ENGINEERED TRANSAMINASE POLYPEPTIDES

[0001] The present application claims priority to US Prov. Pat. Appln. Ser. No. 63/008,047, filed April 10, 2020, which is incorporated by reference in its entirety, for all purposes.

REFERENCE TO SEQUENCE LISTING, TABLE OR COMPUTER PROGRAM

[0002] The official copy of the Sequence Listing is submitted concurrently with the specification as an ASCII formatted text file via EFS-Web, with a filename of "CX2-197WO1_ST25.txt", a creation date of April 5, 2021, and a size of 2.53 megabytes. The Sequence Listing filed via EFS-Web is part of the specification and incorporated in its entirety by reference herein.

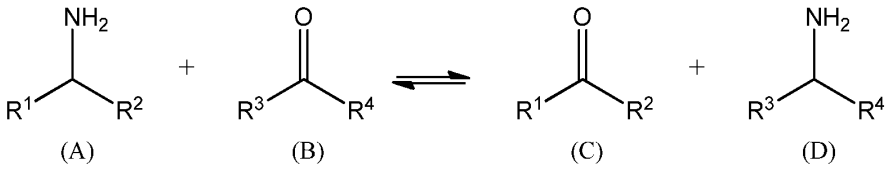
FIELD OF THE INVENTION

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

BACKGROUND

[0004] Transaminases (E.C. 2.6.1) catalyze the transfer of an amino group, a pair of electrons, and a proton from an amino donor compound to the keto group of an amino acceptor compound. Transaminase reactions can result in the formation of a chiral amine product compound. As shown in **Scheme 1,** an amino acceptor compound (B) (which is the keto substrate precursor of a desired chiral amine product (D)) is reacted with an amino donor compound (A) in the presence of a transaminase. The transaminase catalyzes the transfer of the primary amine group of the amino donor compound (A) to the keto group of the amino acceptor compound (B). The transaminase reaction results in a chiral amine product compound (D) (assuming R^3 is not the same as R^4) 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., Biosci. Biotechnol. Biochem., 65:1782-1788 [2001]; Iwasaki et al., Biotech. Lett., 25:1843-1846 [2003]; Iwasaki et al., Appl. Microbiol. Biotech., 69:499-505 [2004]; Yun et al., Appl. Environ. Microbiol., 70:2529-2534 [2004]; and Hwang et al., Enz. Microbiol. Technol., 34:429-426 [2004]); the preparation of intermediates and precursors of pregabalin (e.g., WO 2008/127646); the enzymatic transamination of 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 (See, 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 (See 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; Savile, et al., Sci., 329: 305-9 [2010]; WO2011/159910; and WO2012/024104).

[0007] Wild-type transaminases having the ability to catalyze a reaction of **Scheme 1** have been isolated from various microorganisms, including, but not limited to, *Alcaligenes denitrificans*, *Bordetella bronchiseptica*, *Bordetella parapertussis*, *Brucella melitensis*, *Burkholderia malle*, *Burkholderia pseudomallei*, *Chromobacterium violaceum*, *Oceanicola granulosus* HTCC2516, *Oceanobacter* sp. RED65, *Oceanospirillum sp.* MED92, *Pseudomonas putida*, *Ralstonia solanacearum*, *Rhizobium meliloti*, *Rhizobium sp.* (strain NGR234), *Bacillus thuringensis*, *Klebsiella pneumonia, Vibrio fluvialis* (See e.g., Shin et al., Biosci. Biotechnol, Biochem., 65:1782-1788 [2001]), and *Arthrobacter sp. KNK168* (See e.g., Iwasaki et al., Appl. Microbiol. Biotechnol., 69: 499-505 [2006]; and U.S. Pat. No. 7,169,592). Several of these wild-type transaminase genes and encoded polypeptides have been sequenced (e.g., *Ralstonia solanacearum* [Genbank Acc. No. YP_002257813.1, GI:207739420], *Burkholderia pseudomallei* 1710b [Genbank Acc. No. ABA47738.1, GI:76578263], *Bordetella petrii* [Genbank Acc. No. AM902716.1, GI:163258032], *Vibrio fluvialis* JS17 [Genbank Acc. No. AEA39183.1, GI: 327207066], and *Arthrobacter sp.* KNK168 [GenBank Acc. No. BAK39753.1, GI:336088341]). At least two wild-type transaminases of classes EC 2.6.1.18 and EC 2.6.1-19, have been crystallized and structurally characterized (See e.g., Yonaha et al., Agric. Biol. Chem., 47:2257-2265 [1983]).

[0008] Transaminases are known that have (*R*)-selective or (*S*)-selective stereoselectivity. For example, the wild-type transaminase from *Arthrobacter sp.* KNK168 is considered (*R*)-selective and produces primarily (*R*)-amine compounds from certain substrates (See e.g., Iwasaki et al., Appl. Microbiol. Biotechnol., 69: 499-505 [2006]; and U.S. Pat. No. 7,169,592), whereas the wild-type transaminase from *Vibrio fluvialis* JS17 is considered (*S*)-selective and produces primarily *(S)*-amine compounds from certain substrates (See e.g., Shin et al., Appl. Microbiol. Biotechnol., 61: 463-471 [2003]).

[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; WO2011/005477A1; WO2012/024104; and Savile et al., Sci., 329:305-9 [2010]).

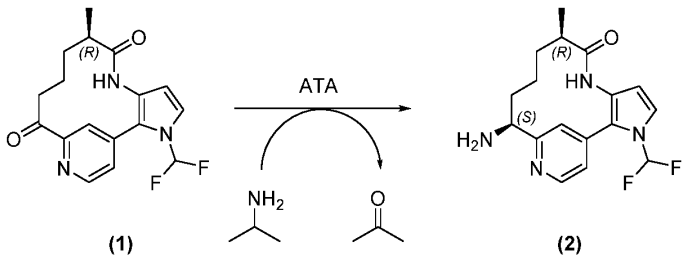
[0010] However, transaminases generally have properties that are undesirable for commercial application in the preparation of chiral amine compounds, such as instability to industrially useful process conditions (e.g., solvent, temperature), poor recognition of, and stereoselectivity for, commercially useful amino acceptor and/or amino donor substrates, and low product yields due to unfavorable reaction equilibrium. Thus, there is a need for engineered transaminases that can be used in industrial processes for preparing chiral amines compounds in an optically active form.

SUMMARY OF THE INVENTION

[0011] The present disclosure provides engineered polypeptides having transaminase activity, polynucleotides encoding the polypeptides, methods of making the polypeptides, and methods of using the polypeptides for the biocatalytic conversion of amino acceptor substrate compounds (i.e., keto group containing compounds) to chiral amine product compounds. The transaminase polypeptides of the present 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: 4) and associated enhanced solvent and thermal stability relative to the transaminase of SEQ ID NO: 4 and the wild-type transaminase of SEQ ID NO: 2. The amino acid residue differences are located at residue positions that result in improvement of various enzyme properties, including among others, activity, stereoselectivity, stability, expression, and product tolerance.

[0012] In particular, the engineered transaminase polypeptides of the present disclosure have been engineered for efficient conversion of the substrate, (*R*)-2^1-(difluoromethyl)-5-methyl-2^1*H*-3-aza-1(4,2)-pyridina-2(2,3)-pyrrolacyclononaphane-4,9-dione (referred to herein as "compound **(1)**") to its corresponding chiral amine product compound, (5*R*,9*S*)-9-amino-2^1-(difluoromethyl)-5-methyl-2^{1}H-3-aza-1(4,2)-pyridina-2(2,3)-pyrrolacyclononaphan-4-one (referred to herein as "compound **(2)**") as shown in **Scheme 2**.

Scheme 2



[0013] In some embodiments, the present disclosure provides engineered transaminases comprising polypeptide sequences having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NOS: 4, 8, 366, and/or 650, or a functional fragment thereof, wherein said engineered transaminases comprise at least one substitution or substitution set in said polypeptide sequences, and wherein the amino acid positions of said polypeptide sequences are numbered with reference to SEQ ID NO: 2, 4, 8, 366, and/or 650. In some embodiments, the engineered transaminase comprises a polypeptide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 4, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 18, 20, 21, 21/23/56/146, 21/23/56/146/432, 21/23/146/417, 21/23/395/417/432, 21/53/56, 21/53/417, 21/56/395, 21/417/432, 23, 23/53, 23/53/56, 23/53/56/146/395, 23/53/395, 23/53/417, 23/53/432, 23/56, 23/56/395, 23/56/395/417, 23/395/417, 23/417, 23/417/432, 53, 53/56, 53/146/417, 53/395, 56, 56/74/241/286/314/316/323, 56/86/163/314/316/383/414/416/422, 56/86/286/314/414/416, 56/86/314/316/323/394/414/422, 56/146/417, 56/146/432, 56/147, 56/163, 56/163/286/316/323/383/394, 56/286/314/316/323/422, 56/286/383, 56/323, 56/323/383, 56/323/383/394, 56/383, 56/395, 74/81/286/316/323/383, 74/85/86/163/286/316/323/394, 74/85/314/316/414/416, 74/86/163/316, 74/86/316/323/383/394, 74/88/286/316/323/383, 74/88/323/383, 74/163/286/316/383/394/416, 74/163/314/316, 74/163/314/316/323/394, 74/163/314/323/383/414/416, 74/286, 74/286/316/323, 74/286/394/416, 74/314/323/383/394/414, 74/316/323/394, 85/86/88/163/323/383/394, 85/86/163/314/323/394/414, 85/286, 85/286/323, 86, 86/88/163/323/383/414/422, 86/383/394, 88, 88/163/286/383, 88/286/316/323, 88/286/316/323/383/414/416, 88/316/323, 146, 146/147/395/417, 146/395, 146/395/417, 146/417, 147/395/417/432, 147/417, 149, 157, 163, 163/222/286/316/323/383/394, 163/286, 163/286/314/316/323/414/416, 163/286/314/323/394, 163/286/316/323/394/416, 163/286/414, 163/314/316/394, 163/314/323/394, 163/314/383, 163/314/414, 163/316/323, 163/323, 163/383, 164, 199/417, 259, 260, 284, 286, 286/314/323/383, 286/314/394, 286/316/323/383/414/416, 286/316/383/394, 286/316/394/414/416, 286/323, 286/323/383/414, 286/323/416, 286/383, 286/416, 314/316, 314/316/323, 314/316/323/383/422, 314/316/323/394, 314/316/394, 314/323/383/394, 314/383, 314/383/414/422, 315, 316, 316/323/383/394, 316/323/394/414/416, 316/414/422, 323, 323/383, 323/383/394/414/416, 323/394, 383, 395, 395/417, 395/417/432, 400, 401, 403, 404, 405, 406, 408, 415, 417, 417/432, 420, and 422, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 4. In some additional embodiments, the engineered transaminase comprises a polypeptide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 4, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 74/81/286/316/323/383, 163/286/314/316/323/414/416, 163/286/314/323/394, 286/314/323/383, 286/316/323/383/414/416, 315, and 408, and wherein the amino

acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 4. In some further embodiments, the engineered transaminase comprises a polypeptide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 8, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 5, 18/23/149/260/383/395/401/416, 18/23/149/383, 18/163/164, 21, 21/163/315/316, 21/163/323/408, 21/408, 23/56/86/149/163/164/383/401/416, 23/86, 23/149/260, 23/149/284/383/395, 23/163/164/383, 23/163/164/401/416, 24, 42, 42/110, 42/187/272, 42/187/324/363/366, 42/187/353, 42/272/291, 42/272/291/363, 42/272/324/363/366, 42/272/363/410, 42/272/410, 42/291/313/363/410, 42/291/363, 42/291/363/366, 42/353, 42/363, 46, 66, 77, 86/149/163/164/383/395/401, 86/149/395, 86/163/164/260/383, 86/383, 107, 110, 110/187, 110/187/253/410, 134, 138, 149/164/260/383/395/401, 149/260/383, 149/416, 163/259/323/408, 163/259/408, 163/315/316, 164/260/401, 164/316/383/401, 167, 186, 187, 187/253/363/366, 187/272/324/363/410, 187/272/363, 187/272/363/366/410, 187/291, 189, 191, 195, 199, 203, 210, 211, 248, 259/307, 260/395/401, 272, 272/353, 272/363/366, 272/410, 277, 291, 305, 309, 315, 342, 343, 351, 354, 358, 361, 362, 363, 363/366, 365, 367, 383, 383/401, 383/416/422, 385, 388, 389, 392, 395, 396, 401, 404, 405, 408, 410, 416, 417, 439, 443, 447, 450, and 451, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 8. In yet some additional embodiments, the engineered transaminase comprises a polypeptide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 8, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 18/23/149/383, 21/163/323/408, 272, 291, and 383, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 8. In some additional embodiments, the engineered transaminase comprises a polypeptide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 366, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 24, 24/42/66/291, 24/42/291/362, 24/66/163/191/362/383/388, 24/66/191/199/260/291/351, 24/66/191/199/291, 24/66/191/260/408, 24/66/260/291/383/388/408, 24/66/291/342/383, 24/66/291/365, 24/66/342/365/388/408, 24/77/291, 24/107/163/191/291/351/383/388, 24/107/291/351/365/388, 24/163/351/383, 24/191/291/365, 24/199/260/351/362/383, 24/199/260/362/383/388, 24/260/362/383/388, 24/291, 24/291/342/351/383, 24/291/362/388, 24/291/408, 24/383/388, 24/388, 25, 28, 33, 42/191/408, 42/199/291/383, 42/291/351/362/365/383/388, 42/291/351/362/383/408, 42/291/383/388, 66/82/291/383, 66/163/191/365/383, 66/199/351/383, 66/291, 66/291/362/365/383, 66/291/383/388, 66/383, 77/291, 77/383/388, 86, 107/191/199/365/383/388, 107/191/291/383, 148, 153, 163/291/362/365/383/388, 163/291/383/388, 163/383, 191/199/365/383/388, 191/260/388, 191/291, 191/291/342/362/365, 191/351/383/388, 199/260/383, 199/291, 260, 260/291/365/383/408, 260/365/383,

291, 291/351/383/388, 291/351/383/388/408, 291/362/365, 291/365/388, 291/383, 314, 315, 316, 319, 342/362, 351/383/388, 362, 362/388, 383, 383/388, 396, 397, 405, 406, 413, 419, and 423, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 366. In some further embodiments, the engineered transaminase comprises a polypeptide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 366, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 24/66/191/199/291, 24/66/291/365, 163/291/362/365/383/388, 163/291/383/388, 191/291/342/362/365, 291, and 291/383, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 366. In yet some further embodiments, the engineered transaminase comprises a polypeptide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 650, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 10, 13, 13/24/108/163, 13/24/108/163/311, 13/24/133/199/311, 13/24/163, 13/24/199/311, 13/108, 13/108/199, 13/108/311, 13/199, 13/311, 14, 14/24/108, 14/24/108/133, 14/24/108/199, 14/24/199, 14/108, 14/108/133/311, 14/108/311, 14/311, 24, 24/163, 24/163/199, 35, 72, 73, 78, 95, 101, 108, 108/199, 114, 154, 163, 169, 175/316, 199, 199/311, 226, 293, 311, 316, 382, 383, and 386, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 650. In still some additional embodiments, the engineered transaminase comprises a polypeptide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 650, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 14/108/133/311, 24/163/199, 72, 78, 316, and 383, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 650.

[0014] In some further embodiments, the engineered transaminase comprises a polypeptide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, identical to the sequence of at least one engineered transaminase variant set forth in Table 2-1, 2-2, 3-1, 3-2, 4-1, 4-2, 5-1, and/or 5-2. In yet some additional embodiments, the engineered transaminase is a variant engineered transaminase provided in Table 2-1, 2-2, 3-1, 3-2, 4-1, 4-2, 5-1, and/or 5-2. In some further embodiments, the engineered transaminase comprises a polypeptide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to the sequence of at least one engineered transaminase variant set forth in SEQ ID NOS: 2, 4, 8, 366, and/or 650. In some additional embodiments, the engineered transaminase comprises a polypeptide sequence comprising SEQ ID NOS: 2, 4, 8, 366, and/or 650. In some further embodiments, the engineered transaminase comprises a polypeptide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to the sequence of at least one engineered transaminase variant set forth in the even numbered sequences of

SEQ ID NOS: 6-936. In yet some additional embodiments, the engineered transaminase comprises a polypeptide sequence set forth in the even numbered sequences of SEQ ID NOS: 6-936. In some further embodiments, the engineered transaminase comprises at least one improved property compared to wild-type *V. fluvialis* transaminase. In some additional embodiments, the improved property of the engineered transaminase comprises improved activity on a substrate. In some further embodiments, the substrate comprises compound **(1)**. In yet some additional embodiments, the improved property of the engineered transaminase comprises improved thermostability. In some additional embodiments, the engineered transaminase is purified. The present disclosure also provides compositions comprising an engineered transaminase provided herein. In some embodiments, the compositions comprise more than one engineered transaminase provided herein.

[0015] The present disclosure also provides polynucleotide sequences encoding at least one engineered transaminase provided herein. In some embodiments, the polynucleotide sequence encodes at least one engineered transaminase, said polynucleotide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NOS: 3, 7, 365, and/or 649, wherein the polynucleotide sequence of said engineered transaminase comprises at least one substitution at one or more positions. In some further embodiments, the polynucleotide sequence encodes at least one engineered transaminase comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NOS: 2, 4, 8, 366, and/or 650, or a functional fragment thereof. In yet some additional embodiments, the polynucleotide sequence is operably linked to a control sequence. In still some further embodiments, the polynucleotide sequence is codon optimized.

[0016] The present disclosure also provides expression vectors comprising at least one polynucleotide sequence encoding an engineered transaminase provided herein. In some embodiments, the expression vector comprises at least one polynucleotide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NOS: 3, 7, 365, and/or 649, wherein the polynucleotide sequence of said engineered transaminase comprises at least one substitution at one or more positions. In some embodiments, the expression vector comprises a polynucleotide sequence encoding at least one engineered transaminase comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NOS: 3, 7, 365, and/or 649, or a functional fragment thereof.

[0017] The present disclosure also provides host cells comprising at least one expression vector provided herein. In some embodiments, the host cell comprises at least one polynucleotide sequence provided herein. In some embodiments, the host cell comprises at least one polynucleotide sequence comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NOS: 3, 7, 365, and/or 649, wherein the polynucleotide sequence encoding the engineered transaminase comprises at least one substitution at one or more positions. In some embodiments, the host cell comprises a polynucleotide sequence encoding at least one

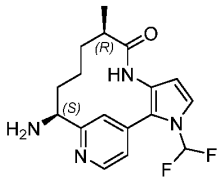
engineered transaminase comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NOS: 2, 4, 8, 366, and/or 650, or a functional fragment thereof. In some embodiments, at least one polynucleotide sequence encoding an engineered transaminase is present in at least one expression vector.

[0018] The present disclosure also provides methods of producing an engineered transaminase in a host cell, comprising culturing the host cell provided herein under suitable conditions, such that at least one engineered transaminase is produced. In some embodiments, the methods further comprise recovering at least one engineered transaminase from the culture and/or host cell. In some additional embodiments, the methods further comprise the step of purifying said at least one engineered transaminase.

[0019] In some embodiments, 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.

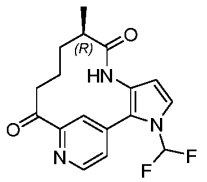
[0020] In some embodiments, the engineered polypeptide having transaminase activity is capable of converting a substrate of compound **(1)** to a product of compound **(2)** under suitable reaction conditions. In some embodiments, the engineered polypeptide is capable of converting compound **(1)** to compound **(2)** with at least 1.2 fold, 2 fold, 5 fold, 10 fold, 20 fold, 25 fold, 50 fold, 75 fold,100 fold, or greater than the activity of are reference sequence SEQ ID NO: 2, 4, 8, 366, and/or 650), under suitable reaction conditions. In some embodiments, the engineered polypeptide is capable of converting compound **(1)** to compound **(2)** with increased activity relative to a reference sequence (e.g., SEQ ID NO: 2, 4, 8, 366, and/or 650), in which the suitable reaction conditions comprise compound **(1)** at a loading of at least 100 g/L, about 1 g/L engineered polypeptide, about 0.5 g/L PLP, about 1 M isopropylamine, about pH 9, and about 50°C.

[0021] In some embodiments, the present disclosure provides a process of preparing compound **(2)**



(2)

comprising a step of contacting a substrate of compound **(1)**



(1)

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

[0022] In some embodiments, the processes using the engineered polypeptides disclosed herein can be carried out wherein the compound **(2)** is produced in at least 90%, 97%, 98%, 99% or greater enantiomeric and diastereomeric excess.

[0023] Any of the processes disclosed herein using the engineered polypeptides for the preparation of compound **(2)** can be carried out under a range of suitable reaction conditions, including but not limited to, ranges of amine donor, pH, temperature, buffer, solvent system, substrate loading, polypeptide loading, cofactor loading, pressure, and reaction time. For example, in some embodiments, the preparation of compound **(2)** can be carried out wherein the suitable reaction conditions comprise: (a) substrate loading of about 10 to 300 g/L of substrate compound (e.g., 50 g/L or 200 g/L of compound **(1)**); (b) of about 0.5 g/L to 60 g/L engineered polypeptide; (c) IPM concentration of about 0.5 to 2 M; (d) PLP cofactor concentration of about 0.1 to 1 g/L; (e) DMSO concentration of about 0% (v/v) to about 20% (v/v); (f) pH of about 8.5 to 11.5; and (g) temperature of about 45°C to 65°C. In some embodiments, the suitable reaction conditions comprise: (a) about 100 g/L of substrate compound (e.g., compound **(1)**); (b) about 1 g/L engineered polypeptide; (c) about 1 M isopropylamine (IPM); (d) about 0.5 g/L pyridoxal phosphate (PLP); (e) about pH 9; and (g) about 50°C.

[0024] In some embodiments, the processes using the engineered polypeptides disclosed herein can be carried out wherein the amino group donor is selected from isopropylamine, alanine, 3-aminobutyric acid, or methylbenzylamine. In some embodiments, the amino group donor is isopropylamine.

DESCRIPTION OF THE INVENTION

[0025] For the descriptions provided herein, the use of the singular includes the plural (and vice versa) unless specifically stated otherwise. For instance, the singular forms “a”, “an” and “the” include plural referents unless the context clearly indicates otherwise. Similarly, “comprise,” “comprises,” “comprising” “include,” “includes,” and “including” are interchangeable and not intended to be limiting.

[0026] 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.”

[0027] Both the foregoing general description, including the drawings, and the following detailed description are exemplary and explanatory only and are not restrictive of this invention. Moreover, the section headings used herein are for organizational purposes only and not to be construed as limiting the subject matter described.

Definitions

[0028] As used herein, the following terms are intended to have the following meanings. In reference to the present invention, the technical and scientific terms used in the descriptions herein will have the

meanings commonly understood by one of ordinary skill in the art, unless specifically defined otherwise. Accordingly, the following terms are intended to have the following meanings. In addition, all patents and publications, including all sequences disclosed within such patents and publications, referred to herein are expressly incorporated by reference.

[0029] Unless otherwise indicated, the practice of the present disclosure involves conventional techniques commonly used in molecular biology, fermentation, microbiology, and related fields, which are known to those of skill in the art. Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present disclosure, the preferred methods and materials are described. Indeed, it is intended that the present invention not be limited to the particular methodology, protocols, and reagents described herein, as these may vary, depending upon the context in which they are used. The headings provided herein are not limitations of the various aspects or embodiments of the present invention that can be had by reference to the specification as a whole. Accordingly, the terms defined below are more fully defined by reference to the specification as a whole.

[0030] Nonetheless, in order to facilitate understanding of the present disclosure, a number of terms are defined below. Numeric ranges are inclusive of the numbers defining the range. Thus, every numerical range disclosed herein is intended to encompass every narrower numerical range that falls within such broader numerical range, as if such narrower numerical ranges were all expressly written herein. It is also intended that every maximum (or minimum) numerical limitation disclosed herein includes every lower (or higher) numerical limitation, as if such lower (or higher) numerical limitations were expressly written herein.

[0031] 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).

[0032] As used herein and in the appended claims, the singular “a”, “an” and “the” include the plural reference unless the context clearly dictates otherwise. Thus, for example, reference to a “host cell” includes a plurality of such host cells.

[0033] Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation and amino acid sequences are written left to right in amino to carboxy orientation, respectively.

[0034] As used herein, the terms “protein,” “polypeptide,” and “peptide” are used interchangeably herein to denote a polymer of at least two amino acids covalently linked by an amide bond, regardless of length or post-translational modification (e.g., glycosylation, phosphorylation, lipidation, myristilation, ubiquitination, etc.). Included within this definition are D- and L-amino acids, and mixtures of D- and L-amino acids.

[0035] The abbreviations used for the genetically encoded amino acids are conventional and are as follows:

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

[0036] 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.

[0037] As used herein, "hydrophilic amino acid or residue" refers to an amino acid or residue having a side chain exhibiting a hydrophobicity of less than zero according to the normalized consensus

hydrophobicity scale of Eisenberg et al., (Eisenberg et al., J. Mol. Biol., 179:125-142 [1984]). 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).

[0038] As used herein, "acidic amino acid or residue" refers to a hydrophilic amino acid or residue having a side chain exhibiting a pK value of less than about 6 when the amino acid is included in a peptide or polypeptide. Acidic amino acids typically have negatively charged side chains at physiological pH due to loss of a hydrogen ion. Genetically encoded acidic amino acids include L-Glu (E) and L-Asp (D).

[0039] As used herein, "basic amino acid or residue" refers to a hydrophilic amino acid or residue having a side chain exhibiting a pK value of greater than about 6 when the amino acid is included in a peptide or polypeptide. Basic amino acids typically have positively charged side chains at physiological pH due to association with hydronium ion. Genetically encoded basic amino acids include L-Arg (R) and L-Lys (K).

[0040] 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).

[0041] 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).

[0042] 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).

[0043] 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.

[0044] 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).

[0045] 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.

[0046] 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).

[0047] 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).

[0048] As used herein, "amino acid difference" and "residue difference" refer to a difference in the amino acid residue at a position of a polypeptide sequence relative to the amino acid residue at a corresponding position in a reference sequence. The positions of amino acid differences generally are referred to herein as "Xn," where n refers to the corresponding position in the reference sequence upon which the residue difference is based. For example, a "residue difference at position X40 as compared to SEQ ID NO:2" refers to a difference of the amino acid residue at the polypeptide position corresponding to position 40 of SEQ ID NO:2. Thus, if the reference polypeptide of SEQ ID NO:2 has a histidine at position 40, then a "residue difference at position X40 as compared to SEQ ID NO:2" refers to an amino acid substitution of any residue other than histidine at the position of the polypeptide corresponding to position 40 of SEQ ID NO:2. In most instances herein, the specific amino acid residue difference at a position is indicated as "XnY" where "Xn" specified the corresponding position as described above, and "Y" is the single letter identifier of the amino acid found in the engineered polypeptide (i.e., the different residue than in the reference polypeptide). In some instances, the present disclosure also provides specific amino acid differences denoted by the conventional notation "AnB", where A is the single letter identifier of the residue in the reference sequence, "n" is the number of the residue position in the reference sequence, and B is the single letter identifier of the residue substitution in the sequence of the engineered polypeptide. In some instances, a polypeptide of the present disclosure can include one or

more amino acid residue differences relative to a reference sequence, which is indicated by a list of the specified positions where residue differences are present relative to the reference sequence. In some embodiments, where more than one amino acid can be used in a specific residue position of a polypeptide, the various amino acid residues that can be used are separated by a "/" (e.g., X192A/G). In some embodiments, in which there are variants with multiple substitutions, the substitutions are separated by either a semicolon (;) or a slash (/) (e.g., Y17V;I259T;E347K or Y17V/I259T/E347K).

[0049] The present disclosure includes engineered polypeptide sequences comprising one or more amino acid differences that include either/or both conservative and non-conservative amino acid substitutions. The amino acid sequences of the specific recombinant carbonic anhydrase polypeptides included in the Sequence Listing of the present disclosure include an initiating methionine (M) residue (i.e., M represents residue position 1). The skilled artisan, however, understands that this initiating methionine residue can be removed by biological processing machinery, such as in a host cell or in vitro translation system, to generate a mature protein lacking the initiating methionine residue, but otherwise retaining the enzyme's properties. Consequently, the term "amino acid residue difference relative to SEQ ID NO:2 at position Xn" as used herein may refer to position "Xn" or to the corresponding position (e.g., position (X-1)n) in a reference sequence that has been processed so as to lack the starting methionine.

[0050] As used herein, the phrase "conservative amino acid substitutions" refers to the interchangeability of residues having similar side chains, and thus typically involves substitution of the amino acid in the polypeptide with amino acids within the same or similar defined class of amino acids. By way of example and not limitation, in some embodiments, an amino acid with an aliphatic side chain is substituted with another aliphatic amino acid (e.g., alanine, valine, leucine, and isoleucine); an amino acid with a hydroxyl side chain is substituted with another amino acid with a hydroxyl side chain (e.g., serine and threonine); an amino acids having aromatic side chains is substituted with another amino acid having an aromatic side chain (e.g., phenylalanine, tyrosine, tryptophan, and histidine); an amino acid with a basic side chain is substituted with another amino acid with a basis side chain (e.g., lysine and arginine); an amino acid with an acidic side chain is substituted with another amino acid with an acidic side chain (e.g., aspartic acid or glutamic acid); and/or a hydrophobic or hydrophilic amino acid is replaced with another hydrophobic or hydrophilic amino acid, respectively. Exemplary conservative substitutions are provided in Table 1.

Table 1. Exemplary Conservative Amino Acid Substitutions

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

[0051] As used herein, the phrase "non-conservative substitution" refers to substitution of an amino acid in the polypeptide with an amino acid with significantly differing side chain properties. Non-conservative substitutions may use amino acids between, rather than within, the defined groups and affects (a) the structure of the peptide backbone in the area of the substitution (e.g., proline for glycine) (b) the charge or hydrophobicity, or (c) the bulk of the side chain. By way of example and not limitation, an exemplary non-conservative substitution can be an acidic amino acid substituted with a basic or aliphatic amino acid; an aromatic amino acid substituted with a small amino acid; and a hydrophilic amino acid substituted with a hydrophobic amino acid.

[0052] As used herein, "deletion" refers to modification of the polypeptide by removal of one or more amino acids from the reference polypeptide. Deletions can comprise removal of 1 or more amino acids, 2 or more amino acids, 5 or more amino acids, 10 or more amino acids, 15 or more amino acids, or 20 or more amino acids, up to 10% of the total number of amino acids, or up to 20% of the total number of amino acids making up the polypeptide while retaining enzymatic activity and/or retaining the improved properties of an engineered enzyme. Deletions can be directed to the internal portions and/or terminal portions of the polypeptide. In various embodiments, the deletion can comprise a continuous segment or can be discontinuous.

[0053] As used herein, "insertion" refers to modification of the polypeptide by addition of one or more amino acids to the reference polypeptide. In some embodiments, the improved engineered transaminase enzymes comprise insertions of one or more amino acids to the naturally occurring transaminase polypeptide as well as insertions of one or more amino acids to engineered transaminase polypeptides. Insertions can be in the internal portions of the polypeptide, or to the carboxy or amino terminus. Insertions as used herein include fusion proteins as is known in the art. The insertion can be a contiguous segment of amino acids or separated by one or more of the amino acids in the naturally occurring polypeptide.

[0054] The term "amino acid substitution set" or "substitution set" refers to a group of amino acid substitutions in a polypeptide sequence, as compared to a reference sequence. A substitution set can have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or more amino acid substitutions. In some embodiments, a substitution set refers to the set of amino acid substitutions that is present in any of the variant transaminases listed in the Tables provided in the Examples. The term "substitution set" is also used in reference to a group of nucleotide substitutions in a polynucleotide sequence, as compared to a reference sequence.

[0055] As used herein, "fragment" refers to a polypeptide that has an amino-terminal and/or carboxy-terminal deletion, but where the remaining amino acid sequence is identical to the corresponding positions in the sequence. Fragments can typically have about 80%, about 90%, about 95%, about 98%, or about 99% of the full-length transaminase polypeptide, for example the polypeptide of SEQ ID NO:2. In some embodiments, the fragment is "biologically active" (i.e., it exhibits the same enzymatic activity as the full-length sequence).

[0056] As used herein, "isolated polypeptide" refers to a polypeptide that is substantially separated from other contaminants that naturally accompany it (e.g., proteins, lipids, and polynucleotides). The term embraces polypeptides which have been removed or purified from their naturally-occurring environment or expression system (e.g., host cell or in vitro synthesis). The improved transaminase enzymes may be present within a cell, present in the cellular medium, or prepared in various forms, such as lysates or isolated preparations. As such, in some embodiments, the engineered transaminase polypeptides of the present disclosure can be an isolated polypeptide.

[0057] As used herein, "substantially pure polypeptide" refers to a composition in which the polypeptide species is the predominant species present (i.e., on a molar or weight basis it is more abundant than any other individual macromolecular species in the composition), and is generally a substantially purified composition when the object species comprises at least about 50 percent of the macromolecular species present by mole or % weight. Generally, a substantially pure engineered transaminase polypeptide composition comprises about 60% or more, about 70% or more, about 80% or more, about 90% or more, about 91% or more, about 92% or more, about 93% or more, about 94% or more, about 95% or more, about 96% or more, about 97% or more, about 98% or more, or about 99% of all macromolecular species by mole or % weight present in the composition. Solvent species, small molecules (<500 Daltons), and elemental ion species are not considered macromolecular species. In some embodiments, the isolated improved transaminase polypeptide is a substantially pure polypeptide composition.

[0058] As used herein, "substantially pure polynucleotide" refers to a composition in which the polynucleotide species is the predominant species present (i.e., on a molar or weight basis it is more abundant than any other individual macromolecular species in the composition), and is generally a substantially purified composition when the object species comprises at least about 50 percent of the macromolecular species present by mole or % weight. Generally, a substantially pure engineered transaminase polynucleotide composition comprises about 60% or more, about 70% or more, about 80% or more, about 90% or more, about 91% or more, about 92% or more, about 93% or more, about 94% or more, about 95% or more, about 96% or more, about 97% or more, about 98% or more, or about 99% of all macromolecular species by mole or % weight present in the composition. In some embodiments, the isolated improved transaminase polypeptide is a substantially pure polynucleotide composition.

[0059] As used herein, "polynucleotide" and "nucleic acid' refer to two or more nucleosides that are covalently linked together. The polynucleotide may be wholly comprised ribonucleosides (i.e., an RNA), wholly comprised of 2' deoxyribonucleotides (i.e., a DNA) or mixtures of ribo- and 2' deoxyribonucleosides. While the nucleosides will typically be linked together via standard phosphodiester linkages, the polynucleotides may include one or more non-standard linkages. The polynucleotide may be single-stranded or double-stranded, or may include both single-stranded regions and double-stranded regions. Moreover, while a polynucleotide will typically be composed of the naturally occurring encoding nucleobases (i.e., adenine, guanine, uracil, thymine, and cytosine), it may

include one or more modified and/or synthetic nucleobases (e.g., inosine, xanthine, hypoxanthine, etc.). Preferably, such modified or synthetic nucleobases will be encoding nucleobases.

[0060] 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.

[0061] As used herein, "hybridization stringency" relates to hybridization conditions, such as washing conditions, in the hybridization of nucleic acids. Generally, hybridization reactions are performed under conditions of lower stringency, followed by washes of varying but higher stringency. The term "moderately stringent hybridization" refers to conditions that permit target-DNA to bind a complementary nucleic acid that has about 60% identity, preferably about 75% identity, about 85% identity to the target DNA; with greater than about 90% identity to target-polynucleotide. Exemplary moderately stringent conditions are conditions equivalent to hybridization in 50% formamide, 5× Denhart's solution, 5×SSPE, 0.2% SDS at 42°C., followed by washing in 0.2×SSPE, 0.2% SDS, at 42°C. "High stringency hybridization" refers generally to conditions that are about 10°C or less from the thermal melting temperature T_m as determined under the solution condition for a defined polynucleotide sequence. In some embodiments, a high stringency condition refers to conditions that permit hybridization of only those nucleic acid sequences that form stable hybrids in 0.018M NaCl at 65°C. (i.e., if a hybrid is not stable in 0.018M NaCl at 65°C, it will not be stable under high stringency conditions, as contemplated herein). High stringency conditions can be provided, for example, by hybridization in conditions equivalent to 50% formamide, 5× Denhart's solution, 5×SSPE, 0.2% SDS at 42°C, followed by washing in 0.1×SSPE, and 0.1% SDS at 65°C. Another high stringency condition is hybridizing in conditions equivalent to hybridizing in 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 known to those of skill in the art.

[0062] As used herein, "coding sequence" refers to that portion of a nucleic acid (e.g., a gene) that encodes an amino acid sequence of a protein.

[0063] As used herein, "codon optimized" refers to changes in the codons of the polynucleotide encoding a protein to those preferentially used in a particular organism such that the encoded protein is efficiently expressed in the organism of interest. In some embodiments, the polynucleotides encoding the transaminase enzymes may be codon optimized for optimal production from the host organism selected for expression. Although the genetic code is degenerate in that most amino acids are represented by several codons, called "synonyms" or "synonymous" codons, it is well known that codon usage by particular organisms is nonrandom and biased towards particular codon triplets. This codon usage bias

may be higher in reference to a given gene, genes of common function or ancestral origin, highly expressed proteins versus low copy number proteins, and the aggregate protein coding regions of an organism's genome. In some embodiments, the polynucleotides encoding the transaminase enzymes may be codon optimized for optimal production from the host organism selected for expression.

[0064] As used herein, "preferred, optimal, high codon usage bias codons" refers interchangeably to codons that are used at higher frequency in the protein coding regions than other codons that code for the same amino acid. The preferred codons may be determined in relation to codon usage in a single gene, a set of genes of common function or origin, highly expressed genes, the codon frequency in the aggregate protein coding regions of the whole organism, codon frequency in the aggregate protein coding regions of related organisms, or combinations thereof. Codons whose frequency increases with the level of gene expression are typically optimal codons for expression. A variety of methods are known for determining the codon frequency (e.g., codon usage, relative synonymous codon usage) and codon preference in specific organisms, including multivariate analysis, for example, using cluster analysis or correspondence analysis, and the effective number of codons used in a gene (See e.g., GCG CodonPreference, Genetics Computer Group Wisconsin Package; CodonW, John Peden, University of Nottingham; McInerney, Bioinform., 14:372-73 [1998]; Stenico et al., Nucleic Acids Res., 222:437-46 [1994]; and Wright, Gene 87:23-29 [1990]). Codon usage tables are available for a growing list of organisms (See e.g., Wada et al., Nucleic Acids Res., 20:2111-2118 [1992]; Nakamura et al., Nucl. Acids Res., 28:292 [2000]; Duret, et al., *supra*; Henaut and Danchin, "*Escherichia coli* and *Salmonella*," Neidhardt, et al. (eds.), ASM Press, Washington D.C., [1996], p. 2047-2066. The data source for obtaining codon usage may rely on any available nucleotide sequence capable of coding for a protein. These data sets include nucleic acid sequences actually known to encode expressed proteins (e.g., complete protein coding sequences-CDS), expressed sequence tags (ESTS), or predicted coding regions of genomic sequences (See e.g., Uberbacher, Meth. Enzymol., 266:259-281 [1996]; Tiwari et al., Comput. Appl. Biosci., 13:263-270 [1997]).

[0065] As used herein, "control sequence" is defined herein to include all components, which are necessary or advantageous for the expression of a polynucleotide and/or polypeptide of the present disclosure. Each control sequence may be native or foreign to the polynucleotide of interest. Such control sequences include, but are not limited to, a leader, polyadenylation sequence, propeptide sequence, promoter, signal peptide sequence, and transcription terminator.

[0066] As used herein, "operably linked" is defined herein as a configuration in which a control sequence is appropriately placed (i.e., in a functional relationship) at a position relative to a polynucleotide of interest such that the control sequence directs or regulates the expression of the polynucleotide and/or polypeptide of interest.

[0067] As used herein, "promoter sequence" refers to a nucleic acid sequence that is recognized by a host cell for expression of a polynucleotide of interest, such as a coding sequence. The control sequence may comprise an appropriate promoter sequence. The promoter sequence contains transcriptional control

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

[0068] As used herein, "naturally occurring" and "wild-type" refers to the form found in nature. For example, a naturally occurring or wild-type polypeptide or polynucleotide sequence is a sequence present in an organism that can be isolated from a source in nature and which has not been intentionally modified by human manipulation.

[0069] As used herein, "non-naturally occurring," "engineered," and "recombinant" when used in the present disclosure with reference to (e.g., a cell, nucleic acid, or polypeptide), refers to a material, or a material corresponding to the natural or native form of the material, that has been modified in a manner that would not otherwise exist in nature. In some embodiments the material is identical to naturally occurring material, but is produced or derived from synthetic materials and/or by manipulation using recombinant techniques. Non-limiting examples include, among others, recombinant cells expressing genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise expressed at a different level.

[0070] As used herein, "percentage of sequence identity," "percent identity," and "percent identical" refer to comparisons between polynucleotide sequences or polypeptide sequences, and are determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which either the identical nucleic acid base or amino acid residue occurs in both sequences or a nucleic acid base or amino acid residue is aligned with a gap to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Determination of optimal alignment and percent sequence identity is performed using the BLAST and BLAST 2.0 algorithms (See e.g., Altschul et al., J. Mol. Biol. 215: 403-410 [1990]; and Altschul et al., Nucl. Acids Res., 25: 3389-3402 [1977]). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information website.

[0071] Briefly, the BLAST analyses involve first identifying high scoring sequence pairs (HSPs) by identifying short words of length Win the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as, the neighborhood word score threshold (Altschul et al., *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward

score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (See e.g., Henikoff and Henikoff, Proc. Natl. Acad. Sci. USA 89:10915 [1989]).

[0072] Numerous other algorithms are available and known in the art that function similarly to BLAST in providing percent identity for two sequences. Optimal alignment of sequences for comparison can be conducted using any suitable method known in the art (e.g., by the local homology algorithm of Smith and Waterman, Adv. Appl. Math., 2:482 [1981]; by the homology alignment algorithm of Needleman and Wunsch, J. Mol. Biol., 48:443 [1970]; by the search for similarity method of Pearson and Lipman, Proc. Natl. Acad. Sci. USA 85:2444 [1988]; and/or by computerized implementations of these algorithms [GAP, BESTFIT, FASTA, and TFASTA in the GCG Wisconsin Software Package]), or by visual inspection, using methods commonly known in the art. Additionally, determination of sequence alignment and percent sequence identity can employ the BESTFIT or GAP programs in the GCG Wisconsin Software package (Accelrys, Madison WI), using the default parameters provided.

[0073] As used herein, "substantial identity" refers to a polynucleotide or polypeptide sequence that has at least 80 percent sequence identity, at least 85 percent identity and 89 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison window of at least 20 residue positions, frequently over a window of at least 30-50 residues, wherein the percentage of sequence identity is calculated by comparing the reference sequence to a sequence that includes deletions or additions which total 20 percent or less of the reference sequence over the window of comparison. In specific embodiments applied to polypeptides, the term "substantial identity" means that two polypeptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap weights, share at least 80 percent sequence identity, preferably at least 89 percent sequence identity, at least 95 percent sequence identity or more (e.g., 99 percent sequence identity). In some preferred embodiments, residue positions that are not identical differ by conservative amino acid substitutions.

[0074] As used herein, "reference sequence" refers to a defined sequence to which another sequence is compared. A reference sequence may be a subset of a larger sequence, for example, a segment of a full-length gene or polypeptide sequence. Generally, a reference sequence is at least 20 nucleotide or amino acid residues in length, at least 25 residues in length, at least 50 residues in length, or the full length of the nucleic acid or polypeptide. Since two polynucleotides or polypeptides may each (1) comprise a

sequence (i.e., a portion of the complete sequence) that is similar between the two sequences, and (2) may further comprise a sequence that is divergent between the two sequences, sequence comparisons between two (or more) polynucleotides or polypeptide are typically performed by comparing sequences of the two polynucleotides over a comparison window to identify and compare local regions of sequence similarity. The term "reference sequence" is not intended to be limited to wild-type sequences, and can include engineered or altered sequences. For example, in some embodiments, a "reference sequence" can be a previously engineered or altered amino acid sequence.

[0075] As used herein, "comparison window" refers to a conceptual segment of at least about 20 contiguous nucleotide positions or amino acids residues wherein a sequence may be compared to a reference sequence of at least 20 contiguous nucleotides or amino acids and wherein the portion of the sequence in the comparison window may comprise additions or deletions (i.e., gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The comparison window can be longer than 20 contiguous residues, and includes, optionally 30, 40, 50, 100, or longer windows.

[0076] As used herein, "corresponding to," "reference to," and "relative to" when used in the context of the numbering of a given amino acid or polynucleotide sequence refers to the numbering of the residues of a specified reference sequence when the given amino acid or polynucleotide sequence is compared to the reference sequence. In other words, the residue number or residue position of a given polymer is designated with respect to the reference sequence rather than by the actual numerical position of the residue within the given amino acid or polynucleotide sequence. For example, a given amino acid sequence, such as that of an engineered transaminase, can be aligned to a reference sequence by introducing gaps to optimize residue matches between the two sequences. In these cases, although the gaps are present, the numbering of the residue in the given amino acid or polynucleotide sequence is made with respect to the reference sequence to which it has been aligned. As used herein, a reference to a residue position, such as "Xn" as further described below, is to be construed as referring to "a residue corresponding to", unless specifically denoted otherwise. Thus, for example, "X94" refers to any amino acid at position 94 in a polypeptide sequence.

[0077] As used herein, when used in reference to a nucleic acid or polypeptide, the term "heterologous" refers to a sequence that is not normally expressed and secreted by an organism (e.g., a wild-type organism). In some embodiments, the term encompasses a sequence that comprises two or more subsequences which are not found in the same relationship to each other as normally found in nature, or is recombinantly engineered so that its level of expression, or physical relationship to other nucleic acids or other molecules in a cell, or structure, is not normally found in nature. For instance, a heterologous nucleic acid is typically recombinantly produced, having two or more sequences from unrelated genes arranged in a manner not found in nature (e.g., a nucleic acid open reading frame (ORF) of the disclosure operatively linked to a promoter sequence inserted into an expression cassette, such as a vector). In some embodiments, "heterologous polynucleotide" refers to any polynucleotide that is introduced into a host

cell by laboratory techniques, and includes polynucleotides that are removed from a host cell, subjected to laboratory manipulation, and then reintroduced into a host cell.

[0078] As used herein, "improved enzyme property" refers to a transaminase that exhibits an improvement in any enzyme property as compared to a reference transaminase. For the engineered transaminase polypeptides described herein, the comparison is generally made to the wild-type transaminase enzyme, although in some embodiments, the reference transaminase can be another improved engineered transaminase. Enzyme properties for which improvement is desirable include, but are not limited to, enzymatic activity (which can be expressed in terms of percent conversion of the substrate at a specified reaction time using a specified amount of transaminase), chemoselectivity, thermal stability, solvent stability, pH activity profile, cofactor requirements, refractoriness to inhibitors (e.g., product inhibition), stereospecificity, and stereoselectivity (including enantioselectivity).

[0079] As used herein, "increased enzymatic activity" and "increased activity" refer to an improved property of an engineered enzyme, which can be represented by an increase in specific activity (e.g., product produced/time/weight protein) or an increase in percent conversion of the substrate to the product (e.g., percent conversion of starting amount of substrate to product in a specified time period using a specified amount of transaminase) as compared to a reference enzyme as described herein. Any property relating to enzyme activity may be affected, including the classical enzyme properties of K_m, V_{max} or k_{cat}, changes of which can lead to increased enzymatic activity. Comparisons of enzyme activities are made using a defined preparation of enzyme, a defined assay under a set condition, and one or more defined substrates, as further described in detail herein. Generally, when enzymes in cell lysates are compared, the numbers of cells and the amount of protein assayed are determined as well as use of identical expression systems and identical host cells to minimize variations in amount of enzyme produced by the host cells and present in the lysates.

[0080] As used herein, "conversion" refers to the enzymatic transformation of a substrate to the corresponding product.

[0081] As used herein "percent conversion" refers to the percent of the substrate that is converted to the product within a period of time under specified conditions. Thus, for example, the "enzymatic activity" or "activity" of a transaminase polypeptide can be expressed as "percent conversion" of the substrate to the product.

[0082] As used herein, "chemoselectivity" refers to the preferential formation in a chemical or enzymatic reaction of one product over another.

[0083] As used herein, "thermostable" and "thermal stable" are used interchangeably to refer to a polypeptide that is resistant to inactivation when exposed to a set of temperature conditions (e.g., 40-80°C) for a period of time (e.g., 0.5-24 hrs) compared to the untreated enzyme, thus retaining a certain level of residual activity (e.g., more than 60% to 80%) after exposure to elevated temperatures.

[0084] As used herein, "solvent stable" refers to the ability of a polypeptide to maintain similar activity (e.g., more than 60% to 80%) after exposure to varying concentrations (e.g., 5-99%) of solvent (e.g.,

isopropyl alcohol, tetrahydrofuran, 2-methyltetrahydrofuran, acetone, toluene, butylacetate, methyl tert-butylether, etc.) for a period of time (e.g., 0.5-24 hrs) compared to the untreated enzyme.

[0085] 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.

[0086] As used herein, "thermo- and solvent stable" refers to a transaminase polypeptide that is both thermostable and solvent stable.

[0087] As used herein, "suitable reaction conditions" refer to those conditions in the biocatalytic reaction solution (e.g., ranges of enzyme loading, substrate loading, cofactor loading, temperature, pH, buffers, co-solvents, etc.) under which a transaminase polypeptide of the present disclosure is capable of transamination. Exemplary "suitable reaction conditions" are provided in the present disclosure and illustrated by the Examples.

[0088] As used herein, "loading," such as in "compound loading," "enzyme loading," or "cofactor loading" refers to the concentration or amount of a component in a reaction mixture at the start of the reaction.

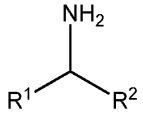
[0089] As used herein, "substrate" in the context of a biocatalyst mediated process refers to the compound or molecule acted on by the biocatalyst.

[0090] As used herein "product" in the context of a biocatalyst mediated process refers to the compound or molecule resulting from the action of the biocatalyst.

[0091] As used herein, "equilibration" as used herein refers to the process resulting in a steady state concentration of chemical species in a chemical or enzymatic reaction (e.g., interconversion of two species A and B), including interconversion of stereoisomers, as determined by the forward rate constant and the reverse rate constant of the chemical or enzymatic reaction.

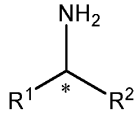
[0092] "Transaminase" or "aminotransferase" are used interchangeably herein to refer to a polypeptide having an enzymatic capability of transferring an amino group ($-NH_2$), a pair of electrons, and a proton from the primary amine of an amine donor compound to the carbonyl group (C=O) of an amine acceptor compound, thereby converting the amine donor compound into its corresponding carbonyl compound and the carbonyl acceptor compound into its corresponding primary amine compound (See e.g., **Scheme 1**). Transaminases as used herein include naturally occurring (wild type) transaminase as well as non-naturally occurring engineered polypeptides generated by human manipulation.

[0093] "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,



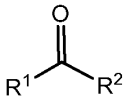
in which each of R^1, and R^2, when taken independently, is an alkyl, an alkylaryl group, or aryl group which is unsubstituted or substituted with one or more enzymatically non-inhibiting groups. R^1 can be the same or different from R^2 in structure or chirality. The groups R^1 and R^2, taken together, 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.

[0094] "Chiral amine" refers to an amino group containing compound having 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 groups. R^1 is different from R^2 in structure causing the carbon bearing the amino group (denoted with a *) to be stereogenic center. The groups R^1 and R^2, taken together, may form a ring that is unsubstituted, substituted, or fused to other rings but is otherwise not symmetrical.

[0095] "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 R^1 and R^2 are defined above for the amino group donor.

[0096] "Amino acceptor" and "amine acceptor," "keto substrate," are used interchangeably herein to refer to a carbonyl group containing compound that accepts the amino group from an amino group donor in a reaction mediated by a transaminase (See e.g., **Scheme 1**). In the context of the present disclosure, the amino acceptor compound for the transaminase can include, among others, compound (**2**).

[0097] "Cofactor," as used herein, refers to a non-protein compound that operates in combination with an enzyme in catalyzing a reaction. As used herein, "cofactor" is intended to encompass the vitamin B_6 family compounds PLP, PN, PL, PM, PNP, and PMP, which are sometimes also referred to as coenzymes.

[0098] "Pyridoxal-phosphate," "PLP," "pyridoxal-5'-phosphate," "PYP," and "P5P" are used interchangeably herein to refer to the compound that acts as a cofactor in transaminase reactions. In some embodiments, pyridoxal phosphate is defined by the structure 1-(4'-formyl-3'-hydroxy-2'-methyl-5'-pyridyl)methoxyphosphonic acid, CAS number [54-47-7]. Pyridoxal-5'-phosphate can be produced *in vivo* by phosphorylation and oxidation of pyridoxol (also known as Vitamin B_6). In transamination

reactions using transaminase enzymes, the amine group of the amino donor is transferred to the cofactor to produce a keto byproduct, while pyridoxal-5'-phosphate is converted to pyridoxamine phosphate. Pyridoxal-5'-phosphate is regenerated by reaction with a different keto compound (the amino acceptor). The transfer of the amine group from pyridoxamine phosphate to the amino acceptor produces an amine and regenerates the cofactor. In some embodiments, the pyridoxal-5'-phosphate can be replaced by other members of the vitamin B_6 family, including pyridoxine (PN), pyridoxal (PL), pyridoxamine (PM), and their phosphorylated counterparts; pyridoxine phosphate (PNP), and pyridoxamine phosphate (PMP).

[0099] "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).

[0100] "Alkenyl" refers to groups of from 2 to 12 carbon atoms, either straight or branched containing at least one double bond but optionally containing more than one double bond.

[0101] "Alkynyl" refers to groups of from 2 to 12 carbon atoms, either straight or branched containing at least one triple bond but optionally containing more than one triple bond, and optionally containing one or more double bonded moieties.

[0102] "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.

[0103] "Arylalkyl" refers to an alkyl substituted with an aryl moiety. Representative arylalkyl groups include benzyl, phenethyl and the like.

[0104] "Arylalkenyl" refers to an alkenyl as defined herein substituted with an aryl group.

[0105] "Arylalkynyl" refers to an alkynyl as defined herein substituted with an aryl group.

[0106] "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., indolizinyl or benzothienyl). For multiple condensed rings, at least one of the rings is aromatic.

[0107] "Heteroarylalkyl" refers to an alkyl substituted with a heteroaryl moiety as defined herein.

[0108] "Heteroarylalkenyl" refers to an alkenyl substituted with a heteroaryl group as defined herein.

[0109] "Heteroarylalkynyl" refers to an alkynyl substituted with a heteroaryl moiety as defined herein.

[0110] "Cycloalkyl" refers to cyclic alkyl groups of from 3 to 12 carbon atoms having a single cyclic ring or multiple condensed rings. Representative cycloalkyl groups include, by way of example, single ring structures such as cyclopropyl, cyclobutyl, cyclopentyl, cyclooctyl, 1-methylcyclopropyl, 2-methylcyclopentyl, 2-methylcyclooctyl, and the like, or multiple ring structures, including bridged ring systems, such as adamantyl, and the like.

[0111] "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., indolinyl, dihydrobenzofuran or quinuclidinyl). Representative heterocycles and heteroaryls include, but are not limited to, furan, thiophene, thiazole, oxazole, pyrrole, imidazole, pyrazole, pyridine, pyrazine, pyrimidine, pyridazine, indolizine, isoindole, indole, indazole, purine, quinolizine, isoquinoline, quinoline, phthalazine, naphthylpyridine, quinoxaline, quinazoline, cinnoline, pteridine, carbazole, carboline, phenanthridine, acridine, phenanthroline, isothiazole, phenazine, isoxazole, phenoxazine, phenothiazine, imidazolidine, imidazoline, piperidine, piperazine, pyrrolidine, indoline and the like.

[0112] "Cycloalkylalkyl" refers to an alkyl substituted with a cycloalkyl moiety as defined herein.

[0113] "Cycloalkylalkenyl" refers to an alkenyl substituted with a cycloalkyl moiety as defined herein.

[0114] "Cycloalkylalkynyl" refers to an alkynyl substituted with a cycloalkyl moiety as defined herein.

[0115] "Heterocycloalkylalkyl" refers to an alkyl substituted with a heterocycloalkyl moiety as defined herein.

[0116] "Heterocycoalkenyl" refers to an alkenyl substituted with a heterocycloalkyl moiety as defined herein.

[0117] "Heterocycloalkylalkynyl" refers to an alkynyl substituted with a heterocycloalkyl moiety as defined herein.

[0118] "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.

[0119] "Amino" refers to the group $-NH_2$. Substituted amino refers to the group –NHR', NR'R', and NR'R'R' , where each R' is independently of the others selected from substituted or unsubstituted alkyl, cycloalkyl, heterocycloalkyl, cycloalkylalkyl, heterocycloalkylalkyl, alkyloxy, aryl, heteroaryl, arylalkyl, heteroarylalkyl, acyl, alkyloxycarbonyl, sulfanyl, sulfinyl, sulfonyl, and the like. Typical amino groups include, but are not limited to, dimethylamino, diethylamino, trimethylammonium, triethylammonium, methylysulfonylamino, furanyl-oxy-sulfamino, and the like.

[0120] "Carboxy" refers to –COOH.

[0121] "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.

[0122] "Hydroxy" refers to –OH.

[0123] "Cyano" refers to –CN.

[0124] "Halogen" or "halo" refers to fluoro, chloro, bromo and iodo.

[0125] "Sulfonyl" refers to $-SO_2-$. Substituted sulfonyl refers to $-SO_2R'$, where R' is a suitable substituent as described below.

[0126] "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.

[0127] "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 disclosure, and is otherwise chemically reasonable.

[0128] "Optional" or "optionally" means that the subsequently described event or circumstance may or may not occur, and that the description includes instances where said event or circumstance occurs and instances in which it does not. One of ordinary skill in the art would understand that with respect to any molecule described as containing one or more optional substituents, only sterically practical and/or synthetically feasible compounds are meant to be included. "Optionally substituted" refers to all subsequent modifiers in a term or series of chemical groups. For example, in the term "optionally substituted arylalkyl, the "alkyl" portion and the "aryl" portion of the molecule may or may not be substituted, and for the series "optionally substituted alkyl, cycloalkyl, aryl and heteroaryl," the alkyl, cycloalkyl, aryl, and heteroaryl groups, independently of the others, may or may not be substituted.

[0129] "Protecting group" refers to a group of atoms that mask, reduce or prevent the reactivity of the functional group when attached to a reactive functional group in a molecule. Typically, a protecting group may be selectively removed as desired during the course of a synthesis. Examples of protecting groups are known in the art (e.g., Wuts and Greene, "Greene's Protective Groups in Organic Synthesis," 4th Ed., Wiley Interscience [2006], and Harrison *et al.*, Compendium of Synthetic Organic Methods, Vols. 1-8, John Wiley & Sons, NY [1971-1976]. Functional groups that can have a protecting group

include, but are not limited to, hydroxy, amino, and carboxy groups. Representative amino protecting groups include, but are not limited to, formyl, acetyl, trifluoroacetyl, benzyl, benzyloxycarbonyl ("CBZ"), tert-butoxycarbonyl ("Boc"), trimethylsilyl ("TMS"), 2-trimethylsilyl-ethanesulfonyl ("SES"), trityl and substituted trityl groups, allyloxycarbonyl, 9-fluorenylmethyloxycarbonyl ("FMOC"), nitro-veratryloxycarbonyl ("NVOC") and the like. 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.

[0130] "Leaving group" generally refers to any atom or moiety that is capable of being displaced by another atom or moiety in a chemical reaction. More specifically, a leaving group refers to an atom or moiety that is readily displaced and substituted by a nucleophile (e.g., an amine, a thiol, an alcohol, or cyanide). Such leaving groups are well known and include carboxylates, N-hydroxysuccinimide ("NHS"), N-hydroxybenzotriazole, a halogen (fluorine, chlorine, bromine, or iodine), and alkyloxy groups. Non-limiting characteristics and examples of leaving groups are known in the art and described in various chemistry texts.

Engineered Transaminase Polypeptides

[0131] The present disclosure provides engineered polypeptides having transaminase activity (also referred to herein as "engineered transaminase polypeptides") useful for the selective transamination of amino acceptor substrate compounds to produce chiral amine products, which, in some embodiments, can include compound **(2).** Accordingly, in one aspect, the present disclosure provides engineered polypeptides having transaminase activity which are capable of converting substrate compound **(1)** to product compound **(2)** as shown in **Scheme 2**. Further, the present disclosure provides polynucleotides encoding the engineered polypeptides, associated vectors and host cells comprising the polynucleotides, methods for making the engineered polypeptides, and methods for using the engineered polypeptides, including suitable reaction conditions.

[0132] The engineered polypeptides of the present 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 *Vibrio fluvialis* JS17 (GenBank Acc. No. AEA39183.1, GI: 327207066; SEQ ID NO:2), and also as compared to the reference engineered transaminase polypeptide of SEQ ID NO: 4, which was used as the starting backbone sequence for the directed evolution of the engineered polypeptides of the present disclosure. The reference engineered transaminase polypeptide of SEQ ID NO:4 has the following 11 amino acid differences relative to the wild-type transaminase of *Vibrio fluvialis* JS17 (SEQ ID NO:2): A9T, N45H, W57L, F86S, R88H, V153A, V177L, R211K, M294V, S324G, and T391A.

[0133] The engineered transaminase polypeptides of the present disclosure were generated by directed evolution of SEQ ID NO: 4 for efficient conversion of compound **(1)** to compound **(2)** under certain industrially relevant conditions and have one or more residue differences as compared to a reference engineered transaminase polypeptide. These residue differences are associated with improvements in various enzyme properties, particularly increased activity, increased stereoselectivity, increased stability, and tolerance of increased substrate and/or product concentration (*e.g.*, decreased product inhibition). Accordingly, in some embodiments, the engineered polypeptides having transaminase activity are capable of converting the substrate compound **(1)** to compound **(2)** with an activity that is increased at least about 1.2 fold, 1.5 fold, 2 fold, 3 fold, 4 fold, 5 fold, 10 fold, 20 fold, 30 fold, 40 fold, 50 fold, 100 fold, or more relative to the activity of a reference polypeptide (e.g., SEQ ID NO:2, 4, 8, 366, and/or 650), under suitable reaction conditions. In some embodiments, the engineered polypeptides having transaminase activity are capable of converting the substrate of compound **(1)** to compound **(2)** with a percent conversion of at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, or at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, in a reaction time of about 48 h, about 36 h, about 24 h, or even a shorter length of time, under suitable reaction conditions. In some embodiments, the engineered polypeptides having transaminase activity are capable of converting compound **(1)** to compound **(2)** in diastereomeric excess of at least 90%, 95%, 97%, 98%, 99%, or greater, under suitable reaction conditions.

[0134] The present disclosure provides numerous exemplary engineered transaminase polypeptides comprising amino acid sequences of the even-numbered sequence identifiers SEQ ID NO: 6 - 936. These exemplary engineered transaminase polypeptides comprise amino acid sequences that include one or more of the following residue differences associated with their improved properties for conversion of compound **(1)** to compound **(2)** as compared to a reference sequence (e.g., SEQ ID NO: 2, 4, 8, 366, and/or 650).

[0135] In some cases, the exemplary engineered polypeptides have an amino acid sequence that further comprises one or more residue differences as compared to a reference sequence (e.g., SEQ ID NO: 2, 4, 8, 366, and/or 650). In some cases, the exemplary engineered polypeptides have an amino acid sequence that further comprises one or more residue differences as compared to a reference sequence (e.g., SEQ ID NO: 2, 4, 8, 366, and/or 650).

[0136] In some embodiments, the engineered polypeptide comprises an amino acid sequence that is at least about 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, identical to a reference sequence selected from SEQ ID NO: 2, 4, 8, 366, and/or 650, where the polypeptide has transaminase activity and one or more of the improved properties as described herein, for example the ability to convert compound **(1)** to product compound **(2)** with increased activity compared to a reference sequence (e.g., the polypeptide of SEQ ID NO: 2, 4, 8, 366, and/or 650). In some embodiments, the reference sequence is SEQ ID NO: 2. In some embodiments, the reference sequence is

SEQ ID NO: 4. In some embodiments, the reference sequence is SEQ ID NO: 8. In some embodiments, the reference sequence is SEQ ID NO: 366. In some embodiments, the reference sequence is SEQ ID NO: 650.

[0137] In some embodiments, the engineered transaminase polypeptide comprising an amino acid sequence has one or more amino acid residue differences as compared to SEQ ID NO: 2, 4, 8, 366, and/or 650. In some embodiments, the present disclosure provides an engineered polypeptide having transaminase activity comprising an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to reference sequence of SEQ ID NO: 2, 4, 8, 366, and/or 650 and (a) at least one amino acid residue difference selected from those substitutions provided herein (See e.g., Tables 2-1, 2-2, 3-1, 3-2, 4-1, 4-2, 5-1, and/or 5-2).

[0138] In some embodiments, the present disclosure provides an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO: 4 selected 18, 20, 21, 21/23/56/146, 21/23/56/146/432, 21/23/146/417, 21/23/395/417/432, 21/53/56, 21/53/417, 21/56/395, 21/417/432, 23, 23/53, 23/53/56, 23/53/56/146/395, 23/53/395, 23/53/417, 23/53/432, 23/56, 23/56/395, 23/56/395/417, 23/395/417, 23/417, 23/417/432, 53, 53/56, 53/146/417, 53/395, 56, 56/74/241/286/314/316/323, 56/86/163/314/316/383/414/416/422, 56/86/286/314/414/416, 56/86/314/316/323/394/414/422, 56/146/417, 56/146/432, 56/147, 56/163, 56/163/286/316/323/383/394, 56/286/314/316/323/422, 56/286/383, 56/323, 56/323/383, 56/323/383/394, 56/383, 56/395, 74/81/286/316/323/383, 74/85/86/163/286/316/323/394, 74/85/314/316/414/416, 74/86/163/316, 74/86/316/323/383/394, 74/88/286/316/323/383, 74/88/323/383, 74/163/286/316/383/394/416, 74/163/314/316, 74/163/314/316/323/394, 74/163/314/323/383/414/416, 74/286, 74/286/316/323, 74/286/394/416, 74/314/323/383/394/414, 74/316/323/394, 85/86/88/163/323/383/394, 85/86/163/314/323/394/414, 85/286, 85/286/323, 86, 86/88/163/323/383/414/422, 86/383/394, 88, 88/163/286/383, 88/286/316/323, 88/286/316/323/383/414/416, 88/316/323, 146, 146/147/395/417, 146/395, 146/395/417, 146/417, 147/395/417/432, 147/417, 149, 157, 163, 163/222/286/316/323/383/394, 163/286, 163/286/314/316/323/414/416, 163/286/314/323/394, 163/286/316/323/394/416, 163/286/414, 163/314/316/394, 163/314/323/394, 163/314/383, 163/314/414, 163/316/323, 163/323, 163/383, 164, 199/417, 259, 260, 284, 286, 286/314/323/383, 286/314/394, 286/316/323/383/414/416, 286/316/383/394, 286/316/394/414/416, 286/323, 286/323/383/414, 286/323/416, 286/383, 286/416, 314/316, 314/316/323, 314/316/323/383/422, 314/316/323/394, 314/316/394, 314/323/383/394, 314/383, 314/383/414/422, 315, 316, 316/323/383/394, 316/323/394/414/416, 316/414/422, 323, 323/383, 323/383/394/414/416, 323/394, 383, 395, 395/417, 395/417/432, 400, 401, 403, 404, 405, 406, 408, 415, 417, 417/432, 420, and 422, wherein the positions are numbered with reference to SEQ ID NO: 4. In some embodiments, the amino acid differences comprise the substitution(s) 18A, 20C, 21H, 21P/23S/56C/146H, 21P/23S/56C/146H/432V, 21P/23S/146H/417V, 21P/23S/395D/417S/432V,

21P/53C/56C, 21P/53C/417S, 21P/56C/395D, 21P/417S/432V, 21R, 23A, 23R, 23S/53C, 23S/53C/56C, 23S/53C/56C/146H/395D, 23S/53C/395D, 23S/53C/417S, 23S/53C/432V, 23S/56C, 23S/56C/395D, 23S/56C/395D/417V, 23S/395D/417S, 23S/417S, 23S/417V, 23S/417V/432V, 53C, 53C/56C, 53C/146H/417S, 53C/395D, 56A, 56A/74T/241V/286S/314R/316W/323T, 56A/86A/163F/314R/316W/383V/414V/416A/422A, 56A/86A/286S/314R/414V/416A, 56A/86A/314R/316W/323T/394G/414V/422A, 56A/163F, 56A/163F/286S/316W/323T/383V/394G, 56A/286S/314R/316W/323T/422A, 56A/286S/383V, 56A/323T, 56A/323T/383V, 56A/323T/383V/394G, 56A/383V, 56C, 56C/146H/417V, 56C/146H/432V, 56C/147R, 56C/395D, 56T, 56V, 74T/81S/286S/316W/323T/383V, 74T/85V/86A/163F/286S/316W/323T/394G, 74T/85V/314R/316W/414V/416A, 74T/86A/163F/316W, 74T/86A/316W/323T/383V/394G, 74T/88R/286S/316W/323T/383V, 74T/88R/323T/383V, 74T/163F/286S/316W/383V/394G/416A, 74T/163F/314R/316W, 74T/163F/314R/316W/323T/394G, 74T/163F/314R/323T/383V/414V/416A, 74T/286S, 74T/286S/316W/323T, 74T/286S/394G/416A, 74T/314R/323T/383V/394G/414V, 74T/316W/323T/394G, 85V/86A/88R/163F/323T/383V/394G, 85V/86A/163F/314R/323T/394G/414V, 85V/286S, 85V/286S/323T, 86A/88R/163F/323T/383V/414V/422A, 86A/383V/394G, 86G, 88R/163F/286S/383V, 88R/286S/316W/323T, 88R/286S/316W/323T/383V/414V/416A, 88R/316W/323T, 88S, 88T, 146H, 146H/147R/395D/417S, 146H/395D, 146H/395D/417S, 146H/417S, 146H/417V, 147R/395D/417S/432V, 147R/417S, 149S, 157A, 163F, 163F/222V/286S/316W/323T/383V/394G, 163F/286S, 163F/286S/314R/316W/323T/414V/416A, 163F/286S/314R/323T/394G, 163F/286S/316W/323T/394G/416A, 163F/286S/414V, 163F/314R/316W/394G, 163F/314R/323T/394G, 163F/314R/383V, 163F/314R/414V, 163F/316W/323T, 163F/323T, 163F/383V, 163L, 163M, 164A, 164D, 164Q, 164S, 199V/417S, 259V, 260T, 284A, 286S, 286S/314R/323T/383V, 286S/314R/394G, 286S/316W/323T/383V/414V/416A, 286S/316W/383V/394G, 286S/316W/394G/414V/416A, 286S/323T, 286S/323T/383V/414V, 286S/323T/416A, 286S/383V, 286S/416A, 314R/316W, 314R/316W/323T, 314R/316W/323T/383V/422A, 314R/316W/323T/394G, 314R/316W/394G, 314R/323T/383V/394G, 314R/383V, 314R/383V/414V/422A, 315G, 315R, 316A, 316F, 316G, 316H, 316L, 316N, 316R, 316V, 316W/323T/383V/394G, 316W/323T/394G/414V/416A, 316W/414V/422A, 323C, 323S, 323T, 323T/383V, 323T/383V/394G/414V/416A, 323T/394G, 383V, 395D, 395D/417S, 395D/417S/432V, 395D/417V, 400D, 401A, 401K, 401S, 403V, 404S, 405H, 405W, 406S, 408F, 408L, 408W, 415G, 415W, 417A, 417S, 417S/432V, 417V, 417V/432V, 420G, 422L, and 422T, and 424R, wherein the positions are numbered with reference to SEQ ID NO: 4. In some additional embodiments, the amino acid differences comprise the substitution(s) G18A, T20C, D21H, D21P/P23S/L56C/R146H, D21P/P23S/L56C/R146H/A432V, D21P/P23S/R146H/L417V, D21P/P23S/G395D/L417S/A432V, D21P/N53C/L56C, D21P/N53C/L417S, D21P/L56C/G395D, D21P/L417S/A432V, D21R, P23A, P23R, P23S/N53C, P23S/N53C/L56C, P23S/N53C/L56C/R146H/G395D, P23S/N53C/G395D, P23S/N53C/L417S, P23S/N53C/A432V, P23S/L56C, P23S/L56C/G395D, P23S/L56C/G395D/L417V,

P23S/G395D/L417S, P23S/L417S, P23S/L417V, P23S/L417V/A432V, N53C, N53C/L56C, N53C/R146H/L417S, N53C/G395D, L56A, L56A/A74T/A241V/N286S/I314R/E316W/A323T, L56A/S86A/K163F/I314R/E316W/A383V/C414V/P416A/V422A, L56A/S86A/N286S/I314R/C414V/P416A, L56A/S86A/I314R/E316W/A323T/D394G/C414V/V422A, L56A/K163F, L56A/K163F/N286S/E316W/A323T/A383V/D394G, L56A/N286S/I314R/E316W/A323T/V422A, L56A/N286S/A383V, L56A/A323T, L56A/A323T/A383V, L56A/A323T/A383V/D394G, L56A/A383V, L56C, L56C/R146H/L417V, L56C/R146H/A432V, L56C/W147R, L56C/G395D, L56T, L56V, A74T/G81S/N286S/E316W/A323T/A383V, A74T/F85V/S86A/K163F/N286S/E316W/A323T/D394G, A74T/F85V/I314R/E316W/C414V/P416A, A74T/S86A/K163F/E316W, A74T/S86A/E316W/A323T/A383V/D394G, A74T/H88R/N286S/E316W/A323T/A383V, A74T/H88R/A323T/A383V, A74T/K163F/N286S/E316W/A383V/D394G/P416A, A74T/K163F/I314R/E316W, A74T/K163F/I314R/E316W/A323T/D394G, A74T/K163F/I314R/A323T/A383V/C414V/P416A, A74T/N286S, A74T/N286S/E316W/A323T, A74T/N286S/D394G/P416A, A74T/I314R/A323T/A383V/D394G/C414V, A74T/E316W/A323T/D394G, F85V/S86A/H88R/K163F/A323T/A383V/D394G, F85V/S86A/K163F/I314R/A323T/D394G/C414V, F85V/N286S, F85V/N286S/A323T, S86A/H88R/K163F/A323T/A383V/C414V/V422A, S86A/A383V/D394G, S86G, H88R/K163F/N286S/A383V, H88R/N286S/E316W/A323T, H88R/N286S/E316W/A323T/A383V/C414V/P416A, H88R/E316W/A323T, H88S, H88T, R146H, R146H/W147R/G395D/L417S, R146H/G395D, R146H/G395D/L417S, R146H/L417S, R146H/L417V, W147R/G395D/L417S/A432V, W147R/L417S, A149S, S157A, K163F, K163F/A222V/N286S/E316W/A323T/A383V/D394G, K163F/N286S, K163F/N286S/I314R/E316W/A323T/C414V/P416A, K163F/N286S/I314R/A323T/D394G, K163F/N286S/E316W/A323T/D394G/P416A, K163F/N286S/C414V, K163F/I314R/E316W/D394G, K163F/I314R/A323T/D394G, K163F/I314R/A383V, K163F/I314R/C414V, K163F/E316W/A323T, K163F/A323T, K163F/A383V, K163L, K163M, P164A, P164D, P164Q, P164S, A199V/L417S, I259V, C260T, S284A, N286S, N286S/I314R/A323T/A383V, N286S/I314R/D394G, N286S/E316W/A323T/A383V/C414V/P416A, N286S/E316W/A383V/D394G, N286S/E316W/D394G/C414V/P416A, N286S/A323T, N286S/A323T/A383V/C414V, N286S/A323T/P416A, N286S/A383V, N286S/P416A, I314R/E316W, I314R/E316W/A323T, I314R/E316W/A323T/A383V/V422A, I314R/E316W/A323T/D394G, I314R/E316W/D394G, I314R/A323T/A383V/D394G, I314R/A383V, I314R/A383V/C414V/V422A, E315G, E315R, E316A, E316F, E316G, E316H, E316L, E316N, E316R, E316V, E316W/A323T/A383V/D394G, E316W/A323T/D394G/C414V/P416A, E316W/C414V/V422A, A323C, A323S, A323T, A323T/A383V, A323T/A383V/D394G/C414V/P416A, A323T/D394G, A383V, G395D, G395D/L417S, G395D/L417S/A432V, G395D/L417V, S400D, E401A, E401K, E401S, I403V, A404S, N405H, N405W, T406S, T408F, T408L, T408W, R415G, R415W, L417A, L417S, L417S/A432V, L417V,

L417V/A432V, S420G, V422L, and V422T, wherein the positions are numbered with reference to SEQ ID NO: 4.

[0139] In some embodiments, the present disclosure provides an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO: 4 selected from 74/81/286/316/323/383, 163/286/314/316/323/414/416, 163/286/314/323/394, 286/314/323/383, 286/316/323/383/414/416, 315, and 408, wherein the positions are numbered with reference to SEQ ID NO: 4. In some embodiments, the amino acid difference(s) comprise the substitution(s) 74T/81S/286S/316W/323T/383V, 163F/286S/314R/316W/323T/414V/416A, 163F/286S/314R/323T/394G, 286S/314R/323T/383V, 286S/316W/323T/383V/414V/416A, 315G, and 408F, wherein the positions are numbered with reference to SEQ ID NO: 4. In some additional embodiments, the amino acid difference(s) comprise the substitution(s) A74T/G81S/N286S/E316W/A323T/A383V, K163F/N286S/I314R/E316W/A323T/C414V/P416A, K163F/N286S/I314R/A323T/D394G, N286S/I314R/A323T/A383V, N286S/E316W/A323T/A383V/C414V/P416A, E315G, and T408F, wherein the positions are numbered with reference to SEQ ID NO: 4.

[0140] In some embodiments, the present disclosure provides an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO: 8 selected from 5, 18/23/149/260/383/395/401/416, 18/23/149/383, 18/163/164, 21, 21/163/315/316, 21/163/323/408, 21/408, 23/56/86/149/163/164/383/401/416, 23/86, 23/149/260, 23/149/284/383/395, 23/163/164/383, 23/163/164/401/416, 24, 42, 42/110, 42/187/272, 42/187/324/363/366, 42/187/353, 42/272/291, 42/272/291/363, 42/272/324/363/366, 42/272/363/410, 42/272/410, 42/291/313/363/410, 42/291/363, 42/291/363/366, 42/353, 42/363, 46, 66, 77, 86/149/163/164/383/395/401, 86/149/395, 86/163/164/260/383, 86/383, 107, 110, 110/187, 110/187/253/410, 134, 138, 149/164/260/383/395/401, 149/260/383, 149/416, 163/259/323/408, 163/259/408, 163/315/316, 164/260/401, 164/316/383/401, 167, 186, 187, 187/253/363/366, 187/272/324/363/410, 187/272/363, 187/272/363/366/410, 187/291, 189, 191, 195, 199, 203, 210, 211, 248, 259/307, 260/395/401, 272, 272/353, 272/363/366, 272/410, 277, 291, 305, 309, 315, 342, 343, 351, 354, 358, 361, 362, 363, 363/366, 365, 367, 383, 383/401, 383/416/422, 385, 388, 389, 392, 395, 396, 401, 404, 405, 408, 410, 416, 417, 439, 443, 447, 450, and 451, wherein the positions are numbered with reference to SEQ ID NO: 8. In some embodiments, the amino acid difference(s) comprise the substitution(s) 5E, 5G, 18A/23R/149S/260T/383V/395D/401S/416P, 18A/23R/149S/383V, 18A/163M/164Q, 21H, 21H/163L/315G/316F, 21H/163L/323C/408F, 21H/408F, 23R/56C/86G/149S/163M/164D/383V/401S/416P, 23R/86G, 23R/149S/260T, 23R/149S/284A/383V/395D, 23R/163M/164Q/383V, 23R/163M/164S/401S/416P, 24K, 24R, 42F, 42F/110K, 42F/187E/272E, 42F/187E/324S/363L/366H, 42F/187E/353T, 42F/272E/291Y, 42F/272E/291Y/363L, 42F/272E/324S/363L/366H, 42F/272E/363L/410H, 42F/272E/410H, 42F/291Y/313V/363L/410H, 42F/291Y/363L, 42F/291Y/363L/366H, 42F/353T, 42F/363L, 46S, 66A,

77M, 86G/149S/163M/164S/383V/395D/401S, 86G/149S/395D, 86G/163M/164S/260T/383V, 86G/383V, 107L, 107S, 107Y, 110K, 110K/187E, 110K/187E/253L/410H, 134V, 138R, 149S/164S/260T/383V/395D/401A, 149S/260T/383V, 149S/416P, 163L/259V/323C/408F, 163L/259V/408F, 163L/315G/316F, 164D/316H/383V/401S, 164S/260T/401S, 167N, 186Q, 187E, 187E/253L/363L/366H, 187E/272E/324S/363L/410H, 187E/272E/363L, 187E/272E/363L/366H/410H, 187E/291Y, 189F, 189S, 189V, 189W, 191D, 191F, 195W, 199Q, 203L, 210A, 210L, 210M, 210V, 210Y, 211R, 248G, 259V/307M, 260T/395D/401S, 272E, 272E/353T, 272E/363L/366H, 272E/410H, 277S, 291Y, 305E, 309A, 309F, 309R, 315G, 342T, 343G, 351L, 354S, 358L, 361R, 362Q, 362V, 363L, 363L/366H, 365L, 365Q, 365R, 365S, 367T, 383V, 383V/401A, 383V/416P/422T, 385L, 385T, 388D, 388L, 388P, 389D, 392A, 392L, 395R, 396P, 396Y, 401Q, 401S, 404M, 405W, 408A, 408E, 408W, 410H, 416P, 417S, 439L, 439S, 443L, 443S, 447S, 447T, 450D, and 451S, wherein the positions are numbered with reference to SEQ ID NO: 8. In some additional embodiments, the amino acid difference(s) comprise the substitution(s) Q5E, Q5G, G18A/P23R/A149S/C260T/A383V/G395D/E401S/A416P, G18A/P23R/A149S/A383V, G18A/F163M/P164Q, D21H, D21H/F163L/E315G/W316F, D21H/F163L/T323C/T408F, D21H/T408F, P23R/L56C/S86G/A149S/F163M/P164D/A383V/E401S/A416P, P23R/S86G, P23R/A149S/C260T, P23R/A149S/S284A/A383V/G395D, P23R/F163M/P164Q/A383V, P23R/F163M/P164S/E401S/A416P, S24K, S24R, V42F, V42F/R110K, V42F/Y187E/V272E, V42F/Y187E/G324S/I363L/R366H, V42F/Y187E/A353T, V42F/V272E/F291Y, V42F/V272E/F291Y/I363L, V42F/V272E/G324S/I363L/R366H, V42F/V272E/I363L/L410H, V42F/V272E/L410H, V42F/F291Y/A313V/I363L/L410H, V42F/F291Y/I363L, V42F/F291Y/I363L/R366H, V42F/A353T, V42F/I363L, G46S, K66A, E77M, S86G/A149S/F163M/P164S/A383V/G395D/E401S, S86G/A149S/G395D, S86G/F163M/P164S/C260T/A383V, S86G/A383V, D107L, D107S, D107Y, R110K, R110K/Y187E, R110K/Y187E/V253L/L410H, A134V, P138R, A149S/P164S/C260T/A383V/G395D/E401A, A149S/C260T/A383V, A149S/A416P, F163L/I259V/T323C/T408F, F163L/I259V/T408F, F163L/E315G/W316F, P164D/W316H/A383V/E401S, P164S/C260T/E401S, S167N, R186Q, Y187E, Y187E/V253L/I363L/R366H, Y187E/V272E/G324S/I363L/L410H, Y187E/V272E/I363L, Y187E/V272E/I363L/R366H/L410H, Y187E/F291Y, E189F, E189S, E189V, E189W, G191D, G191F, E195W, A199Q, R203L, Q210A, Q210L, Q210M, Q210V, Q210Y, K211R, K248G, I259V/L307M, C260T/G395D/E401S, V272E, V272E/A353T, V272E/I363L/R366H, V272E/L410H, T277S, F291Y, K305E, T309A, T309F, T309R, E315G, N342T, E343G, R351L, P354S, E358L, K361R, H362Q, H362V, I363L, I363L/R366H, E365L, E365Q, E365R, E365S, P367T, A383V, A383V/E401A, A383V/A416P/V422T, K385L, K385T, A388D, A388L, A388P, S389D, P392A, P392L, G395R, N396P, N396Y, E401Q, E401S, A404M, N405W, T408A, T408E, T408W, L410H, A416P, L417S, D439L, D439S, K443L, K443S, K447S, K447T, A450D, and E451S, wherein the positions are numbered with reference to SEQ ID NO: 8.

[0141] In some embodiments, the present disclosure provides an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO: 8 selected from 18/23/149/383, 21/163/323/408, 272, 291, and 383, wherein the positions are numbered with reference to SEQ ID NO: 8. In some embodiments, the amino acid difference(s) comprise the substitution(s) 18A/23R/149S/383V, 21H/163L/323C/408F, 272E, 291Y, and 383V, wherein the positions are numbered with reference to SEQ ID NO: 8. In some additional embodiments, the amino acid difference(s) comprise the substitution(s) G18A/P23R/A149S/A383V, D21H/F163L/T323C/T408F, V272E, F291Y, and A383V, wherein the positions are numbered with reference to SEQ ID NO: 8.

[0142] In some embodiments, the present disclosure provides an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO: 366 selected from 24, 24/42/66/291, 24/42/291/362, 24/66/163/191/362/383/388, 24/66/191/199/260/291/351, 24/66/191/199/291, 24/66/191/260/408, 24/66/260/291/383/388/408, 24/66/291/342/383, 24/66/291/365, 24/66/342/365/388/408, 24/77/291, 24/107/163/191/291/351/383/388, 24/107/291/351/365/388, 24/163/351/383, 24/191/291/365, 24/199/260/351/362/383, 24/199/260/362/383/388, 24/260/362/383/388, 24/291, 24/291/342/351/383, 24/291/362/388, 24/291/408, 24/383/388, 24/388, 25, 28, 33, 42/191/408, 42/199/291/383, 42/291/351/362/365/383/388, 42/291/351/362/383/408, 42/291/383/388, 66/82/291/383, 66/163/191/365/383, 66/199/351/383, 66/291, 66/291/362/365/383, 66/291/383/388, 66/383, 77/291, 77/383/388, 86, 107/191/199/365/383/388, 107/191/291/383, 148, 153, 163/291/362/365/383/388, 163/291/383/388, 163/383, 191/199/365/383/388, 191/260/388, 191/291, 191/291/342/362/365, 191/351/383/388, 199/260/383, 199/291, 260, 260/291/365/383/408, 260/365/383, 291, 291/351/383/388, 291/351/383/388/408, 291/362/365, 291/365/388, 291/383, 314, 315, 316, 319, 342/362, 351/383/388, 362, 362/388, 383, 383/388, 396, 397, 405, 406, 413, 419, and 423, wherein the positions are numbered with reference to SEQ ID NO: 366. In some embodiments, the amino acid difference(s) comprise the substitution(s) 24E, 24K, 24K/42F/66A/291Y, 24K/42F/291Y/362Q, 24K/66A/163M/191D/362Q/383V/388D, 24K/66A/191D/199Q/260T/291Y/351L, 24K/66A/191D/199Q/291Y, 24K/66A/191D/260T/408A, 24K/66A/260T/291Y/383V/388D/408A, 24K/66A/291Y/342T/383V, 24K/66A/291Y/365S, 24K/66A/342T/365S/388D/408E, 24K/77M/291Y, 24K/107L/163M/191D/291Y/351L/383V/388D, 24K/107L/291Y/351L/365S/388D, 24K/163M/351L/383V, 24K/191D/291Y/365S, 24K/199Q/260T/351L/362Q/383V, 24K/199Q/260T/362Q/383V/388D, 24K/260T/362Q/383V/388D, 24K/291Y, 24K/291Y/342T/351L/383V, 24K/291Y/362Q/388D, 24K/291Y/408A, 24K/383V/388D, 24K/388D, 25H, 25V, 28S, 28T, 33T, 42F/191D/408E, 42F/199Q/291Y/383V, 42F/291Y/351L/362Q/365S/383V/388D, 42F/291Y/351L/362Q/383V/408A, 42F/291Y/383V/388D, 66A/82H/291Y/383V, 66A/163M/191D/365S/383V, 66A/199Q/351L/383V, 66A/291Y, 66A/291Y/362Q/365S/383V, 66A/291Y/383V/388D, 66A/383V, 77M/291Y, 77M/383V/388D, 86T,

107L/191D/199Q/365S/383V/388D, 107L/191D/291Y/383V, 148G, 153S, 163M/291Y/362Q/365S/383V/388D, 163M/291Y/383V/388D, 163M/383V, 191D/199Q/365S/383V/388D, 191D/260T/388D, 191D/291Y, 191D/291Y/342T/362Q/365S, 191D/351L/383V/388D, 199Q/260T/383V, 199Q/291Y, 260T, 260T/291Y/365S/383V/408A, 260T/365S/383V, 291Y, 291Y/351L/383V/388D, 291Y/351L/383V/388D/408A, 291Y/362Q/365S, 291Y/365S/388D, 291Y/383V, 314K, 315S, 316V, 319S, 342T/362Q, 351L/383V/388D, 362Q, 362Q/388D, 383V, 383V/388D, 396R, 397M, 405A, 406H, 413L, 419S, and 423V, wherein the positions are numbered with reference to SEQ ID NO: 366. In some additional embodiments, the amino acid difference(s) comprise the substitution(s) S24E, S24K, S24K/V42F/K66A/F291Y, S24K/V42F/F291Y/H362Q, S24K/K66A/L163M/G191D/H362Q/A383V/A388D, S24K/K66A/G191D/A199Q/C260T/F291Y/R351L, S24K/K66A/G191D/A199Q/F291Y, S24K/K66A/G191D/C260T/F408A, S24K/K66A/C260T/F291Y/A383V/A388D/F408A, S24K/K66A/F291Y/N342T/A383V, S24K/K66A/F291Y/E365S, S24K/K66A/N342T/E365S/A388D/F408E, S24K/E77M/F291Y, S24K/D107L/L163M/G191D/F291Y/R351L/A383V/A388D, S24K/D107L/F291Y/R351L/E365S/A388D, S24K/L163M/R351L/A383V, S24K/G191D/F291Y/E365S, S24K/A199Q/C260T/R351L/H362Q/A383V, S24K/A199Q/C260T/H362Q/A383V/A388D, S24K/C260T/H362Q/A383V/A388D, S24K/F291Y, S24K/F291Y/N342T/R351L/A383V, S24K/F291Y/H362Q/A388D, S24K/F291Y/F408A, S24K/A383V/A388D, S24K/A388D, L25H, L25V, R28S, R28T, V33T, V42F/G191D/F408E, V42F/A199Q/F291Y/A383V, V42F/F291Y/R351L/H362Q/E365S/A383V/A388D, V42F/F291Y/R351L/H362Q/A383V/F408A, V42F/F291Y/A383V/A388D, K66A/Y82H/F291Y/A383V, K66A/L163M/G191D/E365S/A383V, K66A/A199Q/R351L/A383V, K66A/F291Y, K66A/F291Y/H362Q/E365S/A383V, K66A/F291Y/A383V/A388D, K66A/A383V, E77M/F291Y, E77M/A383V/A388D, S86T, D107L/G191D/A199Q/E365S/A383V/A388D, D107L/G191D/F291Y/A383V, N148G, A153S, L163M/F291Y/H362Q/E365S/A383V/A388D, L163M/F291Y/A383V/A388D, L163M/A383V, G191D/A199Q/E365S/A383V/A388D, G191D/C260T/A388D, G191D/F291Y, G191D/F291Y/N342T/H362Q/E365S, G191D/R351L/A383V/A388D, A199Q/C260T/A383V, A199Q/F291Y, C260T, C260T/F291Y/E365S/A383V/F408A, C260T/E365S/A383V, F291Y, F291Y/R351L/A383V/A388D, F291Y/R351L/A383V/A388D/F408A, F291Y/H362Q/E365S, F291Y/E365S/A388D, F291Y/A383V, R314K, E315S, W316V, H319S, N342T/H362Q, R351L/A383V/A388D, H362Q, H362Q/A388D, A383V, A383V/A388D, N396R, L397M, N405A, T406H, I413L, Q419S, and L423V, wherein the positions are numbered with reference to SEQ ID NO: 366.

[0143] In some embodiments, the present disclosure provides an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO: 366 selected from 24/66/191/199/291, 24/66/291/365, 163/291/362/365/383/388,

163/291/383/388, 191/291/342/362/365, 291, and 291/383, wherein the positions are numbered with reference to SEQ ID NO: 366. In some embodiments, the amino acid difference(s) comprise the substitution(s) 24K/66A/191D/199Q/291Y, 24K/66A/291Y/365S, 163M/291Y/362Q/365S/383V/388D, 163M/291Y/383V/388D, 191D/291Y/342T/362Q/365S, 291Y, and 291Y/383V, wherein the positions are numbered with reference to SEQ ID NO: 366. In some additional embodiments, the amino acid difference(s) comprise the substitution(s) S24K/K66A/G191D/A199Q/F291Y, S24K/K66A/F291Y/E365S, L163M/F291Y/H362Q/E365S/A383V/A388D, L163M/F291Y/A383V/A388D, G191D/F291Y/N342T/H362Q/E365S, F291Y, and F291Y/A383V, wherein the positions are numbered with reference to SEQ ID NO: 366.

[0144] In some embodiments, the present disclosure provides an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO: 650 selected from 10, 13, 13/24/108/163, 13/24/108/163/311, 13/24/133/199/311, 13/24/163, 13/24/199/311, 13/108, 13/108/199, 13/108/311, 13/199, 13/311, 14, 14/24/108, 14/24/108/133, 14/24/108/199, 14/24/199, 14/108, 14/108/133/311, 14/108/311, 14/311, 24, 24/163, 24/163/199, 35, 72, 73, 78, 95, 101, 108, 108/199, 114, 154, 163, 169, 175/316, 199, 199/311, 226, 293, 311, 316, 382, 383, and 386, wherein the positions are numbered with reference to SEQ ID NO: 650. In some embodiments, the amino acid difference(s) comprise the substitution(s) 10E, 13A, 13A/24E/108R/163L/311S, 13A/24E/133R/199Q/311S, 13A/24E/163L, 13A/24K/108R/163L, 13A/24K/199Q/311S, 13A/108R, 13A/108R/199Q, 13A/108R/311S, 13A/199Q, 13A/311S, 14A, 14G, 14H, 14H/24E/108R/133R, 14H/24K/108R, 14H/24K/108R/199Q, 14H/24K/199Q, 14H/108R, 14H/108R/133R/311S, 14H/108R/311S, 14H/311S, 24E, 24E/163L, 24E/163L/199Q, 35E, 72G, 73R, 73S, 78A, 95I, 101L, 108R, 108R/199Q, 114A, 154S, 163H, 163S, 163V, 169C, 169V, 175D/316F, 199Q, 199Q/311S, 226Q, 293A, 311K, 316D, 316E, 316F, 316G, 316H, 316I, 316L, 316N, 316S, 316V, 316Y, 382D, 383L, and 386A, wherein the positions are numbered with reference to SEQ ID NO: 650. In some additional embodiments, the amino acid difference(s) comprise the substitution(s) R10E, T13A, T13A/S24E/S108R/M163L/I311S, T13A/S24E/A133R/A199Q/I311S, T13A/S24E/M163L, T13A/S24K/S108R/M163L, T13A/S24K/A199Q/I311S, T13A/S108R, T13A/S108R/A199Q, T13A/S108R/I311S, T13A/A199Q, T13A/I311S, Y14A, Y14G, Y14H, Y14H/S24E/S108R/A133R, Y14H/S24K/S108R, Y14H/S24K/S108R/A199Q, Y14H/S24K/A199Q, Y14H/S108R, Y14H/S108R/A133R/I311S, Y14H/S108R/I311S, Y14H/I311S, S24E, S24E/M163L, S24E/M163L/A199Q, H35E, A72G, K73R, K73S, R78A, M95I, V101L, S108R, S108R/A199Q, T114A, T154S, M163H, M163S, M163V, F169C, F169V, G175D/W316F, A199Q, A199Q/I311S, M226Q, P293A, I311K, W316D, W316E, W316F, W316G, W316H, W316I, W316L, W316N, W316S, W316V, W316Y, E382D, V383L, and D386A, wherein the positions are numbered with reference to SEQ ID NO: 650.

[0145] In some embodiments, the present disclosure provides an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to

SEQ ID NO: 650 selected from 14/108/133/311, 24/163/199, 72, 78, 316, and 383, wherein the positions are numbered with reference to SEQ ID NO: 650. In some embodiments, the amino acid difference(s) comprise the substitution(s) 14H/108R/133R/311S, 24E/163L/199Q, 72G, 78A, 316L, 316N, 316S, and 383L, wherein the positions are numbered with reference to SEQ ID NO: 650. In some additional embodiments, the amino acid difference(s) comprise the substitution(s) Y14H/S108R/A133R/I311S, S24E/M163L/A199Q, A72G, R78A, W316L, W316N, W316S, and V383L, wherein the positions are numbered with reference to SEQ ID NO: 650.

[0146] In some embodiments, the engineered polypeptides having transaminase activity are capable of converting compound **(1)** to compound **(2)** with increased tolerance for the presence of the substrate relative to the substrate tolerance of a reference polypeptide (e.g., SEQ ID NO: 2, 4, 8, 366, and/or 650), under suitable reaction conditions. Accordingly, in some embodiments the engineered polypeptides are capable of converting the substrate of compound **(1)** to compound **(2)** in the presence of a substrate loading concentration of at least about 1 g/L, 5 g/L, 10 g/L, 20 g/L, about 30 g/L, about 40 g/L, about 50 g/L, about 70 g/L, about 75 g/L, about 100 g/L, with a percent conversion of at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, in a reaction time of about 72h, about 48h, about 36h, about 24 h, or even shorter length of time, under suitable reaction conditions.

[0147] Some suitable reaction conditions under which the above-described improved properties of the engineered polypeptides can be determined with respect concentrations or amounts of polypeptide, substrate, amine donor, cofactor, buffer, co-solvent, pH, and/or conditions including temperature and reaction time are provided herein. In some embodiments, the suitable reaction conditions comprise the HTP, SFP, or DSP assay conditions described below and in the Examples.

[0148] As will be apparent to the skilled artisan, the foregoing residue positions and the specific amino acid residues for each residue position can be used individually or in various combinations to synthesize transaminase polypeptides having desired improved properties, including, among others, enzyme activity, substrate/product preference, stereoselectivity, substrate/product tolerance, and stability under various conditions, such as increased temperature, solvent, and/or pH.

[0149] 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 **(1)** to compound **(2)** 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: 6 - 936.

[0150] In some embodiments, the engineered transaminase polypeptide of the disclosure comprises an amino acid sequence comprising a deletion as compared to any one of the engineered transaminase polypeptide sequences described herein, such as the exemplary engineered polypeptide sequences having the even-numbered sequence identifiers of SEQ ID NO: 6-936. 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.

[0151] 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:6 - 936. 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.

[0152] In some embodiments, the polypeptides of the present disclosure are in the form of fusion polypeptides in which the engineered polypeptides are fused to other polypeptides, such as, by way of example and not limitation, antibody tags (e.g., myc epitope), purification sequences (e.g., His tags for binding to metals), and cell localization signals (e.g., secretion signals). Thus, the engineered polypeptides described herein can be used with or without fusions to other polypeptides.

[0153] 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 (MeIle); phenylglycine (Phg); cyclohexylalanine (Cha); norleucine (Nle); naphthylalanine (Nal); 2-chlorophenylalanine (Ocf); 3-chlorophenylalanine (Mcf); 4-chlorophenylalanine (Pcf); 2-fluorophenylalanine (Off); 3-fluorophenylalanine (Mff); 4-fluorophenylalanine (Pff); 2-bromophenylalanine (Obf); 3-bromophenylalanine (Mbf); 4-bromophenylalanine (Pbf); 2-methylphenylalanine (Omf); 3-methylphenylalanine (Mmf); 4-methylphenylalanine (Pmf); 2-nitrophenylalanine (Onf); 3-nitrophenylalanine (Mnf); 4-nitrophenylalanine (Pnf); 2-cyanophenylalanine (Ocf); 3-cyanophenylalanine (Mcf); 4-cyanophenylalanine (Pcf); 2-trifluoromethylphenylalanine (Otf); 3-trifluoromethylphenylalanine (Mtf); 4-trifluoromethylphenylalanine (Ptf); 4-aminophenylalanine (Paf); 4-iodophenylalanine (Pif); 4-aminomethylphenylalanine (Pamf); 2,4-dichlorophenylalanine (Opef); 3,4-dichlorophenylalanine (Mpcf); 2,4-difluorophenylalanine (Opff); 3,4-difluorophenylalanine (Mpff); pyrid-2-ylalanine (2pAla); pyrid-3-ylalanine (3pAla); pyrid-4-ylalanine (4pAla); naphth-1-ylalanine (1nAla); naphth-2-ylalanine (2nAla); thiazolylalanine (taAla); benzothienylalanine (bAla); thienylalanine (tAla); furylalanine (fAla); homophenylalanine (hPhe); homotyrosine (hTyr); homotryptophan (hTrp); pentafluorophenylalanine (5ff); styrylkalanine (sAla); authrylalanine (aAla); 3,3-diphenylalanine (Dfa); 3-amino-5-phenypentanoic acid (Afp); penicillamine (Pen); 1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid (Tic); β-2-thienylalanine (Thi); methionine sulfoxide (Mso); N(w)-nitroarginine (nArg); homolysine (hLys); phosphonomethylphenylalanine (pmPhe); phosphoserine (pSer); phosphothreonine (pThr); homoaspartic acid (hAsp); homoglutamic acid (hGlu); 1-aminocyclopent-(2 or 3)-ene-4 carboxylic acid; pipecolic acid (PA), azetidine-3-carboxylic acid (ACA); 1-aminocyclopentane-3-carboxylic acid; allylglycine (aOly); propargylglycine (pgGly); homoalanine (hAla); norvaline (nVal); homoleucine (hLeu), homovaline (hVal); homoisoleucine (hIle); homoarginine (hArg); N-acetyl lysine (AcLys); 2,4-diaminobutyric acid (Dbu); 2,3-diaminobutyric acid (Dab); N-methylvaline (MeVal); homocysteine (hCys); homoserine (hSer); hydroxyproline (Hyp) and homoproline (hPro). Additional non-encoded amino acids of which the polypeptides described herein may be comprised will be apparent to those of skill in the art. These amino acids may be in either the L- or D-configuration.

[0154] 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 (nitropyridinesulfenyl), 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).

[0155] 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.

[0156] 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.

[0157] In some embodiments, the engineered polypeptides having transaminase activity are bound or immobilized on the solid support such that they retain their improved activity, enantioselectivity, stereoselectivity, and/or other improved properties relative to a reference polypeptide (e.g., SEQ ID NO: 2, 4, 8, 366, and/or 650). In such embodiments, the immobilized polypeptides can facilitate the biocatalytic conversion of the substrate compound to the desired product, and after the reaction is complete are easily retained (e.g., by retaining beads on which polypeptide is immobilized) and then reused or recycled in subsequent reactions. Such immobilized enzyme processes allow for further efficiency and cost reduction. Accordingly, it is further contemplated that any of the methods of using the engineered transaminase polypeptides of the present disclosure can be carried out using the same transaminase polypeptides bound or immobilized on a solid support.

[0158] The engineered transaminase polypeptide can be bound non-covalently or covalently. Various methods for conjugation and immobilization of enzymes to solid supports (e.g., resins, membranes, beads, glass, etc.) are well known in the art. In particular, PCT publication WO2012/177527 A1 discloses immobilized engineered transaminase polypeptides capable of converting compound **(2)** to compound **(1)**, and methods of preparing the immobilized polypeptides, in which the polypeptide is physically attached to a resin by either hydrophobic interactions or covalent bonds, and is stable in a solvent system that comprises at least up to 100% organic solvent**.** Other methods for conjugation and immobilization of enzymes to solid supports (e.g., resins, membranes, beads, glass, etc.) are well known in the art (See e.g., Yi et al., Proc. Biochem., 42: 895-898 [2007]; Martin et al., Appl. Microbiol. Biotechnol., 76: 843-851 [2007]; Koszelewski et al., J. Mol. Cat. B: Enz., 63: 39-44 [2010]; Truppo et al., Org. Proc. Res. Develop., published online: dx.doi.org/10.1021/op200157c; and Mateo et al., Biotechnol. Prog., 18:629-34 [2002], etc.).

[0159] 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.

[0160] 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.

[0161] 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

[0162] 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.

[0163] In some embodiments, the isolated polynucleotide encoding an improved transaminase polypeptide is manipulated in a variety of ways to provide for improved activity and/or expression of the polypeptide. Manipulation of the isolated polynucleotide prior to its insertion into a vector may be desirable or necessary depending on the expression vector. The techniques for modifying polynucleotides and nucleic acid sequences utilizing recombinant DNA methods are well known in the art.

[0164] Those of ordinary skill in the art understand that due to the degeneracy of the genetic code, a multitude of nucleotide sequences encoding variant transaminase acylase polypeptides of the present disclosure exist. For example, the codons AGA, AGG, CGA, CGC, CGG, and CGU all encode the amino acid arginine. Thus, at every position in the nucleic acids of the disclosure where an arginine is specified

by a codon, the codon can be altered to any of the corresponding codons described above without altering the encoded polypeptide. It is understood that "U" in an RNA sequence corresponds to "T" in a DNA sequence. The disclosure contemplates and provides each and every possible variation of nucleic acid sequence encoding a polypeptide of the disclosure that could be made by selecting combinations based on possible codon choices.

[0165] As indicated above, DNA sequence encoding a transaminase may also be designed for high codon usage bias codons (codons that are used at higher frequency in the protein coding regions than other codons that code for the same amino acid). The preferred codons may be determined in relation to codon usage in a single gene, a set of genes of common function or origin, highly expressed genes, the codon frequency in the aggregate protein coding regions of the whole organism, codon frequency in the aggregate protein coding regions of related organisms, or combinations thereof. A codon whose frequency increases with the level of gene expression is typically an optimal codon for expression. In particular, a DNA sequence can be optimized for expression in a particular host organism. A variety of methods are well-known in the art for determining the codon frequency (e.g., codon usage, relative synonymous codon usage) and codon preference in specific organisms, including multivariate analysis (e.g., using cluster analysis or correspondence analysis,) and the effective number of codons used in a gene. The data source for obtaining codon usage may rely on any available nucleotide sequence capable of coding for a protein. These data sets include nucleic acid sequences actually known to encode expressed proteins (e.g., complete protein coding sequences-CDS), expressed sequence tags (ESTs), or predicted coding regions of genomic sequences, as is well-known in the art. Polynucleotides encoding variant transaminases can be prepared using any suitable methods known in the art. Typically, oligonucleotides are individually synthesized, then joined (e.g., by enzymatic or chemical ligation methods, or polymerase-mediated methods) to form essentially any desired continuous sequence. In some embodiments, polynucleotides of the present disclosure are prepared by chemical synthesis using, any suitable methods known in the art, including but not limited to automated synthetic methods. For example, in the phosphoramidite method, oligonucleotides are synthesized (e.g., in an automatic DNA synthesizer), purified, annealed, ligated and cloned in appropriate vectors. In some embodiments, double stranded DNA fragments are then obtained either by synthesizing the complementary strand and annealing the strands together under appropriate conditions, or by adding the complementary strand using DNA polymerase with an appropriate primer sequence. There are numerous general and standard texts that provide methods useful in the present disclosure are well known to those skilled in the art.

[0166] For example, mutagenesis and directed evolution methods can be readily applied to polynucleotides to generate variant libraries that can be expressed, screened, and assayed. Mutagenesis and directed evolution methods are well known in the art (See e.g., 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 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).

[0167] In some embodiments, the polynucleotide encodes a transaminase polypeptide comprising an amino acid sequence that is at least about 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more identical to a reference sequence selected from the even-numbered sequence identifiers of SEQ ID NO: 2, 4, 8, 366, and/or 650, where the polypeptide has transaminase activity and one or more of the improved properties as described herein, for example the ability to convert compound **(1)** to product compound **(2)** with increased activity compared to a reference sequence (e.g., the polypeptide of SEQ ID NO: 2, 4, 8, 366, and/or 650). In some embodiments, the reference sequence is selected from SEQ ID NO: 2, 4, 8, 366, and/or 650. In some embodiments, the reference sequence is SEQ ID NO: 2. In some embodiments, the reference sequence is SEQ ID NO: 4. In some embodiments, the reference sequence is SEQ ID NO: 8. In some embodiments, the reference sequence is SEQ ID NO: 366. In some embodiments, the reference sequence is SEQ ID NO: 650.

[0168] In some embodiments, the polynucleotide encodes an engineered transaminase polypeptide comprising an amino acid sequence that has the percent identity described above and (a) has one or more amino acid residue differences as compared to SEQ ID NO: 2, 4, 8, 366, and/or 650. In some embodiments, the present disclosure provides an engineered polypeptide having transaminase activity comprising an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%,

93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to reference sequence of SEQ ID NO: 2, 4, 8, 366, and/or 650 and (a) at least one amino acid residue difference selected from those substitutions provided herein (See e.g., Tables Tables 2-1, 2-2, 3-1, 3-2, 4-1, 4-2, 5-1, and/or 5-2).

[0169] In some embodiments, the polynucleotide encoding the engineered transaminase polypeptide comprises a sequence selected from the odd-numbered sequence identifiers of SEQ ID NO: 5-935. In some embodiments, the polynucleotide sequences are selected from SEQ ID NO: 1, 3, 7, 365 and/or 935. In some embodiments, the present disclosure provides engineered polynucleotides encoding polypeptides having transaminase activity, wherein the engineered polypeptides have at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to at least one reference sequence selected from SEQ ID NO: 1, 3, 7, 365 and/or 935.

[0170] In some embodiments, the present disclosure provides a polynucleotide that hybridizes under defined conditions, such as moderately stringent or highly stringent conditions, to a polynucleotide sequence (or complement thereof) encoding an engineered transaminase of the present disclosure. In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a polynucleotide selected from the sequences having the odd-numbered sequence identifiers of SEQ ID NO: 5-935, or a complement thereof, and encodes a polypeptide having transaminase activity with one or more of the improved properties described herein.

[0171] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO:4 selected from 18, 20, 21, 21/23/56/146, 21/23/56/146/432, 21/23/146/417, 21/23/395/417/432, 21/53/56, 21/53/417, 21/56/395, 21/417/432, 23, 23/53, 23/53/56, 23/53/56/146/395, 23/53/395, 23/53/417, 23/53/432, 23/56, 23/56/395, 23/56/395/417, 23/395/417, 23/417, 23/417/432, 53, 53/56, 53/146/417, 53/395, 56, 56/74/241/286/314/316/323, 56/86/163/314/316/383/414/416/422, 56/86/286/314/414/416, 56/86/314/316/323/394/414/422, 56/146/417, 56/146/432, 56/147, 56/163, 56/163/286/316/323/383/394, 56/286/314/316/323/422, 56/286/383, 56/323, 56/323/383, 56/323/383/394, 56/383, 56/395, 74/81/286/316/323/383, 74/85/86/163/286/316/323/394, 74/85/314/316/414/416, 74/86/163/316, 74/86/316/323/383/394, 74/88/286/316/323/383, 74/88/323/383, 74/163/286/316/383/394/416, 74/163/314/316, 74/163/314/316/323/394, 74/163/314/323/383/414/416, 74/286, 74/286/316/323, 74/286/394/416, 74/314/323/383/394/414, 74/316/323/394, 85/86/88/163/323/383/394, 85/86/163/314/323/394/414, 85/286, 85/286/323, 86, 86/88/163/323/383/414/422, 86/383/394, 88, 88/163/286/383, 88/286/316/323, 88/286/316/323/383/414/416, 88/316/323, 146, 146/147/395/417, 146/395, 146/395/417, 146/417, 147/395/417/432, 147/417, 149, 157, 163, 163/222/286/316/323/383/394, 163/286, 163/286/314/316/323/414/416, 163/286/314/323/394, 163/286/316/323/394/416, 163/286/414, 163/314/316/394, 163/314/323/394, 163/314/383, 163/314/414, 163/316/323, 163/323, 163/383, 164, 199/417, 259, 260, 284, 286, 286/314/323/383, 286/314/394, 286/316/323/383/414/416, 286/316/383/394, 286/316/394/414/416, 286/323, 286/323/383/414,

286/323/416, 286/383, 286/416, 314/316, 314/316/323, 314/316/323/383/422, 314/316/323/394, 314/316/394, 314/323/383/394, 314/383, 314/383/414/422, 315, 316, 316/323/383/394, 316/323/394/414/416, 316/414/422, 323, 323/383, 323/383/394/414/416, 323/394, 383, 395, 395/417, 395/417/432, 400, 401, 403, 404, 405, 406, 408, 415, 417, 417/432, 420, and 422, wherein the positions are numbered with reference to SEQ ID NO:4. In some embodiments, the amino acid differences comprise the substitution(s) 18A, 20C, 21H, 21P/23S/56C/146H, 21P/23S/56C/146H/432V, 21P/23S/146H/417V, 21P/23S/395D/417S/432V, 21P/53C/56C, 21P/53C/417S, 21P/56C/395D, 21P/417S/432V, 21R, 23A, 23R, 23S/53C, 23S/53C/56C, 23S/53C/56C/146H/395D, 23S/53C/395D, 23S/53C/417S, 23S/53C/432V, 23S/56C, 23S/56C/395D, 23S/56C/395D/417V, 23S/395D/417S, 23S/417S, 23S/417V, 23S/417V/432V, 53C, 53C/56C, 53C/146H/417S, 53C/395D, 56A, 56A/74T/241V/286S/314R/316W/323T, 56A/86A/163F/314R/316W/383V/414V/416A/422A, 56A/86A/286S/314R/414V/416A, 56A/86A/314R/316W/323T/394G/414V/422A, 56A/163F, 56A/163F/286S/316W/323T/383V/394G, 56A/286S/314R/316W/323T/422A, 56A/286S/383V, 56A/323T, 56A/323T/383V, 56A/323T/383V/394G, 56A/383V, 56C, 56C/146H/417V, 56C/146H/432V, 56C/147R, 56C/395D, 56T, 56V, 74T/81S/286S/316W/323T/383V, 74T/85V/86A/163F/286S/316W/323T/394G, 74T/85V/314R/316W/414V/416A, 74T/86A/163F/316W, 74T/86A/316W/323T/383V/394G, 74T/88R/286S/316W/323T/383V, 74T/88R/323T/383V, 74T/163F/286S/316W/383V/394G/416A, 74T/163F/314R/316W, 74T/163F/314R/316W/323T/394G, 74T/163F/314R/323T/383V/414V/416A, 74T/286S, 74T/286S/316W/323T, 74T/286S/394G/416A, 74T/314R/323T/383V/394G/414V, 74T/316W/323T/394G, 85V/86A/88R/163F/323T/383V/394G, 85V/86A/163F/314R/323T/394G/414V, 85V/286S, 85V/286S/323T, 86A/88R/163F/323T/383V/414V/422A, 86A/383V/394G, 86G, 88R/163F/286S/383V, 88R/286S/316W/323T, 88R/286S/316W/323T/383V/414V/416A, 88R/316W/323T, 88S, 88T, 146H, 146H/147R/395D/417S, 146H/395D, 146H/395D/417S, 146H/417S, 146H/417V, 147R/395D/417S/432V, 147R/417S, 149S, 157A, 163F, 163F/222V/286S/316W/323T/383V/394G, 163F/286S, 163F/286S/314R/316W/323T/414V/416A, 163F/286S/314R/323T/394G, 163F/286S/316W/323T/394G/416A, 163F/286S/414V, 163F/314R/316W/394G, 163F/314R/323T/394G, 163F/314R/383V, 163F/314R/414V, 163F/316W/323T, 163F/323T, 163F/383V, 163L, 163M, 164A, 164D, 164Q, 164S, 199V/417S, 259V, 260T, 284A, 286S, 286S/314R/323T/383V, 286S/314R/394G, 286S/316W/323T/383V/414V/416A, 286S/316W/383V/394G, 286S/316W/394G/414V/416A, 286S/323T, 286S/323T/383V/414V, 286S/323T/416A, 286S/383V, 286S/416A, 314R/316W, 314R/316W/323T, 314R/316W/323T/383V/422A, 314R/316W/323T/394G, 314R/316W/394G, 314R/323T/383V/394G, 314R/383V, 314R/383V/414V/422A, 315G, 315R, 316A, 316F, 316G, 316H, 316L, 316N, 316R, 316V, 316W/323T/383V/394G, 316W/323T/394G/414V/416A, 316W/414V/422A, 323C, 323S, 323T, 323T/383V, 323T/383V/394G/414V/416A, 323T/394G, 383V, 395D, 395D/417S, 395D/417S/432V, 395D/417V, 400D, 401A, 401K, 401S, 403V, 404S, 405H, 405W, 406S, 408F, 408L, 408W, 415G, 415W, 417A, 417S, 417S/432V, 417V, 417V/432V, 420G, 422L, and 422T, wherein the

positions are numbered with reference to SEQ ID NO:4. In some additional embodiments, the amino acid differences comprise the substitution(s) G18A, T20C, D21H, D21P/P23S/L56C/R146H, D21P/P23S/L56C/R146H/A432V, D21P/P23S/R146H/L417V, D21P/P23S/G395D/L417S/A432V, D21P/N53C/L56C, D21P/N53C/L417S, D21P/L56C/G395D, D21P/L417S/A432V, D21R, P23A, P23R, P23S/N53C, P23S/N53C/L56C, P23S/N53C/L56C/R146H/G395D, P23S/N53C/G395D, P23S/N53C/L417S, P23S/N53C/A432V, P23S/L56C, P23S/L56C/G395D, P23S/L56C/G395D/L417V, P23S/G395D/L417S, P23S/L417S, P23S/L417V, P23S/L417V/A432V, N53C, N53C/L56C, N53C/R146H/L417S, N53C/G395D, L56A, L56A/A74T/A241V/N286S/I314R/E316W/A323T, L56A/S86A/K163F/I314R/E316W/A383V/C414V/P416A/V422A, L56A/S86A/N286S/I314R/C414V/P416A, L56A/S86A/I314R/E316W/A323T/D394G/C414V/V422A, L56A/K163F, L56A/K163F/N286S/E316W/A323T/A383V/D394G, L56A/N286S/I314R/E316W/A323T/V422A, L56A/N286S/A383V, L56A/A323T, L56A/A323T/A383V, L56A/A323T/A383V/D394G, L56A/A383V, L56C, L56C/R146H/L417V, L56C/R146H/A432V, L56C/W147R, L56C/G395D, L56T, L56V, A74T/G81S/N286S/E316W/A323T/A383V, A74T/F85V/S86A/K163F/N286S/E316W/A323T/D394G, A74T/F85V/I314R/E316W/C414V/P416A, A74T/S86A/K163F/E316W, A74T/S86A/E316W/A323T/A383V/D394G, A74T/H88R/N286S/E316W/A323T/A383V, A74T/H88R/A323T/A383V, A74T/K163F/N286S/E316W/A383V/D394G/P416A, A74T/K163F/I314R/E316W, A74T/K163F/I314R/E316W/A323T/D394G, A74T/K163F/I314R/A323T/A383V/C414V/P416A, A74T/N286S, A74T/N286S/E316W/A323T, A74T/N286S/D394G/P416A, A74T/I314R/A323T/A383V/D394G/C414V, A74T/E316W/A323T/D394G, F85V/S86A/H88R/K163F/A323T/A383V/D394G, F85V/S86A/K163F/I314R/A323T/D394G/C414V, F85V/N286S, F85V/N286S/A323T, S86A/H88R/K163F/A323T/A383V/C414V/V422A, S86A/A383V/D394G, S86G, H88R/K163F/N286S/A383V, H88R/N286S/E316W/A323T, H88R/N286S/E316W/A323T/A383V/C414V/P416A, H88R/E316W/A323T, H88S, H88T, R146H, R146H/W147R/G395D/L417S, R146H/G395D, R146H/G395D/L417S, R146H/L417S, R146H/L417V, W147R/G395D/L417S/A432V, W147R/L417S, A149S, S157A, K163F, K163F/A222V/N286S/E316W/A323T/A383V/D394G, K163F/N286S, K163F/N286S/I314R/E316W/A323T/C414V/P416A, K163F/N286S/I314R/A323T/D394G, K163F/N286S/E316W/A323T/D394G/P416A, K163F/N286S/C414V, K163F/I314R/E316W/D394G, K163F/I314R/A323T/D394G, K163F/I314R/A383V, K163F/I314R/C414V, K163F/E316W/A323T, K163F/A323T, K163F/A383V, K163L, K163M, P164A, P164D, P164Q, P164S, A199V/L417S, I259V, C260T, S284A, N286S, N286S/I314R/A323T/A383V, N286S/I314R/D394G, N286S/E316W/A323T/A383V/C414V/P416A, N286S/E316W/A383V/D394G, N286S/E316W/D394G/C414V/P416A, N286S/A323T, N286S/A323T/A383V/C414V, N286S/A323T/P416A, N286S/A383V, N286S/P416A, I314R/E316W, I314R/E316W/A323T, I314R/E316W/A323T/A383V/V422A, I314R/E316W/A323T/D394G, I314R/E316W/D394G,

I314R/A323T/A383V/D394G, I314R/A383V, I314R/A383V/C414V/V422A, E315G, E315R, E316A, E316F, E316G, E316H, E316L, E316N, E316R, E316V, E316W/A323T/A383V/D394G, E316W/A323T/D394G/C414V/P416A, E316W/C414V/V422A, A323C, A323S, A323T, A323T/A383V, A323T/A383V/D394G/C414V/P416A, A323T/D394G, A383V, G395D, G395D/L417S, G395D/L417S/A432V, G395D/L417V, S400D, E401A, E401K, E401S, I403V, A404S, N405H, N405W, T406S, T408F, T408L, T408W, R415G, R415W, L417A, L417S, L417S/A432V, L417V, L417V/A432V, S420G, V422L, and V422T, wherein the positions are numbered with reference to SEQ ID NO: 4.

[0172] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO: 4 selected from 74/81/286/316/323/383, 163/286/314/316/323/414/416, 163/286/314/323/394, 286/314/323/383, 286/316/323/383/414/416, 315, and 408, wherein the positions are numbered with reference to SEQ ID NO: 4. In some embodiments, the amino acid differences comprise the substitution(s) 74T/81S/286S/316W/323T/383V, 163F/286S/314R/316W/323T/414V/416A, 163F/286S/314R/323T/394G, 286S/314R/323T/383V, 286S/316W/323T/383V/414V/416A, 315G, and 408F, wherein the positions are numbered with reference to SEQ ID NO: 4. In some additional embodiments, the amino acid differences comprise the substitution(s) A74T/G81S/N286S/E316W/A323T/A383V, K163F/N286S/I314R/E316W/A323T/C414V/P416A, K163F/N286S/I314R/A323T/D394G, N286S/I314R/A323T/A383V, N286S/E316W/A323T/A383V/C414V/P416A, E315G, and T408F, wherein the positions are numbered with reference to SEQ ID NO: 4.

[0173] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO: 8 selected from 5, 18/23/149/260/383/395/401/416, 18/23/149/383, 18/163/164, 21, 21/163/315/316, 21/163/323/408, 21/408, 23/56/86/149/163/164/383/401/416, 23/86, 23/149/260, 23/149/284/383/395, 23/163/164/383, 23/163/164/401/416, 24, 42, 42/110, 42/187/272, 42/187/324/363/366, 42/187/353, 42/272/291, 42/272/291/363, 42/272/324/363/366, 42/272/363/410, 42/272/410, 42/291/313/363/410, 42/291/363, 42/291/363/366, 42/353, 42/363, 46, 66, 77, 86/149/163/164/383/395/401, 86/149/395, 86/163/164/260/383, 86/383, 107, 110, 110/187, 110/187/253/410, 134, 138, 149/164/260/383/395/401, 149/260/383, 149/416, 163/259/323/408, 163/259/408, 163/315/316, 164/260/401, 164/316/383/401, 167, 186, 187, 187/253/363/366, 187/272/324/363/410, 187/272/363, 187/272/363/366/410, 187/291, 189, 191, 195, 199, 203, 210, 211, 248, 259/307, 260/395/401, 272, 272/353, 272/363/366, 272/410, 277, 291, 305, 309, 315, 342, 343, 351, 354, 358, 361, 362, 363, 363/366, 365, 367, 383, 383/401, 383/416/422, 385, 388, 389, 392, 395, 396, 401, 404, 405, 408, 410, 416, 417, 439, 443, 447, 450, and 451, wherein the positions are numbered with reference to SEQ ID NO: 8. In some embodiments, the

amino acid differences comprise the substitution(s) 5E, 5G, 18A/23R/149S/260T/383V/395D/401S/416P, 18A/23R/149S/383V, 18A/163M/164Q, 21H, 21H/163L/315G/316F, 21H/163L/323C/408F, 21H/408F, 23R/56C/86G/149S/163M/164D/383V/401S/416P, 23R/86G, 23R/149S/260T, 23R/149S/284A/383V/395D, 23R/163M/164Q/383V, 23R/163M/164S/401S/416P, 24K, 24R, 42F, 42F/110K, 42F/187E/272E, 42F/187E/324S/363L/366H, 42F/187E/353T, 42F/272E/291Y, 42F/272E/291Y/363L, 42F/272E/324S/363L/366H, 42F/272E/363L/410H, 42F/272E/410H, 42F/291Y/313V/363L/410H, 42F/291Y/363L, 42F/291Y/363L/366H, 42F/353T, 42F/363L, 46S, 66A, 77M, 86G/149S/163M/164S/383V/395D/401S, 86G/149S/395D, 86G/163M/164S/260T/383V, 86G/383V, 107L, 107S, 107Y, 110K, 110K/187E, 110K/187E/253L/410H, 134V, 138R, 149S/164S/260T/383V/395D/401A, 149S/260T/383V, 149S/416P, 163L/259V/323C/408F, 163L/259V/408F, 163L/315G/316F, 164D/316H/383V/401S, 164S/260T/401S, 167N, 186Q, 187E, 187E/253L/363L/366H, 187E/272E/324S/363L/410H, 187E/272E/363L, 187E/272E/363L/366H/410H, 187E/291Y, 189F, 189S, 189V, 189W, 191D, 191F, 195W, 199Q, 203L, 210A, 210L, 210M, 210V, 210Y, 211R, 248G, 259V/307M, 260T/395D/401S, 272E, 272E/353T, 272E/363L/366H, 272E/410H, 277S, 291Y, 305E, 309A, 309F, 309R, 315G, 342T, 343G, 351L, 354S, 358L, 361R, 362Q, 362V, 363L, 363L/366H, 365L, 365Q, 365R, 365S, 367T, 383V, 383V/401A, 383V/416P/422T, 385L, 385T, 388D, 388L, 388P, 389D, 392A, 392L, 395R, 396P, 396Y, 401Q, 401S, 404M, 405W, 408A, 408E, 408W, 410H, 416P, 417S, 439L, 439S, 443L, 443S, 447S, 447T, 450D, and 451S, wherein the positions are numbered with reference to SEQ ID NO: 8. In some additional embodiments, the amino acid differences comprise the substitution(s) Q5E, Q5G, G18A/P23R/A149S/C260T/A383V/G395D/E401S/A416P, G18A/P23R/A149S/A383V, G18A/F163M/P164Q, D21H, D21H/F163L/E315G/W316F, D21H/F163L/T323C/T408F, D21H/T408F, P23R/L56C/S86G/A149S/F163M/P164D/A383V/E401S/A416P, P23R/S86G, P23R/A149S/C260T, P23R/A149S/S284A/A383V/G395D, P23R/F163M/P164Q/A383V, P23R/F163M/P164S/E401S/A416P, S24K, S24R, V42F, V42F/R110K, V42F/Y187E/V272E, V42F/Y187E/G324S/I363L/R366H, V42F/Y187E/A353T, V42F/V272E/F291Y, V42F/V272E/F291Y/I363L, V42F/V272E/G324S/I363L/R366H, V42F/V272E/I363L/L410H, V42F/V272E/L410H, V42F/F291Y/A313V/I363L/L410H, V42F/F291Y/I363L, V42F/F291Y/I363L/R366H, V42F/A353T, V42F/I363L, G46S, K66A, E77M, S86G/A149S/F163M/P164S/A383V/G395D/E401S, S86G/A149S/G395D, S86G/F163M/P164S/C260T/A383V, S86G/A383V, D107L, D107S, D107Y, R110K, R110K/Y187E, R110K/Y187E/V253L/L410H, A134V, P138R, A149S/P164S/C260T/A383V/G395D/E401A, A149S/C260T/A383V, A149S/A416P, F163L/I259V/T323C/T408F, F163L/I259V/T408F, F163L/E315G/W316F, P164D/W316H/A383V/E401S, P164S/C260T/E401S, S167N, R186Q, Y187E, Y187E/V253L/I363L/R366H, Y187E/V272E/G324S/I363L/L410H, Y187E/V272E/I363L, Y187E/V272E/I363L/R366H/L410H, Y187E/F291Y, E189F, E189S, E189V, E189W, G191D, G191F,

E195W, A199Q, R203L, Q210A, Q210L, Q210M, Q210V, Q210Y, K211R, K248G, I259V/L307M, C260T/G395D/E401S, V272E, V272E/A353T, V272E/I363L/R366H, V272E/L410H, T277S, F291Y, K305E, T309A, T309F, T309R, E315G, N342T, E343G, R351L, P354S, E358L, K361R, H362Q, H362V, I363L, I363L/R366H, E365L, E365Q, E365R, E365S, P367T, A383V, A383V/E401A, A383V/A416P/V422T, K385L, K385T, A388D, A388L, A388P, S389D, P392A, P392L, G395R, N396P, N396Y, E401Q, E401S, A404M, N405W, T408A, T408E, T408W, L410H, A416P, L417S, D439L, D439S, K443L, K443S, K447S, K447T, A450D, and E451S, wherein the positions are numbered with reference to SEQ ID NO: 8.

[0174] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO: 8 selected from 18/23/149/383, 21/163/323/408, 272, 291, and 383, wherein the positions are numbered with reference to SEQ ID NO: 8. In some embodiments, the amino acid differences comprise the substitution(s) 18A/23R/149S/383V, 21H/163L/323C/408F, 272E, 291Y, and 383V, wherein the positions are numbered with reference to SEQ ID NO: 8. In some additional embodiments, the amino acid differences comprise the substitution(s) G18A/P23R/A149S/A383V, D21H/F163L/T323C/T408F, V272E, F291Y, and A383V, wherein the positions are numbered with reference to SEQ ID NO: 8.

[0175] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO: 366 selected from 24, 24/42/66/291, 24/42/291/362, 24/66/163/191/362/383/388, 24/66/191/199/260/291/351, 24/66/191/199/291, 24/66/191/260/408, 24/66/260/291/383/388/408, 24/66/291/342/383, 24/66/291/365, 24/66/342/365/388/408, 24/77/291, 24/107/163/191/291/351/383/388, 24/107/291/351/365/388, 24/163/351/383, 24/191/291/365, 24/199/260/351/362/383, 24/199/260/362/383/388, 24/260/362/383/388, 24/291, 24/291/342/351/383, 24/291/362/388, 24/291/408, 24/383/388, 24/388, 25, 28, 33, 42/191/408, 42/199/291/383, 42/291/351/362/365/383/388, 42/291/351/362/383/408, 42/291/383/388, 66/82/291/383, 66/163/191/365/383, 66/199/351/383, 66/291, 66/291/362/365/383, 66/291/383/388, 66/383, 77/291, 77/383/388, 86, 107/191/199/365/383/388, 107/191/291/383, 148, 153, 163/291/362/365/383/388, 163/291/383/388, 163/383, 191/199/365/383/388, 191/260/388, 191/291, 191/291/342/362/365, 191/351/383/388, 199/260/383, 199/291, 260, 260/291/365/383/408, 260/365/383, 291, 291/351/383/388, 291/351/383/388/408, 291/362/365, 291/365/388, 291/383, 314, 315, 316, 319, 342/362, 351/383/388, 362, 362/388, 383, 383/388, 396, 397, 405, 406, 413, 419, and 423, wherein the positions are numbered with reference to SEQ ID NO: 366. In some embodiments, the amino acid differences comprise the substitution(s) 24E, 24K, 24K/42F/66A/291Y, 24K/42F/291Y/362Q, 24K/66A/163M/191D/362Q/383V/388D, 24K/66A/191D/199Q/260T/291Y/351L, 24K/66A/191D/199Q/291Y, 24K/66A/191D/260T/408A, 24K/66A/260T/291Y/383V/388D/408A, 24K/66A/291Y/342T/383V, 24K/66A/291Y/365S, 24K/66A/342T/365S/388D/408E, 24K/77M/291Y,

24K/107L/163M/191D/291Y/351L/383V/388D, 24K/107L/291Y/351L/365S/388D, 24K/163M/351L/383V, 24K/191D/291Y/365S, 24K/199Q/260T/351L/362Q/383V, 24K/199Q/260T/362Q/383V/388D, 24K/260T/362Q/383V/388D, 24K/291Y, 24K/291Y/342T/351L/383V, 24K/291Y/362Q/388D, 24K/291Y/408A, 24K/383V/388D, 24K/388D, 25H, 25V, 28S, 28T, 33T, 42F/191D/408E, 42F/199Q/291Y/383V, 42F/291Y/351L/362Q/365S/383V/388D, 42F/291Y/351L/362Q/383V/408A, 42F/291Y/383V/388D, 66A/82H/291Y/383V, 66A/163M/191D/365S/383V, 66A/199Q/351L/383V, 66A/291Y, 66A/291Y/362Q/365S/383V, 66A/291Y/383V/388D, 66A/383V, 77M/291Y, 77M/383V/388D, 86T, 107L/191D/199Q/365S/383V/388D, 107L/191D/291Y/383V, 148G, 153S, 163M/291Y/362Q/365S/383V/388D, 163M/291Y/383V/388D, 163M/383V, 191D/199Q/365S/383V/388D, 191D/260T/388D, 191D/291Y, 191D/291Y/342T/362Q/365S, 191D/351L/383V/388D, 199Q/260T/383V, 199Q/291Y, 260T, 260T/291Y/365S/383V/408A, 260T/365S/383V, 291Y, 291Y/351L/383V/388D, 291Y/351L/383V/388D/408A, 291Y/362Q/365S, 291Y/365S/388D, 291Y/383V, 314K, 315S, 316V, 319S, 342T/362Q, 351L/383V/388D, 362Q, 362Q/388D, 383V, 383V/388D, 396R, 397M, 405A, 406H, 413L, 419S, and 423V, wherein the positions are numbered with reference to SEQ ID NO: 366. In some additional embodiments, the amino acid differences comprise the substitution(s) S24E, S24K, S24K/V42F/K66A/F291Y, S24K/V42F/F291Y/H362Q, S24K/K66A/L163M/G191D/H362Q/A383V/A388D, S24K/K66A/G191D/A199Q/C260T/F291Y/R351L, S24K/K66A/G191D/A199Q/F291Y, S24K/K66A/G191D/C260T/F408A, S24K/K66A/C260T/F291Y/A383V/A388D/F408A, S24K/K66A/F291Y/N342T/A383V, S24K/K66A/F291Y/E365S, S24K/K66A/N342T/E365S/A388D/F408E, S24K/E77M/F291Y, S24K/D107L/L163M/G191D/F291Y/R351L/A383V/A388D, S24K/D107L/F291Y/R351L/E365S/A388D, S24K/L163M/R351L/A383V, S24K/G191D/F291Y/E365S, S24K/A199Q/C260T/R351L/H362Q/A383V, S24K/A199Q/C260T/H362Q/A383V/A388D, S24K/C260T/H362Q/A383V/A388D, S24K/F291Y, S24K/F291Y/N342T/R351L/A383V, S24K/F291Y/H362Q/A388D, S24K/F291Y/F408A, S24K/A383V/A388D, S24K/A388D, L25H, L25V, R28S, R28T, V33T, V42F/G191D/F408E, V42F/A199Q/F291Y/A383V, V42F/F291Y/R351L/H362Q/E365S/A383V/A388D, V42F/F291Y/R351L/H362Q/A383V/F408A, V42F/F291Y/A383V/A388D, K66A/Y82H/F291Y/A383V, K66A/L163M/G191D/E365S/A383V, K66A/A199Q/R351L/A383V, K66A/F291Y, K66A/F291Y/H362Q/E365S/A383V, K66A/F291Y/A383V/A388D, K66A/A383V, E77M/F291Y, E77M/A383V/A388D, S86T, D107L/G191D/A199Q/E365S/A383V/A388D, D107L/G191D/F291Y/A383V, N148G, A153S, L163M/F291Y/H362Q/E365S/A383V/A388D, L163M/F291Y/A383V/A388D, L163M/A383V, G191D/A199Q/E365S/A383V/A388D, G191D/C260T/A388D, G191D/F291Y, G191D/F291Y/N342T/H362Q/E365S, G191D/R351L/A383V/A388D, A199Q/C260T/A383V, A199Q/F291Y, C260T, C260T/F291Y/E365S/A383V/F408A, C260T/E365S/A383V, F291Y,

F291Y/R351L/A383V/A388D, F291Y/R351L/A383V/A388D/F408A, F291Y/H362Q/E365S, F291Y/E365S/A388D, F291Y/A383V, R314K, E315S, W316V, H319S, N342T/H362Q, R351L/A383V/A388D, H362Q, H362Q/A388D, A383V, A383V/A388D, N396R, L397M, N405A, T406H, I413L, Q419S, and L423V, wherein the positions are numbered with reference to SEQ ID NO: 366.

[0176] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO: 366 selected from 24/66/191/199/291, 24/66/291/365, 163/291/362/365/383/388, 163/291/383/388, 191/291/342/362/365, 291, and 291/383, wherein the positions are numbered with reference to SEQ ID NO: 366. In some embodiments, the amino acid differences comprise the substitution(s) 24K/66A/191D/199Q/291Y, 24K/66A/291Y/365S, 163M/291Y/362Q/365S/383V/388D, 163M/291Y/383V/388D, 191D/291Y/342T/362Q/365S, 291Y, and 291Y/383V, wherein the positions are numbered with reference to SEQ ID NO: 366. In some additional embodiments, the amino acid differences comprise the substitution(s) S24K/K66A/G191D/A199Q/F291Y, S24K/K66A/F291Y/E365S, L163M/F291Y/H362Q/E365S/A383V/A388D, L163M/F291Y/A383V/A388D, G191D/F291Y/N342T/H362Q/E365S, F291Y, and F291Y/A383V, wherein the positions are numbered with reference to SEQ ID NO: 366.

[0177] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO: 650 selected from 10, 13, 13/24/108/163, 13/24/108/163/311, 13/24/133/199/311, 13/24/163, 13/24/199/311, 13/108, 13/108/199, 13/108/311, 13/199, 13/311, 14, 14/24/108, 14/24/108/133, 14/24/108/199, 14/24/199, 14/108, 14/108/133/311, 14/108/311, 14/311, 24, 24/163, 24/163/199, 35, 72, 73, 78, 95, 101, 108, 108/199, 114, 154, 163, 169, 175/316, 199, 199/311, 226, 293, 311, 316, 382, 383, and 386, wherein the positions are numbered with reference to SEQ ID NO: 650. In some embodiments, the amino acid differences comprise the substitution(s) 10E, 13A, 13A/24E/108R/163L/311S, 13A/24E/133R/199Q/311S, 13A/24E/163L, 13A/24K/108R/163L, 13A/24K/199Q/311S, 13A/108R, 13A/108R/199Q, 13A/108R/311S, 13A/199Q, 13A/311S, 14A, 14G, 14H, 14H/24E/108R/133R, 14H/24K/108R, 14H/24K/108R/199Q, 14H/24K/199Q, 14H/108R, 14H/108R/133R/311S, 14H/108R/311S, 14H/311S, 24E, 24E/163L, 24E/163L/199Q, 35E, 72G, 73R, 73S, 78A, 95I, 101L, 108R, 108R/199Q, 114A, 154S, 163H, 163S, 163V, 169C, 169V, 175D/316F, 199Q, 199Q/311S, 226Q, 293A, 311K, 316D, 316E, 316F, 316G, 316H, 316I, 316L, 316N, 316S, 316V, 316Y, 382D, 383L, and 386A ,wherein the positions are numbered with reference to SEQ ID NO: 650. In some additional embodiments, the amino acid differences comprise the substitution(s) R10E, T13A, T13A/S24E/S108R/M163L/I311S, T13A/S24E/A133R/A199Q/I311S, T13A/S24E/M163L, T13A/S24K/S108R/M163L, T13A/S24K/A199Q/I311S, T13A/S108R, T13A/S108R/A199Q, T13A/S108R/I311S, T13A/A199Q,

T13A/I311S, Y14A, Y14G, Y14H, Y14H/S24E/S108R/A133R, Y14H/S24K/S108R, Y14H/S24K/S108R/A199Q, Y14H/S24K/A199Q, Y14H/S108R, Y14H/S108R/A133R/I311S, Y14H/S108R/I311S, Y14H/I311S, S24E, S24E/M163L, S24E/M163L/A199Q, H35E, A72G, K73R, K73S, R78A, M95I, V101L, S108R, S108R/A199Q, T114A, T154S, M163H, M163S, M163V, F169C, F169V, G175D/W316F, A199Q, A199Q/I311S, M226Q, P293A, I311K, W316D, W316E, W316F, W316G, W316H, W316I, W316L, W316N, W316S, W316V, W316Y, E382D, V383L, and D386A, wherein the positions are numbered with reference to SEQ ID NO: 650.

[0178] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered transaminase polypeptide comprising an amino acid sequence that has one or more amino acid residue differences as compared to SEQ ID NO: 650 selected from 14/108/133/311, 24/163/199, 72, 78, 316, and 383, wherein the positions are numbered with reference to SEQ ID NO: 650. In some embodiments, the amino acid differences comprise the substitution(s) 14H/108R/133R/311S, 24E/163L/199Q, 72G, 78A, 316L, 316N, 316S, and 383L, wherein the positions are numbered with reference to SEQ ID NO: 650. In some additional embodiments, the amino acid differences comprise the substitution(s) Y14H/S108R/A133R/I311S, S24E/M163L/A199Q, A72G, R78A, W316L, W316N, W316S, and V383L, wherein the positions are numbered with reference to SEQ ID NO: 650.

[0179] In some embodiments, the variant transaminase of the present disclosure further comprises additional sequences that do not alter the encoded activity of the enzyme. For example, in some embodiments, the variant transaminase is linked to an epitope tag or to another sequence useful in purification.

[0180] In some embodiments, the variant transaminase polypeptides of the present disclosure are secreted from the host cell in which they are expressed (e.g., a yeast or filamentous fungal host cell) and are expressed as a pre-protein including a signal peptide (i.e., an amino acid sequence linked to the amino terminus of a polypeptide and which directs the encoded polypeptide into the cell secretory pathway).

[0181] When the sequence of the engineered polypeptide is known, the polynucleotides encoding the enzyme can be prepared by standard solid-phase methods, according to known synthetic methods. In some embodiments, fragments of up to about 100 bases can be individually synthesized, then joined (e.g., by enzymatic or chemical ligation methods, or polymerase mediated methods) to form any desired continuous sequence. For example, polynucleotides and oligonucleotides of the disclosure can be prepared by chemical synthesis (e.g., using the classical phosphoramidite method described by Beaucage et al., Tet. Lett., 22:1859-69 [1981], or the method described by Matthes et al., EMBO J., 3:801-05 [1984], as it is typically practiced in automated synthetic methods). According to the phosphoramidite method, oligonucleotides are synthesized (e.g., in an automatic DNA synthesizer), purified, annealed, ligated and cloned in appropriate vectors. In addition, essentially any nucleic acid can be obtained from any of a variety of commercial sources (e.g., The Midland Certified Reagent Company, Midland, TX,

The Great American Gene Company, Ramona, CA, ExpressGen Inc. Chicago, IL, Operon Technologies Inc., Alameda, CA, and many others).

[0182] The present disclosure also provides recombinant constructs comprising a sequence encoding at least one variant transaminase, as provided herein. In some embodiments, the present disclosure provides an expression vector comprising a variant transaminase polynucleotide operably linked to a heterologous promoter. In some embodiments, expression vectors of the present disclosure are used to transform appropriate host cells to permit the host cells to express the variant transaminase protein. Methods for recombinant expression of proteins in fungi and other organisms are well known in the art, and a number of expression vectors are available or can be constructed using routine methods. In some embodiments, nucleic acid constructs of the present disclosure comprise a vector, such as, a plasmid, a cosmid, a phage, a virus, a bacterial artificial chromosome (BAC), a yeast artificial chromosome (YAC), and the like, into which a nucleic acid sequence of the disclosure has been inserted. In some embodiments, polynucleotides of the present disclosure are incorporated into any one of a variety of expression vectors suitable for expressing variant transaminase polypeptide(s). Suitable vectors include, but are not limited to chromosomal, nonchromosomal and synthetic DNA sequences (e.g., derivatives of SV40), as well as bacterial plasmids, phage DNA, baculovirus, yeast plasmids, vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, pseudorabies, adenovirus, adeno-associated virus, retroviruses, and many others. Any suitable vector that transduces genetic material into a cell, and, if replication is desired, which is replicable and viable in the relevant host finds use in the present disclosure.

[0183] In some embodiments, the construct further comprises regulatory sequences, including but not limited to a promoter, operably linked to the protein encoding sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art. Indeed, in some embodiments, in order to obtain high levels of expression in a particular host it is often useful to express the variant transaminases of the present disclosure under the control of a heterologous promoter. In some embodiments, a promoter sequence is operably linked to the 5' region of the variant transaminase coding sequence using any suitable method known in the art. Examples of useful promoters for expression of variant transaminases include, but are not limited to promoters from fungi. In some embodiments, a promoter sequence that drives expression of a gene other than a transaminase gene in a fungal strain finds use. As a non-limiting example, a fungal promoter from a gene encoding an endoglucanase may be used. In some embodiments, a promoter sequence that drives the expression of a transaminase gene in a fungal strain other than the fungal strain from which the transaminases were derived finds use. Examples of other suitable promoters useful for directing the transcription of the nucleotide constructs of the present disclosure in a filamentous fungal host cell include, but are not limited to promoters obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *Aspergillus niger* neutral alpha-amylase, *Aspergillus niger* acid stable alpha-amylase, *Aspergillus niger* or *Aspergillus awamori* glucoamylase (glaA), *Rhizomucor miehei* lipase, *Aspergillus oryzae* alkaline protease,

Aspergillus oryzae triose phosphate isomerase, *Aspergillus nidulans* acetamidase, and *Fusarium oxysporum* trypsin-like protease (See e.g., WO 96/00787, incorporated herein by reference), as well as the NA2-tpi promoter (a hybrid of the promoters from the genes for *Aspergillus niger* neutral alpha-amylase and *Aspergillus oryzae* triose phosphate isomerase), promoters such as *cbh*1, *cbh*2, *egl*1, *egl*2, *pep*A, *hfb*1, *hfb*2, *xyn*1, *amy*, and *gla*A (See e.g., Nunberg *et al*., Mol. Cell Biol., 4:2306 -2315 [1984]; Boel *et al*., EMBO J., 3:1581-85 [1984]; and European Patent Appln. 137280, all of which are incorporated herein by reference), and mutant, truncated, and hybrid promoters thereof.

[0184] In yeast host cells, useful promoters include, but are not limited to those from the genes for *Saccharomyces cerevisiae* enolase (*eno-1*), *Saccharomyces cerevisiae* galactokinase (*gal1*), *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (*ADH2/GAP*), and *S. cerevisiae* 3-phosphoglycerate kinase. Additional useful promoters useful for yeast host cells are known in the art (See e.g., Romanos *et al*., Yeast 8:423-488 [1992], incorporated herein by reference). In addition, promoters associated with chitinase production in fungi find use in the present disclosure (See e.g., Blaiseau and Lafay, Gene 120243-248 [1992]; and Limon *et al*., Curr. Genet., 28:478-83 [1995], both of which are incorporated herein by reference).

[0185] For bacterial host cells, suitable promoters for directing transcription of the nucleic acid constructs of the present disclosure, include but are not limited to the promoters obtained from the *E. coli* lac operon, *E. coli* trp operon, bacteriophage lambda, *Streptomyces coelicolor* agarase gene (dagA), *Bacillus subtilis* levansucrase gene (sacB), *Bacillus licheniformis* alpha-amylase gene (amyL), *Bacillus stearothermophilus* maltogenic amylase gene (amyM), *Bacillus amyloliquefaciens* alpha-amylase gene (amyQ), *Bacillus licheniformis* penicillinase gene (penP), *Bacillus subtilis* xylA and xylB genes, and prokaryotic beta-lactamase gene (See e.g., Villa-Kamaroff et al., Proc. Natl. Acad. Sci. USA 75: 3727-3731 [1978]), as well as the *tac* promoter (See e.g., DeBoer et al., Proc. Natl. Acad. Sci. USA 80: 21-25 [1983]).

[0186] In some embodiments, cloned variant transaminases of the present disclosure also have a suitable transcription terminator sequence, a sequence recognized by a host cell to terminate transcription. The terminator sequence is operably linked to the 3' terminus of the nucleic acid sequence encoding the polypeptide. Any terminator that is functional in the host cell of choice finds use in the present disclosure. Exemplary transcription terminators for filamentous fungal host cells include, but are not limited to those obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Aspergillus niger* alpha-glucosidase, and *Fusarium oxysporum* trypsin-like protease (See e.g., US Patent No. 7,399,627, incorporated herein by reference). In some embodiments, exemplary terminators for yeast host cells include those obtained from the genes for *Saccharomyces cerevisiae* enolase, *Saccharomyces cerevisiae* cytochrome C (CYCl), and *Saccharomyces cerevisiae* glyceraldehyde-3-phosphate dehydrogenase. Other useful terminators for yeast host cells are well-known to those skilled in the art (See e.g., Romanos *et al*., Yeast 8:423-88 [1992]).

[0187] In some embodiments, a suitable leader sequence is part of a cloned variant transaminase sequence, which is a nontranslated region of an mRNA that is important for translation by the host cell. The leader sequence is operably linked to the 5' terminus of the nucleic acid sequence encoding the polypeptide. Any leader sequence that is functional in the host cell of choice finds use in the present disclosure. Exemplary leaders for filamentous fungal host cells include, but are not limited to those obtained from the genes for *Aspergillus oryzae* TAKA amylase and *Aspergillus nidulans* triose phosphate isomerase. Suitable leaders for yeast host cells include, but are not limited to those obtained from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* 3-phosphoglycerate kinase, *Saccharomyces cerevisiae* alpha-factor, and *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP).

[0188] In some embodiments, the sequences of the present disclosure also comprise a polyadenylation sequence, which is a sequence operably linked to the 3' terminus of the nucleic acid sequence and which, when transcribed, is recognized by the host cell as a signal to add polyadenosine residues to transcribed mRNA. Any polyadenylation sequence which is functional in the host cell of choice finds use in the present disclosure. Exemplary polyadenylation sequences for filamentous fungal host cells include, but are not limited to those obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Fusarium oxysporum* trypsin-like protease, and *Aspergillus niger* alpha-glucosidase. Useful polyadenylation sequences for yeast host cells are known in the art (See e.g., Guo and Sherman, Mol. Cell. Biol., 15:5983-5990 [1995]).

[0189] In some embodiments, the control sequence comprises a signal peptide coding region encoding an amino acid sequence linked to the amino terminus of a polypeptide and directs the encoded polypeptide into the cell's secretory pathway. The 5' end of the coding sequence of the nucleic acid sequence may inherently contain a signal peptide coding region naturally linked in translation reading frame with the segment of the coding region that encodes the secreted polypeptide. Alternatively, the 5' end of the coding sequence may contain a signal peptide coding region that is foreign to the coding sequence. The foreign signal peptide coding region may be required where the coding sequence does not naturally contain a signal peptide coding region.

[0190] Alternatively, the foreign signal peptide coding region may simply replace the natural signal peptide coding region in order to enhance secretion of the polypeptide. However, any signal peptide coding region which directs the expressed polypeptide into the secretory pathway of a host cell of choice may be used in the present disclosure.

[0191] In some embodiments, the signal peptide is an endogenous *V. fluvialis* transaminase signal peptide. In some additional embodiments, signal peptides from other *V. fluvialis* secreted proteins are used. In some embodiments, other signal peptides find use, depending on the host cell and other factors.

[0192] Effective signal peptide coding regions for bacterial host cells include, but are not limited to the signal peptide coding regions obtained from the genes for *Bacillus* NClB 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]).

[0193] 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.

[0194] Useful signal peptides for yeast host cells include, but are not limited to genes for *Saccharomyces cerevisiae* alpha-factor and *Saccharomyces cerevisiae* invertase. Other useful signal peptide coding regions are known in the art (See e.g., Romanos et al., [1992], supra).

[0195] In some embodiments, the control sequence comprises a propeptide coding region that codes for an amino acid sequence positioned at the amino terminus of a polypeptide. The resultant polypeptide is known as a proenzyme or propolypeptide (or a zymogen in some cases). A propolypeptide is generally inactive and can be converted to a mature active transaminase polypeptide by catalytic or autocatalytic cleavage of the propeptide from the propolypeptide. The propeptide coding region may be obtained from the genes for *Bacillus subtilis* alkaline protease (aprE), *Bacillus subtilis* neutral protease (nprT), *Saccharomyces cerevisiae* alpha-factor, *Rhizomucor miehei* aspartic proteinase, and *Myceliophthora thermophila* lactase (See e.g., WO 95/33836).

[0196] 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.

[0197] In some embodiments, regulatory sequences are also used to allow the regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory systems are those which cause the expression of the gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. In prokaryotic host cells, suitable regulatory sequences include, but are not limited to the lac, tac, and trp operator systems. In yeast host cells, suitable regulatory systems include, as examples, the ADH2 system or GAL1 system. In filamentous fungi, suitable regulatory sequences include the TAKA alpha-amylase promoter, *Aspergillus niger* glucoamylase promoter, and *Aspergillus oryzae* glucoamylase promoter.

[0198] Other examples of regulatory sequences are those which allow for gene amplification. In eukaryotic systems, these include the dihydrofolate reductase gene, which is amplified in the presence of methotrexate, and the metallothionein genes, which are amplified with heavy metals. In these cases, the nucleic acid sequence encoding the transaminase polypeptide of the present disclosure would be operably linked with the regulatory sequence.

[0199] Thus, in additional embodiments, the present disclosure provides recombinant expression vectors comprising a polynucleotide encoding an engineered transaminase polypeptide or a variant thereof, and one or more expression regulating regions such as a promoter and a terminator, a replication origin, etc.,

depending on the type of hosts into which they are to be introduced. In some embodiments, the various nucleic acid and control sequences described above are joined together to produce a recombinant expression vector that may include one or more convenient restriction sites to allow for insertion or substitution of the nucleic acid sequence encoding the polypeptide at such sites. Alternatively, in some embodiments, the nucleic acid sequences are expressed by inserting the nucleic acid sequence or a nucleic acid construct comprising the sequence into an appropriate vector for expression. In creating the expression vector, the coding sequence is located in the vector so that the coding sequence is operably linked with the appropriate control sequences for expression.

[0200] The recombinant expression vector comprises any suitable vector (e.g., a plasmid or virus), that can be conveniently subjected to recombinant DNA procedures and can bring about the expression of the polynucleotide sequence. The choice of the vector typically depends on the compatibility of the vector with the host cell into which the vector is to be introduced. In some embodiments, the vectors are linear or closed circular plasmids.

[0201] In some embodiments, the expression vector is an autonomously replicating vector (i.e., a vector that exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, such as a plasmid, an extrachromosomal element, a minichromosome, or an artificial chromosome). In some embodiments, the vector contains any means for assuring self-replication. Alternatively, in some other embodiments, upon being introduced into the host cell, the vector is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. Furthermore, in additional embodiments, a single vector or plasmid or two or more vectors or plasmids which together contain the total DNA to be introduced into the genome of the host cell, or a transposon find use.

[0202] In some embodiments, the expression vector of the present disclosure contains one or more selectable markers, which permit easy selection of transformed cells. A "selectable marker" is a gene, the product of which provides for biocide or viral resistance, resistance to antimicrobials or heavy metals, prototrophy to auxotrophs, and the like. Any suitable selectable markers for use in a filamentous fungal host cell find use in the present disclosure, including, but are not limited to, amdS (acetamidase), argB (ornithine carbamoyltransferase), bar (phosphinothricin acetyltransferase), hph (hygromycin phosphotransferase), niaD (nitrate reductase), pyrG (orotidine-5'-phosphate decarboxylase), sC (sulfate adenyltransferase), and trpC (anthranilate synthase), as well as equivalents thereof. Additional markers useful in host cells such as *Aspergillus*, include but are not limited to the amdS and pyrG genes of *Aspergillus nidulans* or *Aspergillus oryzae,* and the bar gene of *Streptomyces hygroscopicus*. Suitable markers for yeast host cells include, but are not limited to ADE2, HIS3, LEU2, LYS2, MET3, TRP1, and URA3. Examples of bacterial selectable markers include, but are not limited to the *dal* genes from *Bacillus subtilis* or *Bacillus licheniformis*, or markers, which confer antibiotic resistance such as ampicillin, kanamycin, chloramphenicol, and or tetracycline resistance.

[0203] In some embodiments, the expression vectors of the present disclosure contain an element(s) that permits integration of the vector into the host cell's genome or autonomous replication of the vector in the cell independent of the genome. In some embodiments involving integration into the host cell genome, the vectors rely on the nucleic acid sequence encoding the polypeptide or any other element of the vector for integration of the vector into the genome by homologous or non-homologous recombination.

[0204] In some alternative embodiments, the expression vectors contain additional nucleic acid sequences for directing integration by homologous recombination into the genome of the host cell. The additional nucleic acid sequences enable the vector to be integrated into the host cell genome at a precise location(s) in the chromosome(s). To increase the likelihood of integration at a precise location, the integrational elements preferably contain a sufficient number of nucleotides, such as 100 to 10,000 base pairs, preferably 400 to 10,000 base pairs, and most preferably 800 to 10,000 base pairs, which are highly homologous with the corresponding target sequence to enhance the probability of homologous recombination. The integrational elements may be any sequence that is homologous with the target sequence in the genome of the host cell. Furthermore, the integrational elements may be non-encoding or encoding nucleic acid sequences. On the other hand, the vector may be integrated into the genome of the host cell by non-homologous recombination.

[0205] 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, pACYCl77 (which plasmid has the P15A ori), or pACYC184 permitting replication in *E. coli*, and pUB110, pE194, pTA1060, or pAMβ1 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 its functioning temperature-sensitive in the host cell (See e.g., Ehrlich, Proc. Natl. Acad. Sci. USA 75:1433 [1978]).

[0206] In some embodiments, more than one copy of a nucleic acid sequence of the present disclosure is inserted into the host cell to increase production of the gene product. An increase in the copy number of the nucleic acid sequence can be obtained by integrating at least one additional copy of the sequence into the host cell genome or by including an amplifiable selectable marker gene with the nucleic acid sequence where cells containing amplified copies of the selectable marker gene, and thereby additional copies of the nucleic acid sequence, can be selected for by cultivating the cells in the presence of the appropriate selectable agent.

[0207] Many of the expression vectors for use in the present disclosure are commercially available. Suitable commercial expression vectors include, but are not limited to the p3xFLAGTM™ expression vectors (Sigma-Aldrich Chemicals), which include a CMV promoter and hGH polyadenylation site for expression in mammalian host cells and a pBR322 origin of replication and ampicillin resistance markers for amplification in *E. coli*. Other suitable expression vectors include, but are not limited to pBluescriptII

SK(-) and pBK-CMV (Stratagene), and plasmids derived from pBR322 (Gibco BRL), pUC (Gibco BRL), pREP4, pCEP4 (Invitrogen) or pPoly (See e.g., Lathe et al., Gene 57:193-201 [1987]).

[0208] Thus, in some embodiments, a vector comprising a sequence encoding at least one variant transaminase is transformed into a host cell in order to allow propagation of the vector and expression of the variant transaminase (s). In some embodiments, the variant transaminases are post-translationally modified to remove the signal peptide and, in some cases, may be cleaved after secretion. In some embodiments, the transformed host cell described above is cultured in a suitable nutrient medium under conditions permitting the expression of the variant transaminase(s). Any suitable medium useful for culturing the host cells finds use in the present disclosure, including, but not limited to minimal or complex media containing appropriate supplements. In some embodiments, host cells are grown in HTP media. Suitable media are available from various commercial suppliers or may be prepared according to published recipes (e.g., in catalogues of the American Type Culture Collection).

[0209] In another aspect, the present disclosure provides host cells comprising a polynucleotide encoding an improved transaminase polypeptide provided herein, the polynucleotide being operatively linked to one or more control sequences for expression of the transaminase enzyme in the host cell. Host cells for use in expressing the transaminase 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*, *Bacillus megaterium*, *Lactobacillus kefir*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells (e.g., *Saccharomyces cerevisiae* or *Pichia pastoris* (ATCC Accession No. 201178)); insect cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as CHO, COS, BHK, 293, and Bowes melanoma cells; and plant cells. Appropriate culture media and growth conditions for the above-described host cells are well known in the art.

[0210] 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 are known to those skilled in the art.

[0211] In some embodiments, the host cell is a eukaryotic cell. Suitable eukaryotic host cells include, but are not limited to, fungal cells, algal cells, insect cells, and plant cells. Suitable fungal host cells include, but are not limited to, Ascomycota, Basidiomycota, Deuteromycota, Zygomycota, Fungi imperfecti. In some embodiments, the fungal host cells are yeast cells and filamentous fungal cells. The filamentous fungal host cells of the present disclosure include all filamentous forms of the subdivision Eumycotina and Oomycota. Filamentous fungi are characterized by a vegetative mycelium with a cell wall composed of chitin, cellulose and other complex polysaccharides. The filamentous fungal host cells of the present disclosure are morphologically distinct from yeast.

[0212] In some embodiments of the present disclosure, the filamentous fungal host cells are of any suitable genus and species, including, but not limited to *Achlya*, *Acremonium*, *Aspergillus*, *Aureobasidium*, *Bjerkandera*, *Ceriporiopsis, Cephalosporium*, *Chrysosporium*, *Cochliobolus*,

Corynascus, *Cryphonectria*, *Cryptococcus*, *Coprinus*, *Coriolus*, *Diplodia*, *Endothis*, *Fusarium*, *Gibberella*, *Gliocladium*, *Humicola*, *Hypocrea*, *Myceliophthora*, *Mucor*, *Neurospora*, *Penicillium*, *Podospora*, *Phlebia*, *Piromyces*, *Pyricularia*, *Rhizomucor*, *Rhizopus*, *Schizophyllum*, *Scytalidium*, *Sporotrichum, Talaromyces*, *Thermoascus*, *Thielavia*, *Trametes*, *Tolypocladium*, *Trichoderma*, *Verticillium*, and/or *Volvariella*, and/or teleomorphs, or anamorphs, and synonyms, basionyms, or taxonomic equivalents thereof.

[0213] In some embodiments of the present disclosure, the host cell is a yeast cell, including but not limited to cells of *Candida*, *Hansenula*, *Saccharomyces*, *Schizosaccharomyces*, *Pichia*, *Kluyveromyces*, or *Yarrowia* species. In some embodiments of the present disclosure, the yeast cell is *Hansenula polymorpha*, *Saccharomyces cerevisiae*, *Saccharomyces carlsbergensis*, *Saccharomyces diastaticus*, *Saccharomyces norbensis*, *Saccharomyces kluyveri*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia kodamae*, *Pichia membranaefaciens*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria, Pichia quercuum, Pichia pijperi*, *Pichia stipitis*, *Pichia methanolica*, *Pichia angusta*, *Kluyveromyces lactis*, *Candida albicans*, or *Yarrowia lipolytica*.

[0214] In some embodiments of the disclosure, the host cell is an algal cell such as *Chlamydomonas* (e.g., *C. reinhardtii*) and *Phormidium* (e.g., *Phormidium* sp. ATCC29409).

[0215] In some other embodiments, the host cell is a prokaryotic cell. Suitable prokaryotic cells include, but are not limited to Gram-positive, Gram-negative and Gram-variable bacterial cells. Any suitable bacterial organism finds use in the present disclosure, including but not limited to *Agrobacterium*, *Alicyclobacillus*, *Anabaena*, *Anacystis, Acinetobacter, Acidothermus, Arthrobacter, Azobacter, Bacillus, Bifidobacterium, Brevibacterium, Butyrivibrio, Buchnera, Campestris, Camplyobacter, Clostridium, Corynebacterium, Chromatium, Coprococcus, Escherichia, Enterococcus, Enterobacter, Erwinia, Fusobacterium, Faecalibacterium, Francisella, Flavobacterium, Geobacillus, Haemophilus, Helicobacter, Klebsiella, Lactobacillus, Lactococcus, Ilyobacter, Micrococcus, Microbacterium, Mesorhizobium, Methylobacterium, Methylobacterium, Mycobacterium, Neisseria, Pantoea, Pseudomonas, Prochlorococcus, Rhodobacter, Rhodopseudomonas, Rhodopseudomonas, Roseburia, Rhodospirillum, Rhodococcus, Scenedesmus, Streptomyces, Streptococcus, Synecoccus, Saccharomonospora, Staphylococcus, Serratia, Salmonella, Shigella, Thermoanaerobacterium, Tropheryma, Tularensis, Temecula, Thermosynechococcus, Thermococcus, Ureaplasma, Xanthomonas, Xylella, Yersinia* and *Zymomonas*. In some embodiments, the host cell is a species of *Agrobacterium*, *Acinetobacter*, *Azobacter*, *Bacillus*, *Bifidobacterium*, *Buchnera*, *Geobacillus*, *Campylobacter*, *Clostridium*, *Corynebacterium*, *Escherichia*, *Enterococcus*, *Erwinia*, *Flavobacterium*, *Lactobacillus*, *Lactococcus*, *Pantoea*, *Pseudomonas*, *Staphylococcus*, *Salmonella*, *Streptococcus*, *Streptomyces*, or *Zymomonas*. In some embodiments, the bacterial host strain is non-pathogenic to humans. In some embodiments the bacterial host strain is an industrial strain. Numerous bacterial industrial strains are known and suitable in the present disclosure. In some embodiments of the present disclosure, the bacterial host cell is an *Agrobacterium* species (e.g., *A. radiobacter*, *A. rhizogenes*, and *A. rubi*). In some

embodiments of the present disclosure, the bacterial host cell is an *Arthrobacter* species (e.g., *A. aurescens*, *A. citreus*, *A. globiformis*, *A. hydrocarboglutamicus*, *A. mysorens*, *A. nicotianae*, *A. paraffineus*, *A. protophonniae*, *A. roseoparqffinus*, *A. sulfureus*, and *A. ureafaciens).* In some embodiments of the present disclosure, the bacterial host cell is a *Bacillus* species (e.g., *B. thuringensis*, *B. anthracis*, *B. megaterium*, *B. subtilis*, *B. lentus*, *B. circulans*, *B. pumilus*, *B. lautus*, *B.coagulans, B. brevis, B. firmus, B. alkaophius*, *B. licheniformis*, *B. clausii*, *B. stearothermophilus*, *B. halodurans,* and *B. amyloliquefaciens*). In some embodiments, the host cell is an industrial *Bacillus* strain including but not limited to *B. subtilis*, *B. pumilus*, *B. licheniformis*, *B. megaterium, B. clausii*, *B. stearothermophilus,* or *B. amyloliquefaciens.* In some embodiments, the *Bacillus* host cells are *B. subtilis, B. licheniformis*, *B. megaterium, B. stearothermophilus,* and/or *B. amyloliquefaciens.* In some embodiments, the bacterial host cell is a *Clostridium* species *(*e.g., *C. acetobutylicum*, *C. tetani* E88, *C. lituseburense*, *C. saccharobutylicum*, *C. perfringens*, and *C. beijerinckii*). In some embodiments, the bacterial host cell is a *Corynebacterium* species (e.g., *C. glutamicum* and *C. acetoacidophilum*). In some embodiments the bacterial host cell is an *Escherichia* species (e.g., *E. coli*). In some embodiments, the host cell is *Escherichia coli* W3110. In some embodiments, the bacterial host cell is an *Erwinia* species (e.g., *E. uredovora*, *E. carotovora*, *E. ananas*, *E. herbicola*, *E. punctata*, and *E. terreus*). In some embodiments, the bacterial host cell is a *Pantoea* species (e.g., *P. citrea,* and *P. agglomerans).* In some embodiments the bacterial host cell is a *Pseudomonas* species (e.g., *P. putida*, *P. aeruginosa*, *P. mevalonii*, and *P. sp.* D-0l 10). In some embodiments, the bacterial host cell is a *Streptococcus* species (e.g., *S. equisimiles*, *S. pyogenes*, and *S. uberis*). In some embodiments, the bacterial host cell is a *Streptomyces* species (e.g., *S. ambofaciens*, *S. achromogenes*, *S. avermitilis*, *S. coelicolor*, *S. aureofaciens, S. aureus*, *S. fungicidicus, S. griseus*, and *S. lividans*). In some embodiments, the bacterial host cell is a *Zymomonas* species (e.g., *Z. mobilis*, and *Z. lipolytica*).

[0216] Many prokaryotic and eukaryotic strains that find use in the present disclosure are readily available to the public from a number of culture collections such as American Type Culture Collection (ATCC), Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSM), Centraalbureau Voor Schimmelcultures (CBS), and Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL).

[0217] In some embodiments, host cells are genetically modified to have characteristics that improve protein secretion, protein stability and/or other properties desirable for expression and/or secretion of a protein. Genetic modification can be achieved by genetic engineering techniques and/or classical microbiological techniques (e.g., chemical or UV mutagenesis and subsequent selection). Indeed, in some embodiments, combinations of recombinant modification and classical selection techniques are used to produce the host cells. Using recombinant technology, nucleic acid molecules can be introduced, deleted, inhibited or modified, in a manner that results in increased yields of transaminase variant(s) within the host cell and/or in the culture medium. For example, knockout of Alp1 function results in a cell that is protease deficient, and knockout of pyr5 function results in a cell with a pyrimidine deficient

phenotype. In one genetic engineering approach, homologous recombination is used to induce targeted gene modifications by specifically targeting a gene *in vivo* to suppress expression of the encoded protein. In alternative approaches, siRNA, antisense and/or ribozyme technology find use in inhibiting gene expression. A variety of methods are known in the art for reducing expression of protein in cells, including, but not limited to deletion of all or part of the gene encoding the protein and site-specific mutagenesis to disrupt expression or activity of the gene product. (See e.g., Chaveroche et al., Nucl. Acids Res., 28:22 e97 [2000]; Cho et al., Mol. Plant Mic. Interact., 19:7-15 [2006]; Maruyama and Kitamoto, Biotechnol Lett., 30:1811-1817 [2008]; Takahashi et al., Mol. Gen. Genom., 272: 344-352 [2004]; and You et al. , Arch. Micriobiol.,191:615-622 [2009], all of which are incorporated by reference herein). Random mutagenesis, followed by screening for desired mutations also finds use (See e.g., Combier et al., FEMS Microbiol. Lett., **220**:141–8 [2003]; and Firon et al., Eukary. Cell **2**:247–55 [2003], both of which are incorporated by reference).

[0218] Introduction of a vector or DNA construct into a host cell can be accomplished using any suitable method known in the art, including but not limited to calcium phosphate transfection, DEAE-dextran mediated transfection, PEG-mediated transformation, electroporation, or other common techniques known in the art. In some embodiments, the *Escherichia coli* expression vector pCK100900i (See US Pat. Appln. Publn. 2006/0195947, which is hereby incorporated by reference herein) find use.

[0219] In some embodiments, the engineered host cells (i.e., "recombinant host cells") of the present disclosure are cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants, or amplifying the transaminase polynucleotide. Culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and are well-known to those skilled in the art. As noted, many standard references and texts are available for the culture and production of many cells, including cells of bacterial, plant, animal (especially mammalian) and archebacterial origin.

[0220] In some embodiments, cells expressing the variant transaminase polypeptides of the disclosure are grown under batch or continuous fermentations conditions. Classical "batch fermentation" is a closed system, wherein the compositions of the medium are set at the beginning of the fermentation and is not subject to artificial alternations during the fermentation. A variation of the batch system is a "fed-batch fermentation" which also finds use in the present disclosure. In this variation, the substrate is added in increments as the fermentation progresses. Fed-batch systems are useful when catabolite repression is likely to inhibit the metabolism of the cells and where it is desirable to have limited amounts of substrate in the medium. Batch and fed-batch fermentations are common and well known in the art. "Continuous fermentation" is an open system where a defined fermentation medium is added continuously to a bioreactor and an equal amount of conditioned medium is removed simultaneously for processing. Continuous fermentation generally maintains the cultures at a constant high density where cells are primarily in log phase growth. Continuous fermentation systems strive to maintain steady state growth conditions. Methods for modulating nutrients and growth factors for continuous fermentation processes

as well as techniques for maximizing the rate of product formation are well known in the art of industrial microbiology.

[0221] In some embodiments of the present disclosure, cell-free transcription/translation systems find use in producing variant transaminase(s). Several systems are commercially available, and the methods are well-known to those skilled in the art.

[0222] The present disclosure provides methods of making variant transaminase polypeptides or biologically active fragments thereof. In some embodiments, the method comprises: providing a host cell transformed with a polynucleotide encoding an amino acid sequence that comprises at least about 70% (or at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%) sequence identity to SEQ ID NO: 2, 4, 8, 366, and/or 650, and comprising at least one mutation as provided herein; culturing the transformed host cell in a culture medium under conditions in which the host cell expresses the encoded variant transaminase polypeptide; and optionally recovering or isolating the expressed variant transaminase polypeptide, and/or recovering or isolating the culture medium containing the expressed variant transaminase polypeptide. In some embodiments, the methods further provide optionally lysing the transformed host cells after expressing the encoded transaminase polypeptide and optionally recovering and/or isolating the expressed variant transaminase polypeptide from the cell lysate. The present disclosure further provides methods of making a variant transaminase polypeptide comprising cultivating a host cell transformed with a variant transaminase polypeptide under conditions suitable for the production of the variant transaminase polypeptide and recovering the variant transaminase polypeptide. Typically, recovery or isolation of the transaminase polypeptide is from the host cell culture medium, the host cell or both, using protein recovery techniques that are well known in the art, including those described herein. In some embodiments, host cells are harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including, but not limited to freeze-thaw cycling, sonication, mechanical disruption, and/or use of cell lysing agents, as well as many other suitable methods well known to those skilled in the art.

[0223] Engineered transaminase enzymes produced by a host cell can be recovered from the cells and/or the culture medium using any one or more of the techniques known in the art for protein purification, including, among others, lysozyme treatment, sonication, filtration, salting-out, ultra-centrifugation, and chromatography. Suitable solutions for lysing and the high efficiency extraction of proteins from bacteria, such as *E. coli*, are commercially available under the trade name CelLytic B™ (Sigma-Aldrich). Thus, in some embodiments, the resulting polypeptide is recovered/isolated and optionally purified by any of a number of methods known in the art. For example, in some embodiments, the polypeptide is isolated from the nutrient medium by conventional procedures including, but not limited to, centrifugation, filtration, extraction, spray-drying, evaporation, chromatography (e.g., ion exchange, affinity, hydrophobic interaction, chromatofocusing, and size exclusion), or precipitation. In some

embodiments, protein refolding steps are used, as desired, in completing the configuration of the mature protein. In addition, in some embodiments, high performance liquid chromatography (HPLC) is employed in the final purification steps. For example, in some embodiments, methods known in the art, find use in the present disclosure (See e.g., Parry et al., Biochem. J., 353:117 [2001]; and Hong et al., Appl. Microbiol. Biotechnol., 73:1331 [2007], both of which are incorporated herein by reference). Indeed, any suitable purification methods known in the art find use in the present disclosure.

[0224] Chromatographic techniques for isolation of the transaminase polypeptide include, but are not limited to reverse phase chromatography high performance liquid chromatography, ion exchange chromatography, gel electrophoresis, and affinity chromatography. Conditions for purifying a particular enzyme will depend, in part, on factors such as net charge, hydrophobicity, hydrophilicity, molecular weight, molecular shape, etc., are known to those skilled in the art.

[0225] In some embodiments, affinity techniques find use in isolating the improved transaminase enzymes. For affinity chromatography purification, any antibody which specifically binds the transaminase polypeptide may be used. For the production of antibodies, various host animals, including but not limited to rabbits, mice, rats, etc., may be immunized by injection with the transaminase. The transaminase polypeptide may be attached to a suitable carrier, such as BSA, by means of a side chain functional group or linkers attached to a side chain functional group. Various adjuvants may be used to increase the immunological response, depending on the host species, including but not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, and potentially useful human adjuvants such as BCG (Bacillus Calmette Guerin) and *Corynebacterium parvum*.

[0226] In some embodiments, the transaminase variants are prepared and used in the form of cells expressing the enzymes, as crude extracts, or as isolated or purified preparations. In some embodiments, the transaminase variants are prepared as lyophilisates, in powder form (e.g., acetone powders), or prepared as enzyme solutions. In some embodiments, the transaminase variants are in the form of substantially pure preparations.

[0227] In some embodiments, the transaminase polypeptides are attached to any suitable solid substrate. Solid substrates include but are not limited to a solid phase, surface, and/or membrane. Solid supports include, but are not limited to organic polymers such as polystyrene, polyethylene, polypropylene, polyfluoroethylene, polyethyleneoxy, and polyacrylamide, as well as co-polymers and grafts thereof. A solid support can also be inorganic, such as glass, silica, controlled pore glass (CPG), reverse phase silica or metal, such as gold or platinum. The configuration of the substrate can be in the form of beads, spheres, particles, granules, a gel, a membrane or a surface. Surfaces can be planar, substantially planar, or non-planar. Solid supports can be porous or non-porous, and can have swelling or non-swelling characteristics. A solid support can be configured in the form of a well, depression, or other container,

vessel, feature, or location. A plurality of supports can be configured on an array at various locations, addressable for robotic delivery of reagents, or by detection methods and/or instruments.

[0228] In some embodiments, immunological methods are used to purify transaminase variants. In one approach, antibody raised against a variant transaminase polypeptide (e.g., against a polypeptide comprising an engineered transaminase variant provided herein, including, but not limited to SEQ ID NO: 2, 4, 8, 366, and/or 650, and/or an immunogenic fragment thereof) using conventional methods is immobilized on beads, mixed with cell culture media under conditions in which the variant transaminase is bound, and precipitated. In a related approach, immunochromatography finds use.

[0229] In some embodiments, the variant transaminases are expressed as a fusion protein including a non-enzyme portion. In some embodiments, the variant transaminase sequence is fused to a purification facilitating domain. As used herein, the term "purification facilitating domain" refers to a domain that mediates purification of the polypeptide to which it is fused. Suitable purification domains include, but are not limited to metal chelating peptides, histidine-tryptophan modules that allow purification on immobilized metals, a sequence which binds glutathione (e.g., GST), a hemagglutinin (HA) tag (corresponding to an epitope derived from the influenza hemagglutinin protein; See e.g., Wilson et al., Cell 37:767 [1984]), maltose binding protein sequences, the FLAG epitope utilized in the FLAGS extension/affinity purification system (e.g., the system available from Immunex Corp), and the like. One expression vector contemplated for use in the compositions and methods described herein provides for expression of a fusion protein comprising a polypeptide of the disclosure fused to a polyhistidine region separated by an enterokinase cleavage site. The histidine residues facilitate purification on IMIAC (immobilized metal ion affinity chromatography; See e.g., Porath *et al.*, Prot. Exp. Purif., 3:263-281 [1992]) while the enterokinase cleavage site provides a means for separating the variant transaminase polypeptide from the fusion protein. pGEX vectors (Promega) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to ligand-agarose beads (e.g., glutathione-agarose in the case of GST-fusions) followed by elution in the presence of free ligand.

Methods of Using the Engineered Transaminase Enzymes

[0230] In another aspect, the engineered transaminase polypeptides disclosed herein can be used in a process for the conversion of the substrate compound **(1),** or structural analogs thereof, to the product of compound **(2)** or the corresponding structural analog.

[0231] 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 pH, temperature, buffer, solvent system, substrate loading, mixture of substrate compound stereoisomers, polypeptide loading, cofactor loading, pressure, and reaction time. Further suitable reaction conditions for carrying out the process for biocatalytic conversion of substrate compounds to product compounds using an engineered transaminase polypeptide described herein can be readily

optimized by routine experimentation that includes, but is not limited to, contacting the engineered transaminase polypeptide and substrate compound under experimental reaction conditions of concentration, pH, temperature, solvent conditions, and detecting the product compound, for example, using the methods described in the Examples provided herein.

[0232] 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%, 99%, or more identity to a reference amino acid sequence selected from any one of the even-numbered sequences of SEQ ID NO: 6-936, and an engineered transaminase polypeptide comprising an amino acid sequence that has (a) has one or more amino acid residue differences as compared to a reference sequence (e.g., SEQ ID NO: 2, 4, 8, 366, and/or 650). In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes a transaminase polypeptide that has the percent identity described above and one or more residue differences as compared to a reference sequence (e.g., SEQ ID NO: 2, 4, 8, 366, and/or 650).

[0233] 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 **(1)**, however it also contemplated that the equivalent molar amounts of various hydrates and salts of compound **(1)** also can be used in the process.

[0234] In the processes describes herein, the engineered transaminase polypeptide uses an amino donor to form the product compounds. In some embodiments, the amino donor in the reaction condition comprises a compound selected from isopropylamine (also referred to herein as "IPM") or any other suitable amino donor for the reaction of interest. In some embodiments, the amino donor is IPM. In some embodiments, the suitable reaction conditions comprise the amino donor, in particular IPM, present at a concentration of at least about 0.1 to about 3.0 M, 0.2 to about 2.5 M, about 0.5 to about 2 M or about 1 to about 2 M. In some embodiments, the amino donor is present at a concentration of about 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 1, 1.5, 2, 2.5 or 3 M.

[0235] Suitable reaction conditions for the processes also typically comprise the presence of a cofactor in the reaction mixture. Because the engineered transaminases typically use members of the vitamin B_6 family, the reaction condition can comprise a cofactor selected from-pyridoxal-5'-phosphate (also known

as pyridoxal-phosphate, PLP, P5P), pyridoxine (PN), pyridoxal (PL), pyridoxamine (PM), and their phosphorylated counterparts; pyridoxine phosphate (PNP), and pyridoxamine phosphate (PMP). In some embodiments, the suitable reaction conditions can comprise the presence of a cofactor selected from PLP, PN, PL, PM, PNP, and PMP, at a concentration of about 0.1 g/L to about 10 g/L, about 0.2 g/L to about 5 g/L, about 0.5 g/L to about 2.5 g/L. In some embodiments, the cofactor is PLP. 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.

[0236] 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.

[0237] During the course of the transamination reactions, the pH of the reaction mixture may change. The pH of the reaction mixture may be maintained at a desired pH or within a desired pH range. This may be done by the addition of an acid or a base, before and/or during the course of the reaction. Alternatively, the pH may be controlled by using a buffer. Accordingly, in some embodiments, the reaction condition comprises a buffer. Suitable buffers to maintain desired pH ranges are known in the art and include, by way of example and not limitation, borate, carbonate, phosphate, triethanolamine (TEA) buffer, and the like. In some embodiments, the buffer is TEA. In some embodiments of the process, the suitable reaction conditions comprise a buffer solution of TEA, where the TEA concentration is from about 0.01 to about 0.4 M, 0.05 to about 0.4 M, 0.1 to about 0.3 M, or about 0.1 to about 0.2 M. In some embodiments, the reaction condition comprises a TEA concentration of about 0.01, 0.02, 0.03, 0.04, 0.05, 0.07, 0.1, 0.12, 0.14, 0.16, 0.18, 0.2, 0.3, or 0.4 M. In some embodiments, the reaction conditions comprise water as a suitable solvent with no buffer present.

[0238] In some embodiments of the process, the reaction conditions can comprise a suitable pH. As noted above, the desired pH or desired pH range can be maintained by use of an acid or base, an appropriate buffer, or a combination of buffering and acid or base addition. The pH of the reaction mixture can be controlled before and/or during the course of the reaction. In some embodiments, the suitable reaction conditions comprise a solution pH of about 8 to about 12.5, a pH of about 8 to about 12, a pH of about 9.0 to about 11.5, or a pH of about 9.5 to about 11.0. In some embodiments, the reaction conditions comprise a solution pH of about 8, 8.5, 9, 9.5, 10, 10.5, 11, 11.5, 12 or 12.5.

[0239] In some embodiments of the processes herein, a suitable temperature can be used for the reaction conditions, for example, taking into consideration the increase in reaction rate at higher temperatures, the activity of the enzyme for sufficient duration of the reaction, and as further described below. For

example, the engineered polypeptides of the present disclosure have increased stability relative to naturally occurring transaminase polypeptide, which allows the engineered polypeptides of the present disclosure to be used at higher temperatures for increased conversion rates and improved substrate solubility characteristics for the reaction. Accordingly, in some embodiments, the suitable reaction conditions comprise a temperature of about 10°C to about 70°C, about 10°C to about 65°C, about 15°C to about 60°C, about 20°C to about 60°C, about 20°C to about 55°C, about 30°C to about 55°C, or about 40°C to about 50°C. In some embodiments, the suitable reaction conditions comprise a temperature of about 10°C, 15°C, 20°C, 25°C, 30°C, 35°C, 40°C, 45°C, 50°C, 55°C, 60°C, 65°C, or 70°C. In some embodiments, the temperature during the enzymatic reaction can be maintained at a temperature throughout the course of the reaction. In some embodiments, the temperature during the enzymatic reaction can be adjusted over a temperature profile during the course of the reaction.

[0240] In some embodiments of the process, the suitable reaction conditions can further comprise the presence of the reduced cofactor, nicotinamide adenine dinucleotide (NADH), which can act to limit the inactivation of the transaminase enzyme (See e.g., van Ophem *et al.*, Biochem., 37(9):2879-88 [1998]). In such embodiments where NADH is present, a cofactor regeneration system, such as glucose dehydrogenase (GDH) and glucose or formate dehydrogenase and formate can be used to regenerate the NADH in the reaction medium.

[0241] The processes using the engineered transaminases are generally carried out in a solvent. Suitable solvents include water, aqueous buffer solutions, organic solvents, and/or co-solvent systems, which generally comprise aqueous solvents and organic solvents. The aqueous solvent (water or aqueous co-solvent system) may be pH-buffered or unbuffered. In some embodiments, the processes using the engineered transaminase polypeptides are generally carried out in an aqueous co-solvent system comprising an organic solvent (e.g., ethanol, isopropanol (IPA), dimethyl sulfoxide (DMSO), ethyl acetate, butyl acetate, 1-octanol, heptane, octane, methyl t-butyl ether (MTBE), toluene, and the like), ionic liquids (e.g., 1-ethyl 4-methylimidazolium tetrafluoroborate, 1-butyl-3-methylimidazolium tetrafluoroborate, 1-butyl-3-methylimidazolium hexafluorophosphate, and the like). The organic solvent component of an aqueous co-solvent system may be miscible with the aqueous component, providing a single liquid phase, or may be partly miscible or immiscible with the aqueous component, providing two liquid phases. Exemplary aqueous co-solvent systems comprises water and one or more organic solvent. In general, an organic solvent component of an aqueous co-solvent system is selected such that it does not completely inactivate the transaminase enzyme. Appropriate co-solvent systems can be readily identified by measuring the enzymatic activity of the specified engineered transaminase enzyme with a defined substrate of interest in the candidate solvent system, utilizing an enzyme activity assay, such as those described herein. In some embodiments of the process, the suitable reaction conditions comprise an aqueous co-solvent comprising DMSO at a concentration of about 1% to about 80% (v/v), about 1 to about 70% (v/v), about 2% to about 60% (v/v), about 5% to about 40% (v/v), 10% to about 40% (v/v), 10% to about 30% (v/v), or about 10% to about 20% (v/v). In some embodiments of the process, the

suitable reaction conditions comprise an aqueous co-solvent comprising DMSO at a concentration of at least about 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, or 80% (v/v).

[0242] 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. For example, in some embodiments, the preparation of compound **(2)** can be carried out wherein the suitable reaction conditions comprise: (a) substrate loading of about 10 to 300 g/L of substrate compound (e.g., 50 g/L or 200 g/L of compound **(1)**); (b) of about 0.5 g/L to 60 g/L engineered polypeptide; (c) IPM concentration of about 0.5 to 2 M; (d) PLP cofactor concentration of about 0.1 to 1 g/L; (e) DMSO concentration of about 0% (v/v) to about 20% (v/v); (f) pH of about 8.5 to 11.5; and (g) temperature of about 45°C to 65°C.

[0243] In some embodiments, the suitable reaction conditions comprise: (a) about 100 g/L of substrate compound (e.g., compound **(1)**); (b) about 1 g/L engineered polypeptide; (c) about 1 M isopropylamine (IPM); (d) about 0.5 g/L pyridoxal phosphate (PLP); (e) about pH 9; and (g) about 50°C.

[0244] Further exemplary reaction conditions include the assay conditions provided in the Examples. In carrying out the transamination reactions described herein, the engineered transaminase polypeptide may be added to the reaction mixture in the partially purified or purified enzyme, whole cells transformed with gene(s) encoding the enzyme, and/or as cell extracts and/or lysates of such cells. Whole cells transformed with gene(s) encoding the engineered transaminase enzyme or cell extracts, lysates thereof, and isolated enzymes may be employed in a variety of different forms, including solid (e.g., lyophilized, spray-dried, and the like) or semisolid (e.g., a crude paste). The cell extracts or cell lysates may be partially purified by precipitation (e.g., ammonium sulfate, polyethyleneimine, heat treatment or the like), followed by a desalting procedure (e.g., ultrafiltration, dialysis, and the like) prior to lyophilization. Any of the enzyme preparations may be stabilized by crosslinking using known crosslinking agents, such as, for example, glutaraldehyde, or immobilized to a solid phase material (e.g., resins, beads such as chitosan, Eupergit C, SEPABEADs, and the like).

[0245] 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.

[0246] In some embodiments where the engineered polypeptide can be expressed in the form of a secreted polypeptide, the culture medium containing the secreted polypeptides can be used in the process herein.

[0247] In some embodiments, solid reactants (e.g., enzyme, salts, etc.) may be provided to the reaction in a variety of different forms, including powder (e.g., lyophilized, spray dried, and the like), solution, emulsion, suspension, and the like. The reactants can be readily lyophilized or spray dried using methods and equipment that are known to those having ordinary skill in the art. For example, the protein solution can be frozen at -80°C in small aliquots, then added to a pre-chilled lyophilization chamber, followed by the application of a vacuum.

[0248] In some embodiments, the order of addition of reactants is not critical. The reactants may be added together at the same time to a solvent (e.g., monophasic solvent, biphasic aqueous co-solvent system, and the like), or alternatively, some of the reactants may be added separately, and some together at different time points. For example, the cofactor, transaminase, and transaminase substrate may be added first to the solvent. For improved mixing efficiency when an aqueous co-solvent system is used, the transaminase, and cofactor may be added and mixed into the aqueous phase first. The organic phase may then be added and mixed in, followed by addition of the transaminase substrate. Alternatively, the transaminase substrate may be premixed in the organic phase, prior to addition to the aqueous phase.

[0249] In some embodiments, where the choice of the amino donor results in a carbonyl by-product that has a vapor pressure higher than water (e.g., a low boiling co-product such as a volatile organic carbonyl compound), the carbonyl by-product can be removed by sparging the reaction solution with a non-reactive gas or by applying a vacuum to lower the reaction pressure and removing the carbonyl by-product present in the gas phase. A non-reactive gas is any gas that does not react with the reaction components. Various non-reactive gases include nitrogen and noble gases (e.g., inert gases). In some embodiments, the non-reactive gas is nitrogen gas. In some embodiments, the amino donor used in the process is isopropylamine (IPM), which forms the carbonyl by-product acetone upon transfer of the amino group to the amino group acceptor. The acetone can be removed by sparging with nitrogen gas or applying a vacuum to the reaction solution and removing the acetone from the gas phase by an acetone trap, such as a condenser or other cold trap. Alternatively, the acetone can be removed by reduction to isopropanol using a ketoreductase.

[0250] In some embodiments of the process where the carbonyl by-product is removed, the corresponding amino group donor can be added during the transamination reaction to replenish the amino group donor and/or maintain the pH of the reaction. Replenishing the amino group donor also shifts the equilibrium towards product formation, thereby increasing the conversion of substrate to product. Thus, in some embodiments where the amino group donor is IPM and the acetone product is removed *in situ*, the process can further comprise a step of adding IPM to the reaction solution to replenish the amino group donor lost during the acetone removal and to maintain the pH of the reaction (e.g., at about 8.5 to about pH 11.5).

[0251] 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.

[0252] In some embodiments, the processes using the engineered polypeptides disclosed herein can be carried out wherein the amino group donor is selected from isopropylamine, alanine, 3-aminobutyric acid, or methylbenzylamine. In some embodiments, the amino group donor is isopropylamine.

[0253] 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.

[0254] Various features and embodiments of the invention are illustrated in the following representative examples, which are intended to be illustrative, and not limiting.

EXPERIMENTAL

[0255] Various features and embodiments of the invention are illustrated in the following representative examples, which are intended to be illustrative, and not limiting.

[0256] In the experimental disclosure below, the following abbreviations apply: ppm (parts per million); M (molar); mM (millimolar), uM and μM (micromolar); nM (nanomolar); mol (moles); gm and g (gram); mg (milligrams); ug and μg (micrograms); L and l (liter); ml and mL (milliliter); cm (centimeters); mm (millimeters); um and μm (micrometers); sec. (seconds); min(s) (minute(s)); h(s) and hr(s) (hour(s)); U (units); MW (molecular weight); rpm (rotations per minute); °C (degrees Centigrade); RT (room temperature); CDS (coding sequence); DNA (deoxyribonucleic acid); RNA (ribonucleic acid); aa (amino acid); TB (Terrific Broth; 12 g/L bacto-tryptone, 24 g/L yeast extract, 4 mL/L glycerol, 65 mM potassium phosphate, pH 7.0, 1 mM $MgSO_4$); LB (Luria broth); CAM (chloramphenicol); PMBS (polymyxin B sulfate); IPTG (isopropyl thiogalactoside); ATA (omega-transaminase); TFA (trifluoroacetic acid); TEoA (triethanolamine); borate (sodium tetraborate decahydrate); acetonitrile (MeCN); dimethylsulfoxide (DMSO); HPLC (high performance liquid chromatography); FIOP (fold improvement over positive control); HTP (high throughput); MWD (multiple wavelength detector); UV (ultraviolet); Codexis (Codexis, Inc., Redwood City, CA); Sigma-Aldrich (Sigma-Aldrich, St. Louis, MO); Millipore (Millipore, Corp., Billerica MA); Difco (Difco Laboratories, BD Diagnostic Systems, Detroit, MI); Daicel (Daicel, West Chester, PA); Genetix (Genetix USA, Inc., Beaverton, OR); Molecular Devices (Molecular Devices, LLC, Sunnyvale, CA); Applied Biosystems (Applied Biosystems, part of Life Technologies, Corp., Grand Island, NY), Agilent (Agilent Technologies, Inc., Santa Clara, CA); Thermo Scientific (part of Thermo Fisher Scientific, Waltham, MA); (Infors; Infors-

HT, Bottmingen/Basel, Switzerland); Corning (Corning, Inc., Palo Alto, CA); and Bio-Rad (Bio-Rad Laboratories, Hercules, CA); Microfluidics (Microfluidics Corp., Newton, MA); Waters (Waters Corp., Milford, MA).

EXAMPLE 1
Production of Engineered Polypeptides in pCK110900

[0257] The polynucleotide (SEQ ID NO: 1) encoding the polypeptide from *Vibrio fluvialis* having transaminase activity (SEQ ID NO: 2) was cloned into a pCK110900 vector system (See e.g., US Pat. No. 9,714,437, which is hereby incorporated by reference in its entirety) and subsequently expressed in *E. coli* W3110*fhu*A under the control of the *lac* promoter.

[0258] In a 96-well format, single colonies were picked and grown in 190 µL LB media containing 1% glucose and 30 µg/mL chloramphenicol (CAM), at 30°C, 200 rpm, and 85% humidity. Following overnight growth, 20 µL of the grown cultures were transferred into a deep-well plate containing 380 µL of TB with 30 µg/mL CAM. The cultures were grown at 30°C, 250 rpm, with 85% humidity for approximately 2.5 hours. When the optical density (OD_{600}) of the cultures reached 0.4 - 0.6, expression of the transaminase gene was induced by addition of IPTG to a final concentration of 1 mM. Following induction, growth was continued for 18-20 hours at 30°C, 250 rpm with 85% humidity. Cells were harvested by centrifugation at 4000 rpm at 4°C for 10 minutes, and the media was discarded. The cell pellets were stored at -80°C until ready for use. Prior to performing the assay, cell pellets were resuspended in 400 µL of lysis buffer containing 50 mM triethanolamine-HCl, pH 7.5, with 1 g/L PLP, 1 g/L lysozyme, and 0.5 g/L PMBS. The plates were agitated with medium-speed shaking for 2 hours on a microtiter plate shaker at room temperature. The plates were then centrifuged at 4000 rpm for 15 minutes at 4°C, and the clarified supernatants were used in the HTP assay reaction described below.

[0259] Shake-flask procedures can be used to generate engineered transaminase polypeptide shake-flask powders (SFP), which are useful for secondary screening assays and/or use in the biocatalytic processes described herein. Shake flask powder preparation of enzymes provides a more purified preparation (e.g., up to 30% of total protein) of the engineered enzyme, as compared to the cell lysate used in HTP assays and also allows for the use of more concentrated enzyme solutions. To start the cultures, a single colony of *E. coli* containing a plasmid encoding an engineered polypeptide of interest was inoculated into 5 mL LB with 30 µg/mL CAM and 1% glucose. The culture was grown overnight (at least 16 hours) in an incubator at 37°C, with shaking at 250 rpm. The grown culture was then added to 250 mL of TB with 30 µg/mL CAM, in a 1L shake-flask. The 250 mL culture was grown at 30°C at 250 rpm for 3.5 hours until the OD_{600} reached 0.6-0.8. Expression of the transaminase gene was induced by addition of IPTG to a final concentration of 1 mM, and growth was continued for an additional 18-20 hours. Cells were harvested by transferring the culture into a pre-weighed centrifuge bottle, then centrifuged at 4000 rpm for 20 minutes at 4°C. The supernatant was discarded, and the remaining cell pellet was weighed. In

some embodiments, the cells were stored at -80°C until ready to use. For lysis, the cell pellet was resuspended in 6 mL/g wet cell weight of 25 mM triethanolamine-HCl buffer, pH 7.5 and lysed using a 110L MICROFLUIDIZER® processor system (Microfluidics). Cell debris was removed by centrifugation at 10,000 rpm for 60 minutes at 4°C. The clarified lysate was collected, frozen at -80°C, and then lyophilized, using standard methods known in the art. Lyophilization of frozen clarified lysate provides a dry shake-flask powder comprising crude engineered polypeptide.

EXAMPLE 2

Evolution and Screening of Engineered Polypeptides Derived from SEQ ID NO: 4 for Improved Production of Compound (2)

[0260] The engineered polynucleotide (SEQ ID NO: 3) encoding the polypeptide with transaminase activity of SEQ ID NO: 4 was used to generate the engineered polypeptides of Table 2-1. These polypeptides displayed improved transaminase activity under the desired conditions e.g., the improvement in the formation of the amine product, compound (2), from the ketone substrate, compound (1), as compared to the starting polypeptide. The engineered polypeptides, having the amino acid sequences of even-numbered sequence identifiers were generated from the "backbone" amino acid sequence of SEQ ID NO: 4 as described below, together with the HTP assay and analytical methods described in Tables 6-1 and 6-2.

[0261] Directed evolution began with the polynucleotide set forth in SEQ ID NO: 3. Libraries of engineered polypeptides were generated using various well-known techniques (e.g., saturation mutagenesis, recombination of previously identified beneficial amino acid differences) and screened using HTP assay and analysis methods that measured the polypeptides' ability to produce compound **(2)**.

[0262] The enzyme assay was carried out in 96-well shallow-well (300 µL volume/well) plates, in 100 µL total reaction volume/well. The reaction contained 55 µL of 64-fold diluted ATA lysate (diluted in 50 mM triethanolamine + 1 g/L PLP, pH 7.5), 25 µL of 4 M isopropylamine-HCl, pH 9, and 20 µL of 100 g/L ketone (1) dissolved in DMSO. The reaction plate was heat-sealed and shaken at 600 rpm at 50°C for 22 hours.

[0263] After overnight incubation (~22 hours), 100 µL/well of 50% formic acid in acetonitrile was added to the reaction plate and mixed well. The plates were sealed and centrifuged at 4000 rpm for 10 min. A 20 µL/well aliquot was removed from the quenched plate and diluted into 180 µL of 1:1 acetonitrile:water and analyzed by HPLC to determine activity and selectivity as described in Tables 6-1 and 6-2, respectively.

[0264] Hit variants were grown in 250 mL shake flasks as described in Example 1 to generate lyophilized enzyme powders. The activity of the enzyme powders were evaluated at 0-10 g/L of the ATA shake flask powder, 20-100 g/L ketone, 1 M IPM (4 M pH 9 stock), 0-40% cosolvent (DMSO or

methanol), pH ~ 9, 50°C, for 20-24 hours, using a similar assay as described above and the hit variants are listed in Table 2-2.

Table 2-1. Transaminase Activity in the Production of Compound (2) Relative to SEQ ID NO: 4

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4)	Percent Conversion Fold Improvement (Relative to SEQ ID NO: 4)[1]
7/8	K163F;N286S;I314R;E316W;A323T;C414V;P416A	+++
9/10	N286S;I314R;A323T;A383V	+++
11/12	K163F;N286S;I314R;A323T;D394G	+++
13/14	N286S;E316W;A323T;A383V;C414V;P416A	+++
15/16	A74T;G81S;N286S;E316W;A323T;A383V	+++
17/18	A74T;N286S;E316W;A323T	+++
5/6	T408F	+++
19/20	E315G	+++
21/22	N286S;E316W;D394G;C414V;P416A	+++
23/24	N286S;A323T;A383V;C414V	+++
25/26	K163F;A222V;N286S;E316W;A323T;A383V;D394G	+++
27/28	N286S;A323T	++
29/30	K163F;N286S;C414V	++
31/32	N286S;A323T;P416A	++
33/34	N286S	++
35/36	N286S;E316W;A383V;D394G	++
37/38	A74T;K163F;I314R;A323T;A383V;C414V;P416A	++
39/40	N286S;I314R;D394G	++
41/42	I314R;A323T;A383V;D394G	++
43/44	K163F;I314R;C414V	++
45/46	K163F;N286S	++
47/48	N286S;A383V	++
49/50	K163F;I314R;A383V	++
51/52	A74T;I314R;A323T;A383V;D394G;C414V	++
53/54	K163F;I314R;A323T;D394G	++
55/56	K163F;N286S;E316W;A323T;D394G;P416A	++
57/58	K163F;A323T	++
59/60	A323S	++
61/62	L417S	++
63/64	I314R;E316W;A323T	++
65/66	T408W	++
67/68	A323C	++
69/70	A323T	++
71/72	A74T;K163F;I314R;E316W;A323T;D394G	++
73/74	F85V;N286S;A323T	++

Table 2-1. Transaminase Activity in the Production of Compound (2) Relative to SEQ ID NO: 4		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 4)**	**Percent Conversion Fold Improvement (Relative to SEQ ID NO: 4)[1]**
75/76	A323T;D394G	++
77/78	E316F	++
79/80	I314R;E316W;A323T;D394G	++
81/82	A323T;A383V	+
83/84	N286S;P416A	+
85/86	I259V	+
87/88	A74T;K163F;N286S;E316W;A383V;D394G;P416A	+
89/90	A74T;N286S	+
91/92	F85V;S86A;K163F;I314R;A323T;D394G;C414V	+
93/94	H88R;N286S;E316W;A323T;A383V;C414V;P416A	+
95/96	I314R;E316W;D394G	+
97/98	H88R;K163F;N286S;A383V	+
99/100	I314R;E316W	+
101/102	A74T;K163F;I314R;E316W	+
103/104	L56A;N286S;I314R;E316W;A323T;V422A	+
105/106	I314R;E316W;A323T;A383V;V422A	+
107/108	D21H	+
109/110	A323T;A383V;D394G;C414V;P416A	+
111/112	L56A;N286S;A383V	+
113/114	K163L	+
115/116	P23S;L417S	+
117/118	K163F;A383V	+
119/120	E316G	+
121/122	V422T	+
123/124	G395D;L417S	+
125/126	E316N	+
127/128	R146H;L417S	+
129/130	K163M	+
131/132	K163F	+
133/134	A74T;H88R;N286S;E316W;A323T;A383V	+
135/136	P164Q	+
137/138	G395D;L417S;A432V	+
139/140	L56A;A74T;A241V;N286S;I314R;E316W;A323T	+
141/142	E316H	+
143/144	L56A;S86A;K163F;I314R;E316W;A383V;C414V;P416A;V422A	+
145/146	L56A;A323T;A383V	+
147/148	G395D	+
149/150	E316W;A323T;D394G;C414V;P416A	+
151/152	K163F;I314R;E316W;D394G	+

Table 2-1. Transaminase Activity in the Production of Compound (2) Relative to SEQ ID NO: 4		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 4)**	**Percent Conversion Fold Improvement (Relative to SEQ ID NO: 4)[1]**
153/154	E401A	+
155/156	L56C	+
157/158	A74T;S86A;E316W;A323T;A383V;D394G	+
159/160	A383V	+
161/162	E316R;	+
163/164	N53C;R146H;L417S	+
165/166	I314R;A383V;C414V;V422A	+
167/168	A74T;E316W;A323T;D394G	+
169/170	L56A;K163F	+
171/172	L56A;S86A;I314R;E316W;A323T;D394G;C414V;V422A	+
173/174	L56A;A383V	+
175/176	S86A;A383V;D394G	+
177/178	L56A;A323T;A383V;D394G	+
179/180	D21P;N53C;L417S	+
181/182	F85V;S86A;H88R;K163F;A323T;A383V;D394G	+
183/184	E401S	+
185/186	R146H;G395D;L417S	+
187/188	E316W;A323T;A383V;D394G	+
189/190	E316W;C414V;V422A	+
191/192	G18A	+
193/194	A74T;S86A;K163F;E316W	+
195/196	A74T;F85V;I314R;E316W;C414V;P416A	+
197/198	H88R;E316W;A323T	+
199/200	P23R	+
201/202	S86A;H88R;K163F;A323T;A383V;C414V;V422A	+
203/204	L56C;R146H;A432V	+
205/206	L56T	+
207/208	S86G	+
209/210	A74T;F85V;S86A;K163F;N286S;E316W;A323T;D394G	+
211/212	H88R;N286S;E316W;A323T	+
213/214	A199V;L417S	+
215/216	L56A;K163F;N286S;E316W;A323T;A383V;D394G	+
217/218	P23S;N53C;L417S	+
219/220	S284A	+
221/222	A74T;H88R;A323T;A383V	+
223/224	P164D	+
225/226	A149S	+
227/228	C260T	+

Table 2-1. Transaminase Activity in the Production of Compound (2) Relative to SEQ ID NO: 4		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 4)**	**Percent Conversion Fold Improvement (Relative to SEQ ID NO: 4)[1]**
229/230	K163F;E316W;A323T	+
231/232	F85V;N286S	+
233/234	P164S	+
235/236	L56V	+
237/238	L56C;G395D	+
239/240	A74T;N286S;D394G;P416A	+
241/242	I314R;A383V	+
243/244	A404S	+
245/246	L56A;A323T	+
247/248	L56A;S86A;N286S;I314R;C414V;P416A	+
249/250	N53C;L56C	+
251/252	P23S;G395D;L417S	+
253/254	S157A	+
255/256	P23S;L56C	+
257/258	H88T	+
259/260	I403V	+
261/262	P23A	+
263/264	L417S;A432V	+
265/266	N53C	+
267/268	D21P;P23S;L56C;R146H	+
269/270	P23S;N53C;L56C;R146H;G395D	+
271/272	D21P;P23S;G395D;L417S;A432V	+
273/274	D21P;N53C;L56C	+
275/276	W147R;G395D;L417S;A432V	+
277/278	H88S	+
279/280	V422L	+
281/282	D21P;P23S;R146H;L417V	+
283/284	E315R	+
285/286	D21P;L56C;G395D	+
287/288	P23S;N53C	+
289/290	L417V;A432V	+
291/292	L56C;W147R	+
293/294	E316L	+
295/296	D21P;P23S;L56C;R146H;A432V	+
297/298	L56A	+
299/300	P23S;L56C;G395D	+
301/302	E316A	+
303/304	L417A	+
305/306	T408L	+
307/308	N53C;G395D	+

Table 2-1. Transaminase Activity in the Production of Compound (2) Relative to SEQ ID NO: 4		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 4)**	**Percent Conversion Fold Improvement (Relative to SEQ ID NO: 4)[1]**
309/310	P23S;N53C;G395D	+
311/312	T20C	+
313/314	P23S;N53C;A432V	+
315/316	E401K	+
317/318	L56C;R146H;L417V	+
319/320	P23S;N53C;L56C	+
321/322	L417V	+
323/324	R415G	+
325/326	D21R	+
327/328	D21P;L417S;A432V	+
329/330	P164A	+
331/332	N405W	+
333/334	W147R;L417S	+
335/336	R146H;L417V	+
337/338	R146H	+
339/340	T406S	+
341/342	G395D;L417V	+
343/344	N405H	+
345/346	R415W	+
347/348	S400D	+
349/350	P23S;L417V;A432V	+
351/352	R146H;G395D	+
353/354	S420G	+
355/356	E316V	+
357/358	R146H;W147R;G395D;L417S	+
359/360	P23S;L56C;G395D;L417V	+
361/362	P23S;L417V	+
[1]Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 4 and defined as follows: "+" 1.20 to 3.00, "++" > 3.00, "+++" > 5.00		

Table 2-2. ATA Activity in the Production of Compound (2) Relative to SEQ ID NO: 4		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 4)**	**Percent Conversion Fold Improvement (Relative to SEQ ID NO: 4)[1]**
7/8	K163F;N286S;I314R;E316W;A323T;C414V;P416A	+++
9/10	N286S;I314R;A323T;A383V	++

Table 2-2. ATA Activity in the Production of Compound (2) Relative to SEQ ID NO: 4		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 4)**	**Percent Conversion Fold Improvement (Relative to SEQ ID NO: 4)[1]**
11/12	K163F;N286S;I314R;A323T;D394G	++
13/14	N286S;E316W;A323T;A383V;C414V;P416A	++
15/16	A74T;G81S;N286S;E316W;A323T;A383V	+
5/6	T408F	+
19/20	E315G	+
[1]Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 4 and defined as follows: "+" 5.00 to 7.00, "++" > 7.00, "+++" > 8.00		

EXAMPLE 3
Evolution and Screening of Engineered Polypeptides Derived from SEQ ID NO: 8 for Improved Production of Compound (2)

[0265] The engineered polynucleotide (SEQ ID NO: 7) encoding the polypeptide with transaminase activity of SEQ ID NO: 8 was used to generate the engineered polypeptides of Table 3-1. These polypeptides displayed improved transaminase activity under the desired conditions e.g., the improvement in the formation of the amine product, compound **(2)**, from the ketone substrate, compound **(1)**, as compared to the starting polypeptide. The engineered polypeptides, having the amino acid sequences of even-numbered sequence identifiers were generated from the "backbone" amino acid sequence of SEQ ID NO: 8 as described below.

[0266] Directed evolution began with the polynucleotide set forth in SEQ ID NO: 7. Libraries of engineered polypeptides were generated using various well-known techniques (e.g., saturation mutagenesis, recombination of previously identified beneficial amino acid differences) and screened using HTP assay and analysis methods that measured the polypeptides' ability to produce compound **(2)**.

[0267] The enzyme assay and analysis were carried out as described in Example 2 except that the lysate was diluted 512-fold before adding to the reaction mixture.

[0268] Hit variants were grown in 250 mL shake flasks and enzyme powders generated as described in Example 1. The activity of the enzyme powders was evaluated at 0-32 g/L of the ATA shake flask powder, 20-200 g/L ketone, 1-2.5 M IPM (4 M pH 9 stock), 0.5 g/L PLP, 0-20% DMSO, pH ~ 9, 40-60°C, for 20-24 hours, using similar assay as described above and the hit variants are listed in Table 3-2.

Table 3-1. ATA Activity in the Production of Compound (2) Relative to SEQ ID NO: 8		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 8)**	**Percent Conversion Fold Improvement (Relative to SEQ ID NO: 8)[1]**
363/364	T408A	+++
365/366	D21H;F163L;T323C;T408F	+++
367/368	F291Y	+++
369/370	R351L	+++
371/372	V272E	+++
373/374	T408E	++
375/376	A383V	++
377/378	E365S	++
379/380	H362Q	++
381/382	N342T	++
383/384	S24K	++
385/386	E189V	++
387/388	V42F;I363L	++
389/390	G191D	++
391/392	D439S	++
393/394	A388L	++
395/396	K447S	++
397/398	V42F;V272E;F291Y	++
399/400	K66A	+
401/402	E365L	+
403/404	A388D	+
405/406	E77M	+
407/408	K443S	+
409/410	E365R	+
411/412	D107L	+
413/414	G18A;P23R;A149S;A383V	+
415/416	A199Q	+
417/418	T277S	+
419/420	G46S	+
421/422	E195W	+
423/424	V42F;V272E;F291Y;I363L	+
425/426	L417S	+
427/428	V42F;V272E;I363L;L410H	+
429/430	V42F;V272E;G324S;I363L;R366H	+
431/432	T309F	+
433/434	K385L	+
435/436	Y187E;F291Y	+
437/438	Y187E	+
439/440	E189W	+

Table 3-1. ATA Activity in the Production of Compound (2) Relative to SEQ ID NO: 8		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 8)**	**Percent Conversion Fold Improvement (Relative to SEQ ID NO: 8)[1]**
441/442	K385T	+
443/444	E343G	+
445/446	E365Q	+
447/448	Q210V	+
449/450	A134V	+
451/452	C260T;G395D;E401S	+
453/454	Q5G	+
455/456	A388P	+
457/458	D107S	+
459/460	G18A;P23R;A149S;C260T;A383V;G395D;E401S;A416P	+
461/462	Q5E	+
463/464	S86G;A383V	+
465/466	E358L	+
467/468	E451S	+
469/470	E189S	+
471/472	E189F	+
473/474	R203L	+
475/476	H362V	+
477/478	N396P	+
479/480	K443L	+
481/482	K248G	+
483/484	Q210A	+
485/486	P23R;A149S;S284A;A383V;G395D	+
487/488	S167N	+
489/490	P138R	+
491/492	G395R	+
493/494	P392L	+
495/496	Q210M	+
497/498	A416P	+
499/500	S389D;	+
501/502	F163L;I259V;T323C;T408F	+
503/504	K361R	+
505/506	V42F	+
507/508	D107Y	+
509/510	P354S	+
511/512	F163L;E315G;W316F	+
513/514	P367T	+
515/516	N396Y	+
517/518	K447T	+
519/520	Q210Y	+

Table 3-1. ATA Activity in the Production of Compound (2) Relative to SEQ ID NO: 8		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 8)**	**Percent Conversion Fold Improvement (Relative to SEQ ID NO: 8)[1]**
521/522	K211R	+
523/524	S24R	+
525/526	I363L;R366H	+
527/528	A404M	+
529/530	D439L	+
531/532	T309A	+
533/534	V42F;Y187E;A353T	+
535/536	V42F;V272E;L410H	+
537/538	Y187E;V272E;G324S;I363L;L410H	+
539/540	N405W	+
541/542	G191F	+
543/544	K305E	+
545/546	V272E;A353T	+
547/548	E401Q	+
549/550	Q210L	+
551/552	R186Q	+
553/554	A450D	+
555/556	T309R	+
557/558	P392A	+
559/560	Y187E;V272E;I363L	+
561/562	S86G;F163M;P164S;C260T;A383V	+
563/564	P164D;W316H;A383V;E401S	+
565/566	V42F;F291Y;I363L	+
567/568	G18A;F163M;P164Q	+
569/570	T408W	+
571/572	P23R;S86G	+
573/574	L410H	+
575/576	V272E;L410H	+
577/578	R110K;Y187E	+
579/580	P23R;L56C;S86G;A149S;F163M;P164D;A383V;E401S;A416P	+
581/582	D21H;F163L;E315G;W316F	+
583/584	F163L;I259V;T408F	+
585/586	A149S;A416P	+
587/588	V42F;F291Y;A313V;I363L;L410H	+
589/590	V42F;A353T	+
591/592	I259V;L307M	+
593/594	I363L	+
595/596	P23R;A149S;C260T	+
597/598	S86G;A149S;F163M;P164S;A383V;G395D;E401S	+

Table 3-1. ATA Activity in the Production of Compound (2) Relative to SEQ ID NO: 8		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 8)**	**Percent Conversion Fold Improvement (Relative to SEQ ID NO: 8)[1]**
599/600	A383V;A416P;V422T	+
601/602	V42F;R110K	+
603/604	V42F;F291Y;I363L;R366H	+
605/606	P164S;C260T;E401S	+
607/608	Y187E;V272E;I363L;R366H;L410H	+
609/610	D21H	+
611/612	R110K	+
613/614	A149S;C260T;A383V	+
615/616	S86G;A149S;G395D	+
617/618	V272E;I363L;R366H	+
619/620	V42F;Y187E;G324S;I363L;R366H	+
621/622	R110K;Y187E;V253L;L410H	+
623/624	E315G	+
625/626	P23R;F163M;P164Q;A383V	+
627/628	V42F;Y187E;V272E	+
629/630	P23R;F163M;P164S;E401S;A416P	+
631/632	D21H;T408F	+
633/634	E401S	+
635/636	A383V;E401A	+
637/638	A149S;P164S;C260T;A383V;G395D;E401A	+
639/640	Y187E;V253L;I363L;R366H	+
[1]Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 8 and defined as follows: "+" 1.20 to 1.75, "++" > 1.75, "+++" > 2.00		

Table 3-2. ATA Activity in the Production of Compound (2) Relative to SEQ ID NO: 8		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 8)**	**Percent Conversion Fold Improvement (Relative to SEQ ID NO: 8)[1]**
367/368	F291Y	++
371/372	V272E	+
365/366	D21H;F163L;T323C;T408F	++
375/376	A383V	+
413/414	G18A;P23R;A149S;A383V	+
[1]Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 8 and defined as follows: "+" 1.50 to 2.00, "++" > 2.00		

EXAMPLE 4

Evolution and Screening of Engineered Polypeptides Derived from SEQ ID NO: 366 for improved production of Compound (2)

[0269] The engineered polynucleotide (SEQ ID NO: 365) encoding the polypeptide with transaminase activity of SEQ ID NO: 366 was used to generate the engineered polypeptides of Table 4-1. These polypeptides displayed improved transaminase activity under the desired conditions e.g., the improvement in the formation of the compound **(2)** (amine product) from the ketone substrate (compound **(1)**) as compared to the starting polypeptide. The engineered polypeptides, having the amino acid sequences of even-numbered sequence identifiers were generated from the "backbone" amino acid sequence of SEQ ID NO: 366 as described below together with the HTP assay and analytical methods described in Tables 6-1 and 6-2.

[0270] The enzyme assay and analysis were carried out as described in Example 2 except that the lysate was diluted 1700-fold before adding to the reaction mixture.

[0271] Hit variants were grown in 250 mL shake flasks and enzyme powders generated as described in Example 1. The activity of the enzyme powders was evaluated at 0-60 g/L of the ATA shake flask powder, 20-300 g/L ketone, 1 M IPM (4 M pH 9 stock), 0.5 g/L PLP, 0-20% DMSO, pH ~ 9, 40-60°C, for 20-24 hours, using similar assay as described above and the hit variants are listed in Table 4-2.

Table 4-1. ATA Activity in the Production of Compound (2) Relative to SEQ ID NO: 366

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 366)	Percent Conversion Fold Improvement (Relative to SEQ ID NO: 366)[1]
641/642	S24K;K66A;F291Y;E365S	+++
643/644	S24K;K66A;G191D;A199Q;F291Y	+++
645/646	G191D;F291Y;N342T;H362Q;E365S	++
647/648	F291Y;A383V	++
649/650	L163M;F291Y;A383V;A388D	++
651/652	S24K	++
653/654	L163M;F291Y;H362Q;E365S;A383V;A388D	++
655/656	F291Y	++
657/658	V42F;F291Y;R351L;H362Q;A383V;F408A	++
659/660	S24K;F291Y	+
661/662	S24K;K66A;N342T;E365S;A388D;F408E	++
663/664	K66A;A383V	++
665/666	K66A;F291Y;H362Q;E365S;A383V	++
667/668	S24K;K66A;F291Y;N342T;A383V	++
669/670	V42F;F291Y;A383V;A388D	++

Table 4-1. ATA Activity in the Production of Compound (2) Relative to SEQ ID NO: 366		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 366)**	**Percent Conversion Fold Improvement (Relative to SEQ ID NO: 366)[1]**
671/672	V42F;G191D;F408E	++
673/674	S24K;A199Q;C260T;R351L;H362Q;A383V	++
675/676	S24K;K66A;L163M;G191D;H362Q;A383V;A388D	++
677/678	S24K;F291Y;N342T;R351L;A383V	++
679/680	K66A;F291Y;A383V;A388D	++
681/682	A199Q;F291Y	++
683/684	S24K;A388D	++
685/686	A199Q;C260T;A383V	++
687/688	H362Q	++
689/690	L163M;A383V	+
691/692	S24K;V42F;K66A;F291Y	+
693/694	F291Y;R351L;A383V;A388D;F408A	+
695/696	V42F;A199Q;F291Y;A383V	+
697/698	S24K;A199Q;C260T;H362Q;A383V;A388D	+
699/700	S24K;L163M;R351L;A383V	+
701/702	S24K;C260T;H362Q;A383V;A388D	+
703/704	A383V	+
705/706	S24K;D107L;L163M;G191D;F291Y;R351L;A383V;A388D	+
707/708	F291Y;R351L;A383V;A388D	+
709/710	C260T;F291Y;E365S;A383V;F408A	+
711/712	G191D;C260T;A388D	+
713/714	L25V	+
715/716	V33T	+
717/718	R351L;A383V;A388D	+
719/720	G191D;R351L;A383V;A388D	+
721/722	N148G	+
723/724	S24E	+
725/726	E315S	+
727/728	S24K;K66A;C260T;F291Y;A383V;A388D;F408A	+
729/730	L397M	+
731/732	N405A	+
733/734	Q419S	+
735/736	D107L;G191D;F291Y;A383V	+
737/738	L423V	+
739/740	G191D;A199Q;E365S;A383V;A388D	+
741/742	N342T;H362Q	+
743/744	K66A;L163M;G191D;E365S;A383V	+
745/746	R28S	+
747/748	D107L;G191D;A199Q;E365S;A383V;A388D	+
749/750	S24K;E77M;F291Y	+

Table 4-1. ATA Activity in the Production of Compound (2) Relative to SEQ ID NO: 366		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 366)**	**Percent Conversion Fold Improvement (Relative to SEQ ID NO: 366)[1]**
751/752	K66A;F291Y	+
753/754	S24K;A383V;A388D	+
755/756	H362Q;A388D	+
757/758	S24K;K66A;G191D;C260T;F408A	+
759/760	K66A;A199Q;R351L;A383V	+
761/762	S24K;K66A;G191D;A199Q;C260T;F291Y;R351L	+
763/764	S24K;G191D;F291Y;E365S	+
765/766	R314K	+
767/768	S24K;D107L;F291Y;R351L;E365S;A388D	+
769/770	G191D;F291Y	+
771/772	S24K;F291Y;F408A	+
773/774	F291Y;E365S;A388D	+
775/776	E77M;A383V;A388D	+
777/778	L25H	+
779/780	S24K;V42F;F291Y;H362Q	+
781/782	F291Y;H362Q;E365S	+
783/784	H319S	+
785/786	V42F;F291Y;R351L;H362Q;E365S;A383V;A388D	+
787/788	C260T	+
789/790	C260T;E365S;A383V	+
791/792	E77M;F291Y	+
793/794	A383V;A388D	+
795/796	K66A;Y82H;F291Y;A383V	+
797/798	S24K;F291Y;H362Q;A388D	+
799/800	R28T	+
801/802	A153S	+
803/804	T406H	+
805/806	S86T	+
807/808	W316V	+
809/810	I413L	+
811/812	N396R	+
[1]Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 366 and defined as follows: "+" 1.10 to 1.50, "++" > 1.50, "+++" > 1.80		

Table 4-2. ATA Activity in the Production of Compound (2) Relative to SEQ ID NO: 366		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 366)**	**Percent Conversion Fold Improvement (Relative to SEQ ID NO: 366)[1]**
645/646	G191D;F291Y;N342T;H362Q;E365S	+
647/648	F291Y;A383V	+
649/650	L163M;F291Y;A383V;A388D	+
653/654	L163M;F291Y;H362Q;E365S;A383V;A388D	+
655/656	F291Y	+
641/642	S24K;K66A;F291Y;E365S	++
643/644	S24K;K66A;G191D;A199Q;F291Y	+
[1]Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 366 and defined as follows: "+" 1.70 to 2.00 "++" > 2.00		

EXAMPLE 5

Evolution and Screening of Engineered Polypeptides Derived from SEQ ID NO: 650 for improved production of Compound (2)

[0272] The engineered polynucleotide (SEQ ID NO: 649) encoding the polypeptide with transaminase activity of SEQ ID NO: 650 was used to generate the engineered polypeptides of Table 5-1. These polypeptides displayed improved the transaminase activity under the desired conditions e.g., the improvement in the formation of the compound **(2)** (amine product) from the ketone substrate (compound V) as compared to the starting polypeptide. The engineered polypeptides, having the amino acid sequences of even-numbered sequence identifiers were generated from the "backbone" amino acid sequence of SEQ ID NO: 650 as described below together with the HTP assay and analytical methods described in Tables 6-1 and 6-2.

[0273] Directed evolution began with the polynucleotide set forth in SEQ ID NO: 649. Libraries of engineered polypeptides were generated using various well-known techniques (e.g., saturation mutagenesis, recombination of previously identified beneficial amino acid differences) and screened using HTP assay and analysis methods that measured the polypeptides' ability to production of compound **(2)**.

[0274] The enzyme assay and analysis were carried out as described in Example 2 except that the lysate was diluted 4000-fold before adding to the reaction mixture.

[0275] Hit variants were grown in 250-mL shake flask and enzyme powders generated. The activity of the enzyme powders was evaluated at 0-3 g/L of the ATA shake flask powders, 20-100 g/L ketone, 1 M IPM (4 M pH 9 stock), 0.5 g/L PLP, 0-18% DMSO co-solvent, pH ~ 9, 50°C, for 24 hours, using similar assay as described above and the hit variants are listed in Table 5-2.

Table 5-1. ATA Activity in the Production of Compound (2) Relative to SEQ ID NO: 650		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 650)**	**Percent Conversion Fold Improvement (Relative to SEQ ID NO: 650)[1]**
813/814	V383L	++
815/816	R78A	++
817/818	W316S	++
819/820	W316L	++
821/822	W316N	+
823/824	A72G	+
825/826	M226Q	+
827/828	Y14A	+
829/830	F169V	+
831/832	G175D;W316F	+
833/834	Y14H;S108R;A133R;I311S	+
835/836	T114A	+
837/838	S24E;M163L;A199Q	+
839/840	W316D	+
841/842	M95I	+
843/844	T154S	+
845/846	W316Y	+
847/848	Y14G	+
849/850	E382D	+
851/852	K73R	+
853/854	W316V	+
855/856	Y14H;I311S	+
857/858	I311K	+
859/860	S108R	+
861/862	T13A;I311S	+
863/864	T13A;S24K;A199Q;I311S	+
865/866	T13A;S108R	+
867/868	K73S	+
869/870	D386A	+
871/872	H35E	+
873/874	F169C	+
875/876	V101L	+
877/878	W316E	+
879/880	P293A	+
881/882	S24E;M163L	+
883/884	T13A	+
885/886	R10E	+
887/888	M163V	+
889/890	W316G	+
891/892	Y14H;S24K;S108R;A199Q	+
893/894	T13A;A199Q	+

Table 5-1. ATA Activity in the Production of Compound (2) Relative to SEQ ID NO: 650		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 650)**	**Percent Conversion Fold Improvement (Relative to SEQ ID NO: 650)[1]**
895/896	W316H	+
897/898	W316I	+
899/900	W316F	+
901/902	Y14H;S24K;A199Q	+
903/904	T13A;S24E;S108R;M163L;I311S	+
905/906	Y14H	+
907/908	T13A;S24E;M163L	+
909/910	T13A;S24E;A133R;A199Q;I311S	+
911/912	S108R;A199Q	+
913/914	Y14H;S24E;S108R;A133R	+
915/916	Y14H;S24K;S108R	+
917/918	M163H	+
919/920	S24E	+
921/922	Y14H;S108R	+
923/924	T13A;S108R;I311S	+
925/926	A199Q;I311S	+
927/928	Y14H;S108R;I311S	+
929/930	T13A;S108R;A199Q	+
931/932	A199Q	+
933/934	M163S	+
935/936	T13A;S24K;S108R;M163L	+
[1]Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 650 and defined as follows: "+" 1.00 to 1.35 "++" > 1.35		

Table 5-2. ATA Activity in the Production of Compound (2) Relative to SEQ ID NO: 650		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 650)**	**Percent Conversion Fold Improvement (Relative to SEQ ID NO: 650)[1]**
813/814	V383L	++
815/816	R78A	++
819/820	W316L	+
817/818	W316S	+
821/822	W316N	+
833/834	Y14H;S108R;A133R;I311S	+
837/838	S24E;M163L;A199Q	+
823/824	A72G	+
[1]Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 650 and defined as follows: "+" 1.20 to 1.40 "++" > 1.40		

EXAMPLE 6

Analytical Detection of Conversion of Compound (1) to Compound (2)

[0276] Data described in Examples 2 to 5 were collected using the analytical methods provided in Tables 6-1 and 6-2. The methods provided herein find use in analyzing the variants produced using the present invention. However, it is not intended that present invention be limited to the methods described herein, as there are other suitable methods known in the art that are applicable to the analysis of the variants provided herein and/or produced using the methods provided herein.

Table 6-1. HPLC Parameters	
Method for determining conversion of Compound (1) to Compound (2)	
Instrument	**Shimadzu HPLC**
Column	Cogent Diamond Hydride, 4.6 x 150 mm x 4 μm, (p/n 70000-15P)
Mobile phase	Isocratic: 65% H_2O + 0.1% TFA, 35% MeCN + 0.1% TFA
Flow rate	1.5 mL/min
Run time	1.6 min
Peak retention times	Amine: 1.15 min Ketone: 1.30 min
Column temperature	50°C
Injection volume	5 μL
Detection wavelength	265 nm

Table 6-2. HPLC Parameters	
Method for determining selectivity	
Instrument	**Shimadzu HPLC**
Column	Agilent Poroshell 120 PhenylHexyl, 4.6 x 100 mm x 2.7 μm (p/n 695975-912)
Mobile phase	A: 0.1% TFA in water, B: 0.1% TFA in MeCN
Mobile phase gradient	10% B for 2 min, step change to 100% B at 2.01 min, hold at 100% B to 3 min, step change to 10% B at 3.01 min, stop at 4 min
Flow rate	1.5 mL/min
Run time	4.0 min
Peak retention times	S-amine: 2.4 min R-amine: 2.6 min Ketone: 3.2 min
Column temperature	50°C
Injection volume	5 μL
Detection	265 nm

[0277] All publications, patents, patent applications and other documents cited in this application are hereby incorporated by reference in their entireties for all purposes to the same extent as if each individual publication, patent, patent application or other document were individually indicated to be incorporated by reference for all purposes.

[0278] While various specific embodiments have been illustrated and described, it will be appreciated that various changes can be made without departing from the spirit and scope of the invention(s).

CLAIMS

We claim:

1. An engineered transaminase comprising a polypeptide sequence having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to SEQ ID NOS: 2, 4, 8, 366, and/or 650, or a functional fragment thereof, wherein said engineered transaminase comprises 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, 4, 8, 366, and/or 650.

2. The engineered transaminase of Claim 1, wherein said polypeptide sequence has at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to SEQ ID NO: 4, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 21/163/286/291/314/316/323/383/388/408/414/416, 18, 20, 21, 21/23/56/146, 21/23/56/146/432, 21/23/146/417, 21/23/395/417/432, 21/53/56, 21/53/417, 21/56/395, 21/417/432, 23, 23/53, 23/53/56, 23/53/56/146/395, 23/53/395, 23/53/417, 23/53/432, 23/56, 23/56/395, 23/56/395/417, 23/395/417, 23/417, 23/417/432, 53, 53/56, 53/146/417, 53/395, 56, 56/74/241/286/314/316/323, 56/86/163/314/316/383/414/416/422, 56/86/286/314/414/416, 56/86/314/316/323/394/414/422, 56/146/417, 56/146/432, 56/147, 56/163, 56/163/286/316/323/383/394, 56/286/314/316/323/422, 56/286/383, 56/323, 56/323/383, 56/323/383/394, 56/383, 56/395, 74/81/286/316/323/383, 74/85/86/163/286/316/323/394, 74/85/314/316/414/416, 74/86/163/316, 74/86/316/323/383/394, 74/88/286/316/323/383, 74/88/323/383, 74/163/286/316/383/394/416, 74/163/314/316, 74/163/314/316/323/394, 74/163/314/323/383/414/416, 74/286, 74/286/316/323, 74/286/394/416, 74/314/323/383/394/414, 74/316/323/394, 85/86/88/163/323/383/394, 85/86/163/314/323/394/414, 85/286, 85/286/323, 86, 86/88/163/323/383/414/422, 86/383/394, 88, 88/163/286/383, 88/286/316/323, 88/286/316/323/383/414/416, 88/316/323, 146, 146/147/395/417, 146/395, 146/395/417, 146/417, 147/395/417/432, 147/417, 149, 157, 163, 163/222/286/316/323/383/394, 163/286, 163/286/314/316/323/414/416, 163/286/314/323/394, 163/286/316/323/394/416, 163/286/414, 163/314/316/394, 163/314/323/394, 163/314/383, 163/314/414, 163/316/323, 163/323, 163/383, 164, 199/417, 259, 260, 284, 286, 286/314/323/383, 286/314/394, 286/316/323/383/414/416, 286/316/383/394, 286/316/394/414/416, 286/323, 286/323/383/414, 286/323/416, 286/383, 286/416, 314/316, 314/316/323, 314/316/323/383/422, 314/316/323/394, 314/316/394, 314/323/383/394, 314/383, 314/383/414/422, 315, 316, 316/323/383/394, 316/323/394/414/416, 316/414/422, 323, 323/383, 323/383/394/414/416, 323/394, 383, 395, 395/417, 395/417/432, 400, 401, 403, 404, 405, 406, 408, 415,

417, 417/432, 420, and 422, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 4.

3. The engineered transaminase of Claim 1, wherein said polypeptide sequence has at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to SEQ ID NO: 4, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 74/81/286/316/323/383, 163/286/314/316/323/414/416, 163/286/314/323/394, 286/314/323/383, 286/316/323/383/414/416, 315, and 408, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 4.

4. The engineered transaminase of Claim 1, wherein said polypeptide sequence has at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to SEQ ID NO: 8, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 5, 18/23/149/260/383/395/401/416, 18/23/149/383, 18/163/164, 21, 21/163/315/316, 21/163/323/408, 21/408, 23/56/86/149/163/164/383/401/416, 23/86, 23/149/260, 23/149/284/383/395, 23/163/164/383, 23/163/164/401/416, 24, 42, 42/110, 42/187/272, 42/187/324/363/366, 42/187/353, 42/272/291, 42/272/291/363, 42/272/324/363/366, 42/272/363/410, 42/272/410, 42/291/313/363/410, 42/291/363, 42/291/363/366, 42/353, 42/363, 46, 66, 77, 86/149/163/164/383/395/401, 86/149/395, 86/163/164/260/383, 86/383, 107, 110, 110/187, 110/187/253/410, 134, 138, 149/164/260/383/395/401, 149/260/383, 149/416, 163/259/323/408, 163/259/408, 163/315/316, 164/260/401, 164/316/383/401, 167, 186, 187, 187/253/363/366, 187/272/324/363/410, 187/272/363, 187/272/363/366/410, 187/291, 189, 191, 195, 199, 203, 210, 211, 248, 259/307, 260/395/401, 272, 272/353, 272/363/366, 272/410, 277, 291, 305, 309, 315, 342, 343, 351, 354, 358, 361, 362, 363, 363/366, 365, 367, 383, 383/401, 383/416/422, 385, 388, 389, 392, 395, 396, 401, 404, 405, 408, 410, 416, 417, 439, 443, 447, 450, and 451, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 8.

5. The engineered transaminase of Claim 1, wherein said polypeptide sequence has at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to SEQ ID NO: 8, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 18/23/149/383, 21/163/323/408, 272, 291, and 383, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 8.

6. The engineered transaminase of Claim 1, wherein said polypeptide sequence has at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to SEQ ID NO: 366, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 24, 24/42/66/291, 24/42/291/362, 24/66/163/191/362/383/388, 24/66/191/199/260/291/351, 24/66/191/199/291, 24/66/191/260/408, 24/66/260/291/383/388/408, 24/66/291/342/383, 24/66/291/365, 24/66/342/365/388/408, 24/77/291, 24/107/163/191/291/351/383/388, 24/107/291/351/365/388, 24/163/351/383, 24/191/291/365, 24/199/260/351/362/383, 24/199/260/362/383/388, 24/260/362/383/388, 24/291, 24/291/342/351/383, 24/291/362/388, 24/291/408, 24/383/388, 24/388, 25, 28, 33, 42/191/408, 42/199/291/383, 42/291/351/362/365/383/388, 42/291/351/362/383/408, 42/291/383/388, 66/82/291/383, 66/163/191/365/383, 66/199/351/383, 66/291, 66/291/362/365/383, 66/291/383/388, 66/383, 77/291, 77/383/388, 86, 107/191/199/365/383/388, 107/191/291/383, 148, 153, 163/291/362/365/383/388, 163/291/383/388, 163/383, 191/199/365/383/388, 191/260/388, 191/291, 191/291/342/362/365, 191/351/383/388, 199/260/383, 199/291, 260, 260/291/365/383/408, 260/365/383, 291, 291/351/383/388, 291/351/383/388/408, 291/362/365, 291/365/388, 291/383, 314, 315, 316, 319, 342/362, 351/383/388, 362, 362/388, 383, 383/388, 396, 397, 405, 406, 413, 419, and 423, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 366.

7. The engineered transaminase of Claim 1, wherein said polypeptide sequence has at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 366, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 24/66/191/199/291, 24/66/291/365, 163/291/362/365/383/388, 163/291/383/388, 191/291/342/362/365, 291, and 291/383, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 366.

8. The engineered transaminase of Claim 1, wherein said polypeptide sequence has at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to SEQ ID NO: 650, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 10, 13, 13/24/108/163, 13/24/108/163/311, 13/24/133/199/311, 13/24/163, 13/24/199/311, 13/108, 13/108/199, 13/108/311, 13/199, 13/311, 14, 14/24/108, 14/24/108/133, 14/24/108/199, 14/24/199, 14/108, 14/108/133/311, 14/108/311, 14/311, 24, 24/163, 24/163/199, 35, 72, 73, 78, 95, 101, 108, 108/199, 114, 154, 163, 169, 175/316, 199, 199/311, 226, 293, 311, 316, 382, 383, and 386, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 650.

9. The engineered transaminase of Claim 1, wherein said polypeptide sequence has at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to SEQ ID NO: 650, wherein said engineered transaminase comprises at least one substitution or substitution set in said polypeptide sequence at one or more positions selected from 14/108/133/311, 24/163/199, 72, 78, 316, and 383, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 650.

10. The engineered transaminase of Claim 1, wherein said engineered transaminase comprises a polypeptide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% identical to the sequence of at least one engineered transaminase variant set forth in Table 2-1, 2-2, 3-1, 3-2, 4-1, 4-2, 5-1, and/or 5-2.

11. The engineered transaminase of Claim 1, wherein said engineered transaminase is a variant engineered transaminase provided in Table 2-1, 2-2, 3-1, 3-2, 4-1, 4-2, 5-1, and/or 5-2.

12. The engineered transaminase of Claim 1, wherein said engineered transaminase is a variant engineered transaminase set forth in SEQ ID NOS: 2, 4, 8, 366, and/or 650.

13. The engineered transaminase of Claim 1, wherein said engineered transaminase comprises a polypeptide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% identical to the sequence of at least one engineered transaminase variant set forth in the even numbered sequences of SEQ ID NOS: 6-936.

14. The engineered transaminase of Claim 1, wherein said engineered transaminase comprises a polypeptide sequence set forth in the even numbered sequences of SEQ ID NOS: 6-936.

15. The engineered transaminase of any of Claims 1-14, wherein said engineered transaminase comprises at least one improved property compared to wild-type *V. fluvialis* transaminase.

16. The engineered transaminase of Claim 15, wherein said improved property comprises improved activity on a substrate.

17. The engineered transaminase of Claim 16, wherein said substrate comprises compound **(1)**.

18. The engineered transaminase of any of Claims 15, wherein said improved property comprises improved thermostability.

19. The engineered transaminase of any of Claims 1-18, wherein said engineered transaminase is purified.

20. A composition comprising the engineered transaminase of any of Claims 1-19.

21. A polynucleotide sequence encoding at least one engineered transaminase of any of Claims 1-19.

22. A polynucleotide sequence encoding at least one engineered transaminase of any Claims 1-19, wherein said polynucleotide sequence 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, 3, 7, 365, and/or 649, and wherein the polynucleotide sequence of said engineered transaminase comprises at least one substitution at one or more positions.

23. A polynucleotide sequence encoding at least one engineered transaminase comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NOS: 2, 4, 8, 366, and/or 650, or a functional fragment thereof.

24. The polynucleotide sequence of any of Claims 21-23, wherein said polynucleotide sequence is operably linked to a control sequence.

25. The polynucleotide sequence of any of Claims 21-24, wherein said polynucleotide sequence is codon optimized.

26. An expression vector comprising at least one polynucleotide sequence of any of Claims 21-25.

27. A host cell comprising at least one expression vector of Claim 26.

28. A host cell comprising at least one polynucleotide sequence of any of Claims 21-25.

29. A method of producing an engineered transaminase in a host cell, comprising culturing the host cell of Claim 27 and/or 28, under suitable conditions, such that at least one engineered transaminase is produced.

30. The method of Claim 29, further comprising recovering at least one engineered transaminase from the culture and/or host cell.

31. The method of Claim 29 and/or 30, further comprising the step of purifying said at least one engineered transaminase.