenyl substituted with a heteroaryl group as defined herein.

[0121] “Heteroarylalkynyl” refers to an alkynyl substituted with a heteroaryl moiety as defined herein.

[0122] “Cycloalkyl” refers to cyclic alkyl groups of from 3 to 12 carbon atoms having a single cyclic ring or multiple condensed rings. Representative cycloalkyl groups include, by way of example, single ring structures such as cyclopropyl, cyclobutyl, cyclopentyl, cyclooctyl, 1-methylcyclopropyl, 2-methylcyclopentyl, 2-methylcyclooctyl, and the like, or multiple ring structures, including bridged ring systems, such as adamantyl, and the like.

[0123] “Heterocycle” and interchangeably “heterocycloalkyl” refer to a saturated or unsaturated group having a single ring or multiple condensed rings, from 3 to 14 ring atoms having from 1 to 4 hetero atoms selected from nitrogen, sulfur or oxygen within the ring. Heterocyclic groups can have a single ring (e.g., piperidinyl or tetrahydrofuryl) or multiple condensed rings (e.g., indolinyll,

dihydrobenzofuran or quinuclidinyl). Representative heterocycles and heteroaryls include, but are not limited to, furan, thiophene, thiazole, oxazole, pyrrole, imidazole, pyrazole, pyridine, pyrazine, pyrimidine, pyridazine, indolizine, isoindole, indole, indazole, purine, quinolizine, isoquinoline, quinoline, phthalazine, naphthylpyridine, quinoxaline, quinazoline, cinnoline, pteridine, carbazole, carboline, phenanthridine, acridine, phenanthroline, isothiazole, phenazine, isoxazole, phenoxazine, phenothiazine, imidazolidine, imidazoline, piperidine, piperazine, pyrrolidine, indoline and the like.

[0124] “Cycloalkylalkyl” refers to an alkyl substituted with a cycloalkyl moiety as defined herein.

[0125] “Cycloalkylalkenyl” refers to an alkenyl substituted with a cycloalkyl moiety as defined herein.

[0126] “Cycloalkylalkynyl” refers to an alkynyl substituted with a cycloalkyl moiety as defined herein.

[0127] “Heterocycloalkylalkyl” refers to an alkyl substituted with a heterocycloalkyl moiety as defined herein.

[0128] “Heterocycloalkenyl” refers to an alkenyl substituted with a heterocycloalkyl moiety as defined herein.

[0129] “Heterocycloalkylalkynyl” refers to an alkynyl substituted with a heterocycloalkyl moiety as defined herein.

[0130] “Alkoxy” or “Alkyloxy” refers to the group alkyl-O—wherein the alkyl group is as defined above, including optionally substituted alkyl groups as also defined above.

[0131] “Amino” refers to the group —NH.sub.2. Substituted amino refers to the group —NHR', NR'R', and NR'R'R', where each R' is independently of the others selected from substituted or unsubstituted alkyl, cycloalkyl, heterocycloalkyl, cycloalkylalkyl, heterocycloalkylalkyl, alkyloxy, aryl, heteroaryl, arylalkyl, heteroarylalkyl, acyl, alkyloxycarbonyl, sulfanyl, sulfinyl, sulfonyl, and the like. Typical amino groups include, but are not limited to, dimethylamino, diethylamino, trimethylammonium, triethylammonium, methylsulfonylamino, furanyl-oxy-sulfamino, and the like.

[0132] “Carboxy” refers to —COOH.

[0133] “Carbonyl” refers to —C(O)—, which may have a variety of substituents to form different carbonyl groups including acids, acid halides, aldehydes, amides, esters, and ketones.

[0134] “Hydroxy” refers to —OH.

[0135] “Cyano” refers to —CN.

[0136] “Halogen” or “halo” refers to fluoro, chloro, bromo and iodo.

[0137] “Sulfonyl” refers to —SO.sub.2—. Substituted sulfonyl refers to —SO.sub.2R', where R' is a suitable substituent as described below.

[0138] “Fused” or “fused rings” such as in fused aryl or fused heteroaryl refers to two or more rings joined such that they have at least two ring atoms in common. Fused aryl refers to fused rings in which at least one of the rings is an aryl. Fused heteroaryl refers to fused rings in which at least one of the rings is a heteroaryl.

[0139] “Substituted” unless otherwise specified, refers to replacement of positions occupied by hydrogen in the foregoing groups with substituents exemplified by, but not limited to, hydroxy, oxo, nitro, methoxy, ethoxy, alkyloxy, substituted alkyloxy, trifluoromethoxy, haloalkyloxy, fluoro, chloro, bromo, iodo, halo, methyl, ethyl, propyl, butyl, alkyl, alkenyl, alkynyl, substituted alkyl, trifluoromethyl, haloalkyl, hydroxyalkyl, alkyloxyalkyl, thio, alkylthio, acyl, carboxy, alkyloxycarbonyl, carboxamido, substituted carboxamido, alkylsulfonyl, alkylsulfinyl, alkylsulfonylamino, sulfonamido, substituted sulfonamido, cyano, amino, substituted amino, alkylamino, dialkylamino, aminoalkyl, acylamino, amidino, amidoximo, hydroxamoyl, phenyl, aryl, substituted aryl, aryloxy, arylalkyl, arylalkenyl, arylalkynyl, pyridyl, imidazolyl, heteroaryl, substituted heteroaryl, heteroaryloxy, heteroarylalkyl, heteroarylalkenyl, heteroarylalkynyl, cyclopropyl, cyclobutyl, cyclopentyl, cyclohexyl, cycloalkyl, cycloalkenyl, cycloalkylalkyl, substituted cycloalkyl, cycloalkyloxy, pyrrolidinyl, piperidinyl, morpholino, heterocycle,

(heterocycle)oxy, and (heterocycle)alkyl; and preferred heteroatoms are oxygen, nitrogen, and sulfur. It is understood that where open valences exist on these substituents they can be further substituted with alkyl, cycloalkyl, aryl, heteroaryl, and/or heterocycle groups, that where these open valences exist on carbon they can be further substituted by halogen and by oxygen-, nitrogen-, or sulfur-bonded substituents, and where multiple such open valences exist, these groups can be joined to form a ring, either by direct formation of a bond or by formation of bonds to a new heteroatom, preferably oxygen, nitrogen, or sulfur. It is further understood that the above substitutions can be made provided that replacing the hydrogen with the substituent does not introduce unacceptable instability to the molecules of the present invention, and is otherwise chemically reasonable.

[0140] “Optional” or “optionally” means that the subsequently described event or circumstance may or may not occur, and that the description includes instances where said event or circumstance occurs and instances in which it does not. One of ordinary skill in the art would understand that with respect to any molecule described as containing one or more optional substituents, only sterically practical and/or synthetically feasible compounds are meant to be included. “Optionally substituted” refers to all subsequent modifiers in a term or series of chemical groups. For example, in the term “optionally substituted arylalkyl, the “alkyl” portion and the “aryl” portion of the molecule may or may not be substituted, and for the series “optionally substituted alkyl, cycloalkyl, aryl and heteroaryl,” the alkyl, cycloalkyl, aryl, and heteroaryl groups, independently of the others, may or may not be substituted.

[0141] “Protecting group” refers to a group of atoms that mask, reduce or prevent the reactivity of the functional group when attached to a reactive functional group in a molecule. Typically, a protecting group may be selectively removed as desired during the course of a synthesis. Examples of protecting groups are known in the art (e.g., Wuts and Greene, “Greene's Protective Groups in Organic Synthesis,” 4^{sup}.f Ed., Wiley Interscience [2006], and Harrison et al., Compendium of Synthetic Organic Methods, Vols. 1-8, John Wiley & Sons, NY [1971-1976]. Functional groups that can have a protecting group include, but are not limited to, hydroxy, amino, and carboxy groups. Representative amino protecting groups include, but are not limited to, formyl, acetyl, trifluoroacetyl, benzyl, benzyloxycarbonyl (“CBZ”), tert-butoxycarbonyl (“Boc”), trimethylsilyl (“TMS”), 2-trimethylsilyl-ethanesulfonyl (“SES”), trityl and substituted trityl groups, allyloxycarbonyl, 9-fluorenylmethyloxycarbonyl (“Fmoc”), nitro-veratryloxycarbonyl (“NVOC”) and the like.

[0142] Representative hydroxyl protecting groups include, but are not limited to, those where the hydroxyl group is either acylated (e.g., methyl and ethyl esters, acetate or propionate groups or glycol esters) or alkylated such as benzyl and trityl ethers, as well as alkyl ethers, tetrahydropyranyl ethers, trialkylsilyl ethers (e.g., TMS or TIPPS groups) and allyl ethers. Other protecting groups can be found in the references noted herein.

[0143] “Leaving group” generally refers to any atom or moiety that is capable of being displaced by another atom or moiety in a chemical reaction. More specifically, a leaving group refers to an atom or moiety that is readily displaced and substituted by a nucleophile (e.g., an amine, a thiol, an alcohol, or cyanide). Such leaving groups are well known and include carboxylates, N-hydroxysuccinimide (“NHS”), N-hydroxybenzotriazole, a halogen (fluorine, chlorine, bromine, or iodine), and alkyloxy groups. Non-limiting characteristics and examples of leaving groups are known in the art and described in various chemistry texts.

Engineered Transaminase Polypeptides

[0144] The present invention provides engineered polypeptides having transaminase activity (also referred to herein as “engineered transaminase polypeptides”) useful for the selective transamination of amino acceptor substrate compounds produce chiral amine products, which, in some embodiments, can include compound (1). Accordingly, in one aspect, the present invention provides engineered polypeptides having transaminase activity which are capable of converting

substrate compound (2) to product compound (1) as shown in Scheme 2. Further, the present invention provides polynucleotides encoding the engineered polypeptides, associated vectors and host cells comprising the polynucleotides, methods for making the engineered polypeptides, and methods for using the engineered polypeptides, including suitable reaction conditions.

[0145] The engineered polypeptides of the present invention are non-naturally occurring transaminases engineered to have improved enzyme properties (such as increased stereoselectivity) as compared to the wild-type transaminase polypeptide of *Vibrio fluvialis* JS17 (GenBank Acc. No. AEA39183.1, GI: 327207066; SEQ ID NO:2), and also as compared to the reference engineered transaminase polypeptide of SEQ ID NO:4, which was used as the starting backbone sequence for the directed evolution of the engineered polypeptides of the present invention. The reference engineered transaminase polypeptide of SEQ ID NO:4 has the following 28 amino acid differences relative to the wild-type transaminase of *Vibrio fluvialis* JS17 (SEQ ID NO:2): A9T, N45H, L56A, W57L, A74T, F86S, V153A, K163F, V177L, R211K, N286S, M294V, A323T, S324G, A383V, T391A, and D394G.

[0146] The engineered transaminase polypeptides of the present invention were generated by directed evolution of SEQ ID NO:4 for efficient conversion of compound (2) to compound (1) under certain industrially relevant conditions and have one or more residue differences as compared to a reference engineered transaminase polypeptide. These residue differences are associated with improvements in various enzyme properties, particularly increased activity, increased stereoselectivity, increased stability, and tolerance of increased substrate and/or product concentration (e.g., decreased product inhibition). Accordingly, in some embodiments, the engineered polypeptides having transaminase activity are capable of converting the substrate compound (2) to compound (1) with an activity that is increased at least about 1.2 fold, 1.5 fold, 2 fold, 3 fold, 4 fold, 5 fold, 10 fold, 20 fold, 30 fold, 40 fold, 50 fold, 100 fold, 200 fold, 500 fold, 1000 fold, 10,000 fold, 20,000 fold, 30,000 fold, 40,000 fold, 50,000 fold, 60,000 fold, 70,000 fold, 80,000 fold, 90,000 fold, 95,000 fold, or more relative to the activity of a reference polypeptide (e.g., SEQ ID NO:2, 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156), under suitable reaction conditions. In some embodiments, the engineered polypeptides having transaminase activity are capable of converting the substrate of compound (2) to compound (1) with a percent conversion of at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, or at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, in a reaction time of about 48 h, about 36 h, about 24 h, or even a shorter length of time, under suitable reaction conditions. In some embodiments, the engineered polypeptides having transaminase activity are capable of converting compound (2) to compound (1) in diastereomeric excess of at least 90%, 95%, 97%, 98%, 99%, or greater, under suitable reaction conditions.

[0147] The present invention provides numerous exemplary engineered transaminase polypeptides comprising amino acid sequences of the even-numbered sequence identifiers SEQ ID NO:6-172.

[0148] These exemplary engineered transaminase polypeptides comprise amino acid sequences that include one or more of the following residue differences associated with their improved properties for conversion of compound (2) to compound (1) as compared to a reference sequence (e.g., SEQ ID NO:2, 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156).

[0149] In some cases, the exemplary engineered polypeptides have an amino acid sequence that further comprises one or more residue differences as compared to a reference sequence (e.g., SEQ ID NO: 2, 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156). In some cases, the exemplary engineered polypeptides have an amino acid sequence that further comprises one or more residue differences as compared to a reference sequence (e.g., SEQ ID NO: 2, 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156).

[0150] In some embodiments, the engineered polypeptide comprises an amino acid sequence that is

at least about 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, identical to a reference sequence selected from SEQ ID NO: 2, 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156, where the polypeptide has transaminase activity and one or more of the improved properties as described herein, for example the ability to convert compound (2) to product compound (1) with increased activity compared to a reference sequence (e.g., the polypeptide of SEQ ID NO: 2, 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156). In some embodiments, the reference sequence is SEQ ID NO:4. In some embodiments, the reference sequence is SEQ ID NO: 18.

[0151] In some embodiments, the reference sequence is SEQ ID NO:20. In some embodiments, the reference sequence is SEQ ID NO:30. In some embodiments, the reference sequence is SEQ ID NO:48. In some embodiments, the reference sequence is SEQ ID NO:66. In some embodiments, the reference sequence is SEQ ID NO:70. In some embodiments, the reference sequence is SEQ ID NO:88. In some embodiments, the reference sequence is SEQ ID NO: 110. In some embodiments, the reference sequence is SEQ ID NO:112. In some embodiments, the reference sequence is SEQ ID NO:120. In some embodiments, the reference sequence is SEQ ID NO:138. In some embodiments, the reference sequence is SEQ ID NO:156.

[0152] In some embodiments, the engineered transaminase polypeptide comprising an amino acid sequence has one or more amino acid residue differences as compared to SEQ ID NO: 2, 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156. In some embodiments, the present invention provides an engineered polypeptide having transaminase activity comprising an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to reference sequence of SEQ ID NO: 2, 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156 and (a) at least one amino acid residue difference selected from those substitutions provided herein (See e.g., Tables 6.1, 7.1, 8.1, 9.1, 10.1, 11.1, 12.1, 13.1, 13.2, 14.1, 15.1, 16.1, 17.1, and 18.1).

[0153] In some embodiments, the present invention provides an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:4 selected from 17, 17/259/347, 20, 45/48/78/81/92, 57, 57/319, 85, 86, 157, 158, 161, 203/416, 314, 316, 319, 375/416, 404, 415, 416, 417, 420, 421, 422, and 424, wherein the positions are numbered with reference to SEQ ID NO:4. In some embodiments, the amino acid differences comprise the substitution(s) 17V/259T/347K, 17W, 20R, 45R/48L/78C/81V/92H, 57I, 57T/319E, 85G, 85S, 85V, 86A, 86G, 157C, 157G, 157P, 157T, 158S, 161G, 203H/416G, 203H/416S, 314R, 316K, 316L, 316M, 316S, 319G, 319L, 319V, 375L/416F, 404T, 404V, 415V, 416C, 416G, 416S, 416Y, 417G, 417M, 417P, 420G, 421I, 422A, 422G, 422R, 424A, and 424R, wherein the positions are numbered with reference to SEQ ID NO:4. In some additional embodiments, the amino acid differences comprise the substitution(s) Y17V/I259T/E347K, Y17W, T20R, H45R/R48L/R78C/G81V/Q92H, L57I, L57T/H319E, F85G, F85S, F85V, S86A, S86G, S157C, S157G, S157P, S157T, A158S, T161G, R203H/P416G, R203H/P416S, I314R, E316K, E316L, E316M, E316S, H319G, H319L, H319V, I375L/P416F, A404T, A404V, R415V, P416C, P416G, P416S, P416Y, L417G, L417M, L417P, S420G, V421I, V422A, V422G, V422R, C424A, and C424R, wherein the positions are numbered with reference to SEQ ID NO:4.

[0154] In some embodiments, the present invention provides an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:18 selected from 5/17/86/157/259/314/316/422, 17/86/157/203/259/314/316/422, 17/86/157/314/316/347/422, 17/86/157/316/422, 85/86, 86/157/314/316/347/416/422, 86/157/416, 86/203/316/416, 86/203/316/422, 86/316/416, 139/221/310/316, 157/316, 195/316, 207/316, 227/422, 280/381/416/445, 314, 314/316, 316, 416, 417, and 422, wherein the positions are numbered with reference to SEQ ID NO:18. In some embodiments, the amino acid difference(s) comprise the substitution(s)

5H/17V/86G/157G/314R/316S/422R, 17V/86A/157G/316M/422A, 17V/86G/157G/203H/259T/314R/316L/422R, 17V/86G/157G/316M/422R, 17V/86G/157T/316M/422R, 17W/86A/157T/314R/316M/347K/422R, 85F/86G, 86A/157G/314R/316S/347K/416C/422A, 86A/203H/316M/416G, 86A/203H/316M/422G, 86A/316S/416C, 86G/157T/416G, 139H/221I/310E/316G, 157T/316M, 195K/316S, 207K/316T, 227F/422G, 280E/381V/416G/445R, 314R, 314R/316M, 316C, 316F, 316G, 316M, 316S, 316W, 316Y, 416A, 416G, 416S, 417M, and 422C, wherein the positions are numbered with reference to SEQ ID NO:18. In some additional embodiments, the amino acid difference(s) comprise the substitution(s) Q5H/Y17V/S86G/S157G/I259T/I314R/E316S/V422R, Y17V/S86A/S157G/E316M/V422A, Y17V/S86G/S157G/R203H/I259T/I314R/E316L/V422R, Y17V/S86G/S157G/E316M/V422R, Y17V/S86G/S157T/E316M/V422R, Y17W/S86A/S157T/I314R/E316M/E347K/V422R, V85F/S86G, S86A/S157G/I314R/E316S/E347K/P416C/V422A, S86A/R203H/E316M/P416G, S86A/R203H/E316M/V422G, S86A/E316S/P416C, S86G/S157T/P416G, Q139H/F221I/A310E/E316G, S157T/E316M, E195K/E316S, E207K/E316T, G227F/V422G, A280E/L381V/P416G/L445R, I314R, I314R/E316M, E316C, E316F, E316G, E316M, E316S, E316W, E316Y, P416A, P416G, P416S, L417M, and V422C, wherein the positions are numbered with reference to SEQ ID NO:18.

[0155] In some embodiments, the present invention provides an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:20 selected from 17/139/314/416/417/422, 17/195/316/414/416/422, 17/314/316/414/416, 17/314/316/414/416/417/422, 20, 31, 54, 57, 83, 86, 86/194,121, 139/195/227/314/316/416, 139/202/207/227/314/316/416/422, 139/207/227/314/414/416/422, 139/227/310/414/416/417, 139/316/414/416/417/422, 146/157, 164/207/213/227/310/316/416, 166, 178/314/316/416, 190/195/207/213/244/314/414/416/417/422, 194/213/310/314/316/422, 195/227/314/316/416, 195/227/314/416/417, 195/227/314/422, 195/227/414/416, 195/227/414/416/422, 195/227/416/417, 195/316/416/417/422, 207/208/220/241/291/314/316, 226, 227, 227/416, 228, 258, 259, 271/310/316/416/422, 282, 310/314/316/414/416/417/422, 314/316/414/416/422, 314/316/416, 314/316/416/422, 316, 319, 320, 322, and 417, wherein the positions are numbered with reference to SEQ ID NO:20. In some embodiments, the amino acid difference(s) comprise the substitution(s) 17V/139H/314R/416A/417M/422A, 17V/195K/316W/414V/416S/422C, 17V/314R/316S/414V/416C, 17V/314R/316W/414V/416S/417M/422A, 20K, 20L, 20V, 31A, 31C, 31D, 31E, 31G, 31Q, 54C, 57A, 57C, 83G, 86G, 86G/194K, 86K, 121A, 121C, 121G, 121H, 121L, 121S, 121V, 139H/195K/227F/314R/316W/416C, 139H/202G/207K/227F/314R/316F/416S/422C, 139H/207K/227F/314R/414V/416C/422G, 139H/227F/310E/414V/416C/417M, 139H/316S/414V/416S/417M/422A, 146K/157G, 146V/157G, 164Q/207K/213S/227F/310E/316F/416A, 166T, 178L/314A/316W/416S, 190K/195K/207K/213S/244A/314R/414V/416S/417M/422G, 194D/213C/310E/314R/316W/422C, 195K/227F/314R/316W/416A, 195K/227F/314R/416A/417M, 195K/227F/314R/422C, 195K/227F/414V/416A/422G, 195K/227F/414V/416C, 195K/227F/416S/417M, 195K/316W/416S/417M/422A, 207K/208N/220Y/241V/291S/314R/316S, 226A, 226E, 226T, 227F/416C, 227V, 228G, 258L, 259V, 271R/310E/316W/416S/422A, 282L, 310E/314R/316S/414V/416A/417M/422A, 314A/316W/416A, 314R/316W/414V/416A/422A, 314R/316W/416C/422C, 316A, 316E, 316G, 316H, 316K, 316Q, 316R, 316S, 316T, 316W, 316Y, 319V, 320W, 322S, and 417S, wherein the positions are numbered with reference to SEQ ID NO:20.

[0156] In some additional embodiments, the amino acid difference(s) comprise the substitution(s) Y17V/Q139H/I314R/G416A/L417M/V422A, Y17V/E195K/M316W/C414V/G416S/V422C, Y17V/I314R/M316S/C414V/G416C, Y17V/I314R/M316W/C414V/G416S/L417M/V422A,

T20K, T20L, T20V, V31A, V31C, V31D, V31E, V31Q, S54C, L57A, L57C, H83G, A86G, A86G/E194K, A86K, N121A, N121C, N121G, N121H, N121L, N121S, N121V, Q139H/E195K/G227F/I314R/M316W/G416C, Q139H/A202G/E207K/G227F/I314R/M316F/G416S/V422C, Q139H/E207K/G227F/I314R/C414V/G416C/V422G, Q139H/G227F/A310E/C414V/G416C/L417M, Q139H/M316S/C414V/G416S/L417M/V422A, R146K/S157G, R146V/S157G, P164Q/E207K/G213S/G227F/A310E/M316F/G416A, N166T, H178L/I314A/M316W/G416S, E190K/E195K/E207K/G213S/P244A/I314R/C414V/G416S/L417M/V422G, E194D/G213C/A310E/I314R/M316W/V422C, E195K/G227F/I314R/M316W/G416A, E195K/G227F/I314R/G416A/L417M, E195K/G227F/I314R/V422C, E195K/G227F/C414V/G416A/V422G, E195K/G227F/C414V/G416C, E195K/G227F/G416S/L417M, E195K/M316W/G416S/L417M/V422A, E207K/T208N/F220Y/A241V/F291S/I314R/M316S, M226A, M226E, M226T, G227F/G416C, G227V, A228G, V258L, 1259V, C271R/A310E/M316W/G416S/V422A, I282L, A310E/I314R/M316S/C414V/G416A/L417M/V422A, I314A/M316W/G416A, I314R/M316W/C414V/G416A/V422A, I314R/M316W/G416C/V422C, M316A, M316E, M316G, M316H, M316K, M316Q, M316R, M316S, M316T, M316W, M316Y, H319V, G320W, T322S, and L417S, wherein the positions are numbered with reference to SEQ ID NO:20.

[0157] In some embodiments, the present invention provides an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:30 selected from 21, 21/23/53/146/162/395, 21/23/53/146/162/395/432, 21/23/53/162/395/432, 21/23/53/395/432, 21/23/395, 21/53/159/395, 21/53/163/395/432, 23/53/146/159/395/431, 23/53/146/159/395/432, 23/53/146/162/395/432, 23/53/146/163, 23/146/163/432, 23/162, 53/146/159/395, 53/146/162/395, 53/146/162/432, 53/146/163/395, 53/146/395/432, 53/162/163/395/432, 53/162/395, 53/162/395/432, 53/163/395/432, 53/395, 57/86, 57/86/121/146/157, 57/86/121/157/226/259/282/332, 57/86/121/332, 57/86/157/259, 57/86/157/282/332, 57/86/332, 57/121/417/422, 57/146/226/259, 57/146/226/259/282, 57/259, 86, 86/157, 86/226, 121, 121/146/157/226/332, 121/208/316/417/422, 121/226/332, 121/316/422, 121/332, 121/416/417/422, 121/417/422, 146, 146/162/395/432, 157, 162/432, 226/282/332, 227, 227/417, 259, 282, 282/332, 316/417/422, 332, 395/432, 416, 416/422, 417, and 422, wherein the positions are numbered with reference to SEQ ID NO:30. In some embodiments, the amino acid difference(s) comprise the substitution(s) 21P, 21P/23S/53C/146H/162D/395D, 21P/23S/53C/146H/162D/395D/432V, 21P/23S/53C/162D/395D/432V, 21P/23S/53C/395D/432V, 21P/23S/395D, 21P/53C/159T/395D, 21P/53C/163K/395D/432V, 23S/53C/146H/159T/395D/431V, 23S/53C/146H/159T/395D/432V, 23S/53C/146H/162D/395D/432V, 23S/53C/146H/163K, 23S/146H/163K/432V, 23S/162D, 53C/146H/159T/395D, 53C/146H/162D/395D, 53C/146H/162D/432V, 53C/146H/163K/395D, 53C/146H/395D/432V, 53C/162D/163K/395D/432V, 53C/162D/395D, 53C/162D/395D/432V, 53C/163K/395D/432V, 53C/395D, 57A/86K, 57A/86K/121A/332T, 57A/86K/121C/146K/157G, 57A/86K/157G/259V, 57A/86K/157G/282L/332T, 57A/146K/226T/259V, 57A/146V/226T/259V/282L, 57A/259V, 57C/86K/121A/157G/226T/259V/282L/332T, 57C/86K/332T, 57C/121A/417M/422V, 86K, 86K/157G, 86K/226T, 121A, 121A/208N/316S/417M/422V, 121A/316F/422C, 121A/416S/417M/422C, 121A/417M/422C, 121A/417M/422V, 121C/146K/157G/226A/332T, 121C/226A/332T, 121C/332T, 146H, 146H/162D/395D/432V, 157G, 162D/432V, 226A/282L/332T, 227A, 227F, 227F/417V, 227V, 259V, 282L, 282L/332T, 316S/417M/422V, 332T, 395D/432V, 416S, 416S/422C, 417F, and 422C, wherein the positions are numbered with reference to SEQ ID NO:30. In some additional embodiments, the amino acid difference(s) comprise the substitution(s) D21P, D21P/P23S/N53C/R146H/G162D/G395D, D21P/P23S/N53C/R146H/G162D/G395D/A432V,

D21P/P23S/N53C/G162D/G395D/A432V, D21P/P23S/N53C/G395D/A432V, D21P/P23S/G395D, D21P/N53C/S159T/G395D, D21P/N53C/F163K/G395D/A432V, P23S/N53C/R146H/S159T/G395D/E431V, P23S/N53C/R146H/S159T/G395D/A432V, P23S/N53C/R146H/G162D/G395D/A432V, P23S/N53C/R146H/F163K, P23S/R146H/F163K/A432V, P23S/G162D, N53C/R146H/S159T/G395D, N53C/R146H/G162D/G395D, N53C/R146H/G162D/A432V, N53C/R146H/F163K/G395D, N53C/R146H/G395D/A432V, N53C/G162D/F163K/G395D/A432V, N53C/G162D/G395D, N53C/G162D/G395D/A432V, N53C/F163K/G395D/A432V, N53C/G395D, L57A/A86K, L57A/A86K/N121A/I332T, L57A/A86K/N121C/R146K/S157G, L57A/A86K/S157G/I259V, L57A/A86K/S157G/I282L/I332T, L57A/R146K/M226T/I259V, L57A/R146V/M226T/I259V/I282L, L57A/I259V, L57C/A86K/N121A/S157G/M226T/I259V/I282L/I332T, L57C/A86K/I332T, L57C/N121A/L417M/A422V, A86K, A86K/S157G, A86K/M226T, N121A, N121A/T208N/W316S/L417M/A422V, N121A/W316F/A422C, N121A/A416S/L417M/A422C, N121A/L417M/A422C, N121A/L417M/A422V, N121C/R146K/S157G/M226A/I332T, N121C/M226A/I332T, N121C/I332T, R146H, R146H/G162D/G395D/A432V, S157G, G162D/A432V, M226A/I282L/I332T, G227A, G227F, G227F/L417V, G227V, I259V, I282L, I282L/I332T, W316S/L417M/A422V, I332T, G395D/A432V, A416S, A416S/A422C, L417F, and A422C, wherein the positions are numbered with reference to SEQ ID NO:30.

[0158] In some embodiments, the present invention provides an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:66 selected from 107, 141, 161, 174, 245, 281, 293, 432, and 453, wherein the positions are numbered with reference to SEQ ID NO:66. In some embodiments, the amino acid difference(s) comprise the substitution(s) 107G, 107V, 141L, 161S, 174Q, 245L, 281S, 293A, 432D, 432E, 432V, 453G, and 453T, wherein the positions are numbered with reference to SEQ ID NO:66. In some additional embodiments, the amino acid difference(s) comprise the substitution(s) D107G, D107V, R141L, T161S, P174Q, I245L, I281S, P293A, A432D, A432E, A432V, A453G, and A453T, wherein the positions are numbered with reference to SEQ ID NO:66.

[0159] In some embodiments, the present invention provides an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:70 selected from 2, 4, 5, 5/38/56/346, 13, 17, 17/31, 19, 31, 57, 74, 79, 86, 121, 155, 161, 161/174/432, 163/167, 164, 190/291, 195, 195/408, 203/207/219, 205/419, 207, 207/235/419, 214/228/251, 222, 223, 226/228/230/231/311, 228, 233, 246, 247/271/293, 259, 282, 284, 309, 327, 346, 359, 362, 375, 404, 408, 419, 420, 422, and 443, wherein the positions are numbered with reference to SEQ ID NO:70. In some embodiments, the amino acid difference(s) comprise the substitution(s) 21, 4A, 4C, 4D, 4R, 5E, 5H/38C/56T/346G, 13E, 17A, 17G/31T, 17H, 17I, 17M, 17S, 17V, 19Y, 31C, 31F, 31L, 57A, 57G, 74M, 79L, 86C, 86E, 86G, 86N, 121C, 155C, 161S, 161S/174Q/432V, 163E/167I, 164V, 190K/291C, 195K, 195Q/408E, 203R/207K/219R, 205M/419S, 207I, 207K/235P/419S, 214T/228G/25IT, 222T, 223G, 226D/228C/230R/231C/311R, 228G, 233V, 246V, 247A/271G/293P, 259V, 282L, 284A, 309R, 327E, 346N, 359G, 362Q, 375R, 404V, 408A, 408E, 408L, 408S, 408V, 419D, 419S, 420T, 422C, 443A, and 443I, wherein the positions are numbered with reference to SEQ ID NO:66. In some additional embodiments, the amino acid difference(s) comprise the substitution(s) N21, P4A, P4C, P4D, P4R, Q5E, Q5H/G38C/A56T/A346G, T13E, Y17A, Y17G/V31T, Y17H, Y17I, Y17M, Y17S, Y17V, F19Y, V31C, V31F, V31L, L57A, L57G, T74M, F79L, A86C, A86E, A86G, A86N, N121C, A155C, T161S, T161S/P174Q/A432V, K163E/S167I, P164V, E190K/F291C, E195K, E195Q/T408E, H203R/E207K/G219R, L205M/Q419S, E207I, E207K/A235P/Q419S, A214T/A228G/I25IT, A222T, E223G, M226D/A228C/G230R/V231C/I311R, A228G, P233V, L246V, R247A/C271G/A293P, I259V, I282L, S284A, T309R, P327E, A346N, R359G, H362Q,

I375R, A404V, T408A, T408E, T408S, T408V, Q419D, Q419S, S420T, A422C, K443A, and K443I, wherein the positions are numbered with reference to SEQ ID NO:70.

[0160] In some embodiments, the present invention provides an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:48 selected from 21/23/53/163/293, 21/23/53/163/293/395, 21/53/146, 21/53/146/163/293/395/432, 21/53/163/293/395, 21/53/293/432, 23/53/146/163, 23/53/146/163/293/432, 23/53/146/293/432, 23/53/163, 23/53/163/293/395, 23/53/163/293/432, 23/53/163/395, 23/53/293, 23/53/293/395/432, 23/53/293/432, 23/146/293, 53/146/163/293/395/432, 53/146/163/395/432, 53/146/163/432, 53/146/293, 53/163, 53/163/293/395, 53/163/293/395/432, 53/163/293/432, 53/163/432, 53/293, 53/293/395, 53/395, 53/432, 146, and 293, wherein the positions are numbered with reference to SEQ ID NO:48. In some embodiments, the amino acid difference(s) comprise the substitution(s)

21P/23S/53C/163K/293A, 21P/23S/53C/163K/293A/395D, 21P/53C/146H, 21P/53C/146H/163K/293A/395D/432V, 21P/53C/163K/293A/395D, 21P/53C/293A/432V, 23S/53C/146H/163K, 23S/53C/146H/163K/293A/432V, 23S/53C/146H/293A/432V, 23S/53C/163K, 23S/53C/163K/293A/395D, 23S/53C/163K/293A/432V, 23S/53C/163K/395D, 23S/53C/293A, 23S/53C/293A/395D/432V, 23S/53C/293A/432V, 23S/146H/293A, 53C/146H/163K/293A/395D/432V, 53C/146H/163K/395D/432V, 53C/146H/163K/432V, 53C/146H/293A, 53C/163K, 53C/163K/293A/395D, 53C/163K/293A/395D/432V, 53C/163K/293A/432V, 53C/163K/432V, 53C/293A, 53C/293A/395D, 53C/395D, 53C/432V, 146H, and 293A, wherein the positions are numbered with reference to SEQ ID NO:48. In some additional embodiments, the amino acid difference(s) comprise the substitution(s) D21P/P23S/N53C/F163K/P293A, D21P/P23S/N53C/F163K/P293A/G395D, D21P/N53C/R146H, D21P/N53C/R146H/F163K/P293A/G395D/A432V, D21P/N53C/F163K/P293A/G395D, D21P/N53C/P293A/A432V, P23S/N53C/R146H/F163K, P23S/N53C/R146H/F163K/P293A/A432V, P23S/N53C/R146H/P293A/A432V, P23S/N53C/F163K, P23S/N53C/F163K/P293A/G395D, P23S/N53C/F163K/P293A/A432V, P23S/N53C/F163K/G395D, P23S/N53C/P293A, P23S/N53C/P293A/G395D/A432V, P23S/N53C/P293A/A432V, P23S/R146H/P293A, N53C/R146H/F163K/P293A/G395D/A432V, N53C/R146H/F163K/G395D/A432V, N53C/R146H/F163K/A432V, N53C/R146H/P293A, N53C/F163K, N53C/F163K/P293A/G395D, N53C/F163K/P293A/G395D/A432V, N53C/F163K/P293A/A432V, N53C/F163K/A432V, N53C/P293A, N53C/P293A/G395D, N53C/G395D, N53C/A432V, R146H, and P293A, wherein the positions are numbered with reference to SEQ ID NO:48.

[0161] In some embodiments, the present invention provides an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:88 selected from 157, 163, 163/164, 284, and 332, wherein the positions are numbered with reference to SEQ ID NO:88. In some embodiments, the amino acid difference(s) comprise the substitution(s) 157A, 157C, 163K, 163K/164G, 163M, 284A, and 332V, wherein the positions are numbered with reference to SEQ ID NO:88. In some additional embodiments, the amino acid difference(s) comprise the substitution(s) S157A, S157C, F163K, F163K/P164G, F163M, S284A, and I332V, wherein the positions are numbered with reference to SEQ ID NO:88.

[0162] In some embodiments, the present invention provides an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:88 selected from 31, 86, 149, 191/208/211/286, 192/294, 195/207/208/286, 208/294, 260, 284, 286, 294, 313, 395/396, and 424, wherein the positions are numbered with reference to SEQ ID NO:88. In some embodiments, the amino acid difference(s) comprise the substitution(s) 31M, 86A, 149S, 191S/208P/211R/286A, 192G/294A, 195D/207R/208N/286G, 208N/294S, 208N/294T, 260T, 284A, 286G, 294A, 294S, 313L, 313V,

395D/396K, and 424A, wherein the positions are numbered with reference to SEQ ID NO:88. In some additional embodiments, the amino acid difference(s) comprise the substitution(s) V31M, K86A, A149S, G191S/T208P/K211R/S286A, E192G/V294A, E195D/E207R/T208N/S286G, T208N/V294S, T208N/V294T, C260T, S284A, S286G, V294A, V294S, A313L, A313V, G395D/N396K, and C424A, wherein the positions are numbered with reference to SEQ ID NO:88. [0163] In some embodiments, the present invention provides an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:110 selected from 14/31/208/260/286/294/424, 14/31/260/286/294/424, 31, 31/208, 31/208/241/286/294, 31/208/260/286/294, 31/208/260/286/294/313/424, 31/208/260/286/294/424, 31/208/260/286/313/424, 31/208/260/294/313/424, 31/208/260/294/424, 31/208/286/294, 31/208/286/294/313, 31/208/286/294/313/424, 31/208/286/294/424, 31/208/286/313/424, 31/208/286/424, 31/208/294, 31/208/294/313/424, 31/208/294/424, 31/260/286, 31/260/286/294/313, 31/260/286/294/313/424, 31/260/286/313/424, 31/260/286/424, 31/260/294, 31/260/294/313, 31/286/294, 31/286/294/313, 31/286/294/424, 31/294/313/424, 31/294/424, 208/260/286/294/313/424, 208/260/286/294/424, 208/260/286/424, 208/260/294, 208/260/294/313/424, 208/286/294/424, 260/286/294/424, 260/286/424, 286/294/313/424, and 286/313/424, wherein the positions are numbered with reference to SEQ ID NO:110. In some embodiments, the amino acid difference(s) comprise the substitution(s) 14N/31M/208N/260T/286G/294A/424A, 14N/31M/260T/286G/294A/424A, 31M, 31M/208N, 31M/208N/241V/286G/294A, 31M/208N/260T/286G/294A, 31M/208N/260T/286G/294A/313E/424A, 31M/208N/260T/286G/294A/313L/424A, 31M/208N/260T/286G/294A/424A, 31M/208N/260T/286G/313L/424A, 31M/208N/260T/294A/313L/424A, 31M/208N/260T/294A/424A, 31M/208N/286G/294A, 31M/208N/286G/294A/313L, 31M/208N/286G/294A/313L/424A, 31M/208N/286G/294A/424A, 31M/208N/286G/313L/424A, 31M/208N/286G/313V/424A, 31M/208N/286G/424A, 31M/208N/294A, 31M/208N/294A/313L/424A, 31M/208N/294A/424A, 31M/260T/286G, 31M/260T/286G/294A/313L, 31M/260T/286G/294A/313L/424A, 31M/260T/286G/313L/424A, 31M/260T/286G/424A, 31M/260T/294A, 31M/260T/294A/313L, 31M/286G/294A, 31M/286G/294A/313L, 31M/286G/294A/424A, 31M/294A/313L/424A, 31M/294A/424A, 208N/260T/286G/294A/313L/424A, 208N/260T/286G/294A/424A, 208N/260T/286G/424A, 208N/260T/294A, 208N/260T/294A/313L/424A, 208N/286G/294A/424A, 260T/286G/294A/424A, 260T/286G/424A, 286G/294A/313L/424A, and 286G/313L/424A, wherein the positions are numbered with reference to SEQ ID NO:110. In some additional embodiments, the amino acid difference(s) comprise the substitution(s) Y14N/V31M/T208N/C260T/S286G/V294A/C424A, Y14N/V31M/C260T/S286G/V294A/C424A, V31M, V31M/T208N, V31M/T208N/A241V/S286G/V294A, V31M/T208N/C260T/S286G/V294A, V31M/T208N/C260T/S286G/V294A/A313E/C424A, V31M/T208N/C260T/S286G/V294A/A313L/C424A, V31M/T208N/C260T/S286G/V294A/C424A, V31M/T208N/C260T/S286G/A313L/C424A, V31M/T208N/C260T/V294A/A313L/C424A, V31M/T208N/C260T/V294A/C424A, V31M/T208N/S286G/V294A, V31M/T208N/S286G/V294A/A313L, V31M/T208N/S286G/V294A/A313L/C424A, V31M/T208N/S286G/V294A/C424A, V31M/T208N/S286G/A313L/C424A, V31M/T208N/S286G/A313V/C424A, V31M/T208N/S286G/C424A, V31M/T208N/V294A, V31M/T208N/V294A/A313L/C424A, V31M/T208N/V294A/C424A, V31M/C260T/S286G, V31M/C260T/S286G/V294A/A313L, V31M/C260T/S286G/V294A/A313L/C424A, V31M/C260T/S286G/A313L/C424A, V31M/C260T/S286G/C424A, V31M/C260T/V294A, V31M/C260T/V294A/A313L, V31M/S286G/V294A, V31M/S286G/V294A/A313L, V31M/S286G/V294A/C424A, V31M/V294A/A313L/C424A, V31M/V294A/C424A, T208N/C260T/S286G/V294A/A313L/C424A, T208N/C260T/S286G/V294A/C424A,

T208N/C260T/C424A, T208N/C260T/V294A, T208N/C260T/V294A/A313L/C424A, T208N/S286G/V294A/C424A, C260T/S286G/V294A/C424A, C260T/S286G/C424A, S286G/V294A/A313L/C424A, and S286G/A313L/C424A, wherein the positions are numbered with reference to SEQ ID NO:110.

[0164] In some embodiments, the present invention provides an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:112 selected from 16/272, 17, 19/86, 19/86/321, 20, 21, 22, 22/255, 22/255/362/421, 22/294, 22/362, 24/255/362, 31, 41, 42/187/294, 66, 66/190, 66/394, 84, 114, 153, 157, 161, 163, 164, 164/397, 221, 222, 272, 294, 320, 320/435, and 394, wherein the positions are numbered with reference to SEQ ID NO:112. In some embodiments, the amino acid difference(s) comprise the substitution(s) 16F/272E, 17M, 19H/86N, 19Y/86N/321Y, 20A, 21E, 22L, 22L/255A, 22L/255A/362R/421I, 22L/294M, 22L/362R, 24T/255A/362R, 31T, 41L, 42F/187F/294M, 66P, 66P/190P, 66P/394P, 84N, 84S, 114A, 153S, 157G, 157P, 161R, 163P, 164G, 164R/397P, 221I, 222T, 272E, 294M, 320A, 320A/435N, and 394P, wherein the positions are numbered with reference to SEQ ID NO:112. In some additional embodiments, the amino acid difference(s) comprise the substitution(s) L16F/V272E, Y17M, F19H/K86N, F19Y/K86N/F321Y, T20A, D21E, M22L, M22L/S255A, M22L/S255A/H362R/V421I, M22L/V294M, M22L/H362R, S24T/S255A/H362R, M31T, I41L, V42F/Y187F/V294M, K66P, K66P/E190P, K66P/G394P, A84N, A84S, T114A, A153S, S157G, S157P, T161R, F163P, P164G, P164R/L397P, F221I, A222T, V272E, V294M, G320A, G320A/D435N, and G394P, wherein the positions are numbered with reference to SEQ ID NO:112.

[0165] In some embodiments, the present invention provides an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO: 120 selected from 17, 17/21, 17/21/22, 17/22, 17/22/272, 17/22/272/320, 17/31/320/397, 17/84/164, 17/157/164, 17/157/320/397, 17/221/320, 17/272/320, 17/320, 17/397, 19/22/42/320/394, 19/42, 20/320, 25/320, 66/187/320, 66/272, 66/320, 66/394, 164/320/321, 165, 255, 272, 272/320, and 320, wherein the positions are numbered with reference to SEQ ID NO:120. In some embodiments, the amino acid difference(s) comprise the substitution(s) 17M, 17M/21E, 17M/21E/22L, 17M/22L, 17M/22L/272E, 17M/22L/272E/320A, 17M/31T/320A/397P, 17M/84S/164R, 17M/157G/164R, 17M/157G/320A/397P, 17M/221I/320A, 17M/272E/320A, 17M/320A, 17M/397P, 19Y/22L/42F/320A/394P, 19Y/42F, 20A/320A, 25V/320A, 66P/187F/320A, 66P/272E, 66P/320A, 66P/394P, 164R/320A/321Y, 165W, 255A, 272E, 272E/320A, and 320A, wherein the positions are numbered with reference to SEQ ID NO:120. In some additional embodiments, the amino acid difference(s) comprise the substitution(s) Y17M, Y17M/D21E, Y17M/D21E/M22L, Y17M/M22L, Y17M/M22L/V272E, Y17M/M22L/V272E/G320A, Y17M/M31T/G320A/L397P, Y17M/A84S/P164R, Y17M/S157G/P164R, Y17M/S157G/G320A/L397P, Y17M/F221I/G320A, Y17M/V272E/G320A, Y17M/G320A, Y17M/L397P, H19Y/M22L/V42F/G320A/G394P, H19Y/V42F, T20A/G320A, L25V/G320A, K66P/Y187F/G320A, K66P/V272E, K66P/G320A, K66P/G394P, P164R/G320A/F321Y, Y165W, S255A, V272E, V272E/G320A, and G320A, wherein the positions are numbered with reference to SEQ ID NO:120.

[0166] In some embodiments, the present invention provides an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:138 selected from 24, 24/25, 24/110/167, 25, 25/147/167, 27/190/207/389/396, 27/190/207/396, 41/42/66/165, 42, 42/66/165/394, 42/165/272/394, 66/164, 66/164/272, 66/165, 66/394, 163, 164, 164/165/394, 165, 165/272, 165/394, 166, 167, 190, 190/207, 190/207/210, 190/207/210/385/389, 190/207/385, 190/210, 190/385/396, 190/389, 207, 207/210/243/250/385/389, 207/214/396, 207/243, 207/389, 210, 210/385, 214, 214/396, 272/394, 283, 385, 385/389, 389, 389/396, and 394, wherein the positions are numbered with reference to SEQ ID NO:138. In some embodiments, the amino acid difference(s) comprise the substitution(s)

24T, 24T/25V, 24T/110L/167E, 25V, 25V/147V/167E, 27D/190P/207D/389T/396S, 27D/190P/207D/396S, 41L/42F/66P/165W, 42F, 42F/66P/165W/394P, 42F/165W/272E/394P, 66P/164R, 66P/164R/272E, 66P/165W, 66P/394P, 163L, 164N, 164R/165W/394P, 164S, 165W, 165W/272E, 165W/394P, 166A, 166L, 166T, 167E, 190P, 190P/207D, 190P/207D/210A, 190P/207D/210A/385A/389T, 190P/207D/385A, 190P/210A, 190P/385A/396S, 190P/389T, 207D, 207D/210A/243Q/250G/385A/389T, 207D/214P/396S, 207D/243Q, 207D/389T, 210A, 210A/385A, 214P, 214P/396S, 272E/394P, 283C, 385A, 385A/389T, 389T, 389T/396S, and 394P, wherein the positions are numbered with reference to SEQ ID NO:138.

[0167] In some additional embodiments, the amino acid difference(s) comprise the substitution(s) S24T, S24T/L25V, S24T/R110L/S167E, L25V, L25V/W147V/S167E, Q27D/E190P/E207D/S389T/N396S, Q27D/E190P/E207D/N396S, I41L/V42F/K66P/Y165W, V42F, V42F/K66P/Y165W/G394P, V42F/Y165W/V272E/G394P, K66P/P164R, K66P/P164R/V272E, K66P/Y165W, K66P/G394P, F163L, P164N, P164R/Y165W/G394P, P164S, Y165W, Y165W/V272E, Y165W/G394P, N166A, N166L, N166T, S167E, E190P, E190P/E207D, E190P/E207D/Q210A, E190P/E207D/Q210A/K385A/S389T, E190P/E207D/K385A, E190P/Q210A, E190P/K385A/N396S, E190P/S389T, E207D, E207D/Q210A/L243Q/D250G/K385A/S389T, E207D/A214P/N396S, E207D/L243Q, E207D/S389T, Q210A, Q210A/K385A, A214P, A214P/N396S, V272E/G394P, S283C, K385A, K385A/S389T, S389T, S389T/N396S, and G394P, wherein the positions are numbered with reference to SEQ ID NO:138.

[0168] In some embodiments, the present invention provides an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:156 selected from 12/26/30/291, 12/26/73/77/291, 13/45/107/236/277/346/351/410/429, 22/25/110/291/304, 24/147/164/166/167/243/389, 24/166, 24/166/190, 24/166/207, 26/30, 26/30/168/291, 30/77/291, 30/291, 70/77/108/177/196/247/291/309/361/366, 70/77/168/291, 70/177/291/361/366, 70/177/291/372, 70/186, 70/247/291/309/372, 70/291, 70/291/304/361, 74, 74/211, 77/110/291/308, 77/291/361/366, 107/277/346/410/429, 107/346, 110/235/391, 147/164/166/167, 147/164/166/243/272/283/385/389, 147/164/166/272, 163, 166, 166/167/272/389, 166/190, 166/190/207, 166/207, 177/195/196/291/304/309/366, 177/196/247/304/309/361/366, 177/291/309/361, 196/291/304/361, 196/291/309/361, 198/235/391, 207, 247/304/309/366, 253/391, 277, 291, 291/304/309, and 291/309, wherein the positions are numbered with reference to SEQ ID NO:156. In some embodiments, the amino acid difference(s) comprise the substitution(s) 12D/26K/30P/291Y, 12D/26R/73C/77R/291Y, 13A/45N/107E/236E/277V/346L/351A/410H/429M, 22L/25V/110K/291Y/304A, 24T/147V/164S/166A/167E/243Q/389T, 24T/166A, 24T/166A/190P, 24T/166A/207D, 24T/166L, 26K/30P/168E/291Y, 26R/30P, 30P/77R/291Y, 30P/291Y, 70E/77D/108R/177I/196E/247K/291Y/309A/361A/366H, 70E/77R/168E/291Y, 70E/177I/291Y/361A/366H, 70E/177I/291Y/372W, 70E/186C, 70E/247K/291Y/309A/372W, 70E/291Y, 70E/291Y/304A/361A, 74A, 74A/211R, 77D/110K/291Y/308Q, 77D/291Y/361A/366H, 107E/277V/346L/410H/429M, 107E/346L, 110K/235S/391T, 147V/164R/166A/167E, 147V/164R/166A/243Q/272E/283C/385A/389T, 147V/164R/166A/272E, 163Y, 166A, 166A/167E/272E/389T, 166A/190P, 166A/190P/207D, 166A/207D, 166G, 166S, 166S/207D, 166V, 177I/195A/196E/291Y/304A/309A/366H, 177I/196E/247K/304A/309A/361A/366H, 177I/291Y/309A/361A, 196E/291Y/304A/361A, 196E/291Y/309A/361A, 198T/235S/391T, 207D, 247K/304A/309A/366H, 253L/391T, 277V, 291Y, 291Y/304A/309A, and 291Y/309A, wherein the positions are numbered with reference to SEQ ID NO:156. In some additional embodiments, the amino acid difference(s) comprise the substitution(s) E12D/H26K/T30P/F291Y, E12D/H26R/K73C/E77R/F291Y, T13A/H45N/D107E/K236E/T277V/A346L/R351A/L410H/L429M,

M25L/V110K/F291Y/S304A, S24T/W147V/P164S/N166A/S167E/L243Q/S389T, S24T/N166A, S24T/N166A/E190P, S24T/N166A/E207D, S24T/N166L, H26K/T30P/V168E/F291Y, H26R/T30P, T30P/E77R/F291Y, T30P/F291Y, D70E/E77D/S108R/L177I/Q196E/R247K/F291Y/T309A/K361A/R366H, D70E/E77R/V168E/F291Y, D70E/L177I/F291Y/K361A/R366H, D70E/L177I/F291Y/Y372W, D70E/R186C, D70E/R247K/F291Y/T309A/Y372W, D70E/F291Y, D70E/F291Y/S304A/K361A, T74A, T74A/K211R, E77D/R110K/F291Y/E308Q, E77D/F291Y/K361A/R366H, D107E/T277V/A346L/L410H/L429M, D107E/A346L, R110K/A235S/A391T, W147V/P164R/N166A/S167E, W147V/P164R/N166A/L243Q/V272E/S283C/K385A/S389T, W147V/P164R/N166A/V272E, F163Y, N166A, N166A/S167E/V272E/S389T, N166A/E190P, N166A/E190P/E207D, N166A/E207D, N166G, N166S, N166S/E207D, N166V, L177I/E195A/Q196E/F291Y/S304A/T309A/R366H, L177I/Q196E/R247K/S304A/T309A/K361A/R366H, L177I/F291Y/T309A/K361A, Q196E/F291Y/S304A/K361A, Q196E/F291Y/T309A/K361A, V198T/A235S/A391T, E207D, R247K/S304A/T309A/R366H, V253L/A391T, T277V, F291Y, F291Y/S304A/T309A, and F291Y/T309A, wherein the positions are numbered with reference to SEQ ID NO:156.

[0169] In some embodiments, the engineered polypeptides having transaminase activity are capable of converting compound (2) to compound (1) with increased tolerance for the presence of the substrate relative to the substrate tolerance of a reference polypeptide (e.g., SEQ ID NO: 2, 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156), under suitable reaction conditions.

Accordingly, in some embodiments the engineered polypeptides are capable of converting the substrate of compound (2) to compound (1) in the presence of a substrate loading concentration of at least about 1 g/L, 5 g/L, 10 g/L, 20 g/L, about 30 g/L, about 40 g/L, about 50 g/L, about 70 g/L, about 75 g/L, about 100 g/L, with a percent conversion of at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 91%, at least about 92%, at least about 94%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, in a reaction time of about 72h, about 48h, about 36h, about 24 h, or even shorter length of time, under suitable reaction conditions.

[0170] Some suitable reaction conditions under which the above-described improved properties of the engineered polypeptides can be determined with respect concentrations or amounts of polypeptide, substrate, amine donor, cofactor, buffer, co-solvent, pH, and/or conditions including temperature and reaction time are provided herein. In some embodiments, the suitable reaction conditions comprise the HTP, SFP, or DSP assay conditions described below and in the Examples.

[0171] As will be apparent to the skilled artisan, the foregoing residue positions and the specific amino acid residues for each residue position can be used individually or in various combinations to synthesize transaminase polypeptides having desired improved properties, including, among others, enzyme activity, substrate/product preference, stereoselectivity, substrate/product tolerance, and stability under various conditions, such as increased temperature, solvent, and/or pH.

[0172] In some embodiments, the present invention also provides engineered transaminase polypeptides that comprise a fragment of any of the engineered transaminase polypeptides described herein that retains the functional transaminase activity and/or improved property of that engineered transaminase polypeptide. Accordingly, in some embodiments, the present invention provides a polypeptide fragment having transaminase activity (e.g., capable of converting compound (2) to compound (1) under suitable reaction conditions), wherein the fragment comprises at least about 80%, 90%, 95%, 98%, or 99% of a full-length amino acid sequence of an engineered polypeptide of the present invention, such as an exemplary engineered polypeptide of having the even-numbered sequence identifiers of SEQ ID NO:6-172.

[0173] In some embodiments, the engineered transaminase polypeptide of the invention comprises an amino acid sequence comprising a deletion as compared to any one of the engineered transaminase polypeptide sequences described herein, such as the exemplary engineered

polypeptide sequences having the even-numbered sequence identifiers of SEQ ID NO: 6-172. Thus, for each and every embodiment of the engineered transaminase polypeptides of the invention, the amino acid sequence can comprise deletions of one or more amino acids, 2 or more amino acids, 3 or more amino acids, 4 or more amino acids, 5 or more amino acids, 6 or more amino acids, 8 or more amino acids, 10 or more amino acids, 15 or more amino acids, or 20 or more amino acids, up to 10% of the total number of amino acids, up to 10% of the total number of amino acids, up to 20% of the total number of amino acids, or up to 30% of the total number of amino acids of the transaminase polypeptides, where the associated functional activity and/or improved properties of the engineered transaminase described herein is maintained. In some embodiments, the deletions can comprise, 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-15, 1-20, 1-21, 1-22, 1-23, 1-24, 1-25, 1-30, 1-35, 1-40, 1-45, 1-50, 1-55, or 1-60 amino acid residues. In some embodiments, the number of deletions can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, or 60 amino acid residues. In some embodiments, the deletions can comprise deletions of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 20, 21, 22, 23, 24, 25 or 30 amino acid residues.

[0174] In some embodiments, the present invention provides an engineered transaminase polypeptide having an amino acid sequence comprising an insertion as compared to any one of the engineered transaminase polypeptide sequences described herein, such as the exemplary engineered polypeptide sequences having the even-numbered sequence identifiers of SEQ ID NO:6-172. Thus, for each and every embodiment of the transaminase polypeptides of the invention, the insertions can comprise one or more amino acids, 2 or more amino acids, 3 or more amino acids, 4 or more amino acids, 5 or more amino acids, 6 or more amino acids, 8 or more amino acids, 10 or more amino acids, 15 or more amino acids, or 20 or more amino acids, where the associated functional activity and/or improved properties of the engineered transaminase described herein is maintained. The insertions can be to amino or carboxy terminus, or internal portions of the transaminase polypeptide.

[0175] In some embodiments, the polypeptides of the present invention are in the form of fusion polypeptides in which the engineered polypeptides are fused to other polypeptides, such as, by way of example and not limitation, antibody tags (e.g., myc epitope), purification sequences (e.g., His tags for binding to metals), and cell localization signals (e.g., secretion signals). Thus, the engineered polypeptides described herein can be used with or without fusions to other polypeptides.

[0176] The engineered transaminase polypeptides described herein are not restricted to the genetically encoded amino acids. Thus, in addition to the genetically encoded amino acids, the polypeptides described herein may be comprised, either in whole or in part, of naturally-occurring and/or synthetic non-encoded amino acids. Certain commonly encountered non-encoded amino acids of which the polypeptides described herein may be comprised include, but are not limited to: the D-stereoisomers of the genetically-encoded amino acids; 2,3-diaminopropionic acid (Dpr); α -aminoisobutyric acid (Aib); s-aminohexanoic acid (Aha); 6-aminovaleric acid (Ava); N-methylglycine or sarcosine (MeGly or Sar); ornithine (Orn); citrulline (Cit); t-butylalanine (Bua); t-butylglycine (Bug); N-methylisoleucine (MeIle); phenylglycine (Phg); cyclohexylalanine (Cha); norleucine (Nle); naphthylalanine (Nal); 2-chlorophenylalanine (Ocf); 3-chlorophenylalanine (Mcf); 4-chlorophenylalanine (Pcf); 2-fluorophenylalanine (Off); 3-fluorophenylalanine (Mff); 4-fluorophenylalanine (Pff); 2-bromophenylalanine (Obf); 3-bromophenylalanine (Mbf); 4-bromophenylalanine (Pbf); 2-methylphenylalanine (Omf); 3-methylphenylalanine (Mmf); 4-methylphenylalanine (Pmf); 2-nitrophenylalanine (Onf); 3-nitrophenylalanine (Mnf); 4-nitrophenylalanine (Pnf); 2-cyanophenylalanine (Ocf); 3-cyanophenylalanine (Mcf); 4-cyanophenylalanine (Pcf); 2-trifluoromethylphenylalanine (Otf); 3-trifluoromethylphenylalanine (Mtf); 4-trifluoromethylphenylalanine (Ptf); 4-aminophenylalanine (Paf); 4-iodophenylalanine (Pif); 4-aminomethylphenylalanine (Pamf); 2,4-dichlorophenylalanine (Opef); 3,4-

dichlorophenylalanine (Mpcf); 2,4-difluorophenylalanine (Opff); 3,4-difluorophenylalanine (Mpff); pyrid-2-ylalanine (2pAla); pyrid-3-ylalanine (3pAla); pyrid-4-ylalanine (4pAla); naphth-1-ylalanine (1nAla); naphth-2-ylalanine (2nAla); thiazolylalanine (taAla); benzothienylalanine (bAla); thienylalanine (tAla); furylalanine (fAla); homophenylalanine (hPhe); homotyrosine (hTyr); homotryptophan (hTrp); pentafluorophenylalanine (5ff); styrylalanine (sAla); authrylalanine (aAla); 3,3-diphenylalanine (Dfa); 3-amino-5-phenypentanoic acid (Afp); penicillamine (Pen); 1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid (Tic); β -2-thienylalanine (Thi); methionine sulfoxide (Mso); N(w)-nitroarginine (nArg); homolysine (hLys); phosphonomethylphenylalanine (pmPhe); phosphoserine (pSer); phosphothreonine (pThr); homoaspartic acid (hAsp); homoglutamic acid (hGlu); 1-aminocyclopent-(2 or 3)-ene-4 carboxylic acid; pipecolic acid (PA), azetidine-3-carboxylic acid (ACA); 1-aminocyclopentane-3-carboxylic acid; allylglycine (aOly); propargylglycine (pgGly); homoalanine (hAla); norvaline (nVal); homoleucine (hLeu), homovaline (hVal); homoisoleucine (hIle); homoarginine (hArg); N-acetyl lysine (AcLys); 2,4-diaminobutyric acid (Dbu); 2,3-diaminobutyric acid (Dab); N-methylvaline (MeVal); homocysteine (hCys); homoserine (hSer); hydroxyproline (Hyp) and homoproline (hPro). Additional non-encoded amino acids of which the polypeptides described herein may be comprised will be apparent to those of skill in the art. These amino acids may be in either the L- or D-configuration.

[0177] Those of skill in the art will recognize that amino acids or residues bearing side chain protecting groups may also comprise the polypeptides described herein. Non-limiting examples of such protected amino acids, which in this case belong to the aromatic category, include (protecting groups listed in parentheses), but are not limited to: Arg(tos), Cys(methylbenzyl), Cys(nitropyridinesulfonyl), Glu(6-benzylester), Gln(xanthyl), Asn(N-6-xanthyl), His(bom), His(benzyl), His(tos), Lys(fmoc), Lys(tos), Ser(O-benzyl), Thr (O-benzyl) and Tyr(O-benzyl).

[0178] Non-encoding amino acids that are conformationally constrained of which the polypeptides described herein may be composed include, but are not limited to, N-methyl amino acids (L-configuration); 1-aminocyclopent-(2 or 3)-ene-4-carboxylic acid; pipecolic acid; azetidine-3-carboxylic acid; homoproline (hPro); and 1-aminocyclopentane-3-carboxylic acid.

[0179] In some embodiments, the engineered polypeptides can be provided on a solid support, such as a membrane, resin, solid carrier, or other solid phase material. A solid support can be composed of organic polymers such as polystyrene, polyethylene, polypropylene, polyfluoroethylene, polyethyleneoxy, and polyacrylamide, as well as co-polymers and grafts thereof. A solid support can also be inorganic, such as glass, silica, controlled pore glass (CPG), reverse phase silica or metal, such as gold or platinum. The configuration of a solid support can be in the form of beads, spheres, particles, granules, a gel, a membrane or a surface. Surfaces can be planar, substantially planar, or non-planar.

[0180] Solid supports can be porous or non-porous, and can have swelling or non-swelling characteristics. A solid support can be configured in the form of a well, depression, or other container, vessel, feature, or location.

[0181] In some embodiments, the engineered polypeptides having transaminase activity are bound or immobilized on the solid support such that they retain their improved activity, enantioselectivity, stereoselectivity, and/or other improved properties relative to a reference polypeptide (e.g., SEQ ID NO: 2, 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156). In such embodiments, the immobilized polypeptides can facilitate the biocatalytic conversion of the substrate compound to the desired product, and after the reaction is complete are easily retained (e.g., by retaining beads on which polypeptide is immobilized) and then reused or recycled in subsequent reactions. Such immobilized enzyme processes allow for further efficiency and cost reduction. Accordingly, it is further contemplated that any of the methods of using the engineered transaminase polypeptides of the present invention can be carried out using the same transaminase polypeptides bound or immobilized on a solid support.

[0182] The engineered transaminase polypeptide can be bound non-covalently or covalently.

Various methods for conjugation and immobilization of enzymes to solid supports (e.g., resins, membranes, beads, glass, etc.) are well known in the art. In particular, PCT publication WO2012/177527 A1 discloses immobilized engineered transaminase polypeptides capable of converting compound (2) to compound (1), and methods of preparing the immobilized polypeptides, in which the polypeptide is physically attached to a resin by either hydrophobic interactions or covalent bonds, and is stable in a solvent system that comprises at least up to 100% organic solvent. Other methods for conjugation and immobilization of enzymes to solid supports (e.g., resins, membranes, beads, glass, etc.) are well known in the art (See e.g., Yi et al., *Proc. Biochem.*, 42: 895-898 [2007]; Martin et al., *Appl. Microbiol.*

[0183] *Biotechnol.*, 76: 843-851 [2007]; Koszelewski et al., *J. Mol. Cat. B: Enz.*, 63: 39-44 [2010]; Truppo et al., *Org. Proc. Res. Develop.*, published online: [dx.doi.org/10.1021/op200157c](https://doi.org/10.1021/op200157c); and Mateo et al., *Biotechnol. Prog.*, 18:629-34 [2002], etc.).

[0184] Solid supports useful for immobilizing the engineered transaminase polypeptides of the present invention include but are not limited to beads or resins comprising polymethacrylate with epoxide functional groups, polymethacrylate with amino epoxide functional groups, styrene/DVB copolymer or polymethacrylate with octadecyl functional groups. Exemplary solid supports useful for immobilizing the engineered transaminases of the present invention include, but are not limited to, chitosan beads, Eupergit C, and SEPABEADs (Mitsubishi), including the following different types of SEPABEAD: EC-EP, EC-HFA/S, EXA252, EXE119 and EXE120.

[0185] In some embodiments, the engineered transaminase polypeptides can be provided in the form of an array in which the polypeptides are arranged in positionally distinct locations. In some embodiments, the positionally distinct locations are wells in a solid support such as a 96-well plate. A plurality of supports can be configured on an array at various locations, addressable for robotic delivery of reagents, or by detection methods and/or instruments. Such arrays can be used to test a variety of substrate compounds for conversion by the polypeptides.

[0186] In some embodiments, the engineered polypeptides described herein can be provided in the form of kits. The polypeptides in the kits may be present individually or as a plurality of polypeptides.

[0187] The kits can further include reagents for carrying out enzymatic reactions, substrates for assessing the activity of polypeptides, as well as reagents for detecting the products. The kits can also include reagent dispensers and instructions for use of the kits. In some embodiments, the kits of the present invention include arrays comprising a plurality of different engineered transaminase polypeptides at different addressable position, wherein the different polypeptides are different variants of a reference sequence each having at least one different improved enzyme property. Such arrays comprising a plurality of engineered polypeptides and methods of their use are known (See e.g., WO2009/008908A2).

Polynucleotides, Control Sequences, Expression Vectors, and Host Cells Useful for Preparing Engineered Transaminase Polypeptides

[0188] In another aspect, the present invention provides polynucleotides encoding the engineered polypeptides having transaminase activity described herein. The polynucleotides may be operatively linked to one or more heterologous regulatory sequences that control gene expression to create a recombinant polynucleotide capable of expressing the polypeptide. Expression constructs containing a heterologous polynucleotide encoding the engineered transaminase can be introduced into appropriate host cells to express the corresponding engineered transaminase polypeptide.

[0189] In some embodiments, the isolated polynucleotide encoding an improved transaminase polypeptide is manipulated in a variety of ways to provide for improved activity and/or expression of the polypeptide. Manipulation of the isolated polynucleotide prior to its insertion into a vector may be desirable or necessary depending on the expression vector. The techniques for modifying polynucleotides and nucleic acid sequences utilizing recombinant DNA methods are well known in

the art.

[0190] Those of ordinary skill in the art understand that due to the degeneracy of the genetic code, a multitude of nucleotide sequences encoding variant transaminase acylase polypeptides of the present invention exist. For example, the codons AGA, AGG, CGA, CGC, CGG, and CGU all encode the amino acid arginine. Thus, at every position in the nucleic acids of the invention where an arginine is specified by a codon, the codon can be altered to any of the corresponding codons described above without altering the encoded polypeptide. It is understood that “U” in an RNA sequence corresponds to “T” in a DNA sequence. The invention contemplates and provides each and every possible variation of nucleic acid sequence encoding a polypeptide of the invention that could be made by selecting combinations based on possible codon choices.

[0191] As indicated above, DNA sequence encoding a transaminase may also be designed for high codon usage bias codons (codons that are used at higher frequency in the protein coding regions than other codons that code for the same amino acid). The preferred codons may be determined in relation to codon usage in a single gene, a set of genes of common function or origin, highly expressed genes, the codon frequency in the aggregate protein coding regions of the whole organism, codon frequency in the aggregate protein coding regions of related organisms, or combinations thereof. A codon whose frequency increases with the level of gene expression is typically an optimal codon for expression. In particular, a DNA sequence can be optimized for expression in a particular host organism. A variety of methods are well-known in the art for determining the codon frequency (e.g., codon usage, relative synonymous codon usage) and codon preference in specific organisms, including multivariate analysis (e.g., using cluster analysis or correspondence analysis,) and the effective number of codons used in a gene. The data source for obtaining codon usage may rely on any available nucleotide sequence capable of coding for a protein. These data sets include nucleic acid sequences actually known to encode expressed proteins (e.g., complete protein coding sequences-CDS), expressed sequence tags (ESTs), or predicted coding regions of genomic sequences, as is well-known in the art.

[0192] Polynucleotides encoding variant transaminases can be prepared using any suitable methods known in the art. Typically, oligonucleotides are individually synthesized, then joined (e.g., by enzymatic or chemical ligation methods, or polymerase-mediated methods) to form essentially any desired continuous sequence. In some embodiments, polynucleotides of the present invention are prepared by chemical synthesis using, any suitable methods known in the art, including but not limited to automated synthetic methods. For example, in the phosphoramidite method, oligonucleotides are synthesized (e.g., in an automatic DNA synthesizer), purified, annealed, ligated and cloned in appropriate vectors.

[0193] In some embodiments, double stranded DNA fragments are then obtained either by synthesizing the complementary strand and annealing the strands together under appropriate conditions, or by adding the complementary strand using DNA polymerase with an appropriate primer sequence. There are numerous general and standard texts that provide methods useful in the present invention are well known to those skilled in the art.

[0194] For example, mutagenesis and directed evolution methods can be readily applied to polynucleotides to generate variant libraries that can be expressed, screened, and assayed.

Mutagenesis and directed evolution methods are well known in the art (See e.g., U.S. Pat. Nos. 5,605,793, 5,811,238, 5,830,721, 5,834,252, 5,837,458, 5,928,905, 6,096,548, 6,117,679, 6,132,970, 6,165,793, 6,180,406, 6,251,674, 6,265,201, 6,277,638, 6,287,861, 6,287,862, 6,291,242, 6,297,053, 6,303,344, 6,309,883, 6,319,713, 6,319,714, 6,323,030, 6,326,204, 6,335,160, 6,335,198, 6,344,356, 6,352,859, 6,355,484, 6,358,740, 6,358,742, 6,365,377, 6,365,408, 6,368,861, 6,372,497, 6,337,186, 6,376,246, 6,379,964, 6,387,702, 6,391,552, 6,391,640, 6,395,547, 6,406,855, 6,406,910, 6,413,745, 6,413,774, 6,420,175, 6,423,542, 6,426,224, 6,436,675, 6,444,468, 6,455,253, 6,479,652, 6,482,647, 6,483,011, 6,484,105, 6,489,146, 6,500,617, 6,500,639, 6,506,602, 6,506,603, 6,518,065, 6,519,065, 6,521,453,

6,528,311, 6,537,746, 6,573,098, 6,576,467, 6,579,678, 6,586,182, 6,602,986, 6,605,430, 6,613,514, 6,653,072, 6,686,515, 6,703,240, 6,716,631, 6,825,001, 6,902,922, 6,917,882, 6,946,296, 6,961,664, 6,995,017, 7,024,312, 7,058,515, 7,105,297, 7,148,054, 7,220,566, 7,288,375, 7,384,387, 7,421,347, 7,430,477, 7,462,469, 7,534,564, 7,620,500, 7,620,502, 7,629,170, 7,702,464, 7,747,391, 7,747,393, 7,751,986, 7,776,598, 7,783,428, 7,795,030, 7,853,410, 7,868,138, 7,783,428, 7,873,477, 7,873,499, 7,904,249, 7,957,912, 7,981,614, 8,014,961, 8,029,988, 8,048,674, 8,058,001, 8,076,138, 8,108,150, 8,170,806, 8,224,580, 8,377,681, 8,383,346, 8,457,903, 8,504,498, 8,589,085, 8,762,066, 8,768,871, 9,593,326, and all related non-US counterparts; Ling et al., *Anal. Biochem.*, 254(2):157-78 [1997]; Dale et al., *Meth. Mol. Biol.*, 57:369-74 [1996]; Smith, *Ann. Rev. Genet.*, 19:423-462 [1985]; Botstein et al., *Science*, 229:1193-1201 [1985]; Carter, *Biochem. J.*, 237:1-7 [1986]; Kramer et al., *Cell*, 38:879-887 [1984]; Wells et al., *Gene*, 34:315-323 [1985]; Minshull et al., *Curr. Op. Chem. Biol.*, 3:284-290 [1999]; Christians et al., *Nat. Biotechnol.*, 17:259-264 [1999]; Cramer et al., *Nature*, 391:288-291 [1998]; Cramer et al., *Nat. Biotechnol.*, 15:436-438 [1997]; Zhang et al., *Proc. Nat. Acad. Sci.* [0195] U.S.A., 94:4504-4509 [1997]; Cramer et al., *Nat. Biotechnol.*, 14:315-319 [1996]; Stemmer, *Nature*, 370:389-391 [1994]; Stemmer, *Proc. Nat. Acad. Sci. USA*, 91:10747-10751 [1994]; WO 95/22625; WO 97/0078; WO 97/35966; WO 98/27230; WO 00/42651; WO 01/75767; and WO 2009/152336, all of which are incorporated herein by reference).

[0196] In some embodiments, the polynucleotide encodes a transaminase polypeptide comprising an amino acid sequence that is at least about 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more identical to a reference sequence selected from the even-numbered sequence identifiers of SEQ ID NO: 2, 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156, where the polypeptide has transaminase activity and one or more of the improved properties as described herein, for example the ability to convert compound (2) to product compound (1) with increased activity compared to a reference sequence (e.g., the polypeptide of SEQ ID NO: 2, 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156). In some embodiments, the reference sequence is selected from SEQ ID NO: 2, 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156. In some embodiments, the reference sequence is SEQ ID NO:4. In some embodiments, the reference sequence is SEQ ID NO: 18. In some embodiments, the reference sequence is SEQ ID NO:20. In some embodiments, the reference sequence is SEQ ID NO:30. In some embodiments, the reference sequence is SEQ ID NO:48. In some embodiments, the reference sequence is SEQ ID NO:66. In some embodiments, the reference sequence is SEQ ID NO:70. In some embodiments, the reference sequence is SEQ ID NO:88. In some embodiments, the reference sequence is SEQ ID NO: 110. In some embodiments, the reference sequence is SEQ ID NO: 112. In some embodiments, the reference sequence is SEQ ID NO:120. In some embodiments, the reference sequence is SEQ ID NO:138. In some embodiments, the reference sequence is SEQ ID NO:156.

[0197] In some embodiments, the polynucleotide encodes an engineered transaminase polypeptide comprising an amino acid sequence that has the percent identity described above and (a) has one or more amino acid residue differences as compared to SEQ ID NO: 2, 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156. In some embodiments, the present invention provides an engineered polypeptide having transaminase activity comprising an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to reference sequence of SEQ ID NO: 2, 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156 and (a) at least one amino acid residue difference selected from those substitutions provided herein (See e.g., Tables 6.1, 7.1, 8.1, 9.1, 10.1, 11.1, 12.1, 13.1, 13.2, 14.1, 15.1, 16.1, 17.1, and 18.1).

[0198] In some embodiments, the polynucleotide encoding the engineered transaminase polypeptide comprises a sequence selected from the odd-numbered sequence identifiers of SEQ ID NO:5-171. In some embodiments, the polynucleotide sequences are selected from SEQ ID NO:3,

17, 19, 29, 47, 65, 69, 87, 109, 111, 119, 137, and 155. In some embodiments, the present invention provides engineered polynucleotides encoding polypeptides having transaminase activity, wherein the engineered polypeptides have at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to at least one reference sequence selected from SEQ ID NO: 3, 17, 19, 29, 47, 65, 69, 87, 109, 111, 119, 137, and 155.

[0199] In some embodiments, the present invention provides a polynucleotide that hybridizes under defined conditions, such as moderately stringent or highly stringent conditions, to a polynucleotide sequence (or complement thereof) encoding an engineered transaminase of the present invention. In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a polynucleotide selected from the sequences having the odd-numbered sequence identifiers of SEQ ID NO:5-171, or a complement thereof, and encodes a polypeptide having transaminase activity with one or more of the improved properties described herein.

[0200] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:4 selected from 17, 17/259/347, 20, 45/48/78/81/92, 57, 57/319, 85, 86, 157, 158, 161, 203/416, 314, 316, 319, 375/416, 404, 415, 416, 417, 420, 421, 422, and 424, wherein the positions are numbered with reference to SEQ ID NO:4. In some embodiments, the amino acid differences comprise the substitution(s) 17V/259T/347K, 17W, 20R, 45R/48L/78C/81V/92H, 57I, 57T/319E, 85G, 85S, 85V, 86A, 86G, 157C, 157G, 157P, 157T, 158S, 161G, 203H/416G, 203H/416S, 314R, 316K, 316L, 316M, 316S, 319G, 319L, 319V, 375L/416F, 404T, 404V, 415V, 416C, 416G, 416S, 416Y, 417G, 417M, 417P, 420G, 421I, 422A, 422G, 422R, 424A, and 424R, wherein the positions are numbered with reference to SEQ ID NO:4. In some additional embodiments, the amino acid differences comprise the substitution(s) Y17V/I259T/E347K, Y17W, T20R, H45R/R48L/R78C/G81V/Q92H, L57I, L57T/H319E, F85G, F85S, F85V, S86A, S86G, S157C, S157G, S157P, S157T, A158S, T161G, R203H/P416G, R203H/P416S, I314R, E316K, E316L, E316M, E316S, H319G, H319L, H319V, I375L/P416F, A404T, A404V, R415V, P416C, P416G, P416S, P416Y, L417G, L417M, L417P, S420G, V421I, V422A, V422G, V422R, C424A, and C424R, wherein the positions are numbered with reference to SEQ ID NO:4.

[0201] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:18 selected from 5/17/86/157/259/314/316/422, 17/86/157/203/259/314/316/422, 17/86/157/314/316/347/422, 17/86/157/316/422, 85/86, 86/157/314/316/347/416/422, 86/157/416, 86/203/316/416, 86/203/316/422, 86/316/416, 139/221/310/316, 157/316, 195/316, 207/316, 227/422, 280/381/416/445, 314, 314/316, 316, 416, 417, and 422, wherein the positions are numbered with reference to SEQ ID NO:18. In some embodiments, the amino acid differences comprise the substitution(s) 5H/17V/86G/157G/259T/314R/316S/422R, 17V/86A/157G/316M/422A, 17V/86G/157G/203H/259T/314R/316L/422R, 17V/86G/157G/316M/422R, 17V/86G/157T/316M/422R, 17W/86A/157T/314R/316M/347K/422R, 85F/86G, 86A/157G/314R/316S/347K/416C/422A, 86A/203H/316M/416G, 86A/203H/316M/422G, 86A/316S/416C, 86G/157T/416G, 139H/221I/310E/316G, 157T/316M, 195K/316S, 207K/316T, 227F/422G, 280E/381V/416G/445R, 314R, 314R/316M, 316C, 316F, 316G, 316M, 316S, 316W, 316Y, 416A, 416G, 416S, 417M, and 422C, wherein the positions are numbered with reference to SEQ ID NO:18. In some additional embodiments, the amino acid differences comprise the substitution(s) Q5H/Y17V/S86G/S157G/I259T/I314R/E316S/V422R, Y17V/S86A/S157G/E316M/V422A, Y17V/S86G/S157G/R203H/I259T/I314R/E316L/V422R, Y17V/S86G/S157G/E316M/V422R, Y17V/S86G/S157T/E316M/V422R, Y17W/S86A/S157T/I314R/E316M/E347K/V422R, V85F/S86G, S86A/S157G/I314R/E316S/E347K/P416C/V422A, S86A/R203H/E316M/P416G,

S86A/R203H/E316M/V422G, S86A/E316S/P416C, S86G/S157T/P416G, Q139H/F2211/A310E/E316G, S157T/E316M, E195K/E316S, E207K/E316T, G227F/V422G, A280E/L381V/P416G/L445R, 1314R, 1314R/E316M, E316C, E316F, E316G, E316M, E316S, E316W, E316Y, P416A, P416G, P416S, L417M, and V422C, wherein the positions are numbered with reference to SEQ ID NO:18.

[0202] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:20 selected from 17/139/314/416/417/422, 17/195/316/414/416/422, 17/314/316/414/416, 17/314/316/414/416/417/422, 20, 31, 54, 57, 83, 86, 86/194,121, 139/195/227/314/316/416, 139/202/207/227/314/316/416/422, 139/207/227/314/414/416/422, 139/227/310/414/416/417, 139/316/414/416/417/422, 146/157, 164/207/213/227/310/316/416, 166, 178/314/316/416, 190/195/207/213/244/314/414/416/417/422, 194/213/310/314/316/422, 195/227/314/316/416, 195/227/314/416/417, 195/227/314/422, 195/227/414/416, 195/227/414/416/422, 195/227/416/417, 195/316/416/417/422, 207/208/220/241/291/314/316, 226, 227, 227/416, 228, 258, 259, 271/310/316/416/422, 282, 310/314/316/414/416/417/422, 314/316/414/416/422, 314/316/416, 314/316/416/422, 316, 319, 320, 322, and 417, wherein the positions are numbered with reference to SEQ ID NO:20. In some embodiments, the amino acid differences comprise the substitution(s) 17V/139H/314R/416A/417M/422A, 17V/195K/316W/414V/416S/422C, 17V/314R/316S/414V/416C, 17V/314R/316W/414V/416S/417M/422A, 20K, 20L, 20V, 31A, 31C, 31D, 31E, 31G, 31Q, 54C, 57A, 57C, 83G, 86G, 86G/194K, 86K, 121A, 121C, 121G, 121H, 121L, 121S, 121V, 139H/195K/227F/314R/316W/416C, 139H/202G/207K/227F/314R/316F/416S/422C, 139H/207K/227F/314R/414V/416C/422G, 139H/227F/310E/414V/416C/417M, 139H/316S/414V/416S/417M/422A, 146K/157G, 146V/157G, 164Q/207K/213S/227F/310E/316F/416A, 166T, 178L/314A/316W/416S, 190K/195K/207K/213S/244A/314R/414V/416S/417M/422G, 194D/213C/310E/314R/316W/422C, 195K/227F/314R/316W/416A, 195K/227F/314R/416A/417M, 195K/227F/314R/422C, 195K/227F/414V/416A/422G, 195K/227F/414V/416C, 195K/227F/416S/417M, 195K/316W/416S/417M/422A, 207K/208N/220Y/241V/291S/314R/316S, 226A, 226E, 226T, 227F/416C, 227V, 228G, 258L, 259V, 271R/310E/316W/416S/422A, 282L, 310E/314R/316S/414V/416A/417M/422A, 314A/316W/416A, 314R/316W/414V/416A/422A, 314R/316W/416C/422C, 316A, 316E, 316G, 316H, 316K, 316Q, 316R, 316S, 316T, 316W, 316Y, 319V, 320W, 322S, and 417S, wherein the positions are numbered with reference to SEQ ID NO:20.

[0203] In some additional embodiments, the amino acid differences comprise the substitution(s) Y17V/Q139H/1314R/G416A/L417M/V422A, Y17V/E195K/M316W/C414V/G416S/V422C, Y17V/I314R/M316S/C414V/G416C, Y17V/I314R/M316W/C414V/G416S/L417M/V422A, T20K, T20L, T20V, V31A, V31C, V31D, V31E, V31G, V31Q, S54C, L57A, L57C, H83G, A86G, A86G/E194K, A86K, N121A, N121C, N121G, N121H, N121L, N121S, N121V, Q139H/E195K/G227F/I314R/M316W/G416C, Q139H/A202G/E207K/G227F/1314R/M316F/G416S/V422C, Q139H/E207K/G227F/I314R/C414V/G416C/V422G, Q139H/G227F/A310E/C414V/G416C/L417M, Q139H/M316S/C414V/G416S/L417M/V422A, R146K/S157G, R146V/S157G, P164Q/E207K/G213S/G227F/A310E/M316F/G416A, N166T, H178L/I314A/M316W/G416S, E190K/E195K/E207K/G213S/P244A/I314R/C414V/G416S/L417M/V422G, E194D/G213C/A310E/I314R/M316W/V422C, E195K/G227F/I314R/M316W/G416A, E195K/G227F/I314R/G416A/L417M, E195K/G227F/I314R/V422C, E195K/G227F/C414V/G416A/V422G, E195K/G227F/C414V/G416C, E195K/G227F/G416S/L417M, E195K/M316W/G416S/L417M/V422A,

E207K/T208N/F209Y/F291S/1314R/M316S, M226A, M226E, M226T, G227F/G416C, G227V, A228G, V258L, 1259V, C271R/A310E/M316W/G416S/V422A, 1282L, A310E/1314R/M316S/C414V/G416A/L417M/V422A, 1314A/M316W/G416A, 1314R/M316W/C414V/G416A/V422A, 1314R/M316W/G416C/V422C, M316A, M316E, M316G, M316H, M316K, M316Q, M316R, M316S, M316T, M316W, M316Y, H319V, G320W, T322S, and L417S, wherein the positions are numbered with reference to SEQ ID NO:20.

[0204] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:30 selected from 21, 21/23/53/146/162/395, 21/23/53/146/162/395/432, 21/23/53/162/395/432, 21/23/53/395/432, 21/23/395, 21/53/159/395, 21/53/163/395/432, 23/53/146/159/395/431, 23/53/146/159/395/432, 23/53/146/162/395/432, 23/53/146/163, 23/146/163/432, 23/162, 53/146/159/395, 53/146/162/395, 53/146/162/432, 53/146/163/395, 53/146/395/432, 53/162/163/395/432, 53/162/395, 53/162/395/432, 53/163/395/432, 53/395, 57/86, 57/86/121/146/157, 57/86/121/157/226/259/282/332, 57/86/121/332, 57/86/157/259, 57/86/157/282/332, 57/86/332, 57/121/417/422, 57/146/226/259, 57/146/226/259/282, 57/259, 86, 86/157, 86/226, 121, 121/146/157/226/332, 121/208/316/417/422, 121/226/332, 121/316/422, 121/332, 121/416/417/422, 121/417/422, 146, 146/162/395/432, 157, 162/432, 226/282/332, 227, 227/417, 259, 282, 282/332, 316/417/422, 332, 395/432, 416, 416/422, 417, and 422, wherein the positions are numbered with reference to SEQ ID NO:30. In some embodiments, the amino acid differences comprise the substitution(s) 21P, 21P/23S/53C/146H/162D/395D, 21P/23S/53C/146H/162D/395D/432V, 21P/23S/53C/162D/395D/432V, 21P/23S/53C/395D/432V, 21P/23S/395D, 21P/53C/159T/395D, 21P/53C/163K/395D/432V, 23S/53C/146H/159T/395D/431V, 23S/53C/146H/159T/395D/432V, 23S/53C/146H/162D/395D/432V, 23S/53C/146H/163K, 23S/146H/163K/432V, 23S/162D, 53C/146H/159T/395D, 53C/146H/162D/395D, 53C/146H/162D/432V, 53C/146H/163K/395D, 53C/146H/395D/432V, 53C/162D/163K/395D/432V, 53C/162D/395D, 53C/162D/395D/432V, 53C/163K/395D/432V, 53C/395D, 57A/86K, 57A/86K/121A/332T, 57A/86K/121C/146K/157G, 57A/86K/157G/259V, 57A/86K/157G/282L/332T, 57A/146K/226T/259V, 57A/146V/226T/259V/282L, 57A/259V, 57C/86K/121A/157G/226T/259V/282L/332T, 57C/86K/332T, 57C/121A/417M/422V, 86K, 86K/157G, 86K/226T, 121A, 121A/208N/316S/417M/422V, 121A/316F/422C, 121A/416S/417M/422C, 121A/417M/422C, 121A/417M/422V, 121C/146K/157G/226A/332T, 121C/226A/332T, 121C/332T, 146H, 146H/162D/395D/432V, 157G, 162D/432V, 226A/282L/332T, 227A, 227F, 227F/417V, 227V, 259V, 282L, 282L/332T, 316S/417M/422V, 332T, 395D/432V, 416S, 416S/422C, 417F, and 422C, wherein the positions are numbered with reference to SEQ ID NO:30. In some additional embodiments, the amino acid differences comprise the substitution(s) D21P, D21P/P23S/N53C/R146H/G162D/G395D, D21P/P23S/N53C/R146H/G162D/G395D/A432V, D21P/P23S/N53C/G162D/G395D/A432V, D21P/P23S/N53C/G395D/A432V, D21P/P23S/G395D, D21P/N53C/S159T/G395D, D21P/N53C/F163K/G395D/A432V, P23S/N53C/R146H/S159T/G395D/E431V, P23S/N53C/R146H/S159T/G395D/A432V, P23S/N53C/R146H/G162D/G395D/A432V, P23S/N53C/R146H/F163K, P23S/R146H/F163K/A432V, P23S/G162D, N53C/R146H/S159T/G395D, N53C/R146H/G162D/G395D, N53C/R146H/G162D/A432V, N53C/R146H/F163K/G395D, N53C/R146H/G395D/A432V, N53C/G162D/F163K/G395D/A432V, N53C/G162D/G395D, N53C/G162D/G395D/A432V, N53C/F163K/G395D/A432V, N53C/G395D, L57A/A86K, L57A/A86K/N121A/I332T, L57A/A86K/N121C/R146K/S157G, L57A/A86K/S157G/I259V, L57A/A86K/S157G/I282L/I332T, L57A/R146K/M226T/I259V, L57A/R146V/M226T/I259V/I282L, L57A/I259V, L57C/A86K/N121A/S157G/M226T/I259V/I282L/I332T, L57C/A86K/I332T,

L57C/N121A/L417M/A422V, A86K, A86K/S157G, A86K/M226T, N121A, N121A/T208N/W316S/L417M/A422V, N121A/W316F/A422C, N121A/A416S/L417M/A422C, N121A/L417M/A422C, N121A/L417M/A422V, N121C/R146K/S157G/M226A/I332T, N121C/M226A/I332T, N121C/I332T, R146H, R146H/G162D/G395D/A432V, S157G, G162D/A432V, M226A/I282L/I332T, G227A, G227F, G227F/L417V, G227V, I259V, I282L, I282L/I332T, W316S/L417M/A422V, I332T, G395D/A432V, A416S, A416S/A422C, L417F, and A422C, wherein the positions are numbered with reference to SEQ ID NO:30.

[0205] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:66 selected from 107, 141, 161, 174, 245, 281, 293, 432, and 453, wherein the positions are numbered with reference to SEQ ID NO:66. In some embodiments, the amino acid differences comprise the substitution(s) 107G, 107V, 141L, 161S, 174Q, 245L, 281S, 293A, 432D, 432E, 432V, 453G, and 453T, wherein the positions are numbered with reference to SEQ ID NO:66. In some additional embodiments, the amino acid differences comprise the substitution(s) D107G, D107V, R141L, T161S, P174Q, 1245L, 1281S, P293A, A432D, A432E, A432V, A453G, and A453T, wherein the positions are numbered with reference to SEQ ID NO:66.

[0206] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:70 selected from 2, 4, 5, 5/38/56/346, 13, 17, 17/31, 19, 31, 57, 74, 79, 86, 121, 155, 161, 161/174/432, 163/167, 164, 190/291, 195, 195/408, 203/207/219, 205/419, 207, 207/235/419, 214/228/251, 222, 223, 226/228/230/231/311, 228, 233, 246, 247/271/293, 259, 282, 284, 309, 327, 346, 359, 362, 375, 404, 408, 419, 420, 422, and 443, wherein the positions are numbered with reference to SEQ ID NO:70. In some embodiments, the amino acid differences comprise the substitution(s) 21, 4A, 4C, 4D, 4R, 5E, 5H/38C/56T/346G, 13E, 17A, 17G/31T, 17H, 17I, 17M, 17S, 17V, 19Y, 31C, 31F, 31L, 57A, 57G, 74M, 79L, 86C, 86E, 86G, 86N, 121C, 155C, 161S, 161S/174Q/432V, 163E/1671, 164V, 190K/291C, 195K, 195Q/408E, 203R/207K/219R, 205M/419S, 207I, 207K/235P/419S, 214T/228G/251T, 222T, 223G, 226D/228C/230R/231C/311R, 228G, 233V, 246V, 247A/271G/293P, 259V, 282L, 284A, 309R, 327E, 346N, 359G, 362Q, 375R, 404V, 408A, 408E, 408L, 408S, 408V, 419D, 419S, 420T, 422C, 443A, and 443I, wherein the positions are numbered with reference to SEQ ID NO:66. In some additional embodiments, the amino acid differences comprise the substitution(s) N2I, P4A, P4C, P4D, P4R, Q5E, Q5H/G38C/A56T/A346G, T13E, Y17A, Y17G/V31T, Y17H, Y17I, Y17M, Y17S, Y17V, F19Y, V31C, V31F, V31L, L57A, L57G, T74M, F79L, A86C, A86E, A86G, A86N, N121C, A155C, T161S, T161S/P174Q/A432V, K163E/S1671, P164V, E190K/F291C, E195K, E195Q/T408E, H203R/E207K/G219R, L205M/Q419S, E2071, E207K/A235P/Q419S, A214T/A228G/I251T, A222T, E223G, M226D/A228C/G230R/V231C/I311R, A228G, P233V, L246V, R247A/C271G/A293P, I259V, I282L, S284A, T309R, P327E, A346N, R359G, H362Q, I375R, A404V, T408A, T408E, T408L, T408S, T408V, Q419D, Q419S, S420T, A422C, K443A, and K443I, wherein the positions are numbered with reference to SEQ ID NO:70.

[0207] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:48 selected from 21/23/53/163/293, 21/23/53/163/293/395, 21/53/146, 21/53/146/163/293/395/432, 21/53/163/293/395, 21/53/293/432, 23/53/146/163, 23/53/146/163/293/432, 23/53/146/293/432, 23/53/163, 23/53/163/293/395, 23/53/163/293/432, 23/53/163/395, 23/53/293, 23/53/293/395/432, 23/53/293/432, 23/146/293, 53/146/163/293/395/432, 53/146/163/395/432, 53/146/163/432, 53/146/293, 53/163, 53/163/293/395, 53/163/293/395/432, 53/163/293/432, 53/163/432, 53/293, 53/293/395, 53/395, 53/432, 146, and 293, wherein the positions are numbered with reference to

SEQ ID NO:48. In some embodiments, the amino acid differences comprise the substitution(s) 21P/23S/53C/163K/293A, 21P/23S/53C/163K/293A/395D, 21P/53C/146H, 21P/53C/146H/163K/293A/395D/432V, 21P/53C/163K/293A/395D, 21P/53C/293A/432V, 23S/53C/146H/163K, 23S/53C/146H/163K/293A/432V, 23S/53C/146H/293A/432V, 23S/53C/163K, 23S/53C/163K/293A/395D, 23S/53C/163K/293A/432V, 23S/53C/163K/395D, 23S/53C/293A, 23S/53C/293A/395D/432V, 23S/53C/293A/432V, 23S/146H/293A, 53C/146H/163K/293A/395D/432V, 53C/146H/163K/395D/432V, 53C/146H/163K/432V, 53C/146H/293A, 53C/163K, 53C/163K/293A/395D, 53C/163K/293A/395D/432V, 53C/163K/293A/432V, 53C/163K/432V, 53C/293A, 53C/293A/395D, 53C/395D, 53C/432V, 146H, and 293A, wherein the positions are numbered with reference to SEQ ID NO:48. In some additional embodiments, the amino acid differences comprise the substitution(s) D21P/P23S/N53C/F163K/P293A, D21P/P23S/N53C/F163K/P293A/G395D, D21P/N53C/R146H, D21P/N53C/R146H/F163K/P293A/G395D/A432V, D21P/N53C/F163K/P293A/G395D, D21P/N53C/P293A/A432V, P23S/N53C/R146H/F163K, P23S/N53C/R146H/F163K/P293A/A432V, P23S/N53C/R146H/P293A/A432V, P23S/N53C/F163K, P23S/N53C/F163K/P293A/G395D, P23S/N53C/F163K/P293A/A432V, P23S/N53C/F163K/G395D, P23S/N53C/P293A, P23S/N53C/P293A/G395D/A432V, P23S/N53C/P293A/A432V, P23S/R146H/P293A, N53C/R146H/F163K/P293A/G395D/A432V, N53C/R146H/F163K/G395D/A432V, N53C/R146H/F163K/A432V, N53C/R146H/P293A, N53C/F163K, N53C/F163K/P293A/G395D, N53C/F163K/P293A/G395D/A432V, N53C/F163K/P293A/A432V, N53C/F163K/A432V, N53C/P293A, N53C/P293A/G395D, N53C/G395D, N53C/A432V, R146H, and P293A, wherein the positions are numbered with reference to SEQ ID NO:48.

[0208] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:88 selected from 157, 163, 163/164, 284, and 332, wherein the positions are numbered with reference to SEQ ID NO:88. In some embodiments, the amino acid differences comprise the substitution(s) 157A, 157C, 163K, 163K/164G, 163M, 284A, and 332V, wherein the positions are numbered with reference to SEQ ID NO:88. In some additional embodiments, the amino acid differences comprise the substitution(s) S157A, S157C, F163K, F163K/P164G, F163M, S284A, and I332V, wherein the positions are numbered with reference to SEQ ID NO:88.

[0209] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:88 selected from 31, 86, 149, 191/208/211/286, 192/294, 195/207/208/286, 208/294, 260, 284, 286, 294, 313, 395/396, and 424, wherein the positions are numbered with reference to SEQ ID NO:88. In some embodiments, the amino acid differences comprise the substitution(s) 31M, 86A, 149S, 191S/208P/211R/286A, 192G/294A, 195D/207R/208N/286G, 208N/294S, 208N/294T, 260T, 284A, 286G, 294A, 294S, 313L, 313V, 395D/396K, and 424A, wherein the positions are numbered with reference to SEQ ID NO:88.

[0210] In some additional embodiments, the amino acid differences comprise the substitution(s) V31M, K86A, A149S, G191S/T208P/K211R/S286A, E192G/V294A, E195D/E207R/T208N/S286G, T208N/V294S, T208N/V294T, C260T, S284A, S286G, V294A, V294S, A313L, A313V, G395D/N396K, and C424A, wherein the positions are numbered with reference to SEQ ID NO:88.

[0211] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:110 selected from 14/31/208/260/286/294/424, 14/31/260/286/294/424, 31, 31/208, 31/208/241/286/294,

31/208/260/286/294, 31/208/260/286/294/313/424, 31/208/260/286/294/424, 31/208/260/286/313/424, 31/208/260/294/313/424, 31/208/260/294/424, 31/208/286/294, 31/208/286/294/313, 31/208/286/294/313/424, 31/208/286/294/424, 31/208/286/313/424, 31/208/286/424, 31/208/294, 31/208/294/313/424, 31/208/294/424, 31/260/286, 31/260/286/294/313, 31/260/286/294/313/424, 31/260/286/313/424, 31/260/286/424, 31/260/294, 31/260/294/313, 31/286/294, 31/286/294/313, 31/286/294/424, 31/294/313/424, 31/294/424, 208/260/286/294/313/424, 208/260/286/294/424, 208/260/286/424, 208/260/294, 208/260/294/313/424, 208/286/294/424, 260/286/294/424, 260/286/424, 286/294/313/424, and 286/313/424, wherein the positions are numbered with reference to SEQ ID NO:110. In some embodiments, the amino acid differences comprise the substitution(s) 14N/31M/208N/260T/286G/294A/424A, 14N/31M/260T/286G/294A/424A, 31M, 31M/208N, 31M/208N/241V/286G/294A, 31M/208N/260T/286G/294A, 31M/208N/260T/286G/294A/313E/424A, 31M/208N/260T/286G/294A/313L/424A, 31M/208N/260T/286G/294A/424A, 31M/208N/260T/286G/313L/424A, 31M/208N/260T/294A/313L/424A, 31M/208N/260T/294A/424A, 31M/208N/286G/294A, 31M/208N/286G/294A/313L, 31M/208N/286G/294A/313L/424A, 31M/208N/286G/294A/424A, 31M/208N/286G/313L/424A, 31M/208N/286G/313V/424A, 31M/208N/286G/424A, 31M/208N/294A, 31M/208N/294A/313L/424A, 31M/208N/294A/424A, 31M/260T/286G, 31M/260T/286G/294A/313L, 31M/260T/286G/294A/313L/424A, 31M/260T/286G/313L/424A, 31M/260T/286G/424A, 31M/260T/294A, 31M/260T/294A/313L, 31M/286G/294A, 31M/286G/294A/313L, 31M/286G/294A/424A, 31M/294A/313L/424A, 31M/294A/424A, 208N/260T/286G/294A/313L/424A, 208N/260T/286G/294A/424A, 208N/260T/286G/424A, 208N/260T/294A, 208N/260T/294A/313L/424A, 208N/286G/294A/424A, 260T/286G/294A/424A, 260T/286G/424A, 286G/294A/313L/424A, and 286G/313L/424A, wherein the positions are numbered with reference to SEQ ID NO:110. In some additional embodiments, the amino acid differences comprise the substitution(s) Y14N/V31M/T208N/C260T/S286G/V294A/C424A, Y14N/V31M/C260T/S286G/V294A/C424A, V31M, V31M/T208N, V31M/T208N/A241V/S286G/V294A, V31M/T208N/C260T/S286G/V294A, V31M/T208N/C260T/S286G/V294A/A313E/C424A, V31M/T208N/C260T/S286G/V294A/A313L/C424A, V31M/T208N/C260T/S286G/V294A/C424A, V31M/T208N/C260T/S286G/A313L/C424A, V31M/T208N/C260T/V294A/A313L/C424A, V31M/T208N/C260T/V294A/C424A, V31M/T208N/S286G/V294A, V31M/T208N/S286G/V294A/A313L, V31M/T208N/S286G/V294A/A313L/C424A, V31M/T208N/S286G/V294A/C424A, V31M/T208N/S286G/A313L/C424A, V31M/T208N/S286G/A313V/C424A, V31M/T208N/S286G/C424A, V31M/T208N/V294A, V31M/T208N/V294A/A313L/C424A, V31M/T208N/V294A/C424A, V31M/C260T/S286G, V31M/C260T/S286G/V294A/A313L, V31M/C260T/S286G/V294A/A313L/C424A, V31M/C260T/S286G/A313L/C424A, V31M/C260T/S286G/C424A, V31M/C260T/V294A, V31M/C260T/V294A/A313L, V31M/S286G/V294A, V31M/S286G/V294A/A313L, V31M/S286G/V294A/C424A, V31M/V294A/A313L/C424A, V31M/V294A/C424A, T208N/C260T/S286G/V294A/A313L/C424A, T208N/C260T/S286G/V294A/C424A, T208N/C260T/S286G/C424A, T208N/C260T/V294A, T208N/C260T/V294A/A313L/C424A, T208N/S286G/V294A/C424A, C260T/S286G/V294A/C424A, C260T/S286G/C424A, S286G/V294A/A313L/C424A, and S286G/A313L/C424A, wherein the positions are numbered with reference to SEQ ID NO:110.

[0212] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:112 selected from 16/272, 17, 19/86, 19/86/321, 20, 21, 22, 22/255, 22/255/362/421, 22/294, 22/362, 24/255/362, 31,

41, 42/187/294, 66, 66/190, 66/394, 84, 114, 153, 157, 161, 163, 164, 164/397, 221, 222, 272, 294, 320, 320/435, and 394, wherein the positions are numbered with reference to SEQ ID NO:112. In some embodiments, the amino acid differences comprise the substitution(s) 16F/272E, 17M, 19H/86N, 19Y/86N/321Y, 20A, 21E, 22L, 22L/255A, 22L/255A/362R/4211, 22L/294M, 22L/362R, 24T/255A/362R, 31T, 41L, 42F/187F/294M, 66P, 66P/190P, 66P/394P, 84N, 84S, 114A, 153S, 157G, 157P, 161R, 163P, 164G, 164R/397P, 2211, 222T, 272E, 294M, 320A, 320A/435N, and 394P, wherein the positions are numbered with reference to SEQ ID NO:112. In some additional embodiments, the amino acid differences comprise the substitution(s) L16F/V272E, Y17M, F19H/K86N, F19Y/K86N/F321Y, T20A, D21E, M22L, M22L/S255A, M22L/S255A/H362R/V4211, M22L/V294M, M22L/H362R, S24T/S255A/H362R, M31T, 141L, V42F/Y187F/V294M, K66P, K66P/E190P, K66P/G394P, A84N, A84S, T114A, A153S, S157G, S157P, T161R, F163P, P164G, P164R/L397P, F2211, A222T, V272E, V294M, G320A, G320A/D435N, and G394P, wherein the positions are numbered with reference to SEQ ID NO:112.

[0213] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:120 selected from 17, 17/21, 17/21/22, 17/22, 17/22/272, 17/22/272/320, 17/31/320/397, 17/84/164, 17/157/164, 17/157/320/397, 17/221/320, 17/272/320, 17/320, 17/397, 19/22/42/320/394, 19/42, 20/320, 25/320, 66/187/320, 66/272, 66/320, 66/394, 164/320/321, 165, 255, 272, 272/320, and 320, wherein the positions are numbered with reference to SEQ ID NO:120. In some embodiments, the amino acid differences comprise the substitution(s) 17M, 17M/21E, 17M/21E/22L, 17M/22L, 17M/22L/272E, 17M/22L/272E/320A, 17M/31T/320A/397P, 17M/84S/164R, 17M/157G/164R, 17M/157G/320A/397P, 17M/221I/320A, 17M/272E/320A, 17M/320A, 17M/397P, 19Y/22L/42F/320A/394P, 19Y/42F, 20A/320A, 25V/320A, 66P/187F/320A, 66P/272E, 66P/320A, 66P/394P, 164R/320A/321Y, 165W, 255A, 272E, 272E/320A, and 320A, wherein the positions are numbered with reference to SEQ ID NO:120. In some additional embodiments, the amino acid differences comprise the substitution(s) Y17M, Y17M/D21E, Y17M/D21E/M22L, Y17M/M22L, Y17M/M22L/V272E, Y17M/M22L/V272E/G320A, Y17M/M31T/G320A/L397P, Y17M/A84S/P164R, Y17M/S157G/P164R, Y17M/S157G/G320A/L397P, Y17M/F221I/G320A, Y17M/V272E/G320A, Y17M/G320A, Y17M/L397P, H19Y/M22L/V42F/G320A/G394P, H19Y/V42F, T20A/G320A, L25V/G320A, K66P/Y187F/G320A, K66P/V272E, K66P/G320A, K66P/G394P, P164R/G320A/F321Y, Y165W, S255A, V272E, V272E/G320A, and G320A, wherein the positions are numbered with reference to SEQ ID NO:120.

[0214] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:138 selected from 24, 24/25, 24/110/167, 25, 25/147/167, 27/190/207/389/396, 27/190/207/396, 41/42/66/165, 42, 42/66/165/394, 42/165/272/394, 66/164, 66/164/272, 66/165, 66/394, 163, 164, 164/165/394, 165, 165/272, 165/394, 166, 167, 190, 190/207, 190/207/210, 190/207/210/385/389, 190/207/385, 190/210, 190/385/396, 190/389, 207, 207/210/243/250/385/389, 207/214/396, 207/243, 207/389, 210, 210/385, 214, 214/396, 272/394, 283, 385, 385/389, 389, 389/396, and 394, wherein the positions are numbered with reference to SEQ ID NO:138. In some embodiments, the amino acid differences comprise the substitution(s) 24T, 24T/25V, 24T/110L/167E, 25V, 25V/147V/167E, 27D/190P/207D/389T/396S, 27D/190P/207D/396S, 41L/42F/66P/165W, 42F, 42F/66P/165W/394P, 42F/165W/272E/394P, 66P/164R, 66P/164R/272E, 66P/165W, 66P/394P, 163L, 164N, 164R/165W/394P, 164S, 165W, 165W/272E, 165W/394P, 166A, 166L, 166T, 167E, 190P, 190P/207D, 190P/207D/210A, 190P/207D/210A/385A/389T, 190P/207D/385A, 190P/210A, 190P/385A/396S, 190P/389T, 207D, 207D/210A/243Q/250G/385A/389T, 207D/214P/396S, 207D/243Q, 207D/389T, 210A, 210A/385A, 214P, 214P/396S, 272E/394P, 283C, 385A,

385A/389T, 389T/396S, and 394P, wherein the positions are numbered with reference to SEQ ID NO:138. In some additional embodiments, the amino acid differences comprise the substitution(s) S24T, S24T/L25V, S24T/R110L/S167E, L25V, L25V/W147V/S167E, Q27D/E190P/E207D/S389T/N396S, Q27D/E190P/E207D/N396S, I41L/V42F/K66P/Y165W, V42F, V42F/K66P/Y165W/G394P, V42F/Y165W/V272E/G394P, K66P/P164R, K66P/P164R/V272E, K66P/Y165W, K66P/G394P, F163L, P164N, P164R/Y165W/G394P, P164S, Y165W, Y165W/V272E, Y165W/G394P, N166A, N166L, N166T, S167E, E190P, E190P/E207D, E190P/E207D/Q210A, E190P/E207D/Q210A/K385A/S389T, E190P/E207D/K385A, E190P/Q210A, E190P/K385A/N396S, E190P/S389T, E207D, E207D/Q210A/L243Q/D250G/K385A/S389T, E207D/A214P/N396S, E207D/L243Q, E207D/S389T, Q210A, Q210A/K385A, A214P, A214P/N396S, V272E/G394P, S283C, K385A, K385A/S389T, S389T, S389T/N396S, and G394P, wherein the positions are numbered with reference to SEQ ID NO:138.

[0215] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:156 selected from 12/26/30/291, 12/26/73/77/291, 13/45/107/236/277/346/351/410/429, 22/25/110/291/304, 24/147/164/166/167/243/389, 24/166, 24/166/190, 24/166/207, 26/30, 26/30/168/291, 30/77/291, 30/291, 70/77/108/177/196/247/291/309/361/366, 70/77/168/291, 70/177/291/361/366, 70/177/291/372, 70/186, 70/247/291/309/372, 70/291, 70/291/304/361, 74, 74/211, 77/110/291/308, 77/291/361/366, 107/277/346/410/429, 107/346, 110/235/391, 147/164/166/167, 147/164/166/243/272/283/385/389, 147/164/166/272, 163, 166, 166/167/272/389, 166/190, 166/190/207, 166/207, 177/195/196/291/304/309/366, 177/196/247/304/309/361/366, 177/291/309/361, 196/291/304/361, 196/291/309/361, 198/235/391, 207, 247/304/309/366, 253/391, 277, 291, 291/304/309, and 291/309, wherein the positions are numbered with reference to SEQ ID NO:156. In some embodiments, the amino acid differences comprise the substitution(s) 12D/26K/30P/291Y, 12D/26R/73C/77R/291Y, 13A/45N/107E/236E/277V/346L/351A/410H/429M, 22L/25V/110K/291Y/304A, 24T/147V/164S/166A/167E/243Q/389T, 24T/166A, 24T/166A/190P, 24T/166A/207D, 24T/166L, 26K/30P/168E/291Y, 26R/30P, 30P/77R/291Y, 30P/291Y, 70E/77D/108R/177I/196E/247K/291Y/309A/361A/366H, 70E/77R/168E/291Y, 70E/177I/291Y/361A/366H, 70E/177I/291Y/372W, 70E/186C, 70E/247K/291Y/309A/372W, 70E/291Y, 70E/291Y/304A/361A, 74A, 74A/211R, 77D/110K/291Y/308Q, 77D/291Y/361A/366H, 107E/277V/346L/410H/429M, 107E/346L, 110K/235S/391T, 147V/164R/166A/167E, 147V/164R/166A/243Q/272E/283C/385A/389T, 147V/164R/166A/272E, 163Y, 166A, 166A/167E/272E/389T, 166A/190P, 166A/190P/207D, 166A/207D, 166G, 166S, 166S/207D, 166V, 177I/195A/196E/291Y/304A/309A/366H, 177I/196E/247K/304A/309A/361A/366H, 177I/291Y/309A/361A, 196E/291Y/304A/361A, 196E/291Y/309A/361A, 198T/235S/391T, 207D, 247K/304A/309A/366H, 253L/391T, 277V, 291Y, 291Y/304A/309A, and 291Y/309A, wherein the positions are numbered with reference to SEQ ID NO:156. In some additional embodiments, the amino acid differences comprise the substitution(s) E12D/H26K/T30P/F291Y, E12D/H26R/K73C/E77R/F291Y, T13A/H45N/D107E/K236E/T277V/A346L/R351A/L410H/L429M, M22L/L25V/R110K/F291Y/S304A, S24T/W147V/P164S/N166A/S167E/L243Q/S389T, S24T/N166A, S24T/N166A/E190P, S24T/N166A/E207D, S24T/N166L, H26K/T30P/V168E/F291Y, H26R/T30P, T30P/E77R/F291Y, T30P/F291Y, D70E/E77D/S108R/L177I/Q196E/R247K/F291Y/T309A/K361A/R366H, D70E/E77R/V168E/F291Y, D70E/L177I/F291Y/K361A/R366H, D70E/L177I/F291Y/Y372W, D70E/R186C, D70E/R247K/F291Y/T309A/Y372W, D70E/F291Y, D70E/F291Y/S304A/K361A, T74A, T74A/K211R, E77D/R110K/F291Y/E308Q, E77D/F291Y/K361A/R366H,

D107E/T277V/A346L/L410H/L429M, D107E/A346L, R110K/A235S/A391T, W147V/P164R/N166A/S167E, W147V/P164R/N166A/L243Q/V272E/S283C/K385A/S389T, W147V/P164R/N166A/V272E, F163Y, N166A, N166A/S167E/V272E/S389T, N166A/E190P, N166A/E190P/E207D, N166A/E207D, N166G, N166S, N166S/E207D, N166V, L1771/E195A/Q196E/F291Y/S304A/T309A/R366H, L1771/Q196E/R247K/S304A/T309A/K361A/R366H, L1771/F291Y/T309A/K361A, Q196E/F291Y/S304A/K361A, Q196E/F291Y/T309A/K361A, V198T/A235S/A391T, E207D, R247K/S304A/T309A/R366H, V253L/A391T, T277V, F291Y, F291Y/S304A/T309A, and F291Y/T309A, wherein the positions are numbered with reference to SEQ ID NO:156.

[0216] In some embodiments, the variant transaminase of the present invention further comprises additional sequences that do not alter the encoded activity of the enzyme. For example, in some embodiments, the variant transaminase is linked to an epitope tag or to another sequence useful in purification.

[0217] In some embodiments, the variant transaminase polypeptides of the present invention are secreted from the host cell in which they are expressed (e.g., a yeast or filamentous fungal host cell) and are expressed as a pre-protein including a signal peptide (i.e., an amino acid sequence linked to the amino terminus of a polypeptide and which directs the encoded polypeptide into the cell secretory pathway).

[0218] When the sequence of the engineered polypeptide is known, the polynucleotides encoding the enzyme can be prepared by standard solid-phase methods, according to known synthetic methods. In some embodiments, fragments of up to about 100 bases can be individually synthesized, then joined (e.g., by enzymatic or chemical ligation methods, or polymerase mediated methods) to form any desired continuous sequence. For example, polynucleotides and oligonucleotides of the invention can be prepared by chemical synthesis (e.g., using the classical phosphoramidite method described by Beaucage et al., *Tet. Lett.*, 22:1859-69 [1981], or the method described by Matthes et al., *EMBO J.*, 3:801-05 [1984], as it is typically practiced in automated synthetic methods). According to the phosphoramidite method, oligonucleotides are synthesized (e.g., in an automatic DNA synthesizer), purified, annealed, ligated and cloned in appropriate vectors. In addition, essentially any nucleic acid can be obtained from any of a variety of commercial sources (e.g., The Midland Certified Reagent Company, Midland, TX, The Great American Gene Company, Ramona, CA, ExpressGen Inc. Chicago, IL, Operon Technologies Inc., Alameda, CA, and many others).

[0219] The present invention also provides recombinant constructs comprising a sequence encoding at least one variant transaminase, as provided herein. In some embodiments, the present invention provides an expression vector comprising a variant transaminase polynucleotide operably linked to a heterologous promoter. In some embodiments, expression vectors of the present invention are used to transform appropriate host cells to permit the host cells to express the variant transaminase protein.

[0220] Methods for recombinant expression of proteins in fungi and other organisms are well known in the art, and a number of expression vectors are available or can be constructed using routine methods. In some embodiments, nucleic acid constructs of the present invention comprise a vector, such as, a plasmid, a cosmid, a phage, a virus, a bacterial artificial chromosome (BAC), a yeast artificial chromosome (YAC), and the like, into which a nucleic acid sequence of the invention has been inserted. In some embodiments, polynucleotides of the present invention are incorporated into any one of a variety of expression vectors suitable for expressing variant transaminase polypeptide(s). Suitable vectors include, but are not limited to chromosomal, nonchromosomal and synthetic DNA sequences (e.g., derivatives of SV40), as well as bacterial plasmids, phage DNA, baculovirus, yeast plasmids, vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, pseudorabies, adenovirus, adeno-associated virus, retroviruses, and many others. Any suitable vector that

transduces genetic material into a cell, and, if replication is desired, which is replicable and viable in the relevant host finds use in the present invention.

[0221] In some embodiments, the construct further comprises regulatory sequences, including but not limited to a promoter, operably linked to the protein encoding sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art. Indeed, in some embodiments, in order to obtain high levels of expression in a particular host it is often useful to express the variant transaminases of the present invention under the control of a heterologous promoter. In some embodiments, a promoter sequence is operably linked to the 5' region of the variant transaminase coding sequence using any suitable method known in the art. Examples of useful promoters for expression of variant transaminases include, but are not limited to promoters from fungi. In some embodiments, a promoter sequence that drives expression of a gene other than a transaminase gene in a fungal strain finds use. As a non-limiting example, a fungal promoter from a gene encoding an endoglucanase may be used. In some embodiments, a promoter sequence that drives the expression of a transaminase gene in a fungal strain other than the fungal strain from which the transaminases were derived finds use. Examples of other suitable promoters useful for directing the transcription of the nucleotide constructs of the present invention in a filamentous fungal host cell include, but are not limited to promoters obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *Aspergillus niger* neutral alpha-amylase, *Aspergillus niger* acid stable alpha-amylase, *Aspergillus niger* or *Aspergillus awamori* glucoamylase (glaA), *Rhizomucor miehei* lipase, *Aspergillus oryzae* alkaline protease, *Aspergillus oryzae* triose phosphate isomerase, *Aspergillus nidulans* acetamidase, and *Fusarium oxysporum* trypsin-like protease (See e.g., WO 96/00787, incorporated herein by reference), as well as the NA2-tpi promoter (a hybrid of the promoters from the genes for *Aspergillus niger* neutral alpha-amylase and *Aspergillus oryzae* triose phosphate isomerase), promoters such as cbh1, cbh2, egl1, egl2, pepA, hfb1, hfb2, xyn1, amy, and glaA (See e.g., Nunberg et al., *Mol. Cell Biol.*, 4:2306-2315 [1984]; Boel et al., *EMBO J.*, 3:1581-85 [1984]; and European Patent Appln. 137280, all of which are incorporated herein by reference), and mutant, truncated, and hybrid promoters thereof.

[0222] In yeast host cells, useful promoters include, but are not limited to those from the genes for *Saccharomyces cerevisiae* enolase (eno-1), *Saccharomyces cerevisiae* galactokinase (gal), *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP), and *S. cerevisiae* 3-phosphoglycerate kinase. Additional useful promoters useful for yeast host cells are known in the art (See e.g., Romanos et al., *Yeast* 8:423-488 [1992], incorporated herein by reference). In addition, promoters associated with chitinase production in fungi find use in the present invention (See e.g., Blaiseau and Lafay, *Gene* 120:243-248 [1992]; and Limon et al., *Curr. Genet.*, 28:478-83 [1995], both of which are incorporated herein by reference).

[0223] For bacterial host cells, suitable promoters for directing transcription of the nucleic acid constructs of the present disclosure, include but are not limited to the promoters obtained from the *E. coli* lac operon, *E. coli* trp operon, bacteriophage lambda, *Streptomyces coelicolor* agarase gene (dagA), *Bacillus subtilis* levansucrase gene (sacB), *Bacillus licheniformis* alpha-amylase gene (amyL), *Bacillus stearothermophilus* maltogenic amylase gene (amyM), *Bacillus amyloliquefaciens* alpha-amylase gene (amyQ), *Bacillus licheniformis* penicillinase gene (penP), *Bacillus subtilis* xylA and xylB genes, and prokaryotic beta-lactamase gene (See e.g., Villa-Kamaroff et al., *Proc. Natl. Acad. Sci. USA* 75: 3727-3731 [1978]), as well as the tac promoter (See e.g., DeBoer et al., *Proc. Natl. Acad. Sci. USA* 80: 21-25 [1983]).

[0224] In some embodiments, cloned variant transaminases of the present invention also have a suitable transcription terminator sequence, a sequence recognized by a host cell to terminate transcription. The terminator sequence is operably linked to the 3' terminus of the nucleic acid sequence encoding the polypeptide. Any terminator that is functional in the host cell of choice finds use in the present invention. Exemplary transcription terminators for filamentous fungal host cells include, but are not limited to those obtained from the genes for *Aspergillus oryzae* TAKA amylase,

Aspergillus niger glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Aspergillus niger* alpha-glucosidase, and *Fusarium oxysporum* trypsin-like protease (See e.g., U.S. Pat. No. 7,399,627, incorporated herein by reference). In some embodiments, exemplary terminators for yeast host cells include those obtained from the genes for *Saccharomyces cerevisiae* enolase, *Saccharomyces cerevisiae* cytochrome C (CYC1), and *Saccharomyces cerevisiae* glyceraldehyde-3-phosphate dehydrogenase. Other useful terminators for yeast host cells are well-known to those skilled in the art (See e.g., Romanos et al., *Yeast* 8:423-88 [1992]).

[0225] In some embodiments, a suitable leader sequence is part of a cloned variant transaminase sequence, which is a nontranslated region of an mRNA that is important for translation by the host cell.

[0226] The leader sequence is operably linked to the 5' terminus of the nucleic acid sequence encoding the polypeptide. Any leader sequence that is functional in the host cell of choice finds use in the present invention. Exemplary leaders for filamentous fungal host cells include, but are not limited to those obtained from the genes for *Aspergillus oryzae* TAKA amylase and *Aspergillus nidulans* triose phosphate isomerase. Suitable leaders for yeast host cells include, but are not limited to those obtained from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* 3-phosphoglycerate kinase, *Saccharomyces cerevisiae* alpha-factor, and *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP).

[0227] In some embodiments, the sequences of the present invention also comprise a polyadenylation sequence, which is a sequence operably linked to the 3' terminus of the nucleic acid sequence and which, when transcribed, is recognized by the host cell as a signal to add polyadenosine residues to transcribed mRNA. Any polyadenylation sequence which is functional in the host cell of choice finds use in the present invention. Exemplary polyadenylation sequences for filamentous fungal host cells include, but are not limited to those obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Fusarium oxysporum* trypsin-like protease, and *Aspergillus niger* alpha-glucosidase. Useful polyadenylation sequences for yeast host cells are known in the art (See e.g., Guo and Sherman, *Mol. Cell. Biol.*, 15:5983-5990 [1995]).

[0228] In some embodiments, the control sequence comprises a signal peptide coding region encoding an amino acid sequence linked to the amino terminus of a polypeptide and directs the encoded polypeptide into the cell's secretory pathway. The 5' end of the coding sequence of the nucleic acid sequence may inherently contain a signal peptide coding region naturally linked in translation reading frame with the segment of the coding region that encodes the secreted polypeptide. Alternatively, the 5' end of the coding sequence may contain a signal peptide coding region that is foreign to the coding sequence. The foreign signal peptide coding region may be required where the coding sequence does not naturally contain a signal peptide coding region.

[0229] Alternatively, the foreign signal peptide coding region may simply replace the natural signal peptide coding region in order to enhance secretion of the polypeptide. However, any signal peptide coding region which directs the expressed polypeptide into the secretory pathway of a host cell of choice may be used in the present invention.

[0230] In some embodiments, the signal peptide is an endogenous *V. fluvialis* transaminase signal peptide. In some additional embodiments, signal peptides from other *V. fluvialis* secreted proteins are used. In some embodiments, other signal peptides find use, depending on the host cell and other factors.

[0231] Effective signal peptide coding regions for bacterial host cells include, but are not limited to the signal peptide coding regions obtained from the genes for *Bacillus* NC1B 11837 maltogenic amylase, *Bacillus stearothermophilus* alpha-amylase, *Bacillus licheniformis* subtilisin, *Bacillus licheniformis* beta-lactamase, *Bacillus stearothermophilus* neutral proteases (nprT, nprS, nprM), and *Bacillus subtilis* prsA. Further signal peptides are known in the art (See e.g., Simonen and

Palva, *Microbiol. Rev.*, 57: 109-137 [1993]).

[0232] Effective signal peptide coding regions for filamentous fungal host cells include, but are not limited to the signal peptide coding regions obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* neutral amylase, *Aspergillus niger* glucoamylase, *Rhizomucor miehei* aspartic proteinase, *Humicola insolens* cellulase, and *Humicola lanuginosa* lipase.

[0233] Useful signal peptides for yeast host cells include, but are not limited to genes for *Saccharomyces cerevisiae* alpha-factor and *Saccharomyces cerevisiae* invertase. Other useful signal peptide coding regions are known in the art (See e.g., Romanos et al., [1992], supra).

[0234] In some embodiments, the control sequence comprises a propeptide coding region that codes for an amino acid sequence positioned at the amino terminus of a polypeptide. The resultant polypeptide is known as a proenzyme or propolypeptide (or a zymogen in some cases). A propolypeptide is generally inactive and can be converted to a mature active transaminase polypeptide by catalytic or autocatalytic cleavage of the propeptide from the propolypeptide. The propeptide coding region may be obtained from the genes for *Bacillus subtilis* alkaline protease (aprE), *Bacillus subtilis* neutral protease (nprT), *Saccharomyces cerevisiae* alpha-factor, *Rhizomucor miehei* aspartic proteinase, an dMyceliophthora *thermophila* lactase (See e.g., WO 95/33836).

[0235] Where both signal peptide and propeptide regions are present at the amino terminus of a polypeptide, the propeptide region is positioned next to the amino terminus of a polypeptide and the signal peptide region is positioned next to the amino terminus of the propeptide region.

[0236] In some embodiments, regulatory sequences are also used to allow the regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory systems are those which cause the expression of the gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. In prokaryotic host cells, suitable regulatory sequences include, but are not limited to the lac, tac, and trp operator systems. In yeast host cells, suitable regulatory systems include, as examples, the ADH2 system or GAL1 system. In filamentous fungi, suitable regulatory sequences include the TAKA alpha-amylase promoter, *Aspergillus niger* glucoamylase promoter, and *Aspergillus oryzae* glucoamylase promoter.

[0237] Other examples of regulatory sequences are those which allow for gene amplification. In eukaryotic systems, these include the dihydrofolate reductase gene, which is amplified in the presence of methotrexate, and the metallothionein genes, which are amplified with heavy metals. In these cases, the nucleic acid sequence encoding the transaminase polypeptide of the present invention would be operably linked with the regulatory sequence.

[0238] Thus, in additional embodiments, the present invention provides recombinant expression vectors comprising a polynucleotide encoding an engineered transaminase polypeptide or a variant thereof, and one or more expression regulating regions such as a promoter and a terminator, a replication origin, etc., depending on the type of hosts into which they are to be introduced. In some embodiments, the various nucleic acid and control sequences described above are joined together to produce a recombinant expression vector that may include one or more convenient restriction sites to allow for insertion or substitution of the nucleic acid sequence encoding the polypeptide at such sites.

[0239] Alternatively, in some embodiments, the nucleic acid sequences are expressed by inserting the nucleic acid sequence or a nucleic acid construct comprising the sequence into an appropriate vector for expression. In creating the expression vector, the coding sequence is located in the vector so that the coding sequence is operably linked with the appropriate control sequences for expression.

[0240] The recombinant expression vector comprises any suitable vector (e.g., a plasmid or virus), that can be conveniently subjected to recombinant DNA procedures and can bring about the expression of the polynucleotide sequence. The choice of the vector typically depends on the

compatibility of the vector with the host cell into which the vector is to be introduced. In some embodiments, the vectors are linear or closed circular plasmids.

[0241] In some embodiments, the expression vector is an autonomously replicating vector (i.e., a vector that exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, such as a plasmid, an extrachromosomal element, a minichromosome, or an artificial chromosome). In some embodiments, the vector contains any means for assuring self-replication. Alternatively, in some other embodiments, upon being introduced into the host cell, the vector is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. Furthermore, in additional embodiments, a single vector or plasmid or two or more vectors or plasmids which together contain the total DNA to be introduced into the genome of the host cell, or a transposon find use.

[0242] In some embodiments, the expression vector of the present invention contains one or more selectable markers, which permit easy selection of transformed cells. A “selectable marker” is a gene, the product of which provides for biocide or viral resistance, resistance to antimicrobials or heavy metals, prototrophy to auxotrophs, and the like. Any suitable selectable markers for use in a filamentous fungal host cell find use in the present invention, including, but are not limited to, *amdS* (acetamidase), *argB* (ornithine carbamoyltransferase), *bar* (phosphinothricin acetyltransferase), *hph* (hygromycin phosphotransferase), *niaD* (nitrate reductase), *pyrG* (orotidine-5'-phosphate decarboxylase), *sC* (sulfate adenylyltransferase), and *trpC* (anthranilate synthase), as well as equivalents thereof. Additional markers useful in host cells such as *Aspergillus*, include but are not limited to the *amdS* and *pyrG* genes of *Aspergillus nidulans* or *Aspergillus oryzae*, and the *bar* gene of *Streptomyces hygroscopicus*. Suitable markers for yeast host cells include, but are not limited to ADE2, HIS3, LEU2, LYS2, MET3, TRP1, and URA3. Examples of bacterial selectable markers include, but are not limited to the *dal* genes from *Bacillus subtilis* or *Bacillus licheniformis*, or markers, which confer antibiotic resistance such as ampicillin, kanamycin, chloramphenicol, and or tetracycline resistance.

[0243] In some embodiments, the expression vectors of the present invention contain an element(s) that permits integration of the vector into the host cell's genome or autonomous replication of the vector in the cell independent of the genome. In some embodiments involving integration into the host cell genome, the vectors rely on the nucleic acid sequence encoding the polypeptide or any other element of the vector for integration of the vector into the genome by homologous or non-homologous recombination.

[0244] In some alternative embodiments, the expression vectors contain additional nucleic acid sequences for directing integration by homologous recombination into the genome of the host cell. The additional nucleic acid sequences enable the vector to be integrated into the host cell genome at a precise location(s) in the chromosome(s). To increase the likelihood of integration at a precise location, the integrational elements preferably contain a sufficient number of nucleotides, such as 100 to 10,000 base pairs, preferably 400 to 10,000 base pairs, and most preferably 800 to 10,000 base pairs, which are highly homologous with the corresponding target sequence to enhance the probability of homologous recombination. The integrational elements may be any sequence that is homologous with the target sequence in the genome of the host cell. Furthermore, the integrational elements may be non-encoding or encoding nucleic acid sequences. On the other hand, the vector may be integrated into the genome of the host cell by non-homologous recombination.

[0245] For autonomous replication, the vector may further comprise an origin of replication enabling the vector to replicate autonomously in the host cell in question. Examples of bacterial origins of replication are P15A on or the origins of replication of plasmids pBR322, pUC19, pACYC177 (which plasmid has the P15A ori), or pACYC184 permitting replication in *E. coli*, and pUB110, pE194, pTA1060, or pAM\$1 permitting replication in *Bacillus*. Examples of origins of replication for use in a yeast host cell are the 2 micron origin of replication, ARS1, ARS4, the combination of ARS1 and CEN3, and the combination of ARS4 and CEN6. The origin of

replication may be one having a mutation which makes it's functioning temperature-sensitive in the host cell (See e.g., Ehrlich, Proc. *P Natl. Acad. Sci. USA* 75:1433 [1978]).

[0246] In some embodiments, more than one copy of a nucleic acid sequence of the present invention is inserted into the host cell to increase production of the gene product. An increase in the copy number of the nucleic acid sequence can be obtained by integrating at least one additional copy of the sequence into the host cell genome or by including an amplifiable selectable marker gene with the nucleic acid sequence where cells containing amplified copies of the selectable marker gene, and thereby additional copies of the nucleic acid sequence, can be selected for by cultivating the cells in the presence of the appropriate selectable agent.

[0247] Many of the expression vectors for use in the present invention are commercially available.

[0248] Suitable commercial expression vectors include, but are not limited to the p3xFLAGTM expression vectors (Sigma-Aldrich Chemicals), which include a CMV promoter and hGH polyadenylation site for expression in mammalian host cells and a pBR322 origin of replication and ampicillin resistance markers for amplification in *E. coli*. Other suitable expression vectors include, but are not limited to pBluescriptII SK(-) and pBK-CMV (Stratagene), and plasmids derived from pBR322 (Gibco BRL), pUC (Gibco BRL), pREP4, pCEP4 (Invitrogen) or pPoly (See e.g., Lathe et al., *Gene* 57:193-201 [1987]).

[0249] Thus, in some embodiments, a vector comprising a sequence encoding at least one variant transaminase is transformed into a host cell in order to allow propagation of the vector and expression of the variant transaminase (s). In some embodiments, the variant transaminases are post-translationally modified to remove the signal peptide and in some cases may be cleaved after secretion. In some embodiments, the transformed host cell described above is cultured in a suitable nutrient medium under conditions permitting the expression of the variant transaminase(s). Any suitable medium useful for culturing the host cells finds use in the present invention, including, but not limited to minimal or complex media containing appropriate supplements. In some embodiments, host cells are grown in HTP media. Suitable media are available from various commercial suppliers or may be prepared according to published recipes (e.g., in catalogues of the American Type Culture Collection).

[0250] In another aspect, the present invention provides host cells comprising a polynucleotide encoding an improved transaminase polypeptide provided herein, the polynucleotide being operatively linked to one or more control sequences for expression of the transaminase enzyme in the host cell.

[0251] Host cells for use in expressing the transaminase polypeptides encoded by the expression vectors of the present invention are well known in the art and include but are not limited to, bacterial cells, such as *E. coli*, *Bacillus megaterium*, *Lactobacillus kefir*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells (e.g., *Saccharomyces cerevisiae* or *Pichia pastoris* (ATCC Accession No. 201178)); insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, BHK, 293, and Bowes melanoma cells; and plant cells. Appropriate culture media and growth conditions for the above-described host cells are well known in the art.

[0252] Polynucleotides for expression of the transaminase may be introduced into cells by various methods known in the art. Techniques include among others, electroporation, biolistic particle bombardment, liposome mediated transfection, calcium chloride transfection, and protoplast fusion.

[0253] Various methods for introducing polynucleotides into cells are known to those skilled in the art.

[0254] In some embodiments, the host cell is a eukaryotic cell. Suitable eukaryotic host cells include, but are not limited to, fungal cells, algal cells, insect cells, and plant cells. Suitable fungal host cells include, but are not limited to, Ascomycota, Basidiomycota, Deuteromycota, Zygomycota, Fungi imperfecti. In some embodiments, the fungal host cells are yeast cells and

filamentous fungal cells. The filamentous fungal host cells of the present invention include all filamentous forms of the subdivision Eumycotina and Oomycota. Filamentous fungi are characterized by a vegetative mycelium with a cell wall composed of chitin, cellulose and other complex polysaccharides. The filamentous fungal host cells of the present invention are morphologically distinct from yeast.

[0255] In some embodiments of the present invention, the filamentous fungal host cells are of any suitable genus and species, including, but not limited to *Achlya*, *Acremonium*, *Aspergillus*, *Aureobasidium*, *Bjerkandera*, *Ceriporiopsis*, *Cephalosporium*, *Chrysosporium*, *Cochliobolus*, *Corynascus*, *Cryphonectria*, *Cryptococcus*, *Coprinus*, *Coriolus*, *Diplodia*, *Endothia*, *Fusarium*, *Gibberella*, *Gliocladium*, *Humicola*, *Hypocrea*, *Myceliophthora*, *Mucor*, *Neurospora*, *Penicillium*, *Podospora*, *Phlebia*, *Piromyces*, *Pyricularia*, *Rhizomucor*, *Rhizopus*, *Schizophyllum*, *Scytalidium*, *Sporotrichum*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Trametes*, *Tolypocladium*, *Trichoderma*, *Verticillium*, and/or *Volvariella*, and/or teleomorphs, or anamorphs, and synonyms, basionyms, or taxonomic equivalents thereof.

[0256] In some embodiments of the present invention, the host cell is a yeast cell, including but not limited to cells of *Candida*, *Hansenula*, *Saccharomyces*, *Schizosaccharomyces*, *Pichia*, *Kluyveromyces*, or *Yarrowia* species. In some embodiments of the present invention, the yeast cell is *Hansenula polymorpha*, *Saccharomyces cerevisiae*, *Saccharomyces carlsbergensis*, *Saccharomyces diastaticus*, *Saccharomyces norbensis*, *Saccharomyces kluyveri*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia kodamae*, *Pichia membranaefaciens*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia quercuum*, *Pichia pijperi*, *Pichia stipitis*, *Pichia methanolica*, *Pichia angusta*, *Kluyveromyces lactis*, *Candida albicans*, or *Yarrowia lipolytica*.

[0257] In some embodiments of the invention, the host cell is an algal cell such as *Chlamydomonas* (e.g., *C. reinhardtii*) and *Phormidium* (e.g., *Phormidium* sp. ATCC29409).

[0258] In some other embodiments, the host cell is a prokaryotic cell. Suitable prokaryotic cells include, but are not limited to Gram-positive, Gram-negative and Gram-variable bacterial cells. Any suitable bacterial organism finds use in the present invention, including but not limited to *Agrobacterium*, *Alicyclobacillus*, *Anabaena*, *Anacystis*, *Acinetobacter*, *Acidothermus*, *Arthrobacter*, *Azobacter*, *Bacillus*, *Bifidobacterium*, *Brevibacterium*, *Butyrivibrio*, *Buchnera*, *Campestris*, *Campylobacter*, *Clostridium*, *Corynebacterium*, *Chromatium*, *Coprococcus*, *Escherichia*, *Enterococcus*, *Enterobacter*, *Erwinia*, *Fusobacterium*, *Faecalibacterium*, *Francisella*, *Flavobacterium*, *Geobacillus*, *Haemophilus*, *Helicobacter*, *Klebsiella*, *Lactobacillus*, *Lactococcus*, *Ilyobacter*, *Micrococcus*, *Microbacterium*, *Mesorhizobium*, *Methylobacterium*, *Methylobacterium*, *Mycobacterium*, *Neisseria*, *Pantoea*, *Pseudomonas*, *Prochlorococcus*, *Rhodobacter*, *Rhodopseudomonas*, *Rhodopseudomonas*, *Roseburia*, *Rhodospirillum*, *Rhodococcus*, *Scenedesmus*, *Streptomyces*, *Streptococcus*, *Synecoccus*, *Saccharomonospora*, *Staphylococcus*, *Serratia*, *Salmonella*, *Shigella*, *Thermoanaerobacterium*, *Tropheryma*, *Tularensis*, *Temecula*, *Thermosynechococcus*, *Thermococcus*, *Ureaplasma*, *Xanthomonas*, *Xylella*, *Yersinia* and *Zymomonas*. In some embodiments, the host cell is a species of *Agrobacterium*, *Acinetobacter*, *Azobacter*, *Bacillus*, *Bifidobacterium*, *Buchnera*, *Geobacillus*, *Campylobacter*, *Clostridium*, *Corynebacterium*, *Escherichia*, *Enterococcus*, *Erwinia*, *Flavobacterium*, *Lactobacillus*, *Lactococcus*, *Pantoea*, *Pseudomonas*, *Staphylococcus*, *Salmonella*, *Streptococcus*, *Streptomyces*, or *Zymomonas*. In some embodiments, the bacterial host strain is non-pathogenic to humans. In some embodiments the bacterial host strain is an industrial strain. Numerous bacterial industrial strains are known and suitable in the present invention. In some embodiments of the present invention, the bacterial host cell is an *Agrobacterium* species (e.g., *A. radiobacter*, *A. rhizogenes*, and *A. rubi*). In some embodiments of the present invention, the bacterial host cell is an *Arthrobacter* species (e.g., *A. aureus*, *A. citreus*, *A. globiformis*, *A. hydrocarboglutamicus*, *A. mysorens*, *A. nicotianae*, *A. paraffineus*, *A. protophonniae*, *A. roseoparaffinus*, *A. sulfureus*, and *A. ureafaciens*). In some

embodiments of the present invention, the bacterial host cell is a *Bacillus* species (e.g., *B. thuringensis*, *B. anthracis*, *B. megaterium*, *B. subtilis*, *B. lentus*, *B. circulans*, *B. pumilus*, *B. lautus*, *B. coagulans*, *B. brevis*, *B. firmus*, *B. alkaophilus*, *B. licheniformis*, *B. clausii*, *B. stearothermophilus*, *B. halodurans*, and *B. amyloliquefaciens*). In some embodiments, the host cell is an industrial *Bacillus* strain including but not limited to *B. subtilis*, *B. pumilus*, *B. licheniformis*, *B. megaterium*, *B. clausii*, *B. stearothermophilus*, or *B. amyloliquefaciens*. In some embodiments, the *Bacillus* host cells are *B. subtilis*, *B. licheniformis*, *B. megaterium*, *B. stearothermophilus*, and/or *B. amyloliquefaciens*. In some embodiments, the bacterial host cell is a *Clostridium* species (e.g., *C. acetobutylicum*, *C. tetani* E88, *C. lituseburense*, *C. saccharobutylicum*, *C. perfringens*, and *C. beijerinckii*). In some embodiments, the bacterial host cell is a *Corynebacterium* species (e.g., *C. glutamicum* and *C. acetoacidophilum*). In some embodiments the bacterial host cell is an *Escherichia* species (e.g., *E. coli*). In some embodiments, the host cell is *Escherichia coli* W3110. In some embodiments, the bacterial host cell is an *Erwinia* species (e.g., *E. uredovora*, *E. carotovora*, *E. ananas*, *E. herbicola*, *E. punctata*, and *E. terreus*). In some embodiments, the bacterial host cell is a *Pantoea* species (e.g., *P. citrea*, and *P. agglomerans*). In some embodiments the bacterial host cell is a *Pseudomonas* species (e.g., *P. putida*, *P. aeruginosa*, *P. mevalonii*, and *P. sp. D-01 10*). In some embodiments, the bacterial host cell is a *Streptococcus* species (e.g., *S. equisimilis*, *S. pyogenes*, and *S. uberis*). In some embodiments, the bacterial host cell is a *Streptomyces* species (e.g., *S. ambofaciens*, *S. achromogenes*, *S. avermitilis*, *S. coelicolor*, *S. aureofaciens*, *S. aureus*, *S. fungicidicus*, *S. griseus*, and *S. lividans*). In some embodiments, the bacterial host cell is a *Zymomonas* species (e.g., *Z. mobilis*, and *Z. lipolytica*).

[0259] Many prokaryotic and eukaryotic strains that find use in the present invention are readily available to the public from a number of culture collections such as American Type Culture Collection (ATCC), Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSM), Centraalbureau Voor Schimmelcultures (CBS), and *Agricultural Research Service* Patent Culture Collection, Northern Regional Research Center (NRRL).

[0260] In some embodiments, host cells are genetically modified to have characteristics that improve protein secretion, protein stability and/or other properties desirable for expression and/or secretion of a protein. Genetic modification can be achieved by genetic engineering techniques and/or classical microbiological techniques (e.g., chemical or UV mutagenesis and subsequent selection). Indeed, in some embodiments, combinations of recombinant modification and classical selection techniques are used to produce the host cells. Using recombinant technology, nucleic acid molecules can be introduced, deleted, inhibited or modified, in a manner that results in increased yields of transaminase variant(s) within the host cell and/or in the culture medium. For example, knockout of *AlpI* function results in a cell that is protease deficient, and knockout of *pyr5* function results in a cell with a pyrimidine deficient phenotype. In one genetic engineering approach, homologous recombination is used to induce targeted gene modifications by specifically targeting a gene in vivo to suppress expression of the encoded protein. In alternative approaches, siRNA, antisense and/or ribozyme technology find use in inhibiting gene expression. A variety of methods are known in the art for reducing expression of protein in cells, including, but not limited to deletion of all or part of the gene encoding the protein and site-specific mutagenesis to disrupt expression or activity of the gene product. (See e.g., Chaveroche et al., *Nucl. Acids Res.*, 28:22 e97 [2000]; Cho et al., *Mol. Plant Mic. Interact.*, 19:7-15 [2006]; Maruyama and Kitamoto, *Biotechnol Lett.*, 30:1811-1817 [2008]; Takahashi et al., *Mol. Gen. Genom.*, 272: 344-352 [2004]; and You et al., *Arch. Microbiol.*, 191:615-622 [2009], all of which are incorporated by reference herein). Random mutagenesis, followed by screening for desired mutations also finds use (See e.g., Combier et al., *FEMS Microbiol. Lett.*, 220:141-8 [2003]; and Firon et al., *Eukary. Cell* 2:247-55 [2003], both of which are incorporated by reference).

[0261] Introduction of a vector or DNA construct into a host cell can be accomplished using any suitable method known in the art, including but not limited to calcium phosphate transfection,

DEAE-dextran mediated transfection, PEG-mediated transformation, electroporation, or other common techniques known in the art. In some embodiments, the *Escherichia coli* expression vector pCK100900i (See US Pat. Appln. Pubin. 2006/0195947, which is hereby incorporated by reference herein) find use.

[0262] In some embodiments, the engineered host cells (i.e., “recombinant host cells”) of the present invention are cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants, or amplifying the transaminase polynucleotide. Culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and are well-known to those skilled in the art. As noted, many standard references and texts are available for the culture and production of many cells, including cells of bacterial, plant, animal (especially mammalian) and archebacterial origin.

[0263] In some embodiments, cells expressing the variant transaminase polypeptides of the invention are grown under batch or continuous fermentations conditions. Classical “batch fermentation” is a closed system, wherein the compositions of the medium is set at the beginning of the fermentation and is not subject to artificial alternations during the fermentation. A variation of the batch system is a “fed-batch fermentation” which also finds use in the present invention. In this variation, the substrate is added in increments as the fermentation progresses. Fed-batch systems are useful when catabolite repression is likely to inhibit the metabolism of the cells and where it is desirable to have limited amounts of substrate in the medium. Batch and fed-batch fermentations are common and well known in the art. “Continuous fermentation” is an open system where a defined fermentation medium is added continuously to a bioreactor and an equal amount of conditioned medium is removed simultaneously for processing. Continuous fermentation generally maintains the cultures at a constant high density where cells are primarily in log phase growth. Continuous fermentation systems strive to maintain steady state growth conditions. Methods for modulating nutrients and growth factors for continuous fermentation processes as well as techniques for maximizing the rate of product formation are well known in the art of industrial microbiology.

[0264] In some embodiments of the present invention, cell-free transcription/translation systems find use in producing variant transaminase(s). Several systems are commercially available and the methods are well-known to those skilled in the art.

[0265] The present invention provides methods of making variant transaminase polypeptides or biologically active fragments thereof. In some embodiments, the method comprises: providing a host cell transformed with a polynucleotide encoding an amino acid sequence that comprises at least about 70% (or at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%) sequence identity to SEQ ID NO: 2, 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156, and comprising at least one mutation as provided herein; culturing the transformed host cell in a culture medium under conditions in which the host cell expresses the encoded variant transaminase polypeptide; and optionally recovering or isolating the expressed variant transaminase polypeptide, and/or recovering or isolating the culture medium containing the expressed variant transaminase polypeptide. In some embodiments, the methods further provide optionally lysing the transformed host cells after expressing the encoded transaminase polypeptide and optionally recovering and/or isolating the expressed variant transaminase polypeptide from the cell lysate. The present invention further provides methods of making a variant transaminase polypeptide comprising cultivating a host cell transformed with a variant transaminase polypeptide under conditions suitable for the production of the variant transaminase polypeptide and recovering the variant transaminase polypeptide. Typically, recovery or isolation of the transaminase polypeptide is from the host cell culture medium, the host cell or both, using protein recovery techniques that are well known in the art, including those described herein. In some embodiments, host cells are harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude

extract retained for further purification. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including, but not limited to freeze-thaw cycling, sonication, mechanical disruption, and/or use of cell lysing agents, as well as many other suitable methods well known to those skilled in the art.

[0266] Engineered transaminase enzymes produced by a host cell can be recovered from the cells and/or the culture medium using any one or more of the techniques known in the art for protein purification, including, among others, lysozyme treatment, sonication, filtration, salting-out, ultra-centrifugation, and chromatography. Suitable solutions for lysing and the high efficiency extraction of proteins from bacteria, such as *E. coli*, are commercially available under the trade name CellLytic B™ (Sigma-Aldrich). Thus, in some embodiments, the resulting polypeptide is recovered/isolated and optionally purified by any of a number of methods known in the art. For example, in some embodiments, the polypeptide is isolated from the nutrient medium by conventional procedures including, but not limited to, centrifugation, filtration, extraction, spray-drying, evaporation, chromatography (e.g., ion exchange, affinity, hydrophobic interaction, chromatofocusing, and size exclusion), or precipitation. In some embodiments, protein refolding steps are used, as desired, in completing the configuration of the mature protein. In addition, in some embodiments, high performance liquid chromatography (HPLC) is employed in the final purification steps. For example, in some embodiments, methods known in the art find use in the present invention (See e.g., Parry et al., *Biochem. J.*, 353:117 [2001]; and Hong et al., *Appl. Microbiol. Biotechnol.*, 73:1331 [2007], both of which are incorporated herein by reference). Indeed, any suitable purification methods known in the art find use in the present invention.

[0267] Chromatographic techniques for isolation of the transaminase polypeptide include, but are not limited to reverse phase chromatography high performance liquid chromatography, ion exchange chromatography, gel electrophoresis, and affinity chromatography. Conditions for purifying a particular enzyme will depend, in part, on factors such as net charge, hydrophobicity, hydrophilicity, molecular weight, molecular shape, etc., are known to those skilled in the art.

[0268] In some embodiments, affinity techniques find use in isolating the improved transaminase enzymes. For affinity chromatography purification, any antibody which specifically binds the transaminase polypeptide may be used. For the production of antibodies, various host animals, including but not limited to rabbits, mice, rats, etc., may be immunized by injection with the transaminase. The transaminase polypeptide may be attached to a suitable carrier, such as BSA, by means of a side chain functional group or linkers attached to a side chain functional group. Various adjuvants may be used to increase the immunological response, depending on the host species, including but not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, and potentially useful human adjuvants such as BCG (*Bacillus Calmette Guerin*) and *Corynebacterium parvum*.

[0269] In some embodiments, the transaminase variants are prepared and used in the form of cells expressing the enzymes, as crude extracts, or as isolated or purified preparations. In some embodiments, the transaminase variants are prepared as lyophilisates, in powder form (e.g., acetone powders), or prepared as enzyme solutions. In some embodiments, the transaminase variants are in the form of substantially pure preparations.

[0270] In some embodiments, the transaminase polypeptides are attached to any suitable solid substrate. Solid substrates include but are not limited to a solid phase, surface, and/or membrane. Solid supports include, but are not limited to organic polymers such as polystyrene, polyethylene, polypropylene, polyfluoroethylene, polyethyleneoxy, and polyacrylamide, as well as co-polymers and grafts thereof. A solid support can also be inorganic, such as glass, silica, controlled pore glass (CPG), reverse phase silica or metal, such as gold or platinum. The configuration of the substrate can be in the form of beads, spheres, particles, granules, a gel, a membrane or a surface. Surfaces can be planar, substantially planar, or non-planar. Solid supports can be porous or non-porous, and

can have swelling or non-swelling characteristics. A solid support can be configured in the form of a well, depression, or other container, vessel, feature, or location. A plurality of supports can be configured on an array at various locations, addressable for robotic delivery of reagents, or by detection methods and/or instruments.

[0271] In some embodiments, immunological methods are used to purify transaminase variants. In one approach, antibody raised against a variant transaminase polypeptide (e.g., against a polypeptide comprising an engineered transaminase variant provided herein, including, but not limited to SEQ ID NO: 2, 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156, and/or an immunogenic fragment thereof) using conventional methods is immobilized on beads, mixed with cell culture media under conditions in which the variant transaminase is bound, and precipitated. In a related approach, immunochromatography finds use.

[0272] In some embodiments, the variant transaminases are expressed as a fusion protein including a non-enzyme portion. In some embodiments, the variant transaminase sequence is fused to a purification facilitating domain. As used herein, the term “purification facilitating domain” refers to a domain that mediates purification of the polypeptide to which it is fused. Suitable purification domains include, but are not limited to metal chelating peptides, histidine-tryptophan modules that allow purification on immobilized metals, a sequence which binds glutathione (e.g., GST), a hemagglutinin (HA) tag (corresponding to an epitope derived from the influenza hemagglutinin protein; See e.g., Wilson et al., *Cell* 37:767 [1984]), maltose binding protein sequences, the FLAG epitope utilized in the FLAGS extension/affinity purification system (e.g., the system available from Immunex Corp), and the like.

[0273] One expression vector contemplated for use in the compositions and methods described herein provides for expression of a fusion protein comprising a polypeptide of the invention fused to a polyhistidine region separated by an enterokinase cleavage site. The histidine residues facilitate purification on IMIAC (immobilized metal ion affinity chromatography; See e.g., Porath et al., *Prot. Exp. Purif*, 3:263-281 [1992]) while the enterokinase cleavage site provides a means for separating the variant transaminase polypeptide from the fusion protein. pGEX vectors (Promega) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to ligand-agarose beads (e.g., glutathione-agarose in the case of GST-fusions) followed by elution in the presence of free ligand.

Methods of Using the Engineered Transaminase Enzymes

[0274] In another aspect, the engineered transaminase polypeptides disclosed herein can be used in a process for the conversion of the substrate compound (2), or structural analogs thereof, to the product of compound (1) or the corresponding structural analog.

[0275] As described herein, and illustrated in the Examples, the present invention contemplates ranges of suitable reaction conditions that can be used in the processes herein, including but not limited to ranges of pH, temperature, buffer, solvent system, substrate loading, mixture of substrate compound stereoisomers, polypeptide loading, cofactor loading, pressure, and reaction time. Further suitable reaction conditions for carrying out the process for biocatalytic conversion of substrate compounds to product compounds using an engineered transaminase polypeptide described herein can be readily optimized by routine experimentation that includes, but is not limited to, contacting the engineered transaminase polypeptide and substrate compound under experimental reaction conditions of concentration, pH, temperature, solvent conditions, and detecting the product compound, for example, using the methods described in the Examples provided herein.

[0276] As described above, the engineered polypeptides having transaminase activity for use in the processes of the present invention generally comprise an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more identity to a reference amino acid sequence selected from any one of the even-numbered sequences

of SEQ ID NO:6-172, and an engineered transaminase polypeptide comprising an amino acid sequence that has (a) has one or more amino acid residue differences as compared to a reference sequence (e.g., SEQ ID NO: 2, 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156). In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes a transaminase polypeptide that has the percent identity described above and one or more residue differences as compared to a reference sequence (e.g., SEQ ID NO: 2, 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156).

[0277] Substrate compound in the reaction mixtures can be varied, taking into consideration, for example, the desired amount of product compound, the effect of substrate concentration on enzyme activity, stability of enzyme under reaction conditions, and the percent conversion of substrate to product. In some embodiments of the method, the suitable reaction conditions comprise a substrate compound loading of at least about 0.5 to about 200 g/L, 1 to about 200 g/L, 5 to about 150 g/L, about 10 to about 100 g/L, or about 50 to about 100 g/L. In some embodiments, the suitable reaction conditions comprise a substrate compound loading of at least about 0.5 g/L, at least about 1 g/L, at least about 5 g/L, at least about 10 g/L, at least about 15 g/L, at least about 20 g/L, at least about 30 g/L, at least about 50 g/L, at least about 75 g/L, at least about 100 g/L, at least about 150 g/L or at least about 200 g/L, or even greater. The values for substrate loadings provided herein are based on the molecular weight of compound (2), however it also contemplated that the equivalent molar amounts of various hydrates and salts of compound (2) also can be used in the process.

[0278] In the processes describes herein, the engineered transaminase polypeptide uses an amino donor to form the product compounds. In some embodiments, the amino donor in the reaction condition comprises a compound selected from isopropylamine (also referred to herein as “IPM”) or any other suitable amino donor for the reaction of interest. In some embodiments, the amino donor is IPM. In some embodiments, the suitable reaction conditions comprise the amino donor, in particular IPM, present at a concentration of at least about 0.1 to about 3.0 M, 0.2 to about 2.5 M, about 0.5 to about 2 M or about 1 to about 2 M. In some embodiments, the amino donor is present at a concentration of about 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 1, 1.5, 2, 2.5 or 3 M.

[0279] Suitable reaction conditions for the processes also typically comprise the presence of a cofactor in the reaction mixture. Because the engineered transaminases typically use members of the vitamin B.sub.6 family, the reaction condition can comprise a cofactor selected from—pyridoxal-5'-phosphate (also known as pyridoxal-phosphate, PLP, P5P), pyridoxine (PN), pyridoxal (PL), pyridoxamine (PM), and their phosphorylated counterparts; pyridoxine phosphate (PNP), and pyridoxamine phosphate (PMP). In some embodiments, the suitable reaction conditions can comprise the presence of a cofactor selected from PLP, PN, PL, PM, PNP, and PMP, at a concentration of about 0.1 g/L to about 10 g/L, about 0.2 g/L to about 5 g/L, about 0.5 g/L to about 2.5 g/L. In some embodiments, the cofactor is PLP.

[0280] Accordingly, in some embodiments, the suitable reaction conditions can comprise the presence of the cofactor, PLP, at a concentration of about 0.1 g/L to about 10 g/L, about 0.2 g/L to about 5 g/L, about 0.5 g/L to about 2.5 g/L. In some embodiments, the reaction conditions comprise a PLP concentration of about 10 g/L or less, about 5 g/L or less, about 2.5 g/L or less, about 1.0 g/L or less, about 0.5 g/L or less, or about 0.2 g/L or less.

[0281] In some embodiments of the process (e.g., where whole cells or lysates are used), the cofactor is present naturally in the cell extract and does not need to be supplemented. In some embodiments of the process (e.g., using partially purified, or purified transaminase enzyme), the process can further comprise a step of adding cofactor to the enzyme reaction mixture. In some embodiments, the cofactor is added either at the beginning of the reaction and/or additional cofactor is added during the reaction.

[0282] During the course of the transamination reactions, the pH of the reaction mixture may change.

[0283] The pH of the reaction mixture may be maintained at a desired pH or within a desired pH

range. This may be done by the addition of an acid or a base, before and/or during the course of the reaction.

[0284] Alternatively, the pH may be controlled by using a buffer. Accordingly, in some embodiments, the reaction condition comprises a buffer. Suitable buffers to maintain desired pH ranges are known in the art and include, by way of example and not limitation, borate, carbonate, phosphate, triethanolamine (TEA) buffer, and the like. In some embodiments, the buffer is TEA. In some embodiments of the process, the suitable reaction conditions comprise a buffer solution of TEA, where the TEA concentration is from about 0.01 to about 0.4 M, 0.05 to about 0.4 M, 0.1 to about 0.3 M, or about 0.1 to about 0.2 M. In some embodiments, the reaction condition comprises a TEA concentration of about 0.01, 0.02, 0.03, 0.04, 0.05, 0.07, 0.1, 0.12, 0.14, 0.16, 0.18, 0.2, 0.3, or 0.4 M. In some embodiments, the reaction conditions comprise water as a suitable solvent with no buffer present.

[0285] In some embodiments of the process, the reaction conditions can comprise a suitable pH. As noted above, the desired pH or desired pH range can be maintained by use of an acid or base, an appropriate buffer, or a combination of buffering and acid or base addition. The pH of the reaction mixture can be controlled before and/or during the course of the reaction. In some embodiments, the suitable reaction conditions comprise a solution pH of about 8 to about 12.5, a pH of about 8 to about 12, a pH of about 9.0 to about 11.5, or a pH of about 9.5 to about 11.0. In some embodiments, the reaction conditions comprise a solution pH of about 8, 8.5, 9, 9.5, 10, 10.5, 11, 11.5, 12 or 12.5.

[0286] In some embodiments of the processes herein, a suitable temperature can be used for the reaction conditions, for example, taking into consideration the increase in reaction rate at higher temperatures, the activity of the enzyme for sufficient duration of the reaction, and as further described below. For example, the engineered polypeptides of the present invention have increased stability relative to naturally occurring transaminase polypeptide, which allows the engineered polypeptides of the present invention to be used at higher temperatures for increased conversion rates and improved substrate solubility characteristics for the reaction. Accordingly, in some embodiments, the suitable reaction conditions comprise a temperature of about 10° C. to about 70° C., about 10° C. to about 65° C., about 15° C. to about 60° C., about 20° C. to about 60° C., about 20° C. to about 55° C., about 30° C. to about 55° C., or about 40° C. to about 50° C. In some embodiments, the suitable reaction conditions comprise a temperature of about 10° C., 15° C., 20° C., 25° C., 30° C., 35° C., 40° C., 45° C., 50° C., 55° C., 60° C., 65° C., or 70° C. In some embodiments, the temperature during the enzymatic reaction can be maintained at a temperature throughout the course of the reaction. In some embodiments, the temperature during the enzymatic reaction can be adjusted over a temperature profile during the course of the reaction.

[0287] In some embodiments of the process, the suitable reaction conditions can further comprise the presence of the reduced cofactor, nicotinamide adenine dinucleotide (NADH), which can act to limit the inactivation of the transaminase enzyme (See e.g., van Ophem et al., *Biochem.*, 37(9):2879-88 [1998]). In such embodiments where NADH is present, a cofactor regeneration system, such as glucose dehydrogenase (GDH) and glucose or formate dehydrogenase and formate can be used to regenerate the NADH in the reaction medium.

[0288] The processes using the engineered transaminases are generally carried out in a solvent. Suitable solvents include water, aqueous buffer solutions, organic solvents, and/or co-solvent systems, which generally comprise aqueous solvents and organic solvents. The aqueous solvent (water or aqueous co-solvent system) may be pH-buffered or unbuffered. In some embodiments, the processes using the engineered transaminase polypeptides are generally carried out in an aqueous co-solvent system comprising an organic solvent (e.g., ethanol, isopropanol (IPA), dimethyl sulfoxide (DMSO), ethyl acetate, butyl acetate, 1-octanol, heptane, octane, methyl t-butyl ether (MTBE), toluene, and the like), ionic liquids (e.g., 1-ethyl 4-methylimidazolium tetrafluoroborate, 1-butyl-3-methylimidazolium tetrafluoroborate, 1-butyl-3-methylimidazolium

hexafluorophosphate, and the like). The organic solvent component of an aqueous co-solvent system may be miscible with the aqueous component, providing a single liquid phase, or may be partly miscible or immiscible with the aqueous component, providing two liquid phases. Exemplary aqueous co-solvent systems comprises water and one or more organic solvent. In general, an organic solvent component of an aqueous co-solvent system is selected such that it does not completely inactivate the transaminase enzyme. Appropriate co-solvent systems can be readily identified by measuring the enzymatic activity of the specified engineered transaminase enzyme with a defined substrate of interest in the candidate solvent system, utilizing an enzyme activity assay, such as those described herein. In some embodiments of the process, the suitable reaction conditions comprise an aqueous co-solvent comprising DMSO at a concentration of about 1% to about 80% (v/v), about 1 to about 70% (v/v), about 2% to about 60% (v/v), about 5% to about 40% (v/v), 10% to about 40% (v/v), 10% to about 30% (v/v), or about 10% to about 20% (v/v). In some embodiments of the process, the suitable reaction conditions comprise an aqueous co-solvent comprising DMSO at a concentration of at least about 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, or 80% (v/v).

[0289] The suitable reaction conditions can comprise a combination of reaction parameters that provide for the biocatalytic conversion of the substrate compounds to its corresponding product compounds. Accordingly, in some embodiments of the process, the combination of reaction parameters comprises: (a) substrate loading of about 2 to 200 g/L of substrate compound (e.g. compound (2)); (b) engineered polypeptide concentration of about 0.5 g/L to 5 g/L; (c) IPM concentration of about 0.1 to 3 M; (d) PLP cofactor concentration of about 0.1 to 1 g/L; (e) DMSO concentration of about 0% (v/v) to about 60% (v/v); (f) pH of about 8.5 to 11.5; and (g) temperature of about 45° C. to 60° C.

[0290] In some embodiments, the combination of reaction parameters comprises: (a) about 50 g/L of substrate compound (e.g. compound (2)); (b) about 2 g/L engineered polypeptide; (c) about 10% (v/v) dimethylsulfoxide (DMSO); (d) about 1 M isopropylamine (IPM); (e) about 1 g/L pyridoxal phosphate (PLP); (f) about pH 9; and (g) about 50° C.

[0291] In some embodiments, the combination of reaction parameters comprises: (a) about 50 g/L of substrate compound (e.g. compound (2)); (b) about 1 g/L engineered polypeptide; (c) about 10% (v/v) dimethylsulfoxide (DMSO); (d) about 1 M isopropylamine (IPM); (e) about 1 g/L pyridoxal phosphate (PLP); (f) about pH 10; and (g) about 55° C.

[0292] In some embodiments, the combination of reaction parameters comprises: (a) about 50 g/L of substrate compound (e.g. compound (2)); (b) about 0.5 g/L engineered polypeptide; (c) about 10% (v/v) dimethylsulfoxide (DMSO); (d) about 2 M isopropylamine (IPM); (e) about 1 g/L pyridoxal phosphate (PLP); (f) about pH 10.5; and (g) about 55° C.

[0293] Further exemplary reaction conditions include the assay conditions provided in the Examples. In carrying out the transamination reactions described herein, the engineered transaminase polypeptide may be added to the reaction mixture in the partially purified or purified enzyme, whole cells transformed with gene(s) encoding the enzyme, and/or as cell extracts and/or lysates of such cells. Whole cells transformed with gene(s) encoding the engineered transaminase enzyme or cell extracts, lysates thereof, and isolated enzymes may be employed in a variety of different forms, including solid (e.g., lyophilized, spray-dried, and the like) or semisolid (e.g., a crude paste). The cell extracts or cell lysates may be partially purified by precipitation (e.g., ammonium sulfate, polyethyleneimine, heat treatment or the like), followed by a desalting procedure (e.g., ultrafiltration, dialysis, and the like) prior to lyophilization. Any of the enzyme preparations may be stabilized by crosslinking using known crosslinking agents, such as, for example, glutaraldehyde, or immobilized to a solid phase material (e.g., resins, beads such as chitosan, Eupergit C, SEPABEADs, and the like).

[0294] In some embodiments of the transamination reactions described herein, the reaction is carried out under the suitable reaction conditions described herein, wherein the engineered

transaminase polypeptide is immobilized to a solid support. Solid supports useful for immobilizing the engineered transaminases for carrying out the transamination reactions include but are not limited to beads or resins comprising polymethacrylate with epoxide functional groups, polymethacrylate with amino epoxide functional groups, styrene/DVB copolymer or polymethacrylate with octadecyl functional groups.

[0295] Exemplary solid supports include, but are not limited to, chitosan beads, Eupergit C, and SEPABEADs (Mitsubishi), including the following different types of SEPABEAD: EC-EP, EC-HFA/S, EXA252, EXE119 and EXE120.

[0296] In some embodiments where the engineered polypeptide can be expressed in the form of a secreted polypeptide, the culture medium containing the secreted polypeptides can be used in the process herein.

[0297] In some embodiments, solid reactants (e.g., enzyme, salts, etc.) may be provided to the reaction in a variety of different forms, including powder (e.g., lyophilized, spray dried, and the like), solution, emulsion, suspension, and the like. The reactants can be readily lyophilized or spray dried using methods and equipment that are known to those having ordinary skill in the art. For example, the protein solution can be frozen at -80°C . in small aliquots, then added to a pre-chilled lyophilization chamber, followed by the application of a vacuum.

[0298] In some embodiments, the order of addition of reactants is not critical. The reactants may be added together at the same time to a solvent (e.g., monophasic solvent, biphasic aqueous co-solvent system, and the like), or alternatively, some of the reactants may be added separately, and some together at different time points. For example, the cofactor, transaminase, and transaminase substrate may be added first to the solvent. For improved mixing efficiency when an aqueous co-solvent system is used, the transaminase, and cofactor may be added and mixed into the aqueous phase first. The organic phase may then be added and mixed in, followed by addition of the transaminase substrate.

[0299] Alternatively, the transaminase substrate may be premixed in the organic phase, prior to addition to the aqueous phase.

[0300] In some embodiments, where the choice of the amino donor results in a carbonyl by-product that has a vapor pressure higher than water (e.g., a low boiling co-product such as a volatile organic carbonyl compound), the carbonyl by-product can be removed by sparging the reaction solution with a non-reactive gas or by applying a vacuum to lower the reaction pressure and removing the carbonyl by-product present in the gas phase. A non-reactive gas is any gas that does not react with the reaction components. Various non-reactive gases include nitrogen and noble gases (e.g., inert gases). In some embodiments, the non-reactive gas is nitrogen gas. In some embodiments, the amino donor used in the process is isopropylamine (IPM), which forms the carbonyl by-product acetone upon transfer of the amino group to the amino group acceptor. The acetone can be removed by sparging with nitrogen gas or applying a vacuum to the reaction solution and removing the acetone from the gas phase by an acetone trap, such as a condenser or other cold trap. Alternatively, the acetone can be removed by reduction to isopropanol using a ketoreductase.

[0301] In some embodiments of the process where the carbonyl by-product is removed, the corresponding amino group donor can be added during the transamination reaction to replenish the amino group donor and/or maintain the pH of the reaction. Replenishing the amino group donor also shifts the equilibrium towards product formation, thereby increasing the conversion of substrate to product. Thus, in some embodiments where the amino group donor is IPM and the acetone product is removed in situ, the process can further comprise a step of adding IPM to the reaction solution to replenish the amino group donor lost during the acetone removal and to maintain the pH of the reaction (e.g., at about 8.5 to about pH 11.5).

[0302] In some embodiments, it is also contemplated that the process comprising the biocatalytic conversion of amine acceptor substrate compounds to chiral amine product compounds using transaminase polypeptides of the present invention can further comprise steps of formation of

pharmaceutically acceptable salts or acids, pharmaceutically acceptable formulations, product work-up, extraction, isolation, purification, and/or crystallization, each of which can be carried out under a range of conditions.

[0303] In some embodiments, the processes using the engineered polypeptides disclosed herein can be carried out wherein the amino group donor is selected from isopropylamine, alanine, 3-aminobutyric acid, or methylbenzylamine. In some embodiments, the amino group donor is isopropylamine.

[0304] Methods, techniques, and protocols for extracting, isolating, forming a salt of, purifying, and/or crystallizing aminated product compounds or cyclized compounds from biocatalytic reaction mixtures produced by the above disclosed processes are known to the ordinary artisan and/or accessed through routine experimentation. Additionally, illustrative methods are provided in the Examples below.

[0305] Various features and embodiments of the invention are illustrated in the following representative examples, which are intended to be illustrative, and not limiting.

EXPERIMENTAL

[0306] Various features and embodiments of the invention are illustrated in the following representative examples, which are intended to be illustrative, and not limiting.

[0307] In the experimental invention below, the following abbreviations apply: ppm (parts per million); M (molar); mM (millimolar), μ M and μ M (micromolar); nM (nanomolar); mol (moles); gm and g (gram); mg (milligrams); ug and gg (micrograms); L and l (liter); ml and mL (milliliter); cm (centimeters); mm (millimeters); um and μ m (micrometers); sec. (seconds); min(s) (minute(s)); h(s) and hr(s) (hour(s)); U (units); MW (molecular weight); rpm (rotations per minute); ° C. (degrees Centigrade); RT (room temperature); CDS (coding sequence); DNA (deoxyribonucleic acid); RNA (ribonucleic acid); aa (amino acid); TB (Terrific Broth; 12 g/L bacto-tryptone, 24 g/L yeast extract, 4 mL/L glycerol, 65 mM potassium phosphate, pH 7.0, 1 mM MgSO₄); LB (Luria broth); CAM (chloramphenicol); PMBS (polymyxin B sulfate); IPTG (isopropyl thiogalactoside); ATA (omega-transaminase); TFA (trifluoroacetic acid); TEoA (triethanolamine); borate (sodium tetraborate decahydrate); acetonitrile (MeCN); dimethylsulfoxide (DMSO); HPLC (high performance liquid chromatography); FIOP (fold improvement over positive control); HTP (high throughput); MWD (multiple wavelength detector); UV (ultraviolet); Codexis (Codexis, Inc., Redwood City, CA); Sigma-Aldrich (Sigma-Aldrich, St. Louis, MO); Millipore (Millipore, Corp., Billerica MA); Difco (Difco Laboratories, BD Diagnostic Systems, Detroit, MI); Daicel (Daicel, West Chester, PA); Genetix (Genetix USA, Inc., Beaverton, OR); Molecular Devices (Molecular Devices, LLC, Sunnyvale, CA); Applied Biosystems (Applied Biosystems, part of Life Technologies, Corp., Grand Island, NY), Agilent (Agilent Technologies, Inc., Santa Clara, CA); Thermo Scientific (part of Thermo Fisher Scientific, Waltham, MA); (Infors; Infors-HT, Bottmingen/Basel, Switzerland); Corning (Corning, Inc., Palo Alto, CA); and Bio-Rad (Bio-Rad Laboratories, Hercules, CA); Microfluidics (Microfluidics Corp., Newton, MA); Waters (Waters Corp., Milford, MA).

Example 1

E. coli Expression Hosts Containing Recombinant ATA Genes

[0308] The initial ATA enzyme used to produce the variant enzymes of the present invention was obtained from the *Vibrio fluvialis* transaminase variants included in the Codex® ATA panel plate (Codexis). This ATA panel plate comprises a collection of engineered ATA polypeptides that have improved properties, as compared to the wild-type ATAs. The ATA-encoding genes were cloned into the expression vector pCK110900 (See, U.S. Pat. No. 7,629,157 and US Pat. Appln. Publn. 2016/0244787, both of which are hereby incorporated by reference in their entireties and for all purposes), operatively linked to the lac promoter under control of the lacI repressor. The expression vector also contains the P15a origin of replication and a chloramphenicol resistance gene. The resulting plasmids were transformed into *E. coli* W3110, using standard methods known in the art.

The transformants were isolated by subjecting the cells to chloramphenicol selection, as known in the art (See e.g., U.S. Pat. No. 8,383,346 and WO2010/144103, each of which is incorporated herein by reference in its entirety and for all purposes).

Example 2

Preparation of HTP ATA-Containing Wet Cell Pellets

[0309] *E. coli* cells containing recombinant ATA-encoding genes from monoclonal colonies were inoculated into 180 μ l LB containing 1% glucose and 30 gg/mL chloramphenicol into the wells of 96 well shallow-well microtiter plates. The plates were sealed with O.sub.2-permeable seals and cultures were grown overnight at 30° C., 200 rpm and 85% humidity. Then, 10 μ l of each of the cell cultures were transferred into the wells of 96 well deep-well plates containing 390 mL TB and 30 gg/mL CAM. The deep-well plates were sealed with O.sub.2-permeable seals and incubated at 30° C., 250 rpm and 85% humidity until OD.sub.600 0.6-0.8 was reached. The cell cultures were then induced by IPTG to a final concentration of 1 mM and incubated overnight under the same conditions as originally used. The cells were then pelleted using centrifugation at 4000 rpm for 10 min. The supernatants were discarded and the pellets frozen at -80° C. prior to lysis.

Example 3

Preparation of HTP ATA-Containing Cell Lysates

[0310] First, 200 μ l lysis buffer containing 50 mM triethanolamine-HCl buffer, pH 7.5, 1 mg/mL PLP, 1 mg/mL lysozyme, and 0.5 mg/mL PMBS was added to the cell pellet in each well produced as described in EXAMPLE 2. The cells were lysed at room temperature for 2 hours with shaking on a bench top shaker. The plate was then centrifuged for 15 min at 4000 rpm and 4° C. The clear supernatants were then used in biocatalytic reactions to determine their activity levels.

Example 4

Preparation and Assay of Lyophilized Lysates from Shake Flask (SF) Cultures

[0311] Selected HTP cultures grown as described above were plated onto LB agar plates with 1% glucose and 30 gg/ml CAM, and grown overnight at 37° C. A single colony from each culture was transferred to 50 ml of LB with 1% glucose and 30 gg/ml CAM. The cultures were grown for 18 h at 30° C., 250 rpm, and subcultured approximately 1:20 into 250 ml of TB containing 30 gg/ml CAM, to a final OD.sub.600 of 0.2. The cultures were grown for approximately 3 hours at 30° C., 250 rpm, to an OD.sub.600 between 0.6-0.8, and induced with 1 mM IPTG. The cultures were then grown for 20 h at 30° C., 250 rpm. The cultures were centrifuged 4000 rpm \times 20 min and the supernatants were discarded. The cells were resuspended in 13 ml of 50 mM triethanolamine-HCl pH 7.5 with 1 g/L PLP, and lysed using a One-Shot® processor system (Constant Systems) at 33,000 psi. The lysates were pelleted (10,000 rpm \times 60 min) and the supernatants were frozen and lyophilized to generate shake flake (SF) enzymes.

[0312] The activity of selected shake flask ATA variants was evaluated based on the efficiency of the variants to transaminate compound (2) with isopropylamine as the amine donor to the product, compound (1) and acetone. Reactions using shake flask powders were carried out in 96-well shallow well microtiter plates. In this assay, 0.1 mL reaction solutions consisting of 2.5-75 g/L compound (2), 0-50 g/L shake flask powder, 25 mM borate buffer, 0.5 g/L PLP, and 1 M isopropylamine were prepared adjusted to a final pH of 9. The plates were heat sealed and incubated at 40-65° C. and agitated at 600 RPM in an Infors Thermotron® shaker for 0-72 h. The plate was removed and quenched 1:1 (100 μ L/well) with 10% formic acid in MeCN then diluted 1:10 to 1:100 in 10% acetic acid in MeCN. The samples were analyzed by HPLC using an Astec Chirobiotic™ T column (4.6 \times 100 mm, 5 gm particle size) with a 2.5 mL/min flow rate using an isocratic mobile phase consisting of methanol with 0.2% acetic acid and 0.1% triethylamine. The injection volume was 10 μ L. The total run time was 1.8 min; the ketone eluted at 0.54 min, the (R)-diastereomer at 1.1 min and the (S)-diastereomer at 1.4 min (UV monitoring at 245 nm).

Example 5

Evaluation of the Codex® Transaminase Kit for the Transamination of Compound (2)

[0313] In order to find an ATA with initial activity on compound (2), enzymes from the Codex® Transaminase Kit (Codexis) were screened against compound (2). The screening was performed essentially as described in Example 4. The reaction concentrations were: 2.5 g/L compound (2), 50 g/L ATA, 1 M IPM, 0.5 g/L PLP, 25 mM borate buffer, at 40° C., for 22 h. The only active variant from the ATA Kit was ATA-217 giving about 5% conversion under these conditions. However, it had the opposite of the desired selectivity making almost exclusively the undesired (R)-amine diastereomer.

Example 6

Improvements in the Transamination Activity of Compound (2) Compared to SEQ ID NO: 4 in High Throughput Screening

[0314] SEQ ID NO: 4 was selected as the initial backbone, based on the results described in Example 5. Libraries of engineered genes were produced using well-established techniques (e.g., saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP as described in Example 2, and the soluble lysate was generated as described in Example 3.

[0315] HTP reactions were carried out in 96 well shallow-well plates containing 100 µL of 2.5 g/L compound (2), 1 M IPM, 0.5 g/L PLP, 25 mM borate buffer and 50 v/v % ATA lysate, pH 9. The HTP plates were heat sealed and incubated in a Thermotron® shaker at 40° C., 600 rpm, for overnight (approx. 21 hours). The reactions were then quenched with 100 µL/well 10% acetic acid in MeCN and mixed for 10 minutes using a bench top shaker. The plates were then centrifuged at 4000 rpm for 5 minutes and injected onto an HPLC for analysis using the parameters as described in Example 4.

[0316] The activity relative to SEQ ID NO:4 (FIOP) was calculated as the sum of the peak areas of the products produced by the variant over the sum of the peak areas of the products produced by SEQ ID NO: 4. The results are shown in Table 6.1.

TABLE-US-00003 TABLE 6.1 Activity of ATA Variants Relative to SEQ ID NO: 4 Transamination Amino Acid Product Peak SEQ ID Differences Area (FIOP).sup.1 Var. NO: (Relative to (Relative to (S)- (R)- No: (nt/aa) SEQ ID NO: 4) SEQ ID NO: 4) stereoselectivity.sup.2

stereoselectivity.sup.3 1 15/16 F85S +++++ 2 5/6 H319V + 3 7/8 H319L + + 4 11/12 L57T/H319E ++ 5 13/14 F85G +++++ 6 17/18 F85V ++ ++ 7 9/10 L417G + 8 S86G +++ + 9 P416G + + 10 S86A + + 11 R203H/P416S + + 12 P416S + + 13 P416Y + + 14 R203H/P416G + + 15 V422R + + 16 Y17W + + 17 S157T + + 18 S157C + + 19 I314R + + 20 E316K + + 21 Y17V/I259T/E347K + + 22 A158S + + 23 A404T + + 24 A404V + + 25 V421I + + 26 E316M + + 27 I375L/P416F + + 28 V422A + + 29 S420G + + 30 T20R + + 31 S157G + + 32 L417P + + 33 T161G + + 34 P416C + + 35 V422G + + 36 E316S + + 37 L417M + + 38 H319G + + 39 S157P + + 40 C424R + + 41 R415V + + 42 L57I + + 43 E316L + + 44 H45R/R48L/R78C/ + + G81V/Q92H 45 C424A + + .sup.1Levels of increased activity were determined relative to the reference

polypeptide of SEQ ID NO: 4 and defined as follows:: “+” > than 1.2-fold but less than 2.5-fold increase; “++” > than 2.5-fold but less than 5-fold; “+++” > than 5-fold but less than 10-fold.

.sup.2Levels of S-stereoselectivity were defined as: “+” > than 10% d.e. but less than 50% d.e.; “++” > than 50% d.e. but less than 75% d.e.; “+++” > than 75% d.e. but less than 95% d.e.;

“++++” > than 95% d.e. but less than 100% d.e; “+++++” equal to 100% d.e. .sup.3R-stereoselectivity was defined as “+”

Example 7

Improvements in the Transamination Activity of Compound (2) Compared to SEQ ID NO: 18 in High Throughput Screening

[0317] SEQ ID NO: 18 was selected as the next parent enzyme, based on the results described in Example 6. Libraries of engineered genes were produced using well-established techniques (e.g., saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP as described in Example 2, and the

soluble lysate was generated as described in Example 3. HTP reactions were carried out as described in Example 6.

[0318] The activity relative to SEQ ID NO: 18 (FIOP) was calculated as the sum of the peak areas of the products produced by the variant over the sum of the peak areas of the products produced by SEQ ID NO: 18. The results are shown in Table 7.1

TABLE-US-00004 TABLE 7.1 Activity of ATA Variants Relative to SEQ ID NO: 18

Transamination Amino Acid Product Peak SEQ ID Differences Area (FIOP).sup.1 Var. NO:

(Relative to (Relative to (S)- (R)- No: (nt/aa) SEQ ID NO: 18) SEQ ID NO: 18)

stereoselectivity.sup.2 stereoselectivity.sup.3 46 V85F/S86G +++ + 47 Y17V/S86G/S157T/ +++ +
E316M/V422R 48 Y17V/S86G/S157G/ ++ + E316M/V422R 49 S86G/S157T/P416G ++ + 50
E316Y + + 51 23/24 S86A/R203H/E316M/ ++ + + + V422G 52 I314R/E316M ++ ++ 53 19/20
S86A/R203H/E316M/ ++ + + + P416G 54 Q5H/Y17V/S86G/S157G/ ++ +
I259T/I314R/E316S/V422R 55 E316M + + + + + 56 Y17V/S86G/S157G/R203H/ + +
I259T/I314R/E316L/V422R 57 21/22 S86A/E316S/P416C + + + + + 58 P416A + + + + + 59
G227F/V422G + + + + + 60 S157T/E316M + + + + 61 Y17V/S86A/S157G/ + + + + + E316M/V422A
62 25/26 V422C + + + + + 63 I314R + + + + 64 E316F + + + + + 65 E316C + + + + + 66
S86A/S157G/I314R/ + + + + + E316S/E347K/P416C/ V422A 67 E195K/E316S + + + + + 68 E316S +
+ + + + + 69 Y17W/S86A/S157T/ + + I314R/E316M/E347K/ V422R 70 E316G + + + + + 71
E207K/E316T + + + + + 72 Q139H/F221I/A310E/ + + + + + E316G 73 E316W + + + + + 74 L417M +
+ + + + + 75 P416G + + + + + 76 P416S + + + + + 77 A280E/L381V/P416G/ + + + + + L445R .sup.1Levels
of increased activity or selectivity were determined relative to the reference polypeptide of SEQ ID
NO: 18 and defined as follows: “+” than 1.2-fold but less than 2.5-fold increase; “++” > than 2.5-
fold but less than 5-fold; “+++” > than 5-fold but less than 10-fold. .sup.2Levels of S-
stereoselectivity were defined as: “+” > than 10% d.e. but less than 50% d.e.; “++” > than 50% d.e.
but less than 75% d.e.; “+++” > than 75% d.e. but less than 95% d.e.; “++++” > than 95% d.e. but
less than 100% d.e; “+++++” equal to 100% d.e. .sup.3R-stereoselectivity was defined as “+”

Example 8

Improvements in the Transamination Activity of Compound (2) Compared to SEQ ID NO: 20 in High Throughput Screening

[0319] SEQ ID NO: 20 was selected as the next parent enzyme, based on the results described in Example 7. Libraries of engineered genes were produced using well-established techniques (e.g., saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP as described in Example 2, and the soluble lysate was generated as described Example 3. HTP reactions were carried out as described in Example 6, except that the substrate compound (2) concentration was increased to 10 g/L.

[0320] The activity relative to SEQ ID NO:20 (FIOP) was calculated as the sum of the peak areas of the products produced by the variant over the sum of the peak areas of the products produced by SEQ ID NO: 20. The results are shown in Table 8.1

TABLE-US-00005 TABLE 8.1 Activity of ATA Variants Relative to SEQ ID NO: 20

Transamination SEQ Amino Acid Product Peak ID Differences Area (FIOP).sup.1 Var. NO:

(Relative to (Relative to (S)- (R)- No: (nt/aa) SEQ ID NO: 20) SEQ ID NO: 20)

stereoselectivity.sup.2 stereoselectivity.sup.3 78 31/32 E195K/G227F/I314R/M +++ + + + +
316W/G416A 79 33/34 T322S +++ + + + + 80 Y17V/Q139H/I314R/G4 +++ + + +
16A/L417M/V422A 81 Y17V/E195K/M316W/ ++ + + + + + C414V/G416S/V422C 82
P164Q/E207K/G213S/G ++ ++ 227F/A310E/M316F/G4 16A 83 I314R/M316W/G416C/ ++
+ V422C 84 H178L/I314A/M316W/ ++ + + + + + G416S 85 E194D/G213C/A310E/I ++
+ + + + + 314R/M316W/V422C 86 27/28 Y17V/I314R/M316W/C ++ + + + + +
414V/G416S/L417M/V4 22A 87 Q139H/E195K/G227F/I ++ + + + + + 314R/M316W/G416C
88 A310E/I314R/M316S/C ++ + + + + + 414V/G416A/L417M/V 422A 89
E195K/M316W/G416S/ ++ L417M/V422A 90 29/30 I314R/M316W/C414V/ ++ + + + + +

G416A/V422A 91 I314A/M316W/G416A ++ ++++ 92 M316W ++ ++++ 93 E195K/G227F/C414V/G ++ ++++ 416C 94 G227F/G416C ++ ++++ 95 E195K/G227F/G416S/L ++ ++++ 417M 96 M316S ++ ++++ 97 H83G ++ ++++ 98 M316E ++ ++++ 99 M226E ++ ++++ 100 Q139H/A202G/E207K/ ++ ++++ G227F/1314R/M316F/G 416S/V422C 101 C271R/A310E/M316W/ ++ +++ G416S/V422A 102 Q139H/M316S/C414V/ ++ +++ G416S/L417M/V422A 103 35/36 R146V/S157G ++ +++ 104 L57C ++ +++ 105 E195K/G227F/1314R/G ++ +++ 416A/L417M 106 G227V ++ +++ 107 Q139H/G227F/A310E/C ++ +++ 414V/G416C/L417M 108 E195K/G227F/1314R/V ++ +++ 422C 109 I282L ++ +++ 110 E195K/G227F/C414V/G ++ +++ 416A/V422G 111 R146K/S157G ++ +++ 112 M226A ++ +++ 113 E207K/T208N/F220Y/A ++ +++ 241V/F291S/1314R/M31 6S 114 Y17V/I314R/M316S/C4 ++ +++ 14V/G416C 115 A86K ++ +++ 116 Q139H/E207K/G227F/I ++ +++ 314R/C414V/G416C/V4 22G 117 V31A ++ +++ 118 V31Q ++ +++ 119 E190K/E195K/E207K/G ++ + 213S/P244A/I314R/C41 4V/G416S/L417M/V422 G 120 T20L ++ + 121 M316H + ++++ 122 M316T + ++++ 123 M316G + ++++ 124 M316Y + ++++ 125 M316Q + ++++ 126 M316R + ++++ 127 M316K + ++++ 128 S54C + ++++ 129 V31G + ++++ 130 M316A + +++ 131 N121G + +++ 132 H319V + +++ 133 N121A + +++ 134 N121S + +++ 135 N121C + +++ 136 N121V + +++ 137 N121H + +++ 138 I259V + +++ 139 N166T + +++ 140 M226T ++ + 141 N121L + +++ 142 V31D + +++ 143 L57A + +++ 144 V31C + +++ 145 V31E + +++ 146 V258L + +++ 147 L417S + ++ 148 A228G + ++ 149 A86G/E194K + + 150 A86G + + 151 T20V + + 152 T20K 1 153 G320W + +

.sup.1Levels of increased activity or selectivity were determined relative to the reference polypeptide of SEQ ID NO: 20 and defined as follows: "+" > than 1.2-fold but less than 2.5-fold increase; "++" > than 2.5-fold but less than 5-fold; "+++" > than 5-fold but less than 10-fold. .sup.2Levels of S-stereoselectivity were defined as: "+" > than 10% d.e. but less than 50% d.e. ; "++" > than 50% d.e. but less than 75% d.e. ; "+++" > than 75% d.e. but less than 95% d.e. ; "++++" > than 95% d.e. but less than 100% d.e; "+++++" equal to 100% d.e. .sup.3R-stereoselectivity was defined as "+"

Example 9

Improvements in the Transamination Activity of Compound (2) Compared to SEQ ID NO: 30 in High Throughput Screening

[0321] SEQ ID NO: 30 was selected as the next parent enzyme, based on the results described in Example 8. Libraries of engineered genes were produced using well-established techniques (e.g., saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP as described in Example 2, and the soluble lysate was generated as described in Example 3. HTP reactions were carried out as described in Example 8.

[0322] The activity relative to SEQ ID NO:30 (FIOP) was calculated as the sum of the peak areas of the products produced by the variant over the sum of the peak areas of the products produced by SEQ ID NO: 30. The results are shown in Table 9.1

TABLE-US-00006 TABLE 9.1 Activity of ATA Variants Relative to SEQ ID NO: 30

Transamination Amino Acid Product	Peak SEQ ID Differences	Area (FIOP).sup.1	Var. NO: (Relative to (Relative to (S)- No: (nt/aa) SEQ ID NO: 30) SEQ ID NO: 30) stereoselectivity.sup.2					
154 65/66 N53C/R146H/F163K/G395D	+++ ++++++	155 57/58 N53C/R146H/G395D/A432V	+++ ++++++					
156 53/54 P23S/N53C/R146H/F163K	+++ ++++++	157 59/60 P23S/N53C/R146H/S159T/	+++ ++++++					
G395D/A432V	158 63/64 N53C/R146H/G162D/G395D	+++ ++++++	159 55/56 N53C/R146H/G162D/A432V	++ ++++++				
160 51/52 P23S/N53C/R146H/G162D/	++ ++++++	G395D/A432V	161 61/62 D21P/P23S/N53C/G395D/	++ ++++++				
A432V	162 D21P/N53C/F163K/G395D/	++ ++++++	A432V	163 N53C/G162D/G395D	++ ++++++			
164 N53C/G162D/G395D/A432V	++ ++++++	165 N53C/F163K/G395D/A432V	++ ++++++	166 R146H	++ ++++++			
167 N53C/R146H/S159T/G395D	++ ++++++	168 D21P/P23S/N53C/G162D/	++ ++++++	G395D/A432V	169 D21P/P23S/N53C/R146H/	++ ++++++	G162D/G395D/A432V	170

N53C/G395D ++ +++++ 171 39/40 A86K ++ +++++ 172 47/48 L57A/A86K ++ +++++ 173 45/46 L57A/A86K/N121A/I332T ++ +++++ 174 P23S/N53C/R146H/S159T/ ++ +++++ G395D/E431V 175 43/44 L57A/A86K/S157G/I259V ++ +++++ 176 49/50 N121A ++ +++++ 177 41/42 N121C/I332T ++ +++++ 178 G227F ++ +++ 179 P23S/R146H/F163K/A432V ++ +++++ 180 37/38 L57A/A86K/N121C/R146K/ ++ +++++ S157G 181 L57C/A86K/N121A/S157G/ ++ +++++ M226T/I259V/I282L/I332T 182 N121A/L417M/A422V ++ +++++ 183 N121C/M226A/I332T ++ +++++ 184 L57A/R146V/M226T/I259V/ ++ +++++ I282L 185 G227F/L417V ++ +++ 186 A422C ++ +++++ 187 D21P/P23S/N53C/R146H/ + G162D/G395D 188 L417F + +++++ 189 N53C/G162D/F163K/G395D/ + +++++ A432V 190 N121A/A416S/L417M/A422C + +++++ 191 A416S/A422C + +++++ 192 I282L/I332T + +++++ 193 N121A/L417M/A422C + +++++ 194 M226A/I282L/I332T + +++++ 195 G227V + +++++ 196 A86K/S157G + +++++ 197 N121C/R146K/S157G/M226A/ + +++++ I332T 198 L57A/I259V + +++++ 199 L57A/R146K/M226T/I259V + +++++ 200 L57C/A86K/I332T + +++++ 201 W316S/L417M/A422V + +++ 202 N121A/T208N/W316S/L417M/ + +++++ A422V 203 G227A + +++++ 204 I282L + +++++ 205 I259V + +++++ 206 D21P/P23S/G395D + +++++ 207 A86K/M226T + +++++ 208 N121A/W316F/A422C + +++++ 209 L57C/N121A/L417M/A422V + +++++ 210 L57A/A86K/S157G/I282L/ + +++++ I332T 211 I332T + +++++ 212 R146H/G162D/G395D/A432V + +++++ 213 G162D/A432V + +++++ 214 P23S/G162D + +++++ 215 D21P + +++++ 216 S157G + +++++ 217 D21P/N53C/S159T/G395D + +++++ 218 A416S + +++++ 219 G395D/A432V + +++++ .sup.1Levels of increased activity or selectivity were determined relative to the reference polypeptide of SEQ ID NO: 30 and defined as follows: “+” > than 1.2-fold but less than 2.5-fold increase; “++” > than 2.5-fold but less than 5-fold; “+++” > than 5-fold but less than 10-fold.

.sup.2Levels of S-stereoselectivity were defined as: “+” > than 10% d.e. but less than 50% d.e.; “++” > than 50% d.e. but less than 75% d.e.; “+++” > than 75% d.e. but less than 95% d.e.; “++++” > than 95% d.e. but less than 100% d.e; “+++++” equal to 100% d.e. .sup.3R-

stereoselectivity was defined as “+”
Example 10

Improvements in the Transamination Activity of Compound (2) Compared to SEQ ID NO: 66 in High Throughput Screening

[0323] SEQ ID NO: 66 was selected as the next parent enzyme, based on the results described in Example 9. Libraries of engineered genes were produced using well-established techniques (e.g., saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP as described in Example 2, and the soluble lysate was generated as described in Example 3. HTP reactions were carried out as described in Example 8, with the following differences: the reaction was performed at 50° C., and the 1:1 quenched reaction was diluted 10-fold into 10% acetic acid in MeCN before HPLC analysis.

[0324] The activity relative to SEQ ID NO: 66 (FIOP) was calculated as the sum of the peak areas of the products produced by the variant over the sum of the peak areas of the products produced by SEQ ID NO: 66. The results are shown in Table 10.1

TABLE-US-00007 TABLE 10.1 Activity of ATA Variants Relative to SEQ ID NO: 66

Transamination Amino Acid Product Peak SEQ ID Differences Area (FIOP).sup.1 Var. NO:
(Relative to (Relative to (S)- No: (nt/aa) SEQ ID NO: 66) SEQ ID NO: 66) stereoselectivity.sup.2
220 71/72 I281S + +++++ 221 73/74 R141L + +++++ 222 67/68 D107V + +++++ 223 69/70 P293A + +++++ 224 A453T + +++++ 225 A432V + +++++ 226 I245L + +++++ 227 A432E + +++++ 228 A453G + +++++ 229 A432D + +++++ 230 D107G + +++++ 231 T161S + +++++ 232 P174Q + +++++ .sup.1Levels of increased activity or selectivity were determined relative to the reference polypeptide of SEQ ID NO: 66 and defined as follows: “+” > than 1.2-fold but less than 2.5-fold increase .sup.2Levels of S-stereoselectivity were defined as: “+” > than 10% d.e. but less than 50% d.e.; “++” > than 50% d.e. but less than 75% d.e.; “+++” > than 75% d.e. but less than

95% d.e.; “++++” > than 95% d.e. but less than 100% d.e: “+++++” equal to 100% d.e.

Example 11

Improvements in the Transamination Activity of Compound (2) Compared to SEQ ID NO: 70 in High Throughput Screening

[0325] SEQ ID NO: 70 was selected as the next parent enzyme, based on the results described in Example 10. Libraries of engineered genes were produced using well-established techniques (e.g., saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP as described in Example 2, and the soluble lysate was generated as described in Example 3. HTP reactions were carried out as described in Example 10, except the ATA lysate was diluted 4-fold in 50 mM TEOA+1 g/L PLP, pH 7.5 before it was added to the reaction mixture.

[0326] The activity relative to SEQ ID NO:70 (FIOP) was calculated as the sum of the peak areas of the products produced by the variant over the sum of the peak areas of the products produced by SEQ ID NO: 70. The results are shown in Table 11.1

TABLE-US-00008 TABLE 11.1 Activity of ATA Variants Relative to SEQ ID NO: 70

Transamination Amino Acid Product Peak SEQ ID Differences Area (FIOP).sup.1 Var. NO:

(Relative to (Relative to (S)- No: (nt/aa) SEQ ID NO: 70) SEQ ID NO: 70) stereoselectivity.sup.2
233 75/76 Y17G/V31T + +++++ 234 85/86 T408V + +++++ 235 L57G + +++++ 236 79/80 L246V
+ +++++ 237 77/78 P233V + +++++ 238 83/84 T408A + +++++ 239 S284A + +++++ 240 A86G +
+++ 241 Y17M + +++++ 242 81/82 R359G + +++++ 243 Y17A + +++++ 244 I259V + +++++ 245
A222T + +++++ 246 Y17I + +++++ 247 F19Y + +++++ 248 Y17S + +++++ 249 Q5E + +++++ 250
L57A + +++++ 251 V31L + +++++ 252 R247A/C271G/A293P + +++++ 253 T161S + +++++ 254
T408L + +++++ 255 N2I + +++++ 256 N121C + +++++ 257 Y17H + +++++ 258 I282L + +++++
259 K443I + +++++ 260 Q419D + +++++ 261 A404V + +++++ 262 E207I + +++++ 263 A86E +
+++++ 264 T408E + +++++ 265 P4D + +++++ 266 P4A + +++++ 267 V31C + +++++ 268
E190K/F291C + +++++ 269 K443A + +++++ 270 T309R + +++++ 271 A155C + +++++ 272
Y17V + +++++ 273 Q5H/G38C/A56T/A346G + +++++ 274 V31F + +++++ 275 P4C + +++++ 276
Q419S + +++++ 277 A86N + +++++ 278 P327E + +++++ 279 F79L + +++++ 280 H362Q +
+++++ 281 A86C + +++++ 282 T161S/P174Q/A432V + +++++ 283 P164V + +++++ 284 T408S +
+++++ 285 E207K/A235P/Q419S + +++++ 286 P4R + +++++ 287 A422C + +++++ 288 T13E +
+++++ 289 S420T + +++++ 290 E195Q/T408E + +++++ 291 E195K + +++++ 292
H203R/E207K/G219R + +++++ 293 A214T/A228G/I251T + +++++ 294 A346N + +++++ 295
L205M/Q419S + +++++ 296 I375R + +++++ 297 K163E/S167I + +++++ 298 E223G + +++++
299 M226D/A228C/G230R/ + +++++ V231C/I311R 300 A228G + +++++ 301 T74M + +++++

.sup.1Levels of increased activity or selectivity were determined relative to the reference polypeptide of SEQ ID NO: 70 and defined as follows: “+” > than 1.2-fold but less than 2.5-fold increase. .sup.2Levels of S-stereoselectivity were defined as: “+” > than 10% d.e. but less than 50% d.e.; “++” > than 50% d.e. but less than 75% d.e.; “+++” > than 75% d.e. but less than 95% d.e.; “++++” > than 95% d.e. but less than 100% d.e; “+++++” equal to 100% d.e.

Example 12

Improvements in the Transamination Activity of Compound (2) Compared to SEQ ID NO: 48 in High Throughput Screening

[0327] SEQ ID NO: 48 was selected as the next parent enzyme, based on the results described in Example 9. Libraries of engineered genes were produced using well-established techniques (e.g., saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP as described in Example 2, and the soluble lysate was generated as described in Example 3. HTP reactions were carried out as described in Example 11.

[0328] The activity relative to SEQ ID NO:48 (FIOP) was calculated as the sum of the peak areas of the products produced by the variant over the sum of the peak areas of the products produced by

SEQ ID NO: 48. The results are shown in Table 12.1

TABLE-US-00009 TABLE 12.1 Activity of ATA Variants Relative to SEQ ID NO: 48

Transamination Amino Acid Product Peak SEQ ID Differences Area (FIOP).sup.1 Var. NO:

(Relative to (Relative to (S)- No: (nt/aa) SEQ ID NO: 48) SEQ ID NO: 48) stereoselectivity.sup.2

302 87/88 N53C/R146H/P293A +++ +++++ 303 91/92 P23S/N53C/R146H/P293A/ +++ +++++

A432V 304 95/96 N53C/R146H/F163K/P293A/ +++ +++++ G395D/A432V 305 99/100

P23S/N53C/R146H/F163K/ +++ +++++ P293A/A432V 306 97/98 P23S/N53C/R146H/F163K

+++ +++++ 307 89/90 N53C/R146H/F163K/G395D/ +++ +++++ A432V 308 93/94

N53C/F163K/P293A/G395D +++ +++++ 309 P23S/N53C/F163K/P293A/ +++ +++++ G395D 310

N53C/P293A +++ +++++ 311 N53C/F163K/P293A/G395D/ +++ +++++ A432V 312

N53C/P293A/G395D +++ +++++ 313 N53C/F163K/P293A/A432V +++ +++++ 314 N53C/F163K

+++ +++++ 315 N53C/R146H/F163K/A432V ++ +++++ 316 D21P/N53C/R146H/F163K/ ++

+++++ P293A/G395D/A432V 317 N53C/F163K/A432V ++ +++++ 318 P23S/N53C/F163K ++

+++++ 319 P23S/N53C/F163K/P293A/ ++ +++++ A432V 320 N53C/A432V ++ +++++ 321

P23S/N53C/P293A/A432V ++ +++++ 322 P23S/N53C/F163K/G395D ++ +++++ 323

P23S/N53C/P293A/G395D/ ++ +++++ A432V 324 D21P/P23S/N53C/F163K/ ++ +++++ P293A

325 D21P/N53C/R146H ++ +++++ 326 D21P/P23S/N53C/F163K/ ++ +++++ P293A/G395D 327

N53C/G395D ++ +++++ 328 P23S/N53C/P293A ++ +++++ 329 D21P/N53C/F163K/P293A/ ++

+++++ G395D 330 D21P/N53C/P293A/A432V ++ +++++ 331 P23S/R146H/P293A ++ +++++

332 R146H + +++++ 333 P293A + +++++ .sup.1Levels of increased activity or selectivity were

determined relative to the reference polypeptide of SEQ ID NO: 48 and defined as follows: “+” >

than 1.2-fold but less than 2.5-fold increase; “++” > than 2.5-fold but less than 5-fold; “+++” >

than 5-fold but less than 10-fold. .sup.2Levels of S-stereoselectivity were defined as: “+” > than

10% d.e. but less than 50% d.e.; “++” > than 50% d.e. but less than 75% d.e.; “+++” > than 75%

d.e. but less than 95% d.e.; “++++” > than 95% d.e. but less than 100% d.e.; “+++++” equal to

100% d.e.

Example 13

Improvements in the Transamination Activity of Compound (2) Compared to SEQ ID NO: 88 in

High Throughput Screening

[0329] SEQ ID NO: 88 was selected as the next parent enzyme, based on the results described in

Example 12. Libraries of engineered genes were produced using well-established techniques (e.g.,

saturation mutagenesis and recombination of previously identified beneficial mutations). The

polypeptides encoded by each gene were produced in HTP as described in Example 2, and the

soluble lysate was generated as described in Example 3. HTP reactions were carried out as

described in Example 12, except the ATA lysate was diluted 16-fold in 50 mM TEoA+1 g/L PLP,

pH 7.5 before it was added to the reaction mixture and screened at 50° C. In addition, mutants were

also screened with 4-fold diluted lysate at 58° C.

[0330] The activity relative to SEQ ID NO: 88 (FIOP) was calculated as the sum of the peak areas

of the products produced by the variant over the sum of the peak areas of the products produced by

SEQ ID NO: 88. The results are shown in Table 13.1 for the variants screened at 50° C. and Table

13.2 for the variants screened at 58° C.

TABLE-US-00010 TABLE 13.1 Activity of ATA Variants Relative to SEQ ID NO: 88 (Screened at

50° C.) Transamination Amino Acid Product Peak SEQ Differences Area (FIOP).sup.1 Var. ID NO:

(Relative to (Relative to (S)- No: (nt/aa) SEQ ID NO: 88) SEQ ID NO: 88) stereoselectivity.sup.2

334 109/110 S284A ++ +++++ 335 F163K + +++++ 336 F163K/P164G + +++++ 337 101/102

F163M + +++++ 338 107/108 S157A + +++++ 339 103/104 S157C + +++ 340 105/106 I332V +

+++++ .sup.1Levels of increased activity or selectivity were determined relative to the reference

polypeptide of SEQ ID NO: 88 and defined as follows: “+” > than 1.2-fold but less than 2.5-fold

increase; “++” > than 2.5-fold but less than 5-fold. .sup.2Levels of S-stereoselectivity were defined

as: “+” > than 10% d.e. but less than 50% d.e.; “++” > than 50% d.e. but less than 75% d.e.; “+++”

> than 75% d.e. but less than 95% d.e.; “++++” > than 95% d.e. but less than 100% d.e; “+++++” equal to 100% d.e.

TABLE-US-00011 TABLE 13.2 Activity of ATA Variants Relative to SEQ ID NO: 88 (Screened at 58° C.) Transamination Amino Acid Product Peak SEQ ID Differences Area (FIOP).sup.1 V ar. NO: (Relative to (Relative to No: (nt/aa) SEQ ID NO: 88) SEQ ID NO: 88) 341 S284A ++ 342 V31M ++ 343 G191S/T208P/K211R/S286A ++ 344 C260T ++ 345 C424A + 346 E195D/E207R/T208N/S286G + 347 T208N/V294T + 348 S286G + 349 V294A + 350 A313L + 351 E192G/V294A + 352 V294S + 353 A313V + 354 K86A + 355 T208N/V294S + 356 A149S + 357 G395D/N396K + .sup.1Levels of increased activity or selectivity were determined relative to the reference polypeptide of SEQ ID NO: 88 and defined as follows: “+” > than 1.2-fold but less than 2.5-fold increase; “++” > than 2.5-fold but less than 5-fold.

Example 14

Improvements in the Transamination Activity of Compound (2) Compared to SEQ ID NO: 110 in High Throughput Screening

[0331] SEQ ID NO: 110 was selected as the next parent enzyme, based on the results described in Example 13. Libraries of engineered genes were produced using well-established techniques (e.g., saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP as described in Example 2, and the soluble lysate was generated as described in Example 3. HTP reactions were carried out as described in Example 13, except: the lysate was diluted 50-fold in 50 mM triethanolamine-HCl buffer pH 7.5+1 g/L PLP before adding to the reaction, and some of the libraries were screened at both 50° C. and 58° C.

[0332] The activity relative to SEQ ID NO:110 (FIOP) was calculated as the sum of the peak areas of the products produced by the variant over the sum of the peak areas of the products produced by SEQ ID NO: 110. The results are shown in Table 14.1.

TABLE-US-00012 TABLE 14.1 Activity of ATA Variants Relative to SEQ ID NO: 110

Transamination Transamination Product Peak Product Peak Amino Acid Area (FIOP).sup.1 Area (FIOP).sup.1 SEQ ID Differences at 50° C. at 58° C. Var. NO: (Relative to (Relative to (Relative to (S)- No: (nt/aa) SEQ ID NO: 110) SEQ ID NO: 110) SEQ ID NO: 110) stereoselectivity.sup.2 358 117/118 V31M/T208N/C260T/ ++ +++++ +++++ S286G/V294A/C424A 359 113/114 V31M/T208N/C260T/ ++ N/A +++++ S286G/V294A/A313E/ C424A 360 115/116 V31M/T208N/C260T/ ++ N/A +++++ S286G/V294A/A313L/ C424A 361 111/112 V31M/C260T/S286G/ ++ N/A +++++ A313L/C424A 362 V31M/C260T/S286G + +++ +++++ 363 V31M/T208N/S286G/ + +++ +++++ V294A/A313L/C424A 364 T208N/C260T/V294A/ + +++ +++++ A313L/C424A 365 V31M/T208N/V294A/ + +++ +++++ C424A 366 V31M/V294A/A313L/ + +++ +++++ C424A 367 V31M/C260T/S286G/ + +++ +++++ V294A/A313L 368 V31M/C260T/S286G/ + +++ +++++ C424A 369 V31M/T208N/V294A/ + +++ +++++ A313L/C424A 370 T208N/C260T/S286G/ + +++ +++++ V294A/C424A 371 V31M/V294A/C424A + +++ +++++ 372 V31M/T208N/C260T/ + +++ +++++ S286G/V294A 373 V31M/T208N/C260T/ + +++ +++++ V294A/A313L/C424A 374 V31M/S286G/V294A + +++ +++++ 375 T208N/C260T/S286G + +++ +++++ V294A/A313L/C424A 376 C260T/S286G/C424A + +++ +++++ 377 V31M/T208N/S286GV/ + +++ +++++ 294A/C424A 378 V31M/T208N/V294A + ++ +++++ 379 V31M + ++ +++++ 380 V31M/T208N + ++ +++++ 381 T208N/C260T/V294A + ++ +++++ 382 V31M/T208N/C260T/ N/A +++++ +++++ V294A/C424A 383 V31M/T208N/C260T/ N/A +++++ +++++ S286G/A313L/C424A 384 V31M/S286G/V294A/ N/A +++++ +++++ C424A 385 V31M/C260T/V294A N/A +++++ +++++ 386 V31M/T208N/A241V/ N/A +++++ +++++ S286G/V294A 387 V31M/S286G/V294A/ N/A +++++ +++++ A313L 388 V31M/T208N/S286G/ N/A +++++ +++++ C424A 389 V31M/T208N/S286G/ N/A +++++ +++++ A313L/C424A 390 V31M/T208N/S286G/ N/A +++++ +++++ V294A/A313L 391 V31M/T208N/S286G/ N/A +++++ +++++ V294A 392 S286G/V294A/A313L/ N/A +++++ +++++ C424A 393 S286G/A313L/C424A + N/A +++++ 394

V31M/T208N/S286G/ + N/A +++++ A313V/C424A 395 V31M/C260T/V294A/ + N/A +++++
A313L 396 Y14N/V31M/C260T/ + N/A +++++ S286G/V294A/C424A 397 Y14N/V31M/T208N/
+ N/A +++++ C260T/S286G/V294A/ C424A 398 T208N/S286G/V294A/ + N/A +++++ C424A
399 C260T/S286G/V294A/ + N/A +++++ C424A 400 V31M/C260T/S286G/ + N/A +++++
V294A/A313L/C424A 401 T208N/C260T/S286G/ + N/A +++++ C424A .sup.1Levels of increased
activity or selectivity were determined relative to the reference polypeptide of SEQ ID NO: 110
and defined as follows: “+” > than 1.2-fold but less than 2.5-fold increase; “++” > than 2.5-fold but
less than 5-fold; “+++” > than 5-fold but less than 10-fold; “++++” > than 10-fold but less than 15-
fold; “N/A” no activity data available. .sup.2Levels of S-stereoselectivity were defined as: “+” >
than 10% d.e. but less than 50% d.e.; “++” > than 50% d.e. but less than 75% d.e.; “+++” > than
75% d.e. but less than 95% d.e.; “++++” > than 95% d.e, but less than 100% d.e; “+++++” equal to
100% d.e.

Example 15

Improvements in the Transamination Activity of Compound (2) Compared to SEQ ID NO: 112 in
High Throughput Screening

[0333] SEQ ID NO: 112 was selected as the next parent enzyme(based on the results described in
Example 14. Libraries of engineered genes were produced using well-established techniques (e.g.,
saturation mutagenesis and recombination of previously identified beneficial mutations). The
polypeptides encoded by each gene were produced in HTP as described in Example 2, and the
soluble lysate was generated as described in Example 3. HTP reactions were carried out as
described in Example 14, except that the lysate was diluted 20-fold in 50 mM triethanolamine-HCl
buffer pH 7.5+1 g/L PLP before adding to the reaction, and the libraries were screened at 58° C.
[0334] The activity relative to SEQ ID NO: 112 (FIOP) was calculated as the sum of the peak areas
of the products produced by the variant over the sum of the peak areas of the products produced by
SEQ ID NO: 112. The results are shown in Table 15.1.

TABLE-US-00013 TABLE 15.1 Activity of ATA Variants Relative to SEQ ID NO: 112

Transamination Amino Acid Product Peak SEQ ID Differences Area (FIOP).sup.1 Var. NO:
(Relative to (Relative to (S)- No: (nt/aa) SEQ ID NO: 112) SEQ ID NO: 112)
stereoselectivity.sup.2 402 119/120 F19H/K86N + +++++ 403 M31T + ++ 404 Y17M + ++ 405
D21E + ++ 406 121/122 F19Y/K86N/F321Y + +++++ 407 G320A + + 408 I41L + ++ 409
G320A/D435N + ++ 410 123/124 V294M + +++++ 411 127/128 K66P/G394P + ++++++ 412
125/126 V42F/Y187F/V294M + +++++ 413 M22L/V294M + +++++ 414 V272E + ++++++ 415
M22L/S255A + ++++++ 416 M22L/S255A/H362R/V421I + ++++++ 417 S157P + + 418 M22L +
++++ 419 F221I + + 420 P164R/L397P + + 421 A84S + + 422 A153S + + 423 G394P + +++++ 424
K66P/E190P + ++++++ 425 L16F/V272E + +++++ 426 T161R + ++ 427 S157G + + 428
S24T/S255A/H362R + +++++ 429 T20A + + 430 F163P + ++ 431 M22L/H362R + +++++ 432
T114A + ++ 433 K66P + +++ 434 A84N + ++ 435 A222T + + 436 P164G + + .sup.1Levels of
increased activity or selectivity were determined relative to the reference polypeptide of SEQ ID
NO: 112 and defined as follows: “+” > than 1.15-fold but less than 2.5-fold increase. .sup.2Levels
of S-stereoselectivity were defined as: “+” > than 10% d.e. but less than 50% d.e.; “++” > than
50% d.e. but less than 75% d.e.; “+++” > than 75% d.e. but less than 95% d.e.; “++++” > than 95%
d.e. but less than 100% d.e; “+++++” equal to 100% d.e.

Example 16

Improvements in the Transamination Activity of Compound (2) Compared to SEQ ID NO: 120 in
High Throughput Screening

[0335] SEQ ID NO: 120 was selected as the next parent enzyme, based on the results described in
Example 15. Libraries of engineered genes were produced using well-established techniques (e.g.,
saturation mutagenesis and recombination of previously identified beneficial mutations). The
polypeptides encoded by each gene were produced in HTP as described in Example 2, and the
soluble lysate was generated as described in Example 3. HTP reactions were carried out as

described in Example 15, except that the lysate was diluted 60-fold in 50 mM triethanolamine-HCl buffer pH 7.5+1 g/L PLP before adding to the reaction. In addition, after the reactions were quenched 1:1 with 10% acetic acid in MeCN they were further diluted 5000-fold in MeCN containing 0.1% formic acid and analyzed by the Agilent RapidFire 365 high throughput mass spectrometer, using the manufacturer's protocols.

[0336] The activity relative to SEQ ID NO:120 (FIOP) was calculated as the peak area of the product produced by the variant over peak area of the product produced by SEQ ID NO: 120. The results are shown in Table 16.1.

TABLE-US-00014 TABLE 16.1 Activity of ATA Variants Relative to SEQ ID NO: 120

Transamination Amino Acid Product Peak SEQ ID Differences Area (FIOP).sup.1 Var. NO:

(Relative to (Relative to No: (nt/aa) SEQ ID NO:120) SEQ ID NO: 120) 437 137/138

Y17M/G320A + 438 141/142 V272E/G320A + 439 151/152 Y17M/A84S/P164R + 440 147/148

Y17M/F221I/G320A + 441 135/136 K66P/G320A + 442 149/150 Y17M/S157G/G320A/L397P +

443 129/130 H19Y/M22L/V42F/G320A/ + G394P 444 139/140 Y17M/M22L/V272E + 445

143/144 Y17M/V272E/G320A + 446 131/132 H19Y/V42F + 447 133/134 K66P/V272E + 448

145/146 V17M/L397P + 449 Y17M/M22L/V272E/G320A + 450 Y165W + 451 G320A + 452

Y17M/M31T/G320A/L397P + 453 Y17M + 454 K66P/Y187F/G320A + 455 Y17M/D21E + 456

P164R/G320A/F321Y + 457 Y17M/D21E/M22L + 458 K66P/G394P + 459 V272E + 460

Y17M/M22L + 461 L25V/G320A + 462 S255A + 463 Y17M/S157G/P164R + 464 T20A/G320A +

.sup.1Levels of increased activity or selectivity were determined relative to the reference polypeptide of SEQ ID NO: 120 and defined as follows: "+" > than 1.2-fold but less than 2.5-fold increase.

Example 17

Improvements in the Transamination Activity of Compound (2) Compared to SEQ ID NO: 138 in High Throughput Screening

[0337] SEQ ID NO: 138 was selected as the next parent enzyme, based on the results described in Example 16. Libraries of engineered genes were produced using well-established techniques (e.g., saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP as described in Example 2, and the soluble lysate was generated as described in Example 3. HTP reactions were carried out as described in Example 16, except: the lysate was diluted 120-fold in 50 mM triethanolamine-HCl buffer pH 7.5+1 g/L PLP before adding to the reaction.

[0338] The activity relative to SEQ ID NO: 138 (FIOP) was calculated as the peak area of the product produced by the variant over peak area of the product produced by SEQ ID NO: 138. The results are shown in Table 17.1.

TABLE-US-00015 TABLE 17.1 Activity of ATA Variants Relative to SEQ ID NO: 138

Transamination Amino Acid Product Peak SEQ ID Differences Area (FIOP).sup.1 Var. NO:

(Relative to (Relative to No: (nt/aa) SEQ ID NO: 138) SEQ ID NO: 138) 465 157/158

Y165W/G394P + 466 159/160 E190P/E207D/Q210A + 467 155/156 V42F/K66P/Y165W/G394P +

468 161/162 S24T + 469 E207D + 470 E207D/L243Q + 471 S389T + 472 153/154 K66P/Y165W

+ 473 K66P/G394P + 474 E190P/E207D/K385A + 475 E190P + 476 V272E/G394P + 477 K385A

+ 478 E190P/E207D + 479 S389T/N396S + 480 P164R/Y165W/G394P + 481 A214P + 482

E190P/S389T + 483 V42F + 484 Q27D/E190P/E207D/N396S + 485 Y165W + 486 S167E + 487

141L/V42F/K66P/Y165W + 488 Q27D/E190P/E207D/S389T/ + N396S 489

E207D/Q210A/L243Q/D250G/ + K385A/S389T 490 E190P/E207D/Q210A/K385A/ + S389T 491

S24T/R110L/S167E + 492 A214P/N396S + 493 V42F/Y165W/V272E/G394P + 494 S24T/L25V +

495 K66P/P164R/V272E + 496 Q210A + 497 K385A/S389T + 498 G394P + 499 Y165W/V272E

+ 500 K66P/P164R + 501 E207D/A214P/N396S + 502 E190P/Q210A + 503

E190P/K385A/N396S + 504 Q210A/K385A + 505 E207D/S389T + 506 L25V + 507

L25V/W147V/S167E + 508 N166L + 509 N166T + 510 N166A + 511 P164N + 512 S283C + 513

P164S + 514 F163L + .sup.1Levels of increased activity or selectivity were determined relative to the reference polypeptide of SEQ ID NO: 138 and defined as follows: “+” > than 1.2-fold but less than 2.5-fold increase.

Example 18

Improvements in the Transamination Activity of Compound (2) Compared to SEQ ID NO: 156 in High Throughput Screening

[0339] SEQ ID NO: 156 was selected as the next parent enzyme, based on the results described in Example 17. Libraries of engineered genes were produced using well-established techniques (e.g., saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP as described in Example 2, and the soluble lysate was generated as described in Example 3. HTP reactions were carried out as described in Example 17.

[0340] The activity relative to SEQ ID NO:156 (FIOP) was calculated as the peak area of the product produced by the variant over peak area of the product produced by SEQ ID NO: 156. The results are shown in Table 18.1.

TABLE-US-00016 TABLE 18.1 Activity of ATA Variants Relative to SEQ ID NO: 156

Transamination Amino Acid Product Peak SEQ ID Differences Area (FIOP).sup.1 Var. NO:

(Relative to (Relative to No: (nt/aa) SEQ ID NO: 156) SEQ ID NO: 156) 515 165/166

F291Y/T309A + 516 169/170 E77D/R110K/F291Y/E308Q + 517 167/168 N166A/E207D + 518

W147V/P164R/N166A/S167E + 519 E77D/F291Y/K361A/R366H + 520 D70E/F291Y + 521

M22L/L25V/R110K/F291Y/ + S304A 522 T30P/E77R/F291Y + 523 L177I/F291Y/T309A/K361A

+ 524 V253L/A391T + 525 N166A/E190P/E207D + 526 H26K/T30P/V168E/F291Y + 527

D70E/E77D/S108R/L177I/ + Q196E/R247K/F291Y/T309A/ K361A/R366H 528 163/164 N166G

+ 529 T74A + 530 N166S/E207D + 531 N166A + 532 D70E/R186C + 533 S24T/N166L + 534

D70E/L177I/F291Y/K361A/ + R366H 535 L177I/E195A/Q196E/F291Y/ + S304A/T309A/R366H

536 F291Y/S304A/T309A + 537 D70E/E77R/V168E/F291Y + 538 S24T/N166A/E207D + 539

Q196E/F291Y/T309A/K361A + 540 S24T/W147V/P164S/N166A/ + S167E/L243Q/S389T 541

E207D + 542 171/172 F291Y + 543 N166A/E190P + 544 N166S + 545

W147V/P164R/N166A/V272E + 546 E12D/H26R/K73C/E77R/F291Y + 547 S24T/N166A/E190P

+ 548 N166A/S167E/V272E/S389T + 549 T30P/F291Y + 550 E12D/H26K/T30P/F291Y + 551

T13A/H45N/D107E/K236E/ + T277V/A346L/R351A/L410H/ L429M 552

D70E/F291Y/S304A/K361A + 553 Q196E/F291Y/S304A/K361A + 554 S24T/N166A + 555

T277V + 556 R110K/A235S/A391T + 557 D107E/A346L + 558 V198T/A235S/A391T + 559

R247K/S304A/T309A/R366H + 560 D107E/T277V/A346L/L410H/ + L429M 561 T74A/K211R +

562 D70E/L177I/F291Y/Y372W + 563 F163Y + 564 W147V/P164R/N166A/L243Q/ +

V272E/S283C/K385A/S389T 565 D70E/R247K/F291Y/T309A/ + Y372W 566

L177I/Q196E/R247K/S304A/ + T309A/K361A/R366H 567 N166V + 568 H26R/T30P +

.sup.1Levels of increased activity or selectivity were determined relative to the reference polypeptide of SEQ ID NO: 156 and defined as follows: “+” > than 1.2-fold but less than 2.5-fold increase; “++” > than 2.5-fold but less than 5-fold.

[0341] All publications, patents, patent applications and other documents cited in this application are hereby incorporated by reference in their entireties for all purposes to the same extent as if each individual publication, patent, patent application or other document were individually indicated to be incorporated by reference for all purposes.

[0342] While various specific embodiments have been illustrated and described, it will be appreciated that various changes can be made without departing from the spirit and scope of the invention(s).

Claims

1. An engineered transaminase comprising a polypeptide sequence having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NOS: 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156, or a functional fragment thereof, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 17, 17/259/347, 20, 45/48/78/81/92, 57, 57/319, 85, 86, 157, 158, 161, 203/416, 314, 316, 319, 375/416, 404, 415, 416, 417, 420, 421, 422, and 424, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, or 156.
2. The engineered transaminase of claim 1, wherein said polypeptide sequence has at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 18, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 5/17/86/157/259/314/316/422, 17/86/157/203/259/314/316/422, 17/86/157/314/316/347/422, 17/86/157/316/422, 85/86, 86/157/314/316/347/416/422, 86/157/416, 86/203/316/416, 86/203/316/422, 86/316/416, 139/221/310/316, 157/316, 195/316, 207/316, 227/422, 280/381/416/445, 314, 314/316, 316, 416, 417, and 422, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 18.
3. The engineered transaminase of claim 1, wherein said polypeptide sequence has at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 20, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 17/139/314/416/417/422, 17/195/316/414/416/422, 17/314/316/414/416, 17/314/316/414/416/417/422, 20, 31, 54, 57, 83, 86, 86/194,121, 139/195/227/314/316/416, 139/202/207/227/314/316/416/422, 139/207/227/314/414/416/422, 139/227/310/414/416/417, 139/316/414/416/417/422, 146/157, 164/207/213/227/310/316/416, 166, 178/314/316/416, 190/195/207/213/244/314/414/416/417/422, 194/213/310/314/316/422, 195/227/314/316/416, 195/227/314/416/417, 195/227/314/422, 195/227/414/416, 195/227/414/416/422, 195/227/416/417, 195/316/416/417/422, 207/208/220/241/291/314/316, 226, 227, 227/416, 228, 258, 259, 271/310/316/416/422, 282, 310/314/316/414/416/417/422, 314/316/414/416/422, 314/316/416, 314/316/416/422, 316, 319, 320, 322, and 417, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 20.
4. The engineered transaminase of claim 1, wherein said polypeptide sequence has at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 30, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 21, 21/23/53/146/162/395, 21/23/53/146/162/395/432, 21/23/53/162/395/432, 21/23/53/395/432, 21/23/395, 21/53/159/395, 21/53/163/395/432, 23/53/146/159/395/431, 23/53/146/159/395/432, 23/53/146/162/395/432, 23/53/146/163, 23/146/163/432, 23/162, 53/146/159/395, 53/146/162/395, 53/146/162/432, 53/146/163/395, 53/146/395/432, 53/162/163/395/432, 53/162/395, 53/162/395/432, 53/163/395/432, 53/395, 57/86, 57/86/121/146/157, 57/86/121/157/226/259/282/332, 57/86/121/332, 57/86/157/259, 57/86/157/282/332, 57/86/332, 57/121/417/422, 57/146/226/259, 57/146/226/259/282, 57/259, 86, 86/157, 86/226,121, 121/146/157/226/332, 121/208/316/417/422, 121/226/332, 121/316/422, 121/332, 121/416/417/422, 121/417/422, 146, 146/162/395/432, 157, 162/432, 226/282/332, 227, 227/417, 259, 282, 282/332, 316/417/422, 332, 395/432, 416, 416/422, 417, and 422, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 30.
5. The engineered transaminase of claim 1, wherein said polypeptide sequence has at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 66, wherein said engineered transaminase comprises at least one

substitution or substitution set in said polypeptide sequence at one or more positions selected from 107, 141, 161, 174, 245, 281, 293, 432, and 453, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 66.

6. The engineered transaminase of claim 1, wherein said polypeptide sequence has at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 70, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 2, 4, 5, 5/38/56/346, 13, 17, 17/31, 19, 31, 57, 74, 79, 86, 121, 155, 161, 161/174/432, 163/167, 164, 190/291, 195, 195/408, 203/207/219, 205/419, 207, 207/235/419, 214/228/251, 222, 223, 226/228/230/231/311, 228, 233, 246, 247/271/293, 259, 282, 284, 309, 327, 346, 359, 362, 375, 404, 408, 419, 420, 422, and 443, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 70.

7. The engineered transaminase of claim 1, wherein said polypeptide sequence has at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 88, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 31, 86, 149, 157, 163, 163/164, 191/208/211/286, 192/294, 195/207/208/286, 208/294, 260, 284, 286, 294, 313, 332, 395/396, and 424 and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 88.

8. The engineered transaminase of claim 1, wherein said polypeptide sequence has at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 110, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 14/31/208/260/286/294/424, 14/31/260/286/294/424, 31, 31/208, 31/208/241/286/294, 31/208/260/286/294, 31/208/260/286/294/313/424, 31/208/260/286/294/424, 31/208/260/286/313/424, 31/208/260/294/313/424, 31/208/260/294/424, 31/208/286/294, 31/208/286/294/313, 31/208/286/294/313/424, 31/208/286/294/424, 31/208/286/313/424, 31/208/286/424, 31/208/294, 31/208/294/313/424, 31/208/294/424, 31/260/286, 31/260/286/294/313, 31/260/286/294/313/424, 31/260/286/313/424, 31/260/286/424, 31/260/294, 31/260/294/313, 31/286/294, 31/286/294/313, 31/286/294/424, 31/294/313/424, 31/294/424, 208/260/286/294/313/424, 208/260/286/294/424, 208/260/286/424, 208/260/294, 208/260/294/313/424, 208/286/294/424, 260/286/294/424, 260/286/424, 286/294/313/424, and 286/313/424, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 110.

9. The engineered transaminase of claim 1, wherein said polypeptide sequence has at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 112, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 16/272, 17, 19/86, 19/86/321, 20, 21, 22, 22/255, 22/255/362/421, 22/294, 22/362, 24/255/362, 31, 41, 42/187/294, 66, 66/190, 66/394, 84, 114, 153, 157, 161, 163, 164, 164/397, 221, 222, 272, 294, 320, 320/435, and 394, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 112.

10. The engineered transaminase of claim 1, wherein said polypeptide sequence has at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 120, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 17, 17/21, 17/21/22, 17/22, 17/22/272, 17/22/272/320, 17/31/320/397, 17/84/164, 17/157/164, 17/157/320/397, 17/221/320, 17/272/320, 17/320, 17/397, 19/22/42/320/394, 19/42, 20/320, 25/320, 66/187/320, 66/272, 66/320, 66/394, 164/320/321, 165, 255, 272, 272/320, and 320, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ

ID NO: 120.

11. The engineered transaminase of claim 1, wherein said polypeptide sequence has at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 138, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 24, 24/25, 24/110/167, 25, 25/147/167, 27/190/207/389/396, 27/190/207/396, 41/42/66/165, 42, 42/66/165/394, 42/165/272/394, 66/164, 66/164/272, 66/165, 66/394, 163, 164, 164/165/394, 165, 165/272, 165/394, 166, 167, 190, 190/207, 190/207/210, 190/207/210/385/389, 190/207/385, 190/210, 190/385/396, 190/389, 207, 207/210/243/250/385/389, 207/214/396, 207/243, 207/389, 210, 210/385, 214, 214/396, 272/394, 283, 385, 385/389, 389, 389/396, and 394, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 138.

12. The engineered transaminase of claim 1, wherein said polypeptide sequence has at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 156, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 12/26/30/291, 12/26/73/77/291, 13/45/107/236/277/346/351/410/429, 22/25/110/291/304, 24/147/164/166/167/243/389, 24/166, 24/166/190, 24/166/207, 26/30, 26/30/168/291, 30/77/291, 30/291, 70/77/108/177/196/247/291/309/361/366, 70/77/168/291, 70/177/291/361/366, 70/177/291/372, 70/186, 70/247/291/309/372, 70/291, 70/291/304/361, 74, 74/211, 77/110/291/308, 77/291/361/366, 107/277/346/410/429, 107/346, 110/235/391, 147/164/166/167, 147/164/166/243/272/283/385/389, 147/164/166/272, 163, 166, 166/167/272/389, 166/190, 166/190/207, 166/207, 177/195/196/291/304/309/366, 177/196/247/304/309/361/366, 177/291/309/361, 196/291/304/361, 196/291/309/361, 198/235/391, 207, 247/304/309/366, 253/391, 277, 291, 291/304/309, and 291/309, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 156.

13. The engineered transaminase of claim 1, wherein said engineered transaminase comprises a polypeptide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to the sequence of at least one engineered transaminase variant set forth in SEQ ID NOS: 4, 18, 20, 30, 48, 66, 70, 88, 110, 112, 120, 138, and/or 156.

14. The engineered transaminase of claim 1, wherein said engineered transaminase comprises a polypeptide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to the sequence of at least one engineered transaminase variant set forth in the even numbered sequences of SEQ ID NOS: 6-172.

15. The engineered transaminase of claim 1, wherein said engineered transaminase comprises at least one improved property compared to wild-type *V. fluvialis* transaminase.

16. A composition comprising the engineered transaminase of claim 1.

17. A polynucleotide sequence encoding at least one engineered transaminase of claim 1.

18. A host cell comprising at least one polynucleotide sequence of claim 17.

19. A method of producing an engineered transaminase in a host cell, comprising culturing the host cell of claim 18 under suitable conditions, such that at least one engineered transaminase is produced.

20. The method of claim 19, further comprising recovering at least one engineered transaminase from the culture and/or host cell.
