

ENGINEERED TRANSAMINASE POLYPEPTIDES

[0001] The present application claims priority to US Pat. Appln. Ser. No. 62/792,518, filed on January 15, 2019, which is hereby incorporated by reference in its entirety for all purposes.

FIELD OF THE INVENTION

[0002] The disclosure relates to engineered transaminase polypeptides useful under industrial process conditions for the production of pharmaceutical and fine chemical amine compounds.

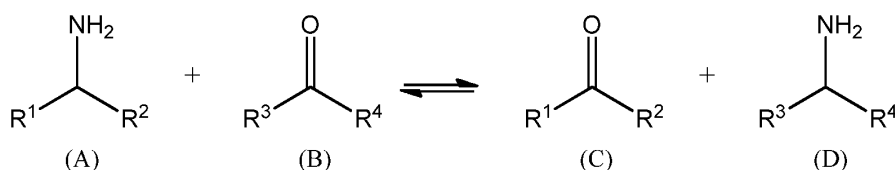
REFERENCE TO SEQUENCE LISTING, TABLE OR COMPUTER PROGRAM

[0003] The official copy of the Sequence Listing is submitted concurrently with the specification as an ASCII formatted text file via EFS-Web, with a file name of "CX2-183WO1_ST25.txt", a creation date of January 10, 2020 and a size of 1.72 megabytes. The Sequence Listing filed via EFS-Web is part of the specification and incorporated in its entirety by reference herein.

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 R¹ is not the same as R²) and a new amino acceptor byproduct (or "carbonyl byproduct") compound (C) which has a keto group.

Scheme 1



[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., 2001, *Biosci. Biotechnol. Biochem.* 65:1782-1788; Iwasaki et al., 2003, *Biotech. Lett.* 25:1843-1846; Iwasaki et al., 2004, *Appl. Microb. Biotech.* 69:499-505; Yun et al., 2004, *Appl. Environ. Microbiol.* 70:2529-2534; and Hwang et al., 2004, *Enzyme Microbiol. Technol.* 34:429-426); the preparation of intermediates and precursors of pregabalin (e.g., WO 2008/127646); the enzymatic transamination of cyclopamine 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 (e.g., US Patent No. US 4,950,606; US Patent No. 5,300,437; and US Patent No. 5,169,780); the production of amino acids and derivatives (e.g., US Patent No. 5,316,943; US Patent No. 4,518,692; US Patent No. 4,826,766; US Patent No. 6,197,558; and US Patent No. 4,600,692); and in the production of the pharmaceutical compounds, sitagliptin, rivastigmine, and vernakalant (See e.g., US Pat. No. 8,293,507 B2, issued Oct. 23, 2012; Savile, *et al.*, 2010, "Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture," *Science* 329(5989): 305-9; WO2011/159910, published Dec. 22, 2011; and WO2012/024104, published Feb. 23, 2012).

[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., 2001, *Biosci. Biotechnol. Biochem.* 65:1782-1788), and *Arthrobacter* sp. *KNK168* (See e.g., Iwasaki et al., *Appl. Microbiol. Biotechnol.*, 2006, 69: 499-505, U.S. Pat. No. 7,169,592). Several of these wild-type transaminase genes and encoded polypeptides have been sequenced, including 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., 1983, *Agric. Biol. Chem.* 47 (10):2257-2265).

[0008] Transaminases are known that have (*R*)-selective or (*S*)-selective stereoselectively. 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. Microbiol. Biotechnol.*, 2006, 69: 499-505, 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.*, "Purification, characterization, and molecular cloning of a novel

amine:pyruvate transaminase from *Vibrio fluvialis JS17*,” *Appl. Microbiol. Biotechnol.* 61 (5-6), 463-471 (2003)).

[0009] 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., US Pat. No. 8,293,507 B2, issued Oct. 23, 2012; WO2011/005477A1, published Jan. 13, 2011; WO2012/024104, published Feb. 23, 2012; and Savile, *et al.*, 2010, “Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture,” *Science* 329(5989): 305-9).

[0010] Wild-type 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.

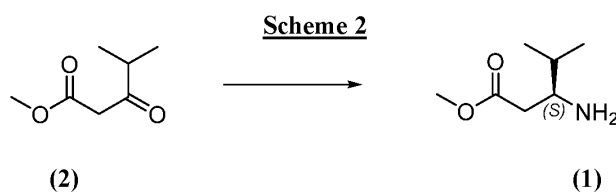
[0011] In particular, transaminase reactions are generally conducted under aqueous conditions at high pH. Typical aqueous reaction conditions often result in low substrate solubility, necessitating an organic co-solvent. Reaction under aqueous conditions is often complicated by product and substrate instability. Additional process steps are often required, such as basification of the reaction mixture and extraction of the amine product with an organic solvent.

SUMMARY OF THE INVENTION

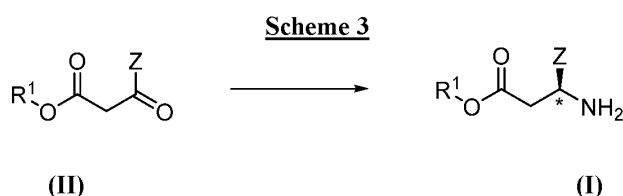
[0012] The present invention provides engineered transaminase polypeptides and compositions thereof, as well as polynucleotides encoding the engineered transaminase polypeptides. Methods for producing transaminase enzymes are also provided. In some embodiments, the engineered transaminase polypeptides are optimized to provide improved enzyme properties under minimal aqueous conditions in wet organic solvents. The transaminase polypeptides of the present disclosure 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 improved enzyme properties relative to previously engineered transaminase polypeptides (See e.g., US 8,293,507 B2, issued Oct. 23, 2012; PCT Publication WO2011005477A1, published Jan. 13, 2011, and PCT publication WO2012024104, published Feb. 23, 2012). The amino residue differences are located at residue positions that result in improvement of various enzyme properties, including among others, activity, stereoselectivity, stability, expression, product tolerance, and thermoactivity in wet organic solvents.

[0013] In particular, the engineered transaminase polypeptides of the present disclosure have been engineered for efficient conversion of the substrate, methyl 4-methyl-3-oxo-pentanoate (referred to herein as “compound (2)”), to its corresponding chiral amine product compound, methyl (*S*)-3-amino-4-

methyl-pentanoate (referred to herein as “compound (1)”) in minimal aqueous conditions in wet organic solvents, as shown in **Scheme 2**.



[0014] The evolved structural features of the engineered transaminase polypeptides of the present disclosure, however, also allow for the biocatalytic conversion of a range of keto ester 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**,



wherein

Z is a linear C₁₋₆ alkyl or a branched C₁₋₆ alkyl, optionally substituted with one or more hydroxy, cyano, or nitro groups ; and

R¹ is C₁₋₆ alkyl, aryl-C₁₋₂ alkyl, heteroaryl-C₁₋₂ alkyl, or a 5- to 6-membered heterocyclic ring system optionally containing an additional heteroatom selected from O, S, and N, the heterocyclic ring being unsubstituted or substituted with one to three substituents independently selected from oxo, hydroxy, halogen, C₁₋₄ alkoxy, and C₁₋₄ alkyl.

[0015] 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.

[0016] The present invention provides engineered polypeptides having transaminase activity 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 NO:2, 6, 332, 518 and/or 656, or a functional fragment thereof, wherein said engineered polypeptide comprises a polypeptide comprising at least one substitution or substitution set in said polypeptide sequence, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:2, 6, 332, 518 and/or 656. In some embodiments, the reference sequence is SEQ ID NO:2. In some embodiments, the

reference sequence is SEQ ID NO:6. In some embodiments, the reference sequence is SEQ ID NO:332. In some embodiments, the reference sequence is SEQ ID NO:518. In some embodiments, the reference sequence is SEQ ID NO:656.

[0017] In some embodiments, the 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:2, and wherein the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from 284 and 284/155, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:2. In some embodiments, the polypeptide sequence of the engineered polypeptide 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:2, and wherein the polypeptide sequence of the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from 284A and 284A/155W, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:2. In some embodiments, the polypeptide sequence of the engineered polypeptide 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:2, and wherein the polypeptide sequence of the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from G284A and G284A/L155W, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:2. In some embodiments, the engineered polypeptide comprises 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 NO:2. In some embodiments, the engineered polypeptide comprises a polypeptide sequence having at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO:2. In some embodiments, the engineered polypeptide comprises a polypeptide sequence having at least 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO:2.

[0018] In some embodiments, the present invention provides engineered polypeptides 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 NO:6, wherein the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from 5/35, 9, 9/20/23/24/134/244, 9/20/23/101/244, 9/20/24/101/244/249, 9/23/24, 9/23/24/29/244, 9/23/24/101/262, 9/23/24/244/262, 9/23/29, 9/24/244, 9/24/244/262, 9/24/249, 9/45/89/93, 9/79/101/244, 9/82/89, 9/89/246, 9/89/271, 9/93/142/246, 9/101/244, 9/110/302, 9/244/249, 9/244/262/264, 9/246, 9/262, 15, 15/16/18/20/89/93/160, 15/24/26/89/93/160, 15/35/273, 15/218/273, 15/249/273, 16, 17/25, 17/25/41, 17/25/160, 17/160, 18, 18/26/246/273, 18/146, 18/146/168, 18/146/262, 18/168, 18/168/244, 18/168/262/324, 18/244, 18/244/246, 18/244/246/284, 18/244/248, 18/244/262/273, 18/244/273, 18/262, 18/273, 20, 20/23/24/55/134/244/262, 20/23/134/244, 20/24/89/197, 20/33/160, 20/324, 22, 22/102, 22/160, 23, 23/25/55/321, 23/25/102/142, 23/102, 23/321, 24, 25, 25/33, 25/104, 25/160, 25/321, 26,

26/273, 29, 29/89, 32, 33, 35, 35/215/218/249, 35/218/249/273, 35/221, 40, 40/168/208/244, 40/244/248, 41, 41/102, 42, 45, 46, 50, 55, 56, 71/102, 74, 79, 82, 82/244, 89, 89/93/156/324, 89/93/321/324, 90, 93, 93/324, 95, 98, 101, 101/160/244/249, 102, 104, 106, 111, 114, 117, 119, 134, 134/244, 134/262, 139, 140/266, 142, 142/321, 145, 146, 146/244, 146/273, 156, 160, 160/215/216/218, 160/273, 168, 168/244/273, 179, 186, 186/200, 197, 203, 207, 208, 208/246, 215, 215/216/249, 216, 216/273, 218, 218/249, 218/313, 234, 236, 244, 244/248, 244/248/284, 244/273, 246, 246/248/273, 248, 249, 249/273, 256, 262, 264, 265, 266, 271, 273, 284, 301, 316, 320, 321, 324, and 329, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:6. In some embodiments, the polypeptide sequence of the engineered polypeptide 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:6, and wherein the polypeptide sequence of the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from 5T/35Y, 9H, 9H/45L/89Q/93H, 9H/82S/89Q, 9H/89Q/246L, 9H/89Q/271C, 9H/93H/142M/246L, 9H/110V/302F, 9H/246L, 9T, 9T/20T/23Q/24F/134S/244G, 9T/20T/23Q/101S/244L, 9T/20T/24F/101S/244G/249G, 9T/23Q/24F, 9T/23Q/24F/29M/244G, 9T/23Q/24F/101S/262V, 9T/23Q/24F/244L/262T, 9T/23Q/29M, 9T/24F/244G, 9T/24F/244G/262T, 9T/24F/249G, 9T/79G/101S/244G, 9T/101S/244G, 9T/244G/249G, 9T/244G/262T/264N, 9T/262T, 15I, 15I/16I/18R/20H/89M/93M/160L, 15I/24P/26P/89M/93M/160L, 15T, 15T/35G/273K, 15T/218R/273K, 15T/249M/273K, 16I, 17I/25R, 17I/25R/41F, 17I/25R/160V, 17I/160V, 18G, 18G/26H/246V/273T, 18G/146L, 18G/146L/262V, 18G/168P, 18G/244G, 18G/244G/248M, 18G/244L/273V, 18G/273T, 18G/273V, 18K, 18K/146L, 18K/146L/168P, 18K/168P, 18K/168P/244L, 18K/168P/262V/324W, 18K/244G/246V, 18K/244G/246V/284G, 18K/244L, 18K/244L/246V, 18K/244T/262V/273T, 18K/262V, 18K/273T, 18R, 20H, 20H/24P/89M/197T, 20H/33L/160L, 20H/324C, 20T, 20T/23Q/24F/55F/134G/244G/262T, 20T/23Q/134S/244G, 22H, 22L, 22L/102L, 22L/160V, 22W, 23H, 23H/25R/55L/321Y, 23H/25R/102L/142L, 23H/102L, 23H/321Y, 23Q, 24D, 24F, 24P, 25R, 25R/33M, 25R/104V, 25R/160V, 25R/321Y, 26H/273K, 26P, 29K, 29K/89Q, 29M, 32L, 32S, 33L, 33M, 33S, 35G, 35G/215L/218R/249T, 35G/218R/249T/273K, 35G/221D, 40M, 40M/168P/208L/244G, 40M/244G/248M, 41F/102L, 41K, 41Y, 42T, 45L, 46C, 50T, 55F, 55G, 55K, 55S, 56R, 71Y/102P, 74A, 79G, 79H, 82P, 82S/244D, 82T, 89L, 89M, 89M/93M/156L/324C, 89M/93M/321V/324C, 89Q, 89W, 90A, 90C, 90S, 90T, 93H, 93M, 93M/324C, 93P, 95P, 95T, 98M, 101L, 101L/160Y/244E/249M, 101S, 102L, 104I, 104S, 106S, 111L, 111T, 114T, 117D, 119A, 134G, 134G/244L, 134G/262V, 134S, 139A, 140N/266N, 142L/321Y, 142T, 145N, 146A, 146A/244D, 146H, 146I, 146L, 146L/244L, 146L/273V, 156E, 156L, 156T, 160A, 160L, 160V, 160Y, 160Y/215L/216G/218R, 160Y/273K, 168P, 168P/244L/273T, 179R, 186R, 186R/200R, 197L, 197T, 203P, 207Y, 208K, 208L, 208L/246V, 208P, 208R, 215L, 215L/216G/249M, 216G, 216P, 216P/273V, 216S, 218N/313H, 218R, 218R/249M, 218S, 234H, 236T, 244E, 244G, 244G/248M, 244G/248M/284G, 244L, 244L/273C, 244L/273V, 244T, 246A, 246L, 246V, 246V/248M/273V, 248M, 248V, 249G, 249M, 249M/273K, 249T, 256C, 262L, 262P,

262T, 262V, 264A, 264E, 264G, 264N, 264P, 264R, 264W, 265R, 266G, 266S, 271C, 273L, 273M, 273T, 273V, 284G, 301E, 301M, 316R, 320R, 320W, 321L, 321V, 321Y, 324C, 324V, 324W, and 329R, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:6. In some embodiments, the polypeptide sequence of the engineered polypeptide 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:6, and wherein the polypeptide sequence of the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from A5T/A35Y, E9H, E9H/F45L/S89Q/S93H, E9H/D82S/S89Q, E9H/S89Q/I246L, E9H/S89Q/E271C, E9H/S93H/K142M/I246L, E9H/L110V/V302F, E9H/I246L, E9T, E9T/Y20T/Y23Q/S24F/T134S/P244G, E9T/Y20T/Y23Q/T101S/P244L, E9T/Y20T/S24F/T101S/P244G/K249G, E9T/Y23Q/S24F, E9T/Y23Q/S24F/D29M/P244G, E9T/Y23Q/S24F/T101S/A262V, E9T/Y23Q/S24F/P244L/A262T, E9T/Y23Q/D29M, E9T/S24F/P244G, E9T/S24F/P244G/A262T, E9T/S24F/K249G, E9T/R79G/T101S/P244G, E9T/T101S/P244G, E9T/P244G/K249G, E9T/P244G/A262T/L264N, E9T/A262T, D15I, D15I/T16I/L18R/Y20H/S89M/S93M/F160L, D15I/S24P/Y26P/S89M/S93M/F160L, D15T, D15T/A35G/Y273K, D15T/L218R/Y273K, D15T/K249M/Y273K, T16I, G17I/D25R, G17I/D25R/I41F, G17I/D25R/F160V, G17I/F160V, L18G, L18G/Y26H/I246V/Y273T, L18G/Q146L, L18G/Q146L/A262V, L18G/H168P, L18G/P244G, L18G/P244G/R248M, L18G/P244L/Y273V, L18G/Y273T, L18G/Y273V, L18K, L18K/Q146L, L18K/Q146L/H168P, L18K/H168P, L18K/H168P/P244L, L18K/H168P/A262V/L324W, L18K/P244G/I246V, L18K/P244G/I246V/A284G, L18K/P244L, L18K/P244L/I246V, L18K/P244T/A262V/Y273T, L18K/A262V, L18K/Y273T, L18R, Y20H, Y20H/S24P/S89M/R197T, Y20H/P33L/F160L, Y20H/L324C, Y20T, Y20T/Y23Q/S24F/I55F/T134G/P244G/A262T, Y20T/Y23Q/T134S/P244G, T22H, T22L, T22L/Q102L, T22L/F160V, T22W, Y23H, Y23H/D25R/I55L/P321Y, Y23H/D25R/Q102L/K142L, Y23H/Q102L, Y23H/P321Y, Y23Q, S24D, S24F, S24P, D25R, D25R/P33M, D25R/E104V, D25R/F160V, D25R/P321Y, Y26H/Y273K, Y26P, D29K, D29K/S89Q, D29M, N32L, N32S, P33L, P33M, P33S, A35G, A35G/G215L/L218R/K249T, A35G/L218R/K249T/Y273K, A35G/E221D, W40M, W40M/H168P/E208L/P244G, W40M/P244G/R248M, I41F/Q102L, I41K, I41Y, G42T, F45L, V46C, E50T, I55F, I55G, I55K, I55S, F56R, H71Y/Q102P, N74A, R79G, R79H, D82P, D82S/P244D, D82T, S89L, S89M, S89M/S93M/Q156L/L324C, S89M/S93M/P321V/L324C, S89Q, S89W, N90A, N90C, N90S, N90T, S93H, S93M, S93M/L324C, S93P, R95P, R95T, P98M, T101L, T101L/F160Y/P244E/K249M, T101S, Q102L, E104I, E104S, K106S, E111L, E111T, A114T, E117D, R119A, T134G, T134G/P244L, T134G/A262V, T134S, D139A, I140N/D266N, K142L/P321Y, K142T, P145N, Q146A, Q146A/P244D, Q146H, Q146I, Q146L, Q146L/P244L, Q146L/Y273V, Q156E, Q156L, Q156T, F160A, F160L, F160V, F160Y, F160Y/G215L/D216G/L218R, F160Y/Y273K, H168P, H168P/P244L/Y273T, P179R, Q186R, Q186R/Q200R, R197L, R197T, H203P, F207Y, E208K, E208L, E208L/I246V, E208P, E208R, G215L, G215L/D216G/K249M, D216G, D216P, D216P/Y273V, D216S,

L218N/R313H, L218R, L218R/K249M, L218S, V234H, R236T, P244E, P244G, P244G/R248M, P244G/R248M/A284G, P244L, P244L/Y273C, P244L/Y273V, P244T, I246A, I246L, I246V, I246V/R248M/Y273V, R248M, R248V, K249G, K249M, K249M/Y273K, K249T, E256C, A262L, A262P, A262T, A262V, L264A, L264E, L264G, L264N, L264P, L264R, L264W, A265R, D266G, D266S, E271C, Y273L, Y273M, Y273T, Y273V, A284G, G301E, G301M, E316R, E320R, E320W, P321L, P321V, P321Y, L324C, L324V, L324W, and Q329R, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:6. In some embodiments, the engineered polypeptide comprises 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 NO:6. In some embodiments, the engineered polypeptide comprises a polypeptide sequence having at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO:6. In some embodiments, the engineered polypeptide comprises a polypeptide sequence having at least 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO:6.

[0019] In some additional embodiments, the present invention provides engineered polypeptides 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 NO:332, wherein the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from 9/89/93/101/160, 9/89/160, 33, 35, 44, 49, 74, 89, 89/93/160/249, 107, 108, 117, 140, 141, 142, 144, 156, 168, 203, 207, 234, 270, 273, 274, 292, 296, 297, 300, 302, 305, 317, 319, 321, 323, 327, and 329, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:332. In some embodiments, the polypeptide sequence of the engineered polypeptide 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:332, and wherein the polypeptide sequence of the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from 9H/89M/93M/101S/160Y, 9H/89Q/160Y, 33N, 35H, 35M, 44T, 49P, 74K, 74R, 89L, 89M/93M/160V/249M, 89V, 89W, 107M, 107R, 107V, 108L, 117D, 140L, 140M, 140S, 141A, 141H, 142N, 142S, 144A, 156C, 156T, 168P, 168R, 203L, 207G, 234I, 234T, 270R, 270S, 273L, 274A, 274G, 292T, 296M, 296V, 296W, 297L, 297T, 300T, 302I, 302L, 302Q, 305Q, 305T, 317W, 319T, 321A, 321S, 321T, 323R, 323V, 327H, 327Q, 327S, 329R, and 329V, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:332. In some embodiments, the polypeptide sequence of the engineered polypeptide 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:332, and wherein the polypeptide sequence of the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from E9H/S89M/S93M/T101S/F160Y, E9H/S89Q/F160Y, P33N, A35H, A35M, A44T, S49P, N74K, N74R, S89L, S89M/S93M/F160V/K249M, S89V, S89W, E107M, E107R, E107V, I108L, E117D, I140L, I140M, I140S, T141A, T141H, K142N, K142S, R144A, Q156C, Q156T, H168P, H168R, H203L,

F207G, V234I, V234T, A270R, A270S, Y273L, D274A, D274G, S292T, N296M, N296V, N296W, S297L, S297T, D300T, V302I, V302L, V302Q, P305Q, P305T, L317W, V319T, P321A, P321S, P321T, S323R, S323V, P327H, P327Q, P327S, Q329R, and Q329V, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:332. In some embodiments, the engineered polypeptide comprises 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 NO:332. In some embodiments, the engineered polypeptide comprises a polypeptide sequence having at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO:332. In some embodiments, the engineered polypeptide comprises a polypeptide sequence having at least 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO:332.

[0020] In some additional embodiments, the present invention provides engineered polypeptides 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 NO:518, wherein the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from 35/273, 46/273, 111/159/273, 159, 159/273, 174, 199, 273, and 287, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:518. In some embodiments, the polypeptide sequence of the engineered polypeptide 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:518, and wherein the polypeptide sequence of the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from 35H/273L, 46L/273L, 111Q/159D/273L, 159D, 159D/273L, 174G, 199L, 273L, and 287P, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:518. In some embodiments, the polypeptide sequence of the engineered polypeptide 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:518, and wherein the polypeptide sequence of the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from A35H/Y273L, V46L/Y273L, E111Q/P159D/Y273L, P159D, P159D/Y273L, S174G, I199L, Y273L, and V287P, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:518. In some embodiments, the engineered polypeptide comprises 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 NO:518. In some embodiments, the engineered polypeptide comprises a polypeptide sequence having at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO:518. In some embodiments, the engineered polypeptide comprises a polypeptide sequence having at least 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO:518.

[0021] In some additional embodiments, the present invention provides engineered polypeptides 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 NO:332, wherein said engineered polypeptides have increased thermoactivity in wet organic solvents compared to the engineered polypeptide of SEQ ID NO:332, and wherein the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from 23/29/55, 34, 44, 74, 92, 108, 117, 140, 141, 144, 156, 160, 168, 215, 256, 269, 270, 274, 292, 296, 302, 319, 323, and 329, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:332. In some embodiments, the polypeptide sequence of the engineered polypeptide 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:332, and wherein the polypeptide sequence of the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from 23Y/29Y/55K, 34Y, 44N, 74K, 74R, 92D, 108L, 117R, 140L, 141H, 144V, 156T, 160V, 160W, 160Y, 168Q, 215A, 215S, 215T, 256G, 269T, 270Q, 270S, 274A, 292T, 296M, 296V, 296Y, 302L, 302Q, 302S, 319R, 323L, 329R, and 329V, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:332. In some embodiments, the polypeptide sequence of the engineered polypeptide 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:332, and wherein the polypeptide sequence of the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from Q23Y/D29Y/I55K, L34Y, A44N, N74K, N74R, E92D, I108L, E117R, I140L, T141H, R144V, Q156T, F160V, F160W, F160Y, H168Q, G215A, G215S, G215T, E256G, P269T, A270Q, A270S, D274A, S292T, N296M, N296V, N296Y, V302L, V302Q, V302S, V319R, S323L, Q329R, and Q329V, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:332. In some embodiments, the engineered polypeptide comprises 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 NO:332. In some embodiments, the engineered polypeptide comprises a polypeptide sequence having at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO:332. In some embodiments, the engineered polypeptide comprises a polypeptide sequence having at least 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO:332.

[0022] In some additional embodiments, the present invention provides engineered polypeptides 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 NO:518, wherein said engineered polypeptides have increased thermoactivity in wet organic solvents compared to the engineered polypeptide of SEQ ID NO:518, and wherein the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from 34/74/92, 35/273/323, 39, 46/273, 48, 55, 74/92, 126, 133, 138, 150, 160/215/329, 178, 199, 207, 273, and 292, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:518. In some embodiments, the polypeptide sequence of the engineered polypeptide 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:518, and wherein the polypeptide sequence of the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from 34Y/74K/92D, 35H/273L/323R, 39G, 46L/273L, 48G, 55V, 74K/92D, 126L, 126V, 133T, 138A, 150W, 160V/215S/329R, 178S, 199L, 207Y, 273L, and 292T, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:518. In some embodiments, the polypeptide sequence of the engineered polypeptide 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:518, and wherein the polypeptide sequence of the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from L34Y/N74K/E92D, A35H/Y273L/S323R, A39G, V46L/Y273L, P48G, I55V, N74K/E92D, T126L, T126V, S133T, R138A, F150W, Y160V/G215S/Q329R, T178S, I199L, F207Y, Y273L, and S292T, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:518. In some embodiments, the engineered polypeptide comprises 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 NO:518. In some embodiments, the engineered polypeptide comprises a polypeptide sequence having at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO:518. In some embodiments, the engineered polypeptide comprises a polypeptide sequence having at least 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO:518.

[0023] In some additional embodiments, the present invention provides engineered polypeptides 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 NO:656, wherein said engineered polypeptides have increased thermoactivity in wet organic solvents compared to the engineered polypeptide of SEQ ID NO:656, and wherein the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from 18, 35/39/55/138/150, 35/39/55/150, 35/39/126/150/273, 35/39/133, 35/39/133/150/273, 35/39/150, 35/39/150/273, 35/133/178, 35/138/150/273, 35/150/273, 35/178/273, 35/273, 39, 39/55/150/178/273, 39/55/150/273, 39/126/150, 39/133/150, 39/133/178/273, 39/133/273, 39/150/178, 39/150/178/207, 39/150/178/207/273, 39/150/273, 39/178, 39/178/273, 39/207, 39/207/273, 39/273, 44, 55, 55/150, 88, 126, 126/133/178, 126/138/178, 126/178/207, 132, 138/150/178, 141, 150, 150/178/273/323, 150/178/323, 150/207/273, 150/273, 168, 178/273, 239, 261, 263, 269, 273, 297, and 312, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:656. In some embodiments, the polypeptide sequence of the engineered polypeptide 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:656, and wherein the polypeptide sequence of the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from 18R, 35H/39G/55V/138A/150W, 35H/39G/55V/150W, 35H/39G/126V/150W/273L, 35H/39G/133T, 35H/39G/133T/150W/273L,

35H/39G/150W, 35H/39G/150W/273L, 35H/133T/178S, 35H/138A/150W/273L, 35H/150W/273L, 35H/178S/273L, 35H/273L, 39G, 39G/55V/150W/178S/273L, 39G/55V/150W/273L, 39G/126V/150W, 39G/133T/150W, 39G/133T/178S/273L, 39G/133T/273L, 39G/150W/178S, 39G/150W/178S/207Y, 39G/150W/178S/207Y/273L, 39G/150W/273L, 39G/178S, 39G/178S/273L, 39G/207Y, 39G/207Y/273L, 39G/273L, 44L, 44Q, 55F, 55V, 55V/150W, 88W, 126V, 126V/133T/178S, 126V/138A/178S, 126V/178S/207Y, 132D, 132E, 138A/150W/178S, 141S, 150W, 150W/178S/273L/323R, 150W/178S/323R, 150W/207Y/273L, 150W/273L, 168D, 178S/273L, 239N, 261R, 263E, 269V, 273L, 297E, 297R, and 312A, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:656. In some embodiments, the polypeptide sequence of the engineered polypeptide 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:656, and wherein the polypeptide sequence of the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from L18R, A35H/A39G/I55V/R138A/F150W, A35H/A39G/I55V/F150W, A35H/A39G/T126V/F150W/Y273L, A35H/A39G/S133T, A35H/A39G/S133T/F150W/Y273L, A35H/A39G/F150W, A35H/A39G/F150W/Y273L, A35H/S133T/T178S, A35H/R138A/F150W/Y273L, A35H/F150W/Y273L, A35H/T178S/Y273L, A35H/Y273L, A39G, A39G/I55V/F150W/T178S/Y273L, A39G/I55V/F150W/Y273L, A39G/T126V/F150W, A39G/S133T/F150W, A39G/S133T/T178S/Y273L, A39G/S133T/Y273L, A39G/F150W/T178S, A39G/F150W/T178S/F207Y, A39G/F150W/T178S/F207Y/Y273L, A39G/F150W/Y273L, A39G/T178S, A39G/T178S/Y273L, A39G/F207Y, A39G/F207Y/Y273L, A39G/Y273L, A44L, A44Q, I55F, I55V, I55V/F150W, F88W, T126V, T126V/S133T/T178S, T126V/R138A/T178S, T126V/T178S/F207Y, S132D, S132E, R138A/F150W/T178S, T141S, F150W, F150W/T178S/Y273L/S323R, F150W/T178S/S323R, F150W/F207Y/Y273L, F150W/Y273L, H168D, T178S/Y273L, G239N, E261R, I263E, P269V, Y273L, S297E, S297R, and R312A, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:656. In some embodiments, the engineered polypeptide comprises 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 NO:656. In some embodiments, the engineered polypeptide comprises a polypeptide sequence having at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO:656. In some embodiments, the engineered polypeptide comprises a polypeptide sequence having at least 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO:656.

[0024] In some additional embodiments, the present invention provides engineered polypeptides 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 NO:2, wherein the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from 9/20/23/39/55/89/134/150/155/160/178/199/244/273/284, 9/20/23/89/134/155/160/199/244/284,

9/20/23/89/134/155/160/244/284, 20/23/134/155/244/284, and 155/284, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:2. In some embodiments, the polypeptide sequence of the engineered polypeptide 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:2, and wherein the polypeptide sequence of the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from

9H/20T/23Q/39G/55V/89Q/134S/150W/155W/160Y/178S/199L/244G/273L/284A,

9H/20T/23Q/89Q/134S/155W/160Y/199L/244G/284A,

9H/20T/23Q/89Q/134S/155W/160Y/244G/284A, 20T/23Q/134S/155W/244G/284A, and 155W/284A,

wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:2. In some embodiments, the polypeptide sequence of the engineered polypeptide 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:2, and wherein the polypeptide sequence of the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from

E9H/Y20T/Y23Q/A39G/I55V/S89Q/T134S/F150W/L155W/F160Y/T178S/I199L/P244G/Y273L/G284

A, E9H/Y20T/Y23Q/S89Q/T134S/L155W/F160Y/I199L/P244G/G284A,

E9H/Y20T/Y23Q/S89Q/T134S/L155W/F160Y/P244G/G284A,

Y20T/Y23Q/T134S/L155W/P244G/G284A, and L155W/G284A, wherein the amino acid positions of

said polypeptide sequence are numbered with reference to SEQ ID NO:2. In some embodiments, the engineered polypeptide comprises 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 NO:2. In some embodiments, the engineered polypeptide comprises a polypeptide sequence having at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO:2. In some embodiments, the engineered polypeptide comprises a polypeptide sequence having at least 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO:2.

[0025] In some embodiments of the engineered polypeptides having transaminase activity of the present disclosure, 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 fold, 2 fold, 3 fold, 5 fold, 7 fold, 10 fold, 12 fold, 15 fold, 20 fold, or greater the activity of SEQ ID NO:2 under suitable reaction conditions. In some embodiments, the engineered polypeptide is capable of converting compound (2) to compound (1) with increased activity relative to SEQ ID NO:2 in which the suitable reaction conditions comprise compound (1) at a loading of at least 20 g/L, wet isopropyl acetate (IPAc), 0.55 M isopropylamine (IPM), and 50°C.

[0026] In some embodiments of the present disclosure, 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.

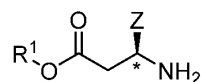
[0027] The present invention also provides polynucleotide sequences encoding at least one engineered polypeptide provided herein. In some embodiments, the polynucleotide sequence encoding at least one engineered polypeptide, comprises a polynucleotide sequence having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NOS: 1, 5, 331, 517 and/or 655. In some embodiments, the polynucleotide sequence encoding at least one engineered polypeptide, comprises a polynucleotide 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: 1, 5, 331, 517 and/or 655, wherein the polynucleotide sequence of said engineered polypeptide comprises at least one substitution at one or more positions. In some further embodiments, the polynucleotide sequence encoding at least one engineered polypeptide comprises 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: 1, 5, 331, 517 and/or 655, or a functional fragment thereof. In some additional embodiments, the polynucleotide sequence comprises a polynucleotide sequence forth in the odd numbered sequences of SEQ ID NOS: 3-845.

[0028] In some additional embodiments, the engineered polynucleotide sequence is operably linked to a control sequence. In some embodiments, the control sequence comprises a promoter. In some additional embodiments, the promoter is a heterologous promoter. In some further embodiments, the control sequence comprises more than one control sequence. In some embodiments, one of the control sequences comprises a promoter and the other control sequence comprises a different control sequence. In some embodiments, the engineered polynucleotide sequence is codon optimized. The present invention also provides expression vectors comprising at least one polynucleotide sequence provided herein. The present invention also provides host cells comprising at least one expression vector provided herein. In some embodiments, the present invention provides host cells comprising at least one polynucleotide sequence provided herein. In some embodiments, the host cell is a prokaryotic cell, while in some other embodiments, the host cell is a eukaryotic cell. In some additional embodiments, the host cell is *E. coli*.

[0029] The present invention also provides methods of producing an engineered transaminase in a host cell, comprising culturing the host cell provided herein, in a culture medium under suitable conditions, such that at least one engineered transaminase provided herein is produced. In some embodiments, the methods further comprise the step of recovering at least one engineered transaminase from the culture medium and/or the host cell. In some additional embodiments, the methods further comprise the step of purifying the at least one transaminase obtained using the methods of the present invention.

[0030] The present disclosure also provides processes for using the engineered transaminase polypeptides disclosed herein for the preparation of a wide range of chiral amine compounds. In some

embodiments, the present disclosure provides a process for preparing a compound of structural Formula (I):

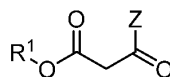


(I)

having the indicated stereochemical configuration at the stereogenic center marked with an *; in an enantiomeric excess of at least 70% over the opposite enantiomer, wherein

Z is a linear C₁₋₆ alkyl or a branched C₁₋₆ alkyl, optionally substituted with one or more hydroxy, cyano, or nitro groups ; and

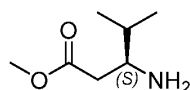
R¹ is C₁₋₆ alkyl, aryl-C₁₋₂ alkyl, heteroaryl-C₁₋₂ alkyl, or a 5- to 6-membered heterocyclic ring system optionally containing an additional heteroatom selected from O, S, and N, the heterocyclic ring being unsubstituted or substituted with one to three substituents independently selected from oxo, hydroxy, halogen, C₁₋₄ alkoxy, and C₁₋₄ alkyl; the process comprising the step of contacting a prochiral keto ester of structural Formula (II):



(II)

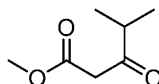
with an engineered polypeptide as disclosed herein in the presence of an amino group donor under suitable reaction conditions.

[0031] In some embodiments, the present disclosure provides a process of preparing compound (1);



(1)

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



(2)

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

[0032] In some embodiments, the processes using the engineered polypeptides disclosed herein can be carried out wherein the chiral amine compound of Formula (I) or the Compound (1), is produced in at least 70%, 80 %, 90%, 97%, 98%, 99% or greater enantiomeric excess.

[0033] Any of the processes disclosed herein using the engineered polypeptides for the preparation of compounds of Formula (I) or compound (1) can be carried out under a range of suitable reaction conditions, including but not limited to, ranges of amino group donor, temperature, wet organic solvent system, substrate loading, polypeptide loading, cofactor loading, and reaction time. In one example, in some embodiments, the preparation of compounds of Formula (I) or compound (1) can be carried out wherein the suitable reaction conditions comprise: (a) substrate loading of about 10 to 220 g/L of substrate compound (*e.g.*, compound (2)); (b) of about 0.5 g/L to 25 g/L engineered polypeptide; (c) IPM concentration of about 0.1 to 3 M; (d) a wet organic solvent; and (e) temperature of about 30°C to 60°C. In some embodiments, the suitable reaction conditions comprise: (a) about 20 g/L of substrate compound (*e.g.*, compound (2)); (b) about 20 g/L engineered polypeptide; (c) wet IPAc; (d) about 0.55 M IPM; and (e) about 50°C.

[0034] In some embodiments, the processes using the engineered polypeptides disclosed herein can be carried out wherein the amino group donor is selected from IPM, alanine, 3-aminobutyric acid, or methylbenzylamine. In some embodiments, the amino group donor is IPM.

[0035] In some embodiments, the wet organic solvent is selected from wet IPAc, wet toluene, wet methyl tertiary-butyl ether (MTBE), wet acetonitrile (MeCN), or wet isopropyl alcohol (IPA). In some embodiments, the wet organic solvent is wet IPAc.

[0036] In some embodiments, the processes using the engineered polypeptides disclosed herein can be carried out wherein the process comprises further steps of isolating the product compounds of Formula (I) or compound (1), from the reaction.

[0037] Further guidance on the choice of engineered polypeptides, their preparation, the choice of substrates, and parameters for carrying out the processes are further described in the more detailed description and Examples that follow.

DESCRIPTION OF THE INVENTION

[0038] The present invention provides engineered transaminase polypeptides and compositions thereof, as well as polynucleotides encoding the engineered transaminase polypeptides. Methods for producing transaminase enzymes are also provided. In some embodiments, the engineered transaminase polypeptides are optimized to provide improved enzyme properties under minimal aqueous conditions in wet organic solvents.

[0039] Unless defined otherwise, all technical and scientific terms used herein generally have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. Generally, the nomenclature used herein and the laboratory procedures of cell culture, molecular genetics, microbiology, organic chemistry, analytical chemistry and nucleic acid chemistry described

below are those well-known and commonly employed in the art. Such techniques are well-known and described in numerous texts and reference works well known to those of skill in the art. Standard techniques, or modifications thereof, are used for chemical syntheses and chemical analyses.

Accordingly, the following terms are intended to have the meanings provided herein.

[0040] Although any suitable methods and materials similar or equivalent to those described herein find use in the practice of the present invention, some methods and materials are described herein. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary, depending upon the context they are used by those of skill in the art. Accordingly, the terms defined immediately below are more fully described by reference to the invention as a whole.

[0041] It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the present invention. The section headings used herein are for organizational purposes only and not to be construed as limiting the subject matter described. 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.

[0042] As used in this specification and the appended claims, the singular forms “a”, “an” and “the” include plural referents unless the context clearly indicates otherwise. Thus, for example, reference to “a polypeptide” includes more than one polypeptide.

[0043] Similarly, “comprise,” “comprises,” “comprising” “include,” “includes,” “including,” “have,” and “having” are interchangeable and not intended to be limiting. Thus, 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).

[0044] 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.”

[0045] All patents, patent applications, articles and publications mentioned herein, both supra and infra, are hereby expressly incorporated herein by reference.

Definitions and Abbreviations

[0046] The abbreviations used for the genetically encoded amino acids are conventional and are as follows: alanine (Ala or A), arginine (Arg or R), asparagine (Asn or N), aspartate (Asp or D), cysteine (Cys or C), glutamate (Glu or E), glutamine (Gln or Q), glycine (Gly or G), histidine (His or H), isoleucine (Ile or I), leucine (Leu or L), lysine (Lys or K), methionine (Met or M), phenylalanine (Phe or

F), proline (Pro or P), serine (Ser or S), threonine (Thr or T), tryptophan (Trp or W), tyrosine (Tyr or Y), and valine (Val or V).

[0047] 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_α). 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.

[0048] 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.

[0049] As used herein, the term “about” means an acceptable error for a particular value. In some instances “about” means within 0.05%, 0.5%, 1.0%, or 2.0%, of a given value range. In some instances, “about” means within 1, 2, 3, or 4 standard deviations of a given value.

[0050] As used herein, “EC” number refers to the Enzyme Nomenclature of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB). The IUBMB biochemical classification is a numerical classification system for enzymes based on the chemical reactions they catalyze.

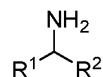
[0051] As used herein, “ATCC” refers to the American Type Culture Collection whose biorepository collection includes genes and strains.

[0052] As used herein, “NCBI” refers to National Center for Biological Information and the sequence databases provided therein.

[0053] “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, as well as polymers comprising D- and L-amino acids, and mixtures of D- and L-amino acids.

[0054] “Transaminase” or “aminotransferase” are used interchangeably herein to refer to a polypeptide having an enzymatic capability of transferring an amino group (-NH_2), a pair of electrons, and a proton from the primary amine of an amino group donor compound to the carbonyl group (C=O) of an amine acceptor compound, thereby converting the amino group 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.

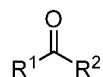
[0055] “Amino group donor” or “amino donor” used interchangeably herein to refer to an amino group containing compound which is capable of donating an amino group to an acceptor carbonyl compound (i.e., an amino group acceptor), thereby becoming a carbonyl by-product. Amino group donors have the general structural formula,



in which each of R^1 , and R^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^1 can be the same or different from R^2 in structure or chirality. The groups R^1 and R^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.

[0056] “Chiral amine” refers to amines of general formula $\text{R}^1\text{-CH(NH}_2\text{)-R}^2$ and is employed herein in its broadest sense, including a wide variety of aliphatic and alicyclic compounds of different, and mixed, functional types, characterized by the presence of a primary amino group bound to a secondary carbon atom which, in addition to a hydrogen atom, carries either (i) a divalent group forming a chiral cyclic structure, or (ii) two substituents (other than hydrogen) differing from each other in structure or chirality. Divalent groups forming a chiral cyclic structure include, for example, 2-methylbutane-1,4-diyl, pentane-1,4-diyl, hexane-1,4-diyl, hexane-1,5-diyl, 2-methylpentane-1,5-diyl. The two different substituents on the secondary carbon atom (R^1 and R^2 above) also can vary widely and include alkyl, arylalkyl, aryl, halo, hydroxy, lower alkyl, lower alkyloxy, lower alkylthio, cycloalkyl, carboxy, carbalkyloxy, carbamoyl, mono- and di-(lower alkyl) substituted carbamoyl, trifluoromethyl, phenyl, nitro, amino, mono- and di-(lower alkyl) substituted amino, alkylsulfonyl, arylsulfonyl, alkylcarboxamido, arylcarboxamido, *etc.*, as well as alkyl, arylalkyl, or aryl substituted by the foregoing.

[0057] “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,



wherein R¹ and R² are defined above for the amino group donor.

[0058] “Amino acceptor” and “amine acceptor,” “keto substrate,” “keto ester 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 disclosure, the amino acceptor compound for the transaminase can include, among others, the compound of Formula (II) and the compound (2), as further described herein.

[0059] “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 B₆ family compounds PLP, PN, PL, PM, PNP, and PMP, which are sometimes also referred to as coenzymes.

[0060] “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 B₆). 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 converted to pyridoxamine phosphate. Pyridoxal-5'-phosphate regenerated by reaction with a different keto compound (the amino acceptor). 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 B₆ family, including pyridoxine (PN), pyridoxal (PL), pyridoxamine (PM), and their phosphorylated counterparts; pyridoxine phosphate (PNP), and pyridoxamine phosphate (PMP).

[0061] “Amino acids” are referred to herein by either their commonly known three-letter symbols or by the one-letter symbols recommended by IUPAC-IUB Biochemical Nomenclature Commission.

Nucleotides, likewise, may be referred to by their commonly accepted single letter codes.

[0062] 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., I, 179:125-142 [1984]). 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).

[0063] As used herein, “acidic amino acid or residue” refers to a hydrophilic amino acid or residue having a side chain exhibiting a pK_a 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).

[0064] As used herein, “basic amino acid or residue” refers to a hydrophilic amino acid or residue having a side chain exhibiting a pK_a 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).

[0065] 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).

[0066] 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).

[0067] 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. 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).

[0068] 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.

[0069] 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).

[0070] 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 disclosure, L-Cys (C) is categorized into its own unique group.

[0071] 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).

[0072] 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).

[0073] As used herein, “polynucleotide” and “nucleic acid” refer to two or more nucleotides that are covalently linked together. The polynucleotide may be wholly comprised of ribonucleotides (i.e., RNA), wholly comprised of 2'-deoxyribonucleotides (i.e., DNA), or comprised of mixtures of ribo- and 2'-deoxyribonucleotides. 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, such as, for example, inosine, xanthine, hypoxanthine, etc. In some embodiments, such modified or synthetic nucleobases are nucleobases encoding amino acid sequences.

[0074] As used herein, “nucleoside” refers to glycosylamines comprising a nucleobase (i.e., a nitrogenous base), and a 5-carbon sugar (e.g., ribose or deoxyribose). Non-limiting examples of nucleosides include cytidine, uridine, adenosine, guanosine, thymidine, and inosine. In contrast, the term “nucleotide” refers to the glycosylamines comprising a nucleobase, a 5-carbon sugar, and one or more phosphate groups. In some embodiments, nucleosides can be phosphorylated by kinases to produce nucleotides.

[0075] As used herein, “nucleoside diphosphate” refers to glycosylamines comprising a nucleobase (i.e., a nitrogenous base), a 5-carbon sugar (e.g., ribose or deoxyribose), and a diphosphate (i.e., pyrophosphate) moiety. In some embodiments herein, “nucleoside diphosphate” is abbreviated as “NDP.” Non-limiting examples of nucleoside diphosphates include cytidine diphosphate (CDP), uridine diphosphate (UDP), adenosine diphosphate (ADP), guanosine diphosphate (GDP), thymidine diphosphate (TDP), and inosine diphosphate. The terms “nucleoside” and “nucleotide” may be used interchangeably in some contexts.

[0076] “Coding sequence” refers to that portion of a nucleic acid (e.g., a gene) that encodes an amino acid sequence of a protein.

[0077] As used herein, the terms “biocatalysis,” “biocatalytic,” “biotransformation,” and “biosynthesis” refer to the use of enzymes to perform chemical reactions on organic compounds.

[0078] “Naturally-occurring” or “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, “recombinant,” “engineered,” “non-naturally occurring,” and “variant,” when used with reference to 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 cell, nucleic acid or polypeptide is identical a naturally occurring cell, nucleic acid or polypeptide, 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] The term “percent (%) sequence identity” is used herein to refer to comparisons among polynucleotides or polypeptides, 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 may be calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences 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. Alternatively, the percentage may be 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. Those of skill in the art appreciate that there are many established algorithms available to align two sequences. Optimal alignment of sequences for comparison can be conducted by any suitable method, including, but not limited to the local homology algorithm of Smith and Waterman (Smith and Waterman, *Adv. Appl. Math.*, 2:482 [1981]), by the homology alignment algorithm of Needleman and Wunsch (Needleman and Wunsch, *J. Mol. Biol.*, 48:443 [1970]), by the search for similarity method of Pearson and Lipman (Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444 [1988]), by computerized implementations of these algorithms (e.g., GAP, BESTFIT, FASTA, and TFASTA in the GCG Wisconsin Software Package), or by visual inspection, as known in the art. Examples of algorithms that are suitable for determining percent sequence identity and sequence similarity include, but are not limited to the BLAST and BLAST 2.0 algorithms, which are described by Altschul et al. (See

Altschul et al., *J. Mol. Biol.*, 215: 403-410 [1990]; and Altschul et al., *Nucl. Acids Res.*, 3389-3402 [1977], respectively). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information website. This algorithm involves 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 (See, 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 word length (*W*) of 3, an expectation (*E*) of 10, and the BLOSUM62 scoring matrix (See, Henikoff and Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 [1989]). Exemplary determination of sequence alignment and % sequence identity can employ the BESTFIT or GAP programs in the GCG Wisconsin Software package (Accelrys, Madison WI), using default parameters provided.

[0081] “Reference sequence” refers to a defined sequence used as a basis for a sequence comparison. 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 or polypeptides over a “comparison window” to identify and compare local regions of sequence similarity. In some embodiments, a “reference sequence” can be based on a primary amino acid sequence, where the reference sequence is a sequence that can have one or more changes in the primary sequence. For instance, a “reference sequence based on SEQ ID NO:2 having at the residue corresponding to X9 a histidine” refers to a reference sequence in which the corresponding residue at X9 in SEQ ID NO:2, which is a tyrosine, has been changed to histidine.

[0082] “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.

[0083] “Corresponding to”, “reference to” or “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.

[0084] 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. In some cases, the reference sequence has a histidine tag, but the numbering is maintained relative to the equivalent reference sequence without the histidine tag. 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 X₉₃ as compared to SEQ ID NO:4” refers to a difference of the amino acid residue at the polypeptide position corresponding to position 93 of SEQ ID NO:4. Thus, if the reference polypeptide of SEQ ID NO:4 has a serine at position 93, then a “residue difference at position X₉₃ as compared to SEQ ID NO:4” an amino acid substitution of any residue other than serine at the position of the polypeptide corresponding to position 93 of SEQ ID NO:4. In most instances herein, the specific amino acid residue difference at a position is indicated as “X_nY” where “X_n” 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 (*e.g.*, in the Tables presented in the Examples), the present invention also provides specific amino acid differences denoted by the conventional notation “A_nB”, 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., X307H/X307P or X307H/P). The slash may also be used to indicate multiple substitutions within a given variant (i.e., there is more than one substitution present in a given sequence, such as in a combinatorial variant). In some embodiments, the present invention includes engineered polypeptide sequences comprising one or more amino acid differences comprising conservative or non-conservative amino acid substitutions. In some additional embodiments, the present invention provides engineered polypeptide sequences comprising both conservative and non-conservative amino acid substitutions.

[0085] As used herein, “conservative amino acid substitution” refers to a substitution of a residue with a different residue having a similar side chain, 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. 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 an hydroxyl side chain is substituted with another amino acid with an hydroxyl side chain (e.g., serine and threonine); an amino acid 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 basic 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.

[0086] “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.

[0087] “Deletion” refers to modification to 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 reference enzyme while retaining enzymatic activity and/or retaining the improved properties of an engineered transaminase 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. Deletions are typically indicated by “-” in amino acid sequences.

[0088] “Insertion” refers to modification to the polypeptide by addition of one or more amino acids from 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 other improved 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.

[0089] 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.

[0090] A “functional fragment” and “biologically active fragment” are used interchangeably herein to refer to a polypeptide that has an amino-terminal and/or carboxy-terminal deletion(s) and/or internal deletions, but where the remaining amino acid sequence is identical to the corresponding positions in the sequence to which it is being compared (e.g., a full-length engineered transaminase of the present invention) and that retains substantially all of the activity of the full-length polypeptide.

[0091] “Isolated polypeptide” refers to a polypeptide which is substantially separated from other contaminants that naturally accompany it, e.g., protein, 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 improved transaminase enzyme can be an isolated polypeptide.

[0092] “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. However, in some embodiments, the composition comprising transaminase comprises transaminase that is less than 50% pure (e.g., about 10%, about 20%, about 30%, about 40%, or about 50%). Generally, a substantially pure transaminase composition will comprise about 60% or more, about 70% or more, about 80% or more, about 90% or more, about 95% or more, and about 98% or more of all macromolecular species by mole or % weight present in the composition. In some embodiments, the object species is purified to essential homogeneity (*i.e.*, contaminant species cannot be detected in the composition by conventional detection methods) wherein the composition consists

essentially of a single macromolecular species. Solvent species, small molecules (<500 Daltons), and elemental ion species are not considered macromolecular species. In some embodiments, the isolated improved transaminase polypeptide is a substantially pure polypeptide composition.

[0093] “Stereoselectivity” refers to the preferential formation in a chemical or enzymatic reaction of one stereoisomer over another. Stereoselectivity can be partial, where the formation of one stereoisomer is favored over the other, or it may be complete where only one stereoisomer is formed. When the stereoisomers are enantiomers, the stereoselectivity is referred to as enantioselectivity, the fraction (typically reported as a percentage) of one enantiomer in the sum of both. It is commonly alternatively reported in the art (typically as a percentage) as the enantiomeric excess (e.e.) calculated therefrom according to the formula $[\text{major enantiomer} - \text{minor enantiomer}] / [\text{major enantiomer} + \text{minor enantiomer}]$. Where the stereoisomers are diastereoisomers, the stereoselectivity is referred to as diastereoselectivity, the fraction (typically reported as a percentage) of one diastereomer in a mixture of two diastereomers, commonly alternatively reported as the diastereomeric excess (d.e.). Where a mixture contains more than two diastereomers it is common to report the ratio of diastereomers or “diastereomeric ratio” rather than diastereomeric excess. Enantiomeric excess and diastereomeric excess are types of stereomeric excess. “Highly stereoselective” refers to a transaminase polypeptide that is capable of converting the substrate to the corresponding chiral amine product with at least about 85% stereomeric excess.

[0094] As used herein, “regioselectivity” and “regioselective reaction” refer to a reaction in which one direction of bond making or breaking occurs preferentially over all other possible directions. Reactions can be completely (100%) regioselective if the discrimination is complete, substantially regioselective (at least 75%), or partially regioselective (x%, wherein the percentage is set dependent upon the reaction of interest), if the product of reaction at one site predominates over the product of reaction at other sites.

[0095] As used herein, “chemoselectivity” refers to the preferential formation in a chemical or enzymatic reaction of one product over another.

[0096] As used herein, “improved enzyme property” refers to at least one improved property of an enzyme. In some embodiments, the present invention provides engineered transaminase polypeptides that exhibit an improvement in any enzyme property as compared to a reference transaminase polypeptide and/or a wild-type transaminase polypeptide, and/or another engineered transaminase polypeptide. Thus, the level of “improvement” can be determined and compared between various transaminase polypeptides, including wild-type, as well as engineered transaminase. Improved properties include, but are not limited, to such properties as increased protein expression, increased thermoactivity, increased thermostability, increased pH activity, increased stability, increased enzymatic activity, increased substrate specificity or affinity, increased specific activity, increased resistance to substrate or end-product inhibition, increased chemical stability, improved chemoselectivity, improved solvent stability, increased tolerance to acidic pH, increased tolerance to proteolytic activity (i.e., reduced sensitivity to proteolysis), reduced aggregation, increased solubility, and altered temperature profile. In

additional embodiments, the term is used in reference to the at least one improved property of transaminase enzymes. In some embodiments, the present invention provides engineered transaminase polypeptides that exhibit an improvement in any enzyme property as compared to a reference transaminase polypeptide and/or a wild-type transaminase polypeptide, and/or another engineered transaminase polypeptide. Thus, the level of “improvement” can be determined and compared between various transaminase polypeptides, including wild-type, as well as engineered transaminase.

[0097] “Increased enzymatic activity” and “enhanced catalytic activity” refer to an improved property of the engineered 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 enzyme) as compared to the reference enzyme. In some embodiments, the terms refer to an improved property of engineered transaminase polypeptides provided herein, 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. In some embodiments, the terms are used in reference to improved transaminase enzymes provided herein. Exemplary methods to determine enzyme activity of the engineered transaminase of the present invention are provided in the Examples. Any property relating to enzyme activity may be affected, including the classical enzyme properties of K_m , V_{max} or k_{cat} , changes of which can lead to increased enzymatic activity. For example, improvements in enzyme activity can be from about 1.1 fold the enzymatic activity of the corresponding wild-type enzyme, to as much as 2-fold, 5-fold, 10-fold, 20-fold, 25-fold, 50-fold, 75-fold, 100-fold, 150-fold, 200-fold or more enzymatic activity than the naturally occurring transaminase or another engineered transaminase from which the transaminase polypeptides were derived.

[0098] “Increased thermoactivity” or “improved thermoactivity” is used herein to refer to a polypeptide with increased catalytic activity relative to a reference or wild-type protein under high temperature conditions.

[0099] “Conversion” refers to the enzymatic conversion of the substrate(s) to the corresponding product(s). “Percent conversion” refers to the percent of the substrate that is converted to the product within a period of time under specified conditions. Thus, the “enzymatic activity” or “activity” of a transaminase polypeptide can be expressed as “percent conversion” of the substrate to the product.

[0100] “Thermostable” refers to a transaminase polypeptide that maintains similar activity (more than 60% to 80% for example) after exposure to elevated temperatures (e.g., 40-80°C) for a period of time (e.g., 0.5-24 hrs) compared to the wild-type enzyme.

[0101] “Solvent stable” refers to a transaminase polypeptide that maintains similar activity (more than e.g., 60% to 80%) after exposure to varying concentrations (e.g., 5-99%) of solvent (ethanol, IPA,

dimethylsulfoxide (DMSO), tetrahydrofuran, 2-methyltetrahydrofuran, acetone, toluene, butyl acetate, methyl tert-butyl ether, etc.) for a period of time (e.g., 0.5-24 hrs) compared to the wild-type enzyme.

[0102] 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.

[0103] The term “stringent hybridization conditions” is used herein to refer to conditions under which nucleic acid hybrids are stable. As known to those of skill in the art, the stability of hybrids is reflected in the melting temperature (T_m) of the hybrids. In general, the stability of a hybrid is a function of ion strength, temperature, G/C content, and the presence of chaotropic agents. The T_m values for polynucleotides can be calculated using known methods for predicting melting temperatures (See e.g., Baldino et al., *Meth. Enzymol.*, 168:761-777 [1989]; Bolton et al., *Proc. Natl. Acad. Sci. USA* 48:1390 [1962]; Bresslauer et al., *Proc. Natl. Acad. Sci. USA* 83:8893-8897 [1986]; Freier et al., *Proc. Natl. Acad. Sci. USA* 83:9373-9377 [1986]; Kierzek et al., *Biochem.*, 25:7840-7846 [1986]; Rychlik et al., *Nucl. Acids Res.*, 18:6409-6412 [1990] (erratum, *Nucl. Acids Res.*, 19:698 [1991]); Sambrook et al., supra; Suggs et al., 1981, in *Developmental Biology Using Purified Genes*, Brown et al. [eds.], pp. 683-693, Academic Press, Cambridge, MA [1981]; and Wetmur, *Crit. Rev. Biochem. Mol. Biol.* 26:227-259 [1991]). In some embodiments, the polynucleotide encodes the polypeptide disclosed herein and hybridizes under defined conditions, such as moderately stringent or highly stringent conditions, to the complement of a sequence encoding an engineered transaminase enzyme of the present invention.

[0104] “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 5X SSC containing 0.1% (w:v) SDS at 65°C and washing in 0.1x SSC containing 0.1% SDS at 65°C. Other

high stringency hybridization conditions, as well as moderately stringent conditions, are described in the references cited above.

[0105] “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 more efficiently expressed in the organism of interest. 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 transaminases enzymes may be codon optimized for optimal production from the host organism selected for expression.

[0106] As used herein, “preferred,” “optimal,” and “high codon usage bias” codons when used alone or in combination refer(s) 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, Peden, University of Nottingham; McInerney, *Bioinform.*, 14:372-73 [1998]; Stenico et al., *Nucl. Acids Res.*, 22:2437-46 [1994]; and Wright, *Gene* 87:23-29 [1990]). Codon usage tables are available for many different organisms (See e.g., Wada et al., *Nucl. Acids Res.*, 20:2111-2118 [1992]; Nakamura et al., *Nucl. Acids Res.*, 28:292 [2000]; Duret, et al., *supra*; Henaut and Danchin, in *Escherichia coli and Salmonella*, Neidhardt, et al. (eds.), ASM Press, Washington D.C., p. 2047-2066 [1996]). 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., Mount, *Bioinformatics: Sequence and Genome Analysis*, Chapter 8, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. [2001]; Uberbacher, *Meth. Enzymol.*, 266:259-281 [1996]; and Tiwari et al., *Comput. Appl. Biosci.*, 13:263-270 [1997]).

[0107] “Control sequence” refers herein to include all components, which are necessary or advantageous for the expression of a polynucleotide and/or polypeptide of the present disclosure. Each control sequence may be native or foreign to the nucleic acid sequence encoding the polypeptide. Such control sequences include, but are not limited to, a leader, polyadenylation sequence, propeptide sequence,

promoter, signal peptide sequence, initiation sequence, and transcription terminator. At a minimum, the control sequences include a promoter, and transcriptional and translational stop signals. The control sequences may be provided with linkers for the purpose of introducing specific restriction sites facilitating ligation of the control sequences with the coding region of the nucleic acid sequence encoding a polypeptide.

[0108] “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.

[0109] “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 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.

[0110] “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.*, (C1-C4) alkyl refers to an alkyl of 1 to 4 carbon atoms.

[0111] “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.

[0112] “Arylalkyl” refers to an alkyl substituted with an aryl moiety. Representative arylalkyl groups include benzyl, phenethyl and the like.

[0113] “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.*, indoliziny or benzothienyl). For multiple condensed rings, at least one of the rings is aromatic.

[0114] “Heteroarylalkyl” refers to an alkyl substituted with a heteroaryl moiety as defined herein.

[0115] “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.*, indoliny, dihydrobenzofuran or quinuclidiny). 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,

[0116] “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.

[0117] “Amino” refers to the group -NH₂. 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 limited to, dimethylamino, diethylamino, trimethylammonium, triethylammonium, methylsulfonylamino, furanyl-oxy-sulfamino, and the like.

[0118] “Carboxy” refers to -COOH.

[0119] “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.

[0120] “Hydroxy” refers to -OH.

[0121] “Cyano” refers to -CN.

[0122] “Halogen” or “halo” refers to fluoro, chloro, bromo and iodo.

[0123] “Nitro” refers to -NO₂

[0124] “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.

[0125] “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.

[0126] "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.

[0127] "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.

[0128] "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 can be found in Wuts and Greene, *Greene's Protective Groups in Organic Synthesis*, 4th Ed., Wiley Interscience (2006), and Harrison *et al.*, *Compendium of Synthetic Organic Methods*, Vols. 1-8, 1971-1996, John Wiley & Sons, NY. 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. 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.

[0129] "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 can be found, for example in *Organic Chemistry*, 2d ed., Francis Carey (1992), pages 328-331; *Introduction to Organic Chemistry*, 2d ed., Andrew Streitwieser and Clayton Heathcock (1981), pages 169-171; and *Organic Chemistry*, 5th

Ed., John McMurry, Brooks/Cole Publishing (2000), pages 398 and 408; all of which are incorporated herein by reference.

[0130] “Suitable reaction conditions” refers 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 disclosure capable of converting a substrate to the desired amino product compound, *e.g.*, converting compound (2) to compound (1). Exemplary “suitable reaction conditions” are provided in the present disclosure and illustrated by the Examples.

[0131] “Loading”, such as in “compound loading” or “enzyme loading” refers to the concentration or amount of a component in a reaction mixture at the start of the reaction.

[0132] “Substrate” in the context of a biocatalyst mediated process refers to the compound or molecule acted on by the biocatalyst.

[0133] “Product” in the context of a biocatalyst mediated process refers to the compound or molecule resulting from the action of the biocatalyst. For example, an exemplary product for the transaminase biocatalyst in the processes disclosed herein is compound (1).

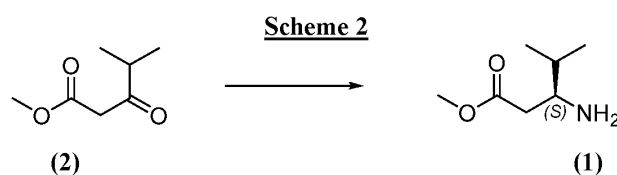
[0134] “Wet organic solvent” is defined as an organic solvent that includes up to the maximum amount of water that can be dissolved in that organic solvent. For example, the solubility of water in IPAc is acknowledged by those skilled in the art to be about 1.9 wt%, and wet IPAc containing up to about 1.9 wt% can be generated as described in the Examples, below, or by other suitable methods. Wet organic solvents, as used in this application, create minimal aqueous conditions without the addition of any water to a chemical reaction, such as the screening assay reactions described in the Examples.

[0135] “Minimal aqueous conditions” are those conditions created by use of wet organic solvents, without the addition of extra water. Minimal aqueous conditions use an organic solvent, containing up to the maximum amount of water that can be dissolved in that organic solvent, as the reaction condition or solvent system. As compared to aqueous conditions, minimal aqueous conditions do not include reaction conditions or solvent systems where water is added to the reaction or solvent system. Rather, the minimal aqueous conditions are created by first dissolving up to the maximum amount of water that can be dissolved in an organic solvent, next removing any undissolved water, and then using the wet organic solvent in a solvent system that does not include any additional water that is not dissolved in the wet organic solvent.

DETAILED DESCRIPTION OF THE INVENTION

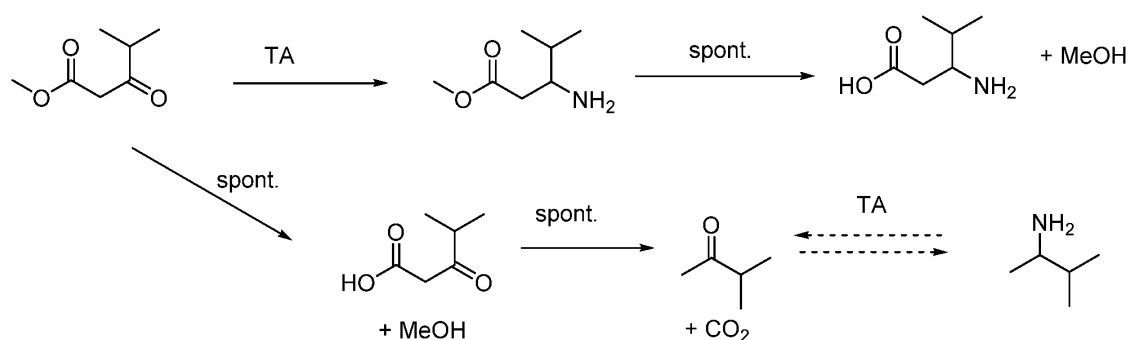
[0136] The present invention provides engineered transaminase polypeptides and compositions thereof, as well as polynucleotides encoding the engineered transaminase polypeptides. Methods for producing transaminase enzymes are also provided. In some embodiments, the engineered transaminase polypeptides are optimized to provide improved enzyme properties under minimal aqueous conditions in wet organic solvents.

[0137] The transamination of a methyl keto ester substrate (methyl 4-methyl-3-oxo-pentanoate) to the corresponding S-homovaline ester product (methyl (S)-3-amino-4-methyl-pentanoate) is depicted in Scheme 2.



[0138] Product yields under typical aqueous conditions are limited by spontaneous hydrolysis of the amino ester product and spontaneous hydrolysis of the keto ester followed by decomposition. See Scheme 4, below.

Scheme 4



[0139] One approach to overcome the limitations of asymmetric enzymatic synthesis for industrial processes is use of wet organic solvents. However, this approach has been limited due to the lack of enzymes that demonstrate activity in organic solvents. Recently, preliminary work demonstrated activity of ω -transaminases from *Arthrobacter* species in wet organic solvents without the need for immobilization (Mutti & Kroutil, *Adv. Synth. Catal.*, 2012, 354, 3409-4313).

[0140] The polypeptides of the current invention have been engineered for high enantioselectivity and to be catalytically active under minimal aqueous conditions in wet organic solvents. This allows for the direct conversion of the methyl keto ester to the S-homovaline ester in wet organic solvent, which greatly minimizes spontaneous hydrolysis of both the methyl keto ester and the S-homovaline ester. Additionally, the engineered polypeptides of the current invention have been engineered to have increased thermoactivity, allowing reactions at elevated temperatures in the presence of wet organic solvents.

[0141] Engineered transaminases with improved activity in wet organic solvents and improved activity at elevated temperatures in wet organic solvents can be used in industrial processes for preparing chiral amines compounds in an optically active form.

Engineered Polypeptides Having Transaminase Activity

[0142] The present invention provides engineered transaminase polypeptides and compositions thereof, as well as polynucleotides encoding the engineered transaminase polypeptides. Methods for producing transaminase enzymes are also provided. In some embodiments, the engineered transaminase polypeptides are optimized to provide improved enzyme properties under minimal aqueous conditions in wet organic solvents.

[0143] The engineered polypeptides of the disclosure are non-naturally occurring transaminases engineered to have improved enzyme properties (such as increased stereoselectivity) as compared to the wild-type transaminase polypeptide of *Arthrobacter* sp. KNK168 (GenBank Acc. No. BAK39753.1, GI:336088341), and also as compared to the reference engineered transaminase polypeptide of SEQ ID NO:2, which was used as the starting backbone sequence for the directed evolution of the engineered polypeptides of the present disclosure. The reference engineered transaminase polypeptide of SEQ ID NO:2 has the following 30 amino acid differences relative to the wild-type *Arthrobacter* sp. KNK168 polypeptide sequence: S8P, E42G, S54P, Y60F, L61Y, H62T, V65A, V69T, D81G, M94I, I96L, F122M, S124T, S126T, G136F, Y150F, V152S, Q155L, W156Q, A169L, V199I, A209L, G217N, S223P, L269P, L273Y, T282S, A284G, P297S, and S321P.

[0144] The engineered transaminase polypeptides of the present disclosure were generated by directed evolution of SEQ ID NO:2 for efficient conversion of compound (2) to compound (1) under certain industrially relevant conditions and have one or more residue differences as compared to the reference engineered transaminase polypeptide of SEQ ID NO:2. 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) under minimal aqueous conditions in wet organic solvents. 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 to 2 fold, 1.5 to 2.5 fold, 2 to 3 fold, 2.5 to 3.5 fold, 3 to 4 fold, 3.5 to 4.5 fold, 4 to 5 fold, 4.5 to 5 fold, 5 to 10 fold, 10 to 15 fold, or more relative to the activity of the reference polypeptide of SEQ ID NO:2, 6, 332, 518 and/or 656, 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 1%, at least about 2%, at least about 3%, at least about 10%, at least about 15%, or at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, in a reaction time of about 72h, about 48h, about 36h, about 24h, about 20h, 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 enantiomeric excess of at least 90%, 95%, 97%, 98%, 99%, or greater, under suitable reaction conditions.

[0145] The suitable reaction conditions under which the above-described improved properties of the engineered polypeptides can be determined with respect to concentrations or amounts of polypeptide, substrate, amino group donor, cofactor, and/or conditions including temperature and reaction time. In some embodiments, the suitable reaction conditions include minimal aqueous conditions in a wet organic solvent. In some embodiments, the suitable reaction conditions comprise the HTP or SFP assay conditions described below and in the Examples.

[0146] In some embodiments, additional reaction components or additional techniques are utilized to supplement the reaction conditions. In some embodiments, these include taking measures to stabilize or prevent inactivation of the enzyme, reduce product inhibition, or shift reaction equilibrium to desired product formation.

[0147] In some further embodiments, any of the above described processes for the conversion of substrate compound to product compound can further comprise one or more steps selected from: extraction, isolation, purification, crystallization, filtration, and/or lyophilization of product compound(s). Methods, techniques, and protocols for extracting, isolating, purifying, and/or crystallizing the product(s) from biocatalytic reaction mixtures produced by the processes provided herein are known to the ordinary artisan and/or accessed through routine experimentation. Additionally, illustrative methods are provided in the Examples below.

[0148] Structure and function information for exemplary non-naturally occurring, engineered transaminase polypeptides of the present disclosure are shown above in Tables 1, 2, 3, 4, 5, 6, 7, and 8. The odd numbered sequence identifiers (*i.e.*, SEQ ID NO) refer to the nucleotide (nt) sequence encoding the amino acid (aa) sequence provided by the even numbered SEQ ID NOs, and the sequences are provided in the electronic sequence listing file accompanying this disclosure, which are hereby incorporated by reference herein. The amino acid residue differences are based on comparison to the reference polypeptide sequence of SEQ ID NO:2, 6, 332, 518 and/or 656. The activity of each engineered polypeptide was determined using a high-throughput (HTP) assay (as a primary screen), and, in some cases, a secondary shake-flask powder (SFP) assay. Further details of the HTP and SFP preparations and assays are described in the Examples.

[0149] 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 minimal aqueous conditions in wet organic solvents.

[0150] In light of the guidance provided herein, it is further contemplated that any of the exemplary engineered polypeptides having the even-numbered sequence identifiers of SEQ ID NO:4 – 846 can be used as the starting amino acid sequence for synthesizing other engineered transaminase polypeptides, for example by subsequent rounds of evolution by adding in new combinations of various amino acid differences from other polypeptides in Tables 1, 2, 3, 4, 5, 6, 7, and 8, and other residue positions

described herein. Further improvements may be generated by including amino acid differences at positions that had been maintained as unchanged throughout earlier rounds of evolution.

[0151] In some embodiments, the present disclosure 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 disclosure 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 disclosure, such as an exemplary engineered polypeptide of having the even-numbered sequence identifiers of SEQ ID NO:4 - 846.

[0152] In some embodiments, the engineered transaminase polypeptide of the disclosure can have 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:4 - 846. Thus, for each and every embodiment of the engineered transaminase polypeptides of the disclosure, 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, 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.

[0153] In some embodiments, the present disclosure 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:4 - 846. Thus, for each and every embodiment of the transaminase polypeptides of the disclosure, 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.

[0154] In some embodiments, the present disclosure provides a engineered polypeptides having transaminase activity, which comprise an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to the sequences having the even-numbered sequence identifiers of SEQ ID NO:4 - 846, with the proviso that the amino acid sequence is not identical to (that is, it excludes) any of the exemplary engineered transaminase polypeptides amino acid sequences disclosed in US Pat. No. 8,293,507 B2, issued Oct. 23, 2012, PCT Publication WO2011005477A1, published Jan. 13, 2011, PCT publication WO2012024104, published Feb. 23, 2012, and PCT Appl. No. PCT/US12/54300, filed Sep. 7, 2012, each of which is hereby incorporated by reference herein.

[0155] In the above embodiments, the suitable reaction conditions for the engineered polypeptides are those described in Tables 1, 2, 3, 4, 5, 6, 7, and 8. Accordingly, in some embodiments, the preparation of compounds of Formula (I) or compound (1) can be carried out wherein the suitable reaction conditions comprise: (a) substrate loading of about 10 to 220 g/L of substrate compound (*e.g.*, compound (2)); (b) of about 0.5 g/L to 25 g/L engineered polypeptide; (c) IPM concentration of about 0.1 to 3 M; (d) a wet organic solvent; and (e) temperature of about 30°C to 60°C. In some embodiments, the suitable reaction conditions comprise: (a) about 20 g/L of substrate compound (*e.g.*, compound (2)); (b) about 20 g/L engineered polypeptide; (c) wet IPAc; (d) about 0.55 M IPM; and (e) about 50°C. Guidance for use of these reaction conditions and the transaminase polypeptides are provided in, among others, Tables 1, 2, 3, 4, 5, 6, 7, and 8, and the Examples.

[0156] In some embodiments, the polypeptides of the disclosure can be 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.

[0157] 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); ϵ -aminohexanoic acid (Aha); δ -aminovaleric acid (Ava); N-methylglycine or sarcosine (MeGly or Sar); ornithine (Orn); citrulline (Cit); t-butylalanine (Bua); t-butylglycine (Bug); N-methylisoleucine (Melle); 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 (see, *e.g.*, the various amino acids provided in Fasman, 1989, *CRC Practical Handbook of Biochemistry and Molecular Biology*, CRC Press, Boca Raton, FL, at pp. 3-70 and the references cited therein, all of which are incorporated by reference). These amino acids may be in either the L- or D-configuration.

[0158] 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(δ -benzylester), Gln(xanthyl), Asn(N- δ -xanthyl), His(bom), His(benzyl), His(tos), Lys(fmoc), Lys(tos), Ser(O-benzyl), Thr (O-benzyl) and Tyr(O-benzyl).

[0159] 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.

[0160] 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. 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.

[0161] 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 the reference polypeptide of SEQ ID NO:2. In such embodiments, the immobilized polypeptides can facilitate the biocatalytic conversion of the substrate compound of Formula (II) or compound (2) to a corresponding amine product compound of Formula (I) or compound (1) 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 disclosure can be carried out using the same transaminase polypeptides bound or immobilized on a solid support.

[0162] 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 immobilized engineered transaminase polypeptides capable of converting compound (2) to compound (1) (including the reference polypeptide of SEQ ID NO:2), 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 and described in *e.g.*, Yi *et al.*, "Covalent immobilization of ω -transaminase from *Vibrio fluvialis* JS17 on chitosan beads," *Process Biochemistry* 42(5): 895-898 (May 2007); Martin *et al.*, "Characterization of free and immobilized (S)-aminotransferase for acetophenone production," *Applied Microbiology and Biotechnology* 76(4): 843-851 (Sept. 2007); Koszelewski *et al.*, "Immobilization of ω -transaminases by encapsulation in a sol-gel/celite matrix," *Journal of Molecular Catalysis B: Enzymatic*, 63: 39-44 (Apr. 2010); Truppo *et al.*, "Development of an Improved Immobilized CAL-B for the Enzymatic Resolution of a Key Intermediate to Odanacatib," *Organic Process Research & Development*, published online: [dx.doi.org/10.1021/op200157c](https://doi.org/10.1021/op200157c); Hermanson, G.T., *Bioconjugate Techniques*, Second Edition, Academic Press (2008); Mateo *et al.*,

“Epoxy sepabeads: a novel epoxy support for stabilization of industrial enzymes via very intense multipoint covalent attachment,” *Biotechnology Progress* 18(3):629-34 (2002); and Bioconjugation Protocols: Strategies and Methods, *In Methods in Molecular Biology*, C.M. Niemeyer ed., Humana Press (2004); the disclosures of each which are incorporated by reference herein.

[0163] Solid supports useful for immobilizing the engineered transaminase polypeptides of the present disclosure 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 disclosure 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.

[0164] 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.

[0165] 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. 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 disclosure 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

[0166] In another aspect, the present disclosure 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.

[0167] As will be apparent to the skilled artisan, availability of a protein sequence and the knowledge of the codons corresponding to the various amino acids provide a description of all the polynucleotides

capable of encoding the subject protein sequence. The degeneracy of the genetic code, where the same amino acids are encoded by alternative or synonymous codons allows an extremely large number of nucleic acids to be made, all of which encode the improved transaminase enzymes disclosed herein. Thus, having identified a particular amino acid sequence, those skilled in the art could make any number of different nucleic acids by simply modifying the sequence of one or more codons in a way which does not change the amino acid sequence of the protein. In this regard, the present disclosure specifically contemplates each and every possible variation of polynucleotides that could be made by selecting combinations based on the possible codon choices, and all such variations are to be considered specifically disclosed for any polypeptide disclosed herein, including the amino acid sequences of the exemplary engineered polypeptides provided in Tables 1, 2, 3, 4, 5, 6, 7, and 8, and disclosed in the sequence listing incorporated by reference herein as the sequences of the even-numbered sequence identifiers of SEQ ID NO:4 – 846.

[0168] In various embodiments, the codons are preferably selected to fit the host cell in which the protein is being produced. For example, preferred codons used in bacteria are used to express the gene in bacteria; preferred codons used in yeast are used for expression in yeast; and preferred codons used in mammals are used for expression in mammalian cells. In some embodiments, all codons need not be replaced to optimize the codon usage of the transaminases since the natural sequence will comprise preferred codons and because use of preferred codons may not be required for all amino acid residues. Consequently, codon optimized polynucleotides encoding the transaminase enzymes may contain preferred codons at about 40%, 50%, 60%, 70%, 80%, or greater than 90% of codon positions of the full length coding region.

[0169] 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:4 – 846, 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 the polypeptide of SEQ ID NO:2. In some embodiments, the reference sequence is selected from SEQ ID NO:6, 332, 518, and 656.

[0170] In some embodiments, the polynucleotide encoding the engineered transaminase polypeptide comprises a sequence selected from the odd-numbered sequence identifiers of SEQ ID NO:3 - 845. In some embodiments, the polynucleotide sequences are selected from SEQ ID NO:5, 331, 517, and 655.

[0171] In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a reference polynucleotide sequence selected from any polynucleotide sequence provided herein, or a complement thereof, or a polynucleotide sequence encoding any of the variant enzyme polypeptides provided herein. In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an enzyme polypeptide comprising an amino acid sequence that has one or more residue differences as compared to a reference sequence.

[0172] In some embodiments, the polynucleotides encode the polypeptides described herein but have about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more sequence identity at the nucleotide level to a reference polynucleotide encoding the engineered transaminase polypeptide. In some embodiments, the reference polynucleotide sequence is selected from the sequences having the odd-numbered sequence identifiers of SEQ ID NO:3 - 845.

[0173] In some embodiments, an isolated polynucleotide encoding any of the engineered enzyme polypeptides herein is manipulated in a variety of ways to facilitate expression of the enzyme polypeptide. In some embodiments, the polynucleotides encoding the enzyme polypeptides comprise expression vectors where one or more control sequences is present to regulate the expression of the enzyme polynucleotides and/or polypeptides. Manipulation of the isolated polynucleotide prior to its insertion into a vector may be desirable or necessary depending on the expression vector utilized. Techniques for modifying polynucleotides and nucleic acid sequences utilizing recombinant DNA methods are well known in the art. In some embodiments, the control sequences include among others, promoters, leader sequences, polyadenylation sequences, propeptide sequences, signal peptide sequences, and transcription terminators. In some embodiments, suitable promoters are selected based on the host cell selection. For bacterial host cells, suitable promoters for directing transcription of the nucleic acid constructs of the present disclosure, include, but are not limited to promoters obtained from the *E. coli* lac operon, *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]). Exemplary promoters for filamentous fungal host cells, include, but are not limited to promoters obtained from the genes for *Aspergillus oryzae* TKA 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), 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), and mutant, truncated, and hybrid promoters thereof. Exemplary yeast cell promoters can be from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* galactokinase (GAL1), *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP), and *Saccharomyces cerevisiae* 3-phosphoglycerate kinase. Other useful promoters for yeast host cells are known in the art (See e.g., Romanos et al., *Yeast* 8:423-488 [1992]).

[0174] In some embodiments, the control sequence is also a suitable transcription terminator sequence (i.e., a sequence recognized by a host cell to terminate transcription). In some embodiments, the terminator sequence is operably linked to the 3' terminus of the nucleic acid sequence encoding the enzyme polypeptide. Any suitable terminator which is functional in the host cell of choice finds use in the present invention. Exemplary transcription terminators for filamentous fungal host cells can be 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. Exemplary terminators for yeast host cells can be 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 known in the art (See e.g., Romanos et al., supra).

[0175] In some embodiments, the control sequence is also a suitable leader sequence (i.e., a non-translated region of an mRNA that is important for translation by the host cell). In some embodiments, the leader sequence is operably linked to the 5' terminus of the nucleic acid sequence encoding the enzyme polypeptide. Any suitable leader sequence that is functional in the host cell of choice finds use in the present invention. Exemplary leaders for filamentous fungal host cells are obtained from the genes for *Aspergillus oryzae* TAKA amylase, and *Aspergillus nidulans* triose phosphate isomerase. Suitable leaders for yeast host cells are 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).

[0176] In some embodiments, the control sequence is also a polyadenylation sequence (i.e., 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 suitable 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 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 (See e.g., Guo and Sherman, *Mol. Cell. Bio.*, 15:5983-5990 [1995]). Exemplary mammalian polyadenylation sequences can be found in Zhang et al., 2005, *Nucleic Acids Res.* 33:D116–D120.

[0177] In some embodiments, the control sequence is also a signal peptide (i.e., a coding region that codes for an amino acid sequence linked to the amino terminus of a polypeptide and directs the encoded polypeptide into the cell's secretory pathway). In some embodiments, the 5' end of the coding sequence of the nucleic acid sequence inherently contains a signal peptide coding region naturally linked in translation reading frame with the segment of the coding region that encodes the secreted polypeptide. Alternatively, in some embodiments, the 5' end of the coding sequence contains a signal peptide coding

region that is foreign to the coding sequence. Any suitable signal peptide coding region which directs the expressed polypeptide into the secretory pathway of a host cell of choice finds use for expression of the engineered polypeptide(s). Effective signal peptide coding regions for bacterial host cells are the signal peptide coding regions include, but are not limited to those obtained from the genes for *Bacillus* NCIB 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]). In some embodiments, 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. Useful signal peptides for yeast host cells include, but are not limited to those from the genes for *Saccharomyces cerevisiae* alpha-factor and *Saccharomyces cerevisiae* invertase.

[0178] In some embodiments, the control sequence is also a propeptide coding region that codes for an amino acid sequence positioned at the amino terminus of a polypeptide. The resultant polypeptide is referred to as a “proenzyme,” “propeptide,” or “zymogen.” A propeptide can be converted to a mature active polypeptide by catalytic or autocatalytic cleavage of the propeptide from the propeptide. The propeptide coding region may be obtained from any suitable source, including, but not limited to the genes for *Bacillus subtilis* alkaline protease (aprE), *Bacillus subtilis* neutral protease (nprT), *Saccharomyces cerevisiae* alpha-factor, *Rhizomucor miehei* aspartic proteinase, and *Myceliophthora thermophila* lactase (See e.g., WO 95/33836). 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. In some embodiments, regulatory sequences are also utilized. These sequences facilitate the regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory systems are those that 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, but are not limited to the ADH2 system or GAL1 system. In filamentous fungi, suitable regulatory sequences include, but are not limited to the TAKA alpha-amylase promoter, *Aspergillus niger* glucoamylase promoter, and *Aspergillus oryzae* glucoamylase promoter.

[0179] In another aspect, the present invention is directed to a recombinant expression vector comprising a polynucleotide encoding an engineered enzyme polypeptide, 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 herein are joined together to produce recombinant expression vectors which include

one or more convenient restriction sites to allow for insertion or substitution of the nucleic acid sequence encoding the enzyme polypeptide at such sites. Alternatively, in some embodiments, the nucleic acid sequence of the present invention is expressed by inserting the nucleic acid sequence or a nucleic acid construct comprising the sequence into an appropriate vector for expression. In some embodiments involving the creation of 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.

[0180] The recombinant expression vector may be any suitable vector (e.g., a plasmid or virus), that can be conveniently subjected to recombinant DNA procedures and bring about the expression of the enzyme 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. The vectors may be linear or closed circular plasmids.

[0181] In some embodiments, the expression vector is an autonomously replicating vector (i.e., a vector that exists as an extra-chromosomal entity, the replication of which is independent of chromosomal replication, such as a plasmid, an extra-chromosomal element, a minichromosome, or an artificial chromosome). The vector may contain any means for assuring self-replication. In some alternative embodiments, the vector is one in which, when introduced into the host cell, it is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. Furthermore, in some 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, and/or a transposon is utilized.

[0182] In some embodiments, the expression vector 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 heavy metals, prototrophy to auxotrophs, and the like. 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 or tetracycline resistance. Suitable markers for yeast host cells include, but are not limited to ADE2, HIS3, LEU2, LYS2, MET3, TRP1, and URA3. Selectable markers for use in filamentous fungal host cells include, but are not limited to, *amdS* (acetamidase; e.g., from *A. nidulans* or *A. oryzae*), *argB* (ornithine carbamoyltransferases), *bar* (phosphinothricin acetyltransferase; e.g., from *S. hygrosopicus*), *hph* (hygromycin phosphotransferase), *niaD* (nitrate reductase), *pyrG* (orotidine-5'-phosphate decarboxylase; e.g., from *A. nidulans* or *A. oryzae*), *sC* (sulfate adenylyltransferase), and *trpC* (anthranilate synthase), as well as equivalents thereof.

[0183] The expression vectors of the present disclosure also can include 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. For integration into the host cell genome, the vector may 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.

[0184] Alternatively, the expression vector may 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 should preferably contain a sufficient number of nucleic acids, 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.

[0185] 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 ori 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 β 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 function in a temperature-sensitive manner in the host cell (see, *e.g.*, Ehrlich, 1978, *Proc Natl Acad Sci. USA* 75:1433).

[0186] More than one copy of a nucleic acid sequence of the present disclosure may be 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.

[0187] Many expression vectors useful with the embodiments of the present disclosure are commercially available. Suitable commercial expression vectors include p3xFLAGTM expression vectors from Sigma-Aldrich Chemicals, which includes 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 are pBluescriptII SK(-) and pBK-CMV, which are commercially available from Stratagene, LaJolla CA, and plasmids which are derived from pBR322 (Gibco BRL), pUC (Gibco BRL), pREP4, pCEP4 (Invitrogen) or pPoly (Lathe *et al.*, 1987, *Gene* 57:193-201).

[0188] An exemplary expression vector can be prepared by operatively linking a polynucleotide encoding an improved transaminase into the plasmid pCK110900I which contains the lac promoter under

control of the lacI repressor. The expression vector also contains the P15a origin of replication and the chloramphenicol resistance gene.

[0189] In another aspect, the present disclosure provides a host cell comprising a polynucleotide encoding an improved transaminase polypeptide of the present disclosure, the polynucleotide being operatively linked to one or more control sequences for expression of the transaminase enzyme in the host cell. Host cells for use in expressing the polypeptides encoded by the expression vectors of the present disclosure are well known in the art and include but are not limited to, bacterial cells, such as *E. coli*, *Arthrobacter sp.* KNK168, *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. An exemplary host cell is *Escherichia coli* W3110 (Δ fhuA). Appropriate culture mediums and growth conditions for the above-described host cells are well known in the art.

[0190] 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. Various methods for introducing polynucleotides into cells will be apparent to the skilled artisan.

Methods of Generating Engineered Transaminase Polypeptides

[0191] In some embodiments, to make the improved engineered polynucleotides and engineered polypeptides of the present disclosure, the naturally-occurring transaminase enzyme that catalyzes the transamination reaction is obtained (or derived) from *Arthrobacter sp.* KNK168. In some embodiments, the parent polynucleotide sequence is codon optimized to enhance expression of the transaminase in a specified host cell. The parental polynucleotide sequence encoding the wild-type polypeptide of *Arthrobacter sp.* KNK168 has been described (See e.g., Iwasaki *et al.*, *Appl. Microbiol. Biotechnol.*, 2006, 69: 499-505). Preparations of engineered transaminases based on this parental sequence are also described in US patent publication no. 2010/0285541A1 and published International application WO2010/099501.

[0192] The engineered transaminases can be produced using any suitable methods known in the art. Genes encoding the wild-type polypeptide of interest can be cloned in vectors, such as plasmids, and expressed in desired hosts, such as *E. coli*, etc. Variants of recombinant polypeptides can be generated by various methods known in the art. Indeed, there is a wide variety of different mutagenesis techniques well known to those skilled in the art. In addition, mutagenesis kits are also available from many commercial molecular biology suppliers. Methods are available to make specific substitutions at defined amino acids (site-directed), specific or random mutations in a localized region of the gene (region-specific), or random mutagenesis over the entire gene (e.g., saturation mutagenesis). Numerous suitable methods are known to those in the art to generate enzyme variants, including but not limited to site-

directed mutagenesis of single-stranded DNA or double-stranded DNA using PCR, cassette mutagenesis, gene synthesis, error-prone PCR, shuffling, and chemical saturation mutagenesis, or any other suitable method known in the art. Mutagenesis and directed evolution methods can be readily applied to enzyme-encoding polynucleotides to generate variant libraries that can be expressed, screened, and assayed. Any suitable mutagenesis and directed evolution methods find use in the present invention and are well known in the art (See e.g., US Patent 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 US, as well as PCT and 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]; Crameri et al., *Nature*, 391:288-291 [1998]; Crameri, et al., *Nat. Biotechnol.*, 15:436-438 [1997]; Zhang et al., *Proc. Nat. Acad. Sci. U.S.A.*, 94:4504-4509 [1997]; Crameri 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).

[0193] The clones obtained following mutagenesis treatment can be screened for engineered transaminases having a desired improved enzyme property. Measuring enzyme activity from the expression libraries can be performed using the standard biochemistry techniques, such as HPLC analysis following OPA derivatization of the product amine.

[0194] Where the improved enzyme property desired is thermostability, enzyme activity may be measured after subjecting the enzyme preparations to a defined temperature and measuring the amount of enzyme activity remaining after heat treatments. Clones containing a polynucleotide encoding a

transaminase are then isolated, sequenced to identify the nucleotide sequence changes (if any), and used to express the enzyme in a host cell.

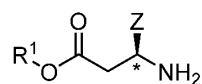
[0195] Where the sequence of the engineered polypeptide is known, the polynucleotides encoding the enzyme can be prepared by standard solid-phase methods, according to known synthetic methods. In some embodiments, fragments of up to about 100 bases can be individually synthesized, then joined (*e.g.*, by enzymatic or chemical ligation methods, or polymerase mediated methods) to form any desired continuous sequence. For example, polynucleotides and oligonucleotides of the disclosure can be prepared by chemical synthesis using, *e.g.*, the classical phosphoramidite method described by Beaucage *et al.*, 1981, *Tet Lett* 22:1859-69, or the method described by Matthes *et al.*, 1984, *EMBO J.* 3:801-05, *e.g.*, 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.

[0196] In some embodiments, the present disclosure also provides methods for preparing or manufacturing the engineered transaminase polypeptides, wherein the methods comprise culturing a host cell capable of expressing a polynucleotide encoding the engineered polypeptide under culture conditions suitable for expression of the polypeptide. In some embodiments, the method for preparation of the polypeptide further comprises isolating the polypeptide. The engineered polypeptides can be expressed in appropriate cells (as described above), and isolated (or recovered) from the host cells and/or the culture medium using any one or more of the well-known techniques used for protein purification, including, among others, lysozyme treatment, sonication, filtration, salting-out, ultra-centrifugation, and chromatography. Chromatographic techniques for isolation of the polypeptide include, among others, reverse phase chromatography high performance liquid chromatography, ion exchange chromatography, gel electrophoresis, and affinity chromatography. Conditions for purifying a particular engineered polypeptide will depend, in part, on factors such as net charge, hydrophobicity, hydrophilicity, molecular weight, molecular shape, etc., and will be apparent to those having skill in the art. The isolated polypeptide has to be lyophilized after purification.

Methods of Using the Engineered Transaminase Enzymes and Compounds Prepared Therewith

[0197] 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. Generally the structural analogs of compound (1) are encompassed within structural Formula (I).

[0198] In some embodiments, the present disclosure provides a process for preparing a compound of structural Formula (I):

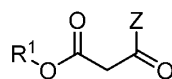


(I)

having the indicated stereochemical configuration at the stereogenic center marked with an *; in an enantiomeric excess of at least 70% over the opposite enantiomer, wherein

Z is a linear C₁₋₆ alkyl or a branched C₁₋₆ alkyl, optionally substituted with one or more hydroxy, cyano, or nitro groups ; and

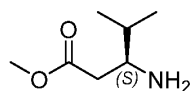
R¹ is C₁₋₆ alkyl, aryl-C₁₋₂ alkyl, heteroaryl-C₁₋₂ alkyl, or a 5- to 6-membered heterocyclic ring system optionally containing an additional heteroatom selected from O, S, and N, the heterocyclic ring being unsubstituted or substituted with one to three substituents independently selected from oxo, hydroxy, halogen, C₁₋₄ alkoxy, and C₁₋₄ alkyl; the process comprising the step of contacting a prochiral keto ester of structural Formula (II):



(II)

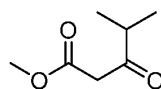
with an engineered polypeptide as disclosed herein in the presence of an amino group donor in a suitable organic solvent under suitable reaction conditions.

[0199] In some embodiments, the present disclosure provides a process of preparing compound (1);



(1)

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



(2)

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

[0200] In some embodiments of the process for preparing a compound of structural Formula (I), the compound of Formula (II) specifically excludes compound (2) and the compound of Formula (I) prepared by the method specifically excludes compound (1).

[0201] As described herein, and illustrated in the Examples, the present disclosure contemplates ranges of suitable reaction conditions that can be used in the processes herein, including but not limited to ranges of temperature, wet organic solvent system, substrate loading, mixture of substrate compound stereoisomers, polypeptide loading, cofactor loading, 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 temperature, wet organic solvent conditions, and detecting the product compound, for example, using the methods described in the Examples provided herein.

[0202] As described above, the engineered polypeptides having transaminase activity for use in the processes of the present disclosure generally comprise an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to a reference amino acid sequence selected from any one of the even-numbered sequences of SEQ ID NO:4 - 846, and an engineered transaminase polypeptide comprising an amino acid sequence that has (a) has one or more amino acid residue differences as compared to SEQ ID NO:2. 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 SEQ ID NO:2.

[0203] 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. In addition, substrate compounds covered by Formula (II) can also be used in appropriate amounts, in light of the amounts used for compound (2).

[0204] In the processes described 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 IPM, putrescine, L-lysine, α -phenethylamine, D-alanine, L-alanine, D,L-alanine, D,L-ornithine, 3-aminobutyric acid, or methylbenzylamine. In some embodiments, the amino donor is selected from the group consisting of isopropylamine, alanine, 3-aminobutyric acid, or methylbenzylamine. 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.

[0205] Suitable reaction conditions for the processes also typically comprise the presence of a cofactor. Because the engineered transaminases typically use members of the vitamin B₆ 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. 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. In some embodiments, the cofactor is added to the cell lysate rather than the reaction mixture. In some embodiments, the cofactor is added to the cell lysate prior to lyophilization. In some embodiments, the cofactor, PLP, is added to the cell lysate 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 prior to lyophilization.

[0206] 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.

[0207] In some embodiments, the reaction mixture comprises a wet organic solvent. Accordingly, in some embodiments, the pH of the reaction mixture is not monitored or maintained. In some embodiments, the wet organic solvent is wet IPAc, wet toluene, wet MTBE, wet MeCN, or wet IPA.

[0208] In the 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 herein, the use of minimal aqueous conditions in wet organic solvents. For example, the engineered polypeptides of the present disclosure have increased thermoactivity in wet organic solvents relative to naturally occurring transaminase polypeptide, and the engineered polypeptide of SEQ ID NO:2, which allow the engineered polypeptides of the present disclosure to be used at higher temperatures for increased conversion rates. 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.

[0209] In some embodiments of the process, the processes using the engineered transaminase polypeptides are conducted in a wet organic solvent with no aqueous component. In some embodiments of the process, the wet organic solvent is prepared by mixing 200 mL of water with 800 mL of dry wet organic solvent on top, then mixed with a magnetic stir bar overnight and allowed to separate before taking from the organic layer.

[0210] 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 10 to 220 g/L of substrate compound (*e.g.*, compound **(2)**); (b) of about 0.5 g/L to 25 g/L engineered polypeptide; (c) IPM concentration of about 0.1 to 3 M; (d) a wet organic solvent; and (e) temperature of about 30°C to 60°C.

[0211] In some embodiments, the combination of reaction parameters comprises: (a) about 20 g/L of substrate compound (*e.g.*, compound **(2)**); (b) about 20 g/L engineered polypeptide; (c) wet IPAc; (d) about 0.55 M IPM; and (e) about 50°C.

[0212] Further exemplary reaction conditions include the assay conditions provided in Tables 1, 2, 3, 4, 5, 6, 7, and 8, and in the Examples.

[0213] 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).

[0214] 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. 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.

[0215] 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). 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.

[0216] 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 disclosure 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.

[0217] In some embodiments, the processes using the engineered polypeptides disclosed herein can be carried out wherein the process further comprises the step of isolating the compound of Formula (I) or the compound (1) from the reaction.

[0218] In some embodiments, the processes using the engineered polypeptides disclosed herein can be carried out wherein the amino group donor is selected from IPM, alanine, 3-aminobutyric acid, or methylbenzylamine. In some embodiments, the amino group donor is IPM.

[0219] 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.

[0220] Various features and embodiments of the disclosure are illustrated in the following representative examples, which are intended to be illustrative, and not limiting.

EXAMPLES

Example 1: *E. coli* Expression Hosts Containing Recombinant Transaminase Genes

[0221] A codon-optimized and engineered transaminase gene (SEQ ID NO:1) encoding the reference engineered polypeptide of SEQ ID NO:2 was used as the starting backbone for directed evolution to generate the genes encoding the engineered polypeptides having transaminase activity of the even numbered sequence identifiers of SEQ ID NO:4- 846, each of which is capable of converting the substrate compound (2) to the product compound (1) with improved enzyme properties relative to it

and/or the reference polypeptide of SEQ ID NO:2. The gene of SEQ ID NO:1 and polypeptide of SEQ ID NO:2 of the present disclosure correspond to SEQ ID NO:129 and 130 of US 9,617,573 B2. The engineered transaminase polypeptide of SEQ ID NO:2 has the following 30 amino acid differences relative to the wild-type *Arthrobacter* sp. KNK168 polypeptide sequence (GenBank accession: BAK39753.1;GI:336088341): S8P, E42G, S54P, Y60F, L61Y, H62T, V65A, V69T, D81G, M94I, I96L, F122M, S124T, S126T, G136F, Y150F, V152S, Q155L, W156Q, A169L, V199I, A209L, G217N, S223P, L269P, L273Y, T282S, A284G, P297S, and S321P. Cloning of SEQ ID NO:1 in the pCK110900 vector system (See e.g., US Patent Application Publication 2006/0195947A1) and subsequent expression in *E.coli* W3110fhuA was as described in US 8,293,507 B2. Briefly, the *E. coli* W3110 expresses the transaminase polypeptides as an intracellular protein under the control of the lac promoter. The polypeptide accumulates primarily as a soluble cytosolic, active enzyme. Standard methods of directed evolution via iterative variant library generation by gene synthesis followed by screening and sequencing of hits to generate the engineered derivatives of the gene sequence SEQ ID NO:1 disclosed herein.

Example 2: Preparation of HTP Transaminase Containing Wet Cell Pellets

[0222] *E. coli* cells containing recombinant transaminase encoding genes from monoclonal colonies were inoculated into 180 µl Luria Broth (LB) containing 1% glucose and 30 µg/mL chloramphenicol (CAM) in the wells of 96-well shallow well microtiter plates. The plates were sealed with O₂-permeable seals and cultures were grown overnight at 30°C, 200 rpm and 85% humidity. Then, 10 µl of each of the cell cultures were transferred into the wells of 96-well deep well plates containing 390 µl Terrific Broth (TB) and 30 µg/mL CAM. The deep-well plates were sealed with O₂-permeable seals and incubated at 30°C, 250 rpm and 85% humidity until OD₆₀₀ 0.6-0.8 was reached. The cell cultures were then induced by isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM and incubated overnight under the same conditions as originally used. The plates were centrifuged at 4,000 rpm for 10 minutes, the supernatants discarded and pellets frozen at -80°C before lysis as described in Example 3. Alternatively the cells from two identical plates were pooled using wide bore tips and pelleted using centrifugation at 4000 rpm for 10 minutes. The supernatants were discarded and the double pellets frozen at -80°C prior to lysis as described in Example 5.

Example 3: Preparation of HTP Transaminase-Containing Cell Lysates

[0223] First, 200 µl lysis buffer containing 100 mM triethanolamine buffer, pH 7.5, 0.25 mg/mL lysozyme, 0.2 mg/mL polymyxin b sulfate (PMBS), and 100 uM pyrodoxal-5-phosphate (PLP) was added to the cell paste 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 10 minutes at 4,000 rpm and 4°C. The clear supernatants were then used in biocatalytic reactions to determine their stereoselectivity as described in Example 4.

Example 4: Improvements over SEQ ID NO:2 in Stereoselectivity

[0224] SEQ ID NO:2 was selected as the parent enzyme. 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.

Each reaction well contained 80 μ L of HTP lysate prepared as explained in Example 3 and the following final conditions in a 200 μ L volume per well: 20 g/L substrate compound (2), DMSO 40 vol%, 0.2M borate with 1.5M IPM 50 vol% and 1 g/L PLP; the pH was adjusted to pH = 8.5. The plates were heat sealed and shaken at 60°C overnight (18-22 hours). The reactions were quenched by adding 200 μ L per well of acetonitrile. The plates were heat sealed and shaken for 15 minutes at 1,000 rpm and then centrifuged at 4,000 rpm for 10 minutes. 200 μ L per well of cleared supernatant were transferred to a filter plate. The filter plate was placed on top of a 96-well round bottom Costar plate and centrifuged for 5 minutes at 4,000 rpm. The 96-well round bottom Costar plate was heat sealed and analyzed using Chiral LC-MS (Method 3, Example 10).

[0225] Stereoselectivity is calculated as % enantiomeric excess (%ee): desired (*S*) product peak area – undesired (*R*) product peak area / desired (*S*) product peak area + undesired (*R*) product peak area * 100.

Table 1: HTP Stereoselectivity

SEQ ID NO: (nt/aa)	Amino Acid Differences (compared to SEQ ID NO:2)	% ee
1/2	None	12
3/4	G284A;	97
5/6	G284A; L155W	99

Example 5: Preparation of Lyophilized HTP Lysates from Transaminase Containing Wet Cell Lysates

[0226] HTP lysis and lyophilization of HTP lysates: Three conditions for HTP lysis of transaminase containing wet cell pellets (generated as double pellets, as described in Example 2) and lyophilization of generated HTP lysates were used: A, B, and C.

[0227] Condition A: First, 300 μ L lysis buffer, containing 0.25 mg/mL (1 mg/mL Condition B and C) lysozyme; 0.2 mg/mL (0.5 mg/mL Condition B and C) PMBS; in 100mM (25 mM Condition B and C) triethanolamine buffer pH 7.5; 10 μ L of 2,000 U/mL endonuclease per plate, was added to the double pellet cell paste 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 plates were then centrifuged for 20 minutes at 4,000 rpm and 4°C. 220 μ L (150 μ L Condition C) of clarified lysate was transferred into 96-well glass coated deep well plates. In preliminary experiments, glass-coated deep well and polypropylene plates warped under certain reaction conditions. Therefore, glass plates or vials were necessary for Condition B and C (Freeslate/Chemglass glass vial plates). The plates were covered with airpore seal and placed in -80°C, with direct contact of 96-well samples with freezer shelf for 2 hours. The plates

were placed in the lyophilizer. The lyophilizer vacuum was brought down to 200 bar and the following program was run: pre-chill at -40°C, thermal treatment at -40°C for 120 minutes, 100 mTorr; drying step: 10°C, 100 mTorr 1,000 – 1,200 minutes. The plates were removed from the lyophilizer and assayed almost immediately. For Condition B and C: The plates were frozen in the lyophilizer for ~2hr at -40°C with the frost seal set to on ~ 600-700 mTorr. The lyophilizer vacuum was brought to 200 mTorr and the following program run: Pre-Chill: -40°C; thermal treatment: -40°C, 100 minutes, 100 mTorr; Drying Steps: 1. -40 to -15°C (1°C/min), 100 mTorr; 2. -15°C, 600 minutes, 100 mTorr; 3. -15 to 0°C (1°C/min), 100 mTorr, 0°C, 900 minutes, 100 mTorr (method complete after overnight and stopped after 1,000 – 1,200 minutes of Step 4).

Example 6: Preparation of Lyophilized Lysates from Shake Flask (SF) Cultures

[0228] Selected HTP cultures grown as described above were plated onto LB agar plates with 1% glucose and 30 µg/ml CAM and grown overnight at 37°C. A single colony from each culture was transferred to 6 ml of LB with 1% glucose and 30µg/ml CAM. The cultures were grown for 18 hours at 30°C, 250 rpm, and subcultured approximately 1:50 into 250 ml of TB containing 30 µg/ml CAM, to a final OD₆₀₀ of 0.05. The cultures were grown for approximately 195 minutes at 30°C, 250 rpm, to an OD₆₀₀ between 0.6-0.8 and induced with 1 mM IPTG. The cultures were then grown for 20 hours at 30°C, 250 rpm. The cultures were centrifuged 4000 rpm for 20 minutes. The supernatant was discarded, and the pellets were resuspended in 30 ml of 25 mM triethanolamine pH 7.5 and lysed using a Microfluidizer system (Microfluidics) at 18,000 psi. The lysates were pelleted (10,000 rpm for 60 minutes) and the supernatants were frozen and lyophilized to generate shake flake (SF) enzymes.

Example 7: HTP Screening for Activity Improvements

[0229] HTP screening, conversion assay: Wet IPAc was prepared by mixing 200 mL of water with 800 mL of dry IPAc on top, which was mixed with a magnetic stir bar overnight and allowed to separate before taking from the organic layer. The ester substrate was premixed in wet IPAc with 4 equivalencies of IPM relative to substrate concentration. The final reaction parameters in the conversion HTP assay were: 20 g/L ester substrate (2), 4 equivalencies of IPM, and HTP lyophilized cake prepared as explained in Conditions A – C in Example 5. First, 200 µl/well of 20 g/L ester substrate and 4 equivalencies IPM in wet IPAc, were added to 96-well glass coated well plates. The glass coated plates were heat sealed and incubated at 30°C, 400 rpm for 24 hours. The reactions in 96-well glass coated plates were quenched by adding 1000 µl/well of dry IPAc. The plates were heat sealed, shaken for 1 minute and then centrifuged at 4°C to 4,000 rpm for 5 minutes. The 6-fold IPAc diluted samples were transferred to 96-well Costar round bottom plates for GC-FID analysis. Alternatively 50 µl of 6-fold IPAc diluted samples were transferred into 950 µl of water in deep well plates. Then, 5 µl of 20-fold water diluted IPAc quenched reaction was added to 195 µl of water in 96-well Costar round bottom plates for RapidFire MS analysis.

Key for Tables 2 - 8	
Indicator	FIOPC
+	1 to 2
++	2 to 3
+++	3 to 4
++++	4 to 5
+++++	5 to 10

Table 2: HTP Activity in Organic Solvent of HTP Lyophilized Powders Produced in Condition A, Example 5

SEQ ID NO: (nt/aa)	Amino Acid Differences (compared to SEQ ID NO:6)	Activity Fold-Improvement ¹
5/6	None	/
7/8	E9H;	+
9/10	H71Y;Q102P;	+
11/12	I140N;D266N;	+
13/14	I55S;	+
15/16	N32S;	+
17/18	E104S;	+
19/20	Q146A;	+
21/22	R119A;	+
23/24	I55G;	+
25/26	P145N;	+
27/28	E111L;	+
29/30	Y23Q;	+
31/32	E104I;	+
33/34	T16I;	+
35/36	R79H;	+
37/38	N90A;	+
39/40	S24P;	+
41/42	Y20H;	+
43/44	D29K;	+
45/46	S24F;	+
47/48	T134S;	+
49/50	I41Y;	+
51/52	Q146I;	+
53/54	P33M;	+
55/56	S89M;	+
57/58	P33L;	+
59/60	Y20T;	+
61/62	T134G;	+
63/64	I55F;	+
65/66	I55K;	+++
67/68	K142T;	+
69/70	S93P;	+
71/72	W40M;	+
73/74	Q156E;	+
75/76	F160A;	+
77/78	D139A;	+

SEQ ID NO: (nt/aa)	Amino Acid Differences (compared to SEQ ID NO:6)	Activity Fold-Improvement ¹
79/80	Q146L;	+
81/82	N32L;	+
83/84	N90C;	+
85/86	E111T;	+
87/88	D82P;	+
89/90	Q156T;	+
91/92	A114T;	+
93/94	V46C;	+
95/96	I41K;	+
97/98	S24D;	+
99/100	T101L;	+
101/102	F45L;	+
103/104	E50T;	+
105/106	N90T;	+
107/108	F160V;	+
109/110	A35G;	+
111/112	L18G;	+
113/114	F160Y;	+
115/116	Q146H;	+
117/118	T22W;	+
119/120	R95P;	+
121/122	A35G;E221D;	+
123/124	D15T;	+
125/126	E117D;	+
127/128	P98M;	+
129/130	T22H;	+
131/132	K106S;	+
133/134	S89W;	++
135/136	N90S;	+
137/138	Y23H;	+
139/140	S89L;	+
141/142	R95T;	+
143/144	D82T;	+
145/146	S93M;	+
147/148	S89Q;	+
149/150	Y26P;	+
151/152	F160L;	+
153/154	P33S;	+
155/156	S93H;	+
157/158	E9T;	+
159/160	T101S;	+
161/162	D15I;	+
163/164	D29M;	+
165/166	A5T;A35Y;	+
167/168	L18R;	+
169/170	L264N;	+
171/172	P244L;	++++
173/174	A262V;	+
175/176	L324C;	+
177/178	E271C;	+
179/180	A262T;	+

SEQ ID NO: (nt/aa)	Amino Acid Differences (compared to SEQ ID NO:6)	Activity Fold- Improvement ¹
181/182	D266S;	+
183/184	R197T;	+
185/186	P244G;	+
187/188	G215L;	+
189/190	L218S;	+
191/192	K249M;	+
193/194	Y26H;Y273K;	+
195/196	H168P;	+
197/198	P244T;	+
199/200	E208L;	+
201/202	E208P;	+
203/204	A265R;	+
205/206	L218N;R313H;	+
207/208	P244L;Y273C;	++
209/210	Y273L;	+
211/212	V234H;	+
213/214	R197L;	+
215/216	E316R;	+
217/218	I246V;	+
219/220	R248M;	+
221/222	E320W;	+
223/224	P179R;	+
225/226	D216P;	+
227/228	P244E;	+
229/230	L264P;	+
231/232	E208R;	+
233/234	H203P;	+
235/236	A262L;	+
237/238	Q186R;Q200R;	+
239/240	K249T;	+
241/242	Q186R;	+
243/244	R248V;	+
245/246	G301M;	+
247/248	L264R;	++
249/250	D216S;	+
251/252	P321L;	+
253/254	F207Y;	+
255/256	G301E;	+
257/258	L264G;	+
259/260	Q329R;	+
261/262	R236T;	+
263/264	E320R;	+
265/266	Y273M;	+
267/268	L264W;	+
269/270	E256C;	+
271/272	L218R;	+
273/274	A262P;	+
275/276	I246A;	+
277/278	L264E;	+
279/280	D266G;	+
281/282	E208K;	+

SEQ ID NO: (nt/aa)	Amino Acid Differences (compared to SEQ ID NO:6)	Activity Fold-Improvement ¹
283/284	D216G;	+
285/286	I246L;	+
287/288	L264A;	+
289/290	L324V;	+
291/292	L324W;	+
293/294	E9T;P244G;A262T;L264N;	++
295/296	E9T;P244G;K249G;	++
297/298	T134G;P244L;	+++
299/300	E9T;Y23Q;S24F;	++
301/302	E9T;T101S;P244G;	+++
303/304	E9T;Y23Q;S24F;T101S;A262V;	+
305/306	E9T;S24F;P244G;A262T;	++
307/308	E9T;Y23Q;D29M;	+
309/310	E9T;Y20T;Y23Q;T101S;P244L;	+++
311/312	E9T;Y23Q;S24F;D29M;P244G;	+
313/314	R79G;	+
315/316	E9T;S24F;K249G;	+
317/318	E9T;S24F;P244G;	++
319/320	K249G;	+
321/322	E9T;Y23Q;S24F;P244L;A262T;	+
323/324	E9T;Y20T;S24F;T101S;P244G;K249G;	+
325/326	T134G;A262V;	++
327/328	Y20T;Y23Q;S24F;I55F;T134G;P244G;A262T;	++++
329/330	E9T;Y20T;Y23Q;S24F;T134S;P244G;	++
331/332	Y20T;Y23Q;T134S;P244G;	+++
333/334	E9T;R79G;T101S;P244G;	++
335/336	E9T;A262T;	+
337/338	Q156L;	+
339/340	N74A;	+
341/342	P321V;	+
343/344	D15I;S24P;Y26P;S89M;S93M;F160L;	+
345/346	Y20H;P33L;F160L;	+
347/348	S89M;S93M;Q156L;L324C;	++
349/350	D15I;T16I;L18R;Y20H;S89M;S93M;F160L;	+
351/352	S89M;S93M;P321V;L324C;	+
353/354	S93M;L324C;	++
355/356	Y20H;L324C;	+
357/358	Y20H;S24P;S89M;R197T;	++
359/360	D82S;P244D;	+
361/362	E9H;I246L;	++
363/364	E9H;S89Q;E271C;	+
365/366	Q146A;P244D;	+
367/368	E9H;S89Q;I246L;	++
369/370	E9H;D82S;S89Q;	++
371/372	E9H;L110V;V302F;	+
373/374	E9H;F45L;S89Q;S93H;	++
375/376	D29K;S89Q;	+
377/378	E9H;S93H;K142M;I246L;	+
379/380	G17I;D25R;F160V;	+
381/382	G17I;D25R;	+
383/384	Y23H;D25R;I55L;P321Y;	+

SEQ ID NO: (nt/aa)	Amino Acid Differences (compared to SEQ ID NO:6)	Activity Fold- Improvement ¹
385/386	D25R;	+
387/388	K142L;P321Y;	+
389/390	G17I;D25R;I41F;	+
391/392	G42T;	+
393/394	D25R;E104V;	+
395/396	T22L;	+
397/398	T22L;F160V;	+
399/400	Q102L;	+
401/402	Y23H;D25R;Q102L;K142L;	+
403/404	F56R;	+
405/406	Y23H;Q102L;	++
407/408	D25R;P321Y;	+
409/410	D25R;F160V;	++
411/412	Y23H;P321Y;	+
413/414	D25R;P33M;	+
415/416	I41F;Q102L;	+
417/418	P321Y;	+
419/420	T22L;Q102L;	+
421/422	G17I;F160V;	+
423/424	L18G;Y273T;	+
425/426	L18K;H168P;P244L;	+
427/428	E208L;I246V;	+
429/430	L18G;Q146L;A262V;	+
431/432	P244G;R248M;A284G;	++
433/434	L18K;P244G;I246V;	+
435/436	W40M;P244G;R248M;	+
437/438	L18K;Q146L;H168P;	+
439/440	L18G;Y273V;	++
441/442	L18G;P244G;R248M;	+
443/444	L18K;P244G;I246V;A284G;	++
445/446	A284G;	++
447/448	L18K;H168P;A262V;L324W;	+
449/450	Y273V;	++
451/452	L18K;	+
453/454	L18K;P244T;A262V;Y273T;	++
455/456	L18G;Y26H;I246V;Y273T;	++
457/458	Y273T;	++
459/460	L18K;Y273T;	++
461/462	L18G;H168P;	++
463/464	L18K;A262V;	++
465/466	P244G;R248M;	++
467/468	L18K;P244L;I246V;	++
469/470	L18G;P244L;Y273V;	+
471/472	L18G;Q146L;	+
473/474	L18K;P244L;	++
475/476	P244L;Y273V;	++
477/478	L18K;Q146L;	++
479/480	L18G;P244G;	++
481/482	Q146L;P244L;	+
483/484	Q146L;Y273V;	++
485/486	L18K;H168P;	++

SEQ ID NO: (nt/aa)	Amino Acid Differences (compared to SEQ ID NO:6)	Activity Fold-Improvement ¹
487/488	W40M;H168P;E208L;P244G;	+
489/490	D216P;Y273V;	+++
491/492	I246V;R248M;Y273V;	++
493/494	H168P;P244L;Y273T;	++
495/496	A35G;G215L;L218R;K249T;	+
497/498	F160Y;G215L;D216G;L218R;	+
499/500	T101L;F160Y;P244E;K249M;	+
501/502	D15T;L218R;Y273K;	+
503/504	L218R;K249M;	+
505/506	G215L;D216G;K249M;	+
507/508	K249M;Y273K;	+
509/510	A35G;L218R;K249T;Y273K;	++
511/512	D15T;K249M;Y273K;	+
513/514	D15T;A35G;Y273K;	+
515/516	F160Y;Y273K;	+

¹ **Activity Fold-Improvement** was calculated as the ratio of the product compound (1) produced by the specified engineered polypeptide under the activity assay reaction conditions (noted in Example 7) per the product produced, under the same reaction conditions by the engineered polypeptide of SEQ ID NO:6. Amount of product produced was determined by GC-FID or RapidFire-MS analysis of the quenched assay samples prepared as noted in Example 10.

Table 3: HTP Activity in Organic Solvent of HTP Lyophilized Powders Produced in Condition A, Example 5

SEQ ID NO: (nt/aa)	Amino Acid Differences (compared to SEQ ID NO:332)	Activity Fold-Improvement ¹
331/332	None	/
517/518	E9H;S89Q;F160Y;	+
519/520	S89M;S93M;F160V;K249M;	+
521/522	E9H;S89M;S93M;T101S;F160Y;	+
523/524	A35M;	+
525/526	L317W;	+
527/528	T141A;	+
529/530	K142N;	+
531/532	I140M;	+
533/534	P321A;	+
535/536	E107R;	+
537/538	S323R;	+
539/540	S297T;	+
541/542	H168R;	+
543/544	N296V;	+
545/546	S292T;	+
547/548	I140S;	+
549/550	P327S;	+
551/552	P321S;	+
553/554	V302Q;	+
555/556	A270S;	+
557/558	R144A;	+
559/560	P327Q;	+

SEQ ID NO: (nt/aa)	Amino Acid Differences (compared to SEQ ID NO:332)	Activity Fold-Improvement ¹
561/562	T141H;	+
563/564	V302L;	+
565/566	P327H;	+
567/568	Q156T;	+
569/570	S323V;	+
571/572	N74K;	+
573/574	Q156C;	+
575/576	Y273L;	+
577/578	P321T;	+
579/580	Q329R;	+
581/582	D300T;	+
583/584	I140L;	+
585/586	N74R;	+
587/588	S89L;	+
589/590	E117D;	+
591/592	F207G;	+
593/594	Q329V;	+
595/596	H168P;	+
597/598	D274G;	+
599/600	S49P;	+
601/602	S297L;	+
603/604	S89W;	+
605/606	V234T;	+
607/608	N296W;	+
609/610	V234I;	+
611/612	P305T;	+
613/614	E107V;	+
615/616	P33N;	+
617/618	P305Q;	+
619/620	K142S;	+
621/622	D274A;	+
623/624	V319T;	+
625/626	S89V;	+
627/628	E107M;	+
629/630	N296M;	+
631/632	H203L;	+
633/634	A44T;	+
635/636	A270R;	+
637/638	I108L;	+
639/640	A35H;	+
641/642	V302I;	+

¹ Activity Fold-Improvement was calculated as the ratio of the product compound (1) produced by the specified engineered polypeptide under the activity assay reaction conditions (noted in Example 7) per the product produced, under the same reaction conditions by the engineered polypeptide of SEQ ID NO:332. Amount of product produced was determined by GC-FID or RapidFire-MS analysis of the quenched assay samples prepared as noted in Example 10.

Table 4: HTP Activity in Organic Solvent of HTP Lyophilized Powders Produced in Condition B, Example 5

SEQ ID NO: (nt/aa)	Amino Acid Differences (compared to SEQ ID NO:518)	Activity Fold-Improvement ¹
517/518	None	/
643/644	V46L;Y273L;	+
645/646	Y273L;	+
647/648	P159D;	+
649/650	E111Q;P159D;Y273L;	+
651/652	P159D;Y273L;	+
653/654	A35H;Y273L;	+
655/656	I199L;	+
657/658	S174G;	+
659/660	V287P;	+

¹ Activity Fold-Improvement was calculated as the ratio of the product compound (1) produced by the specified engineered polypeptide under the activity assay reaction conditions (noted in Example 7) per the product produced, under the same reaction conditions by the engineered polypeptide of SEQ ID NO:518. Amount of product produced was determined by GC-FID or RapidFire-MS analysis of the quenched assay samples prepared as noted in Example 10.

Example 8: HTP Screening for Thermoactivity Improvements

[0230] HTP screening, thermoactivity assay: Wet IPAc was prepared by mixing 200 mL of water with 800 mL of dry IPAc on top, which was mixed with a magnetic stir bar overnight and allowed to separate before taking from the organic layer. The ester substrate was premixed in wet IPAc with 4 equivalencies of IPM relative to substrate concentration. The final reaction parameters in the conversion HTP assay were: 20 g/L ester substrate (2), 4 equivalencies of IPM, and HTP lyophilized cake prepared as explained in Conditions A – C in Example 5. First, 200 µl/well of 20 g/L ester substrate and 4 equivalencies IPM in wet IPAc, were added to Freeslate glass vial plates. The Freeslate glass vial plates were evenly torqued down with screws and incubated at 50°C at 400 rpm for 24 hours. The reactions in Freeslate 96 glass vial plates were quenched by adding 400 µl/well of dry IPAc. Next, 100 µl/well were transferred from Freeslate 96 glass vial plates into 96-well Costar plates containing 100 µl/well of dry IPAc and submitted for GC-FID analysis.

Table 5: HTP Thermoactivity in Organic Solvent of HTP Lyophilized Powders Produced in Condition A, Example 5

SEQ ID NO: (nt/aa)	Amino Acid Differences (compared to SEQ ID NO:332)	Activity Fold-Improvement ¹
331/332	None	/
543/544	N296V;	+
545/546	S292T;	+
553/554	V302Q;	+
555/556	A270S;	+
561/562	T141H;	+
563/564	V302L;	+
567/568	Q156T;	+

SEQ ID NO: (nt/aa)	Amino Acid Differences (compared to SEQ ID NO:332)	Activity Fold-Improvement ¹
571/572	N74K;	+
579/580	Q329R;	+
583/584	I140L;	+
585/586	N74R;	+
593/594	Q329V;	+
621/622	D274A;	+
629/630	N296M;	+
637/638	I108L;	+
661/662	F160Y;	+
663/664	Q23Y;D29Y;I55K;	+
665/666	V319R;	+
667/668	H168Q;	+
669/670	G215A;	+
671/672	G215T;	+
673/674	A270Q;	+
675/676	S323L;	+
677/678	L34Y;	+
679/680	F160W;	+
681/682	R144V;	+
683/684	V302S;	+
685/686	E92D;	+
687/688	F160V;	+
689/690	A44N;	+
691/692	E117R;	+
693/694	E256G;	+
695/696	N296Y;	+
697/698	P269T;	+
699/700	G215S;	+

¹ Activity Fold-Improvement was calculated as the ratio of the product compound (1) produced by the specified engineered polypeptide under the activity assay reaction conditions (noted in Example 8) per the product produced, under the same reaction conditions by the engineered polypeptide of SEQ ID NO:332. Amount of product produced was determined by GC-FID analysis of the quenched assay samples prepared as noted in Example 10.

Table 6: HTP Thermoactivity in Organic Solvent of HTP Lyophilized Powders Produced in Condition B, Example 5

SEQ ID NO: (nt/aa)	Amino Acid Differences (compared to SEQ ID NO:518)	Activity Fold-Improvement ¹
517/518	None	/
643/644	V46L;Y273L;	+
645/646	Y273L;	+
655/656	I199L;	+
701/702	N74K;E92D;	+
703/704	Y160V;G215S;Q329R;	+
705/706	L34Y;N74K;E92D;	+
707/708	S292T;	+
709/710	A35H;Y273L;S323R;	+
711/712	A39G;	+

SEQ ID NO: (nt/aa)	Amino Acid Differences (compared to SEQ ID NO:518)	Activity Fold-Improvement ¹
713/714	F207Y;	+
715/716	T126V;	+
717/718	P48G;	+
719/720	R138A;	+
721/722	T178S;	+
723/724	I55V;	+
725/726	F150W;	+
727/728	T126L;	+
729/730	S133T;	+

¹ Activity Fold-Improvement was calculated as the ratio of the product compound (1) produced by the specified engineered polypeptide under the activity assay reaction conditions (noted in Example 8) per the product produced, under the same reaction conditions by the engineered polypeptide of SEQ ID NO:518. Amount of product produced was determined by GC-FID analysis of the quenched assay samples prepared as noted in Example 10.

Table 7: HTP Thermoactivity in Organic Solvent of HTP Lyophilized Powders Produced in Condition C, Example 5

SEQ ID NO: (nt/aa)	Amino Acid Differences (compared to SEQ ID NO:656)	Activity Fold-Improvement ¹
655/656	None	/
731/732	A39G;F207Y;Y273L;	+
733/734	Y273L;	+
735/736	A39G;T178S;Y273L;	+
737/738	A39G;F150W;Y273L;	+
739/740	T126V;S133T;T178S;	+
741/742	R138A;F150W;T178S;	+
743/744	A39G;S133T;T178S;Y273L;	+
745/746	A39G;F150W;T178S;F207Y;Y273L;	+
747/748	A35H;F150W;Y273L;	+
749/750	F150W;F207Y;Y273L;	+
751/752	F150W;T178S;S323R;	+
753/754	F150W;Y273L;	+
755/756	F150W;	+
757/758	A35H;R138A;F150W;Y273L;	+
759/760	T126V;R138A;T178S;	+
761/762	I55V;	+
763/764	T126V;T178S;F207Y;	+
765/766	A35H;Y273L;	+
767/768	A35H;T178S;Y273L;	+
769/770	T178S;Y273L;	+
771/772	A39G;Y273L;	+
773/774	A39G;T178S;	+
775/776	A39G;F150W;T178S;F207Y;	+
777/778	I55V;F150W;	+
779/780	A35H;A39G;I55V;R138A;F150W;	+
781/782	T126V;	+
783/784	A35H;A39G;T126V;F150W;Y273L;	+
785/786	A35H;A39G;S133T;F150W;Y273L;	+
787/788	A35H;A39G;F150W;	+

SEQ ID NO: (nt/aa)	Amino Acid Differences (compared to SEQ ID NO:656)	Activity Fold-Improvement ¹
789/790	A39G;F207Y;	+
791/792	A39G;I55V;F150W;Y273L;	++
793/794	A35H;A39G;F150W;Y273L;	+
795/796	A39G;S133T;F150W;	+
797/798	A39G;	+
799/800	F150W;T178S;Y273L;S323R;	+
801/802	A35H;A39G;S133T;	+
803/804	A39G;T126V;F150W;	+
805/806	A35H;S133T;T178S;	+
807/808	A39G;I55V;F150W;T178S;Y273L;	++
809/810	A39G;F150W;T178S;	+
811/812	A35H;A39G;I55V;F150W;	+
813/814	A39G;S133T;Y273L;	+
815/816	A44Q;	+
817/818	P269V;	+
819/820	G239N;	+
821/822	F88W;	+
823/824	S132E;	+
825/826	R312A;	+
827/828	I55F;	+
829/830	I263E;	+
831/832	A44L;	+
833/834	S297E;	+
835/836	T141S;	+
837/838	H168D;	+
839/840	S132D;	+
841/842	E261R;	+
843/844	L18R;	+
845/846	S297R;	+

¹ **Activity Fold-Improvement** was calculated as the ratio of the product compound (1) produced by the specified engineered polypeptide under the activity assay reaction conditions (noted in Example 8) per the product produced, under the same reaction conditions by the engineered polypeptide of SEQ ID NO:656. Amount of product produced was determined by GC-FID analysis of the quenched assay samples prepared as noted in Example 10.

Example 9: Shake Flask Powder Testing

[0231] Wet IPAc (or MTBE or toluene) was prepared by mixing 200mL of water with 800mL of dry IPAc on top, which was mixed with a magnetic stir bar overnight and allowed to separate before taking from the organic layer. The ester substrate was premixed in wet IPAc (or MTBE or toluene) with 4 equivalencies of IPM relative to substrate concentration. The final reaction parameters were: 20 g/L ester substrate, 4 equivalencies of IPM, 20 mg of lyophilized shake flask powder prepared as explained in Example 6. Next, 20 mg of lyophilized shake flask powder was weighed into a 4 mL glass vial. Then, 1 mL of reaction mix was added to the vial, and the vial sealed with a Teflon seal. The vials were incubated at 50°C for 24 hours. The reactions were quenched by adding 2 mL of dry IPAc. Then, 500 µL of quenched reaction was diluted into 500 µL of dry IPAc in a GC glass vial. The vial was submitted for GC-FID analysis.

Table 8: Thermoactivity in Organic Solvent of Shake Flask Powders Produced as Described in Example 6

SEQ ID NO: (nt/aa)	Amino Acid Differences (compared to SEQ ID NO:2)	Activity Fold-Improvement in wet IPAc ¹	Activity Fold-Improvement in wet MTBE ²	Activity Fold-Improvement in wet toluene ³
6/7	L155W;G284A;	/	/	/
331/332	Y20T;Y23Q;T134S;L155W;P244G;G284A;	+	+	+++++
517/518	E9H;Y20T;Y23Q;S89Q;T134S;L155W;F160Y;P244G;G284A;	+	+	+++++
655/656	E9H;Y20T;Y23Q;S89Q;T134S;L155W;F160Y;I199L;P244G;G284A;	++	/	+++++
807/808	E9H;Y20T;Y23Q;A39G;I55V;S89Q;T134S;F150W;L155W;F160Y;T178S;I199L;P244G;Y273L;G284A;	++++	/	+++++

¹ Activity Fold-Improvement in wet IPAc was calculated as the ratio of the product compound (1) produced by the specified engineered polypeptide under the activity assay reaction conditions (noted in Example 9) per the product produced, under the same reaction conditions by the engineered polypeptide of SEQ ID NO:6. Amount of product produced was determined by GC-FID analysis of the quenched assay samples prepared as noted in Example 10.

² Activity Fold-Improvement in wet MTBE was calculated as the ratio of the product compound (1) produced by the specified engineered polypeptide under the activity assay reaction conditions (noted in Example 9) per the product produced, under the same reaction conditions by the engineered polypeptide of SEQ ID NO:6. Amount of product produced was determined by GC-FID analysis of the quenched assay samples prepared as noted in Example 10.

³ Activity Fold-Improvement in wet toluene was calculated as the ratio of the product compound (1) produced by the specified engineered polypeptide under the activity assay reaction conditions (noted in Example 9) per the product produced, under the same reaction conditions by the engineered polypeptide of SEQ ID NO:6. Amount of product produced was determined by GC-FID analysis of the quenched assay samples prepared as noted in Example 10.

Example 10: Analytical Methods

Method 1 – Achiral GC-FID Method

[0232] The quenched reactions, prepared as explained in Examples 7, 8 and 9 were subject to GC-FID analysis under the following conditions:

Instrument	Agilent Technologies 6890N Network GC system
Column	Rtx-5 amine 0.25 μ m 30m X 0.25mm
Gas Flow rate	H ₂ ; 110°C for 2.22min, ramp to 150°C @ 18°C/min, then ramp to 200°C @ 25°C/min, 4.82min hold 200°C for a total 4.82 min run.
Detection	FID: 300°C, 40mL/min H₂; 400mL air; make-up (Helium): 5mL/min
Injector Temp.	50°C injector
Injection Volume	100:1 split; 0.2 μ L injection; 12.2 psi
Runtime	9 min
Peak identification details	1.831 min Ester; 2.096 min Product (varies for each GC)

Method 2 – Achiral RapidFire-MS Method

[0233] The quenched reactions, prepared as explained in Example 7 were subject to RapidFire-MS analysis under the following conditions

Agilent RapidFire conditions						
Pump1 Buffer	MS grade water; 1.5 mL/min flow rate					
Pump2 Buffer	90% MS grade acetonitrile 10% water; 1.25 mL/min flow rate					
Pump3 Buffer	90% MS grade acetonitrile 10% water; 1.25 mL/min flow rate					
Aqueous wash	Water					
Organic wash	Acetonitrile					
SPE cartridge	D, hypercarb					
RF state 1 Aspirate	600 ms					
RF state 2 Load/Wash	3000 ms					
RF state 3 Extra Wash	0					
RF state 4 Elute	4000 ms					
RF state 5 Reequilibrate	2000 ms					
Agilent Jet Stream source parameters						
Drying gas temperature	300 °C					
Drying gas flow	10 L/min					
Nebulizer pressure	45 psi					
Sheath gas temperature	350 °C					
Sheath gas flow	11 L/min					
Capillary voltage	2500 V					
Nozzle voltage	1000 V					
Agilent 6470 Triple Quadrupole MRM parameters						
Compound	Q1	Q3	Dwell	Fragmentor	CE	CAV
Amino ester product	146	128.9	60	70	5	3
Amino ester product	146	72	60	70	10	3
Amino ester product	146	55	60	70	29	3
Amino acid product	132	72	60	70	10	3

Amino acid product	132	55.1	60	70	27	3
--------------------	-----	------	----	----	----	---

Method 3 – Chiral LC-MS Method

[0234] The quenched reactions, prepared as explained in Example 4 were subject to Chiral LC-MS analysis under the following conditions:

Instrument	Agilent 1290 – Single Quad MS
Column	Zwix (-) 3 μ m 4 X 150mm
Mobile Phase	A: 49:49:2, ACN:MeoH:water, 30mM ammonium formate (1.89 g/L), 25mM formic acid (950 μ L/L)
Flow rate	1.2mL/min
Detection	UV= 220nm (Resp Time= 40Hz) MS2 ESI
MS Conditions	MODE: ESI (pos) Drying Gas=12 L/min;Nebulizer Press= 35 psi;Drying Gas T = 350C;CAP v (pos)=3000; SIM mode 132.2 and 146.2;Frag= 70v
Column Temp.	45°C
Injection Volume	1 μ L
Runtime	6 min
Peak identification details	MS – Selected Ion Monitoring @ 132.2 and 146.4 m/z: 1.70 min Ester: 3.40 min R Product, 3.90 min S product

What is claimed is:

1. An engineered polypeptide having transaminase activity 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 NO: 2, 6, 332, 518 and/or 656, or a functional fragment thereof, wherein the polypeptide sequence of said engineered polypeptide comprises at least one substitution or substitution set and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 808, 2, 6, 332, 518 and/or 656.
2. The engineered polypeptide 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:2, and wherein the polypeptide sequence of said engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from 20, 23, 244, 284 and 284/155, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:2.
3. The engineered polypeptide 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:6, and wherein the polypeptide sequence of said engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from 5/35, 9, 9/20/23/24/134/244, 9/20/23/101/244, 9/20/24/101/244/249, 9/23/24, 9/23/24/29/244, 9/23/24/101/262, 9/23/24/244/262, 9/23/29, 9/24/244, 9/24/244/262, 9/24/249, 9/45/89/93, 9/79/101/244, 9/82/89, 9/89/246, 9/89/271, 9/93/142/246, 9/101/244, 9/110/302, 9/244/249, 9/244/262/264, 9/246, 9/262, 15, 15/16/18/20/89/93/160, 15/24/26/89/93/160, 15/35/273, 15/218/273, 15/249/273, 16, 17/25, 17/25/41, 17/25/160, 17/160, 18, 18/26/246/273, 18/146, 18/146/168, 18/146/262, 18/168, 18/168/244, 18/168/262/324, 18/244, 18/244/246, 18/244/246/284, 18/244/248, 18/244/262/273, 18/244/273, 18/262, 18/273, 20, 20/23/24/55/134/244/262, 20/23/134/244, 20/24/89/197, 20/33/160, 20/324, 22, 22/102, 22/160, 23, 23/25/55/321, 23/25/102/142, 23/102, 23/321, 24, 25, 25/33, 25/104, 25/160, 25/321, 26, 26/273, 29, 29/89, 32, 33, 35, 35/215/218/249, 35/218/249/273, 35/221, 40, 40/168/208/244, 40/244/248, 41, 41/102, 42, 45, 46, 50, 55, 56, 71/102, 74, 79, 82, 82/244, 89, 89/93/156/324, 89/93/321/324, 90, 93, 93/324, 95, 98, 101, 101/160/244/249, 102, 104, 106, 111, 114, 117, 119, 134, 134/244, 134/262, 139, 140/266, 142, 142/321, 145, 146, 146/244, 146/273, 156, 160, 160/215/216/218, 160/273, 168, 168/244/273, 179, 186, 186/200, 197, 203, 207, 208, 208/246, 215, 215/216/249, 216, 216/273, 218, 218/249, 218/313, 234, 236, 244, 244/248, 244/248/284, 244/273, 246, 246/248/273, 248, 249, 249/273,

256, 262, 264, 265, 266, 271, 273, 284, 301, 316, 320, 321, 324, and 329, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:6.

4. The engineered polypeptide 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:332, and wherein the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from 9/89/93/101/160, 9/89/160, 33, 35, 44, 49, 74, 89, 89/93/160/249, 107, 108, 117, 140, 141, 142, 144, 156, 168, 203, 207, 234, 270, 273, 274, 292, 296, 297, 300, 302, 305, 317, 319, 321, 323, 327, and 329, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:332.
5. The engineered polypeptide 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:518, and wherein the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from 35/273, 46/273, 111/159/273, 159, 159/273, 174, 199, 273, and 287, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:518.
6. The engineered polypeptide 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:332, wherein said engineered polypeptide has increased thermoactivity in wet organic solvents compared to the engineered polypeptide of SEQ ID NO:332, and wherein the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from 23/29/55, 34, 44, 74, 92, 108, 117, 140, 141, 144, 156, 160, 168, 215, 256, 269, 270, 274, 292, 296, 302, 319, 323, and 329, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:332.
7. The engineered polypeptide 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:518, wherein said engineered polypeptide has increased thermoactivity in wet organic solvents compared to the engineered polypeptide of SEQ ID NO:518, and wherein the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from 34/74/92, 35/273/323, 39, 46/273, 48, 55, 74/92, 126, 133, 138, 150, 160/215/329, 178, 199, 207, 273, and

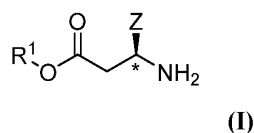
292, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:518.

8. The engineered polypeptide 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:656, wherein said engineered polypeptide has increased thermoactivity in wet organic solvents compared to the engineered polypeptide of SEQ ID NO:656, and wherein the engineered polypeptide comprises at least one substitution or substitution set at one or more positions in said polypeptide sequence selected from 18, 35/39/55/138/150, 35/39/55/150, 35/39/126/150/273, 35/39/133, 35/39/133/150/273, 35/39/150, 35/39/150/273, 35/133/178, 35/138/150/273, 35/150/273, 35/178/273, 35/273, 39, 39/55/150/178/273, 39/55/150/273, 39/126/150, 39/133/150, 39/133/178/273, 39/133/273, 39/150/178, 39/150/178/207, 39/150/178/207/273, 39/150/273, 39/178, 39/178/273, 39/207, 39/207/273, 39/273, 44, 55, 55/150, 88, 126, 126/133/178, 126/138/178, 126/178/207, 132, 138/150/178, 141, 150, 150/178/273/323, 150/178/323, 150/207/273, 150/273, 168, 178/273, 239, 261, 263, 269, 273, 297, and 312, wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO:656.
9. The engineered polypeptide of any of Claims 6-8, wherein said engineered polypeptide has increased thermoactivity in a wet organic solvent relative to the engineered polypeptide of SEQ ID NO:6.
10. The engineered polypeptide of Claim 1, wherein said engineered polypeptide 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 polypeptide variant set forth in Table 1, 2, 3, 4, 5, 6, 7, and/or 8.
11. The engineered polypeptide of Claim 1, wherein said engineered polypeptide 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 SEQ ID NO:2, 6, 332, 518, and/or 656.
12. The engineered polypeptide of Claim 1, wherein said engineered polypeptide comprises a variant engineered polypeptide set forth in SEQ ID NO:6, 332, 518, and/or 656.
13. The engineered polypeptide of Claim 1, wherein said engineered polypeptide 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 polypeptide variant set forth in the even numbered sequences of SEQ ID NOS: 4-846.

14. The engineered polypeptide of Claim 1, wherein said engineered polypeptide comprises a polypeptide sequence forth in at least one of the even numbered sequences of SEQ ID NOS: 4-846.
15. A polynucleotide sequence encoding at least one engineered polypeptide of any of Claims 1-14.
16. A polynucleotide sequence encoding at least one engineered polypeptide, 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: 1, 5, 331, 517, and/or 655, wherein the polynucleotide sequence of said engineered polypeptide comprises at least one substitution at one or more positions.
17. A polynucleotide sequence encoding at least one engineered polypeptide 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: 1, 5, 331, 517, and/or 655, or a functional fragment thereof.
18. The polynucleotide sequence of any of Claims 15-17, wherein said polynucleotide sequence is operably linked to a control sequence.
19. The polynucleotide sequence of any of Claims 15-17, wherein said polynucleotide sequence is codon optimized.
20. The polynucleotide sequence of any of Claims 15-17, wherein said polynucleotide sequence comprises a polynucleotide sequence forth in the odd numbered sequences of SEQ ID NOS: 3-845.
21. An expression vector comprising at least one polynucleotide sequence of any of Claims 15-20.
22. A host cell comprising at least one expression vector of Claim 21.
23. A host cell comprising at least one polynucleotide sequence of any of Claims 15-20.

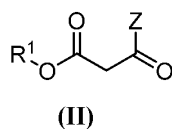
24. A method of producing an engineered polypeptide in a host cell, comprising culturing the host cell of Claim 22 and/or 23, under suitable conditions, such that at least one engineered polypeptide is produced.
25. The method of Claim 24, further comprising recovering at least one engineered polypeptide from the culture and/or host cell.
26. The method of Claim 24 and/or 25, further comprising the step of purifying said at least one engineered polypeptide.
27. A process for preparing a compound of structural 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, wherein

Z is a linear C₁₋₆ alkyl or a branched C₁₋₆ alkyl, optionally substituted with one or more hydroxy, cyano, or nitro groups ; and

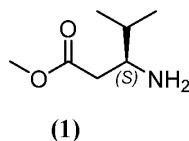
R¹ is C₁₋₆ alkyl, aryl-C₁₋₂ alkyl, heteroaryl-C₁₋₂ alkyl, or a 5- to 6-membered heterocyclic ring system optionally containing an additional heteroatom selected from O, S, and N, the heterocyclic ring being unsubstituted or substituted with one to three substituents independently selected from oxo, hydroxy, halogen, C₁₋₄ alkoxy, and C₁₋₄ alkyl; the process comprising the step of contacting a compound of structural Formula (II):



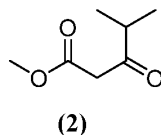
with an engineered polypeptide of any one of Claims 1 to 14 in the presence of an amino group donor under suitable reaction conditions.

28. The process of Claim 27, wherein the compound of structural Formula (II) excludes compound (2) and the compound of structural Formula (I) excludes compound (1).

29. A process of preparing compound (1),



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



with an engineered polypeptide of any one of Claims 1 to 14, in the presence of an amino group donor under suitable reaction conditions.

30. The process of any one of Claims 27 to 29, wherein the compound of Formula (I) or the compound (1) is produced in at least 90% enantiomeric excess.
31. The process of any one of Claims 27 to 29, wherein the compound of Formula (I) or the compound (1) is produced in at least 99% enantiomeric excess.
32. The process of any one of Claims 27 to 29, wherein the amino group donor is selected from isopropylamine, alanine, 3-aminobutyric acid, or methylbenzylamine.
33. The process of any one of Claims 27 to 29, wherein the amino group donor is isopropylamine, optionally at a concentration of 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.
34. The process of any one of Claims 27 to 29, wherein the suitable reaction conditions comprise a temperature of about 45°C to about 60°C.
35. The process of any one of Claims 27 to 29, wherein the suitable reaction conditions comprise a wet organic solvent.
36. The process of Claim 35, wherein the wet organic solvent is selected from wet isopropyl acetate, wet toluene, wet methyl tertiary-butyl ether, wet acetonitrile, or wet isopropyl alcohol.

37. The process of Claim 36, wherein the wet organic solvent is isopropyl acetate.
38. The process of any one of Claims 27 to 29, wherein the suitable reaction conditions comprise the substrate compound at a loading of about 5 g/L to about 200 g/L, about 10 g/L to about 150 g/L, or about 50 g/L to about 100 g/L.
39. The process of any one of Claims 27 to 29, wherein the suitable reaction conditions comprise the engineered polypeptide at a concentration of from about 0.5 g/L to about 5 g/L, about 0.5 g/L to about 3 g/L, about 0.5 g/L to about 2 g/L, or from about 0.5 g/L to about 1 g/L.
40. The process of Claim 29, wherein the suitable reaction conditions comprise: (a) substrate loading of about 10 to 220 g/L of substrate compound **(2)**; (b) engineered polypeptide concentration of about 0.5 g/L to 25 g/L; (c) isopropylamine concentration of about 0.1 to 3 M; (d) a wet organic solvent; and (e) temperature of about 30°C to 60°C.
41. The process of Claim 29, wherein the suitable reaction conditions comprise: (a) about 20 g/L of substrate compound **(2)**; (b) about 20 g/L engineered polypeptide; (c) wet isopropyl acetate; (d) about 0.55 M isopropylamine; and (e) about 50°C.
42. The process of any one of Claims 27 to 29, further comprising the step of isolating the compound of Formula **(I)** or the compound **(1)** from the reaction.