(19) DANMARK

(10) DK/EP 3630798 T3

(12) Oversættelse af europæisk patentskrift

Patent- og Varemærkestyrelsen

(51) Int.Cl.: *C 12 P 13/04 (2006.01)* *C 12 N 9/88 (2006.01)*

(45) Oversættelsen bekendtgjort den: **2023-02-20**

(80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2022-11-23**

(86) Europæisk ansøgning nr.: **18810766.8**

(86) Europæisk indleveringsdag: **2018-05-09**

(87) Den europæiske ansøgnings publiceringsdag: **2020-04-08**

(86) International ansøgning nr.: **CN2018086227**

(87) Internationalt publikationsnr.: **WO2018219107**

(30) Prioritet: **2017-05-27 CN 201710386994**

(84) Designerede stater: **AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR**

(73) Patenthaver: **Enzymaster (Ningbo) Bio-Engineering Co., Ltd., (6-21)-(6-33) , Block 015 , 333 North Century Avenue, Ningbo, Zhejiang 315000, Kina**

(72) Opfinder: **CHEN, Haibin, (6-21)-(6-33), Block 015, 333 North Century Avenue, Ningbo, Zhejiang 315000, Kina**
**BONG, Yong Koy, (6-21)-(6-33), Block 015, 333 North Century Avenue, Ningbo, Zhejiang 315000, Kina**
**XU, Qing, (6-21)-(6-33), Block 015, 333 North Century Avenue, Ningbo, Zhejiang 315000, Kina**
**ZHOU, Ameng, (6-21)-(6-33), Block 015, 333 North Century Avenue, Ningbo, Zhejiang 315000, Kina**
**SHEN, Tianran, (6-21)-(6-33), Block 015, 333 North Century Avenue, Ningbo, Zhejiang 315000, Kina**
**YANG, Jiadong, (6-21)-(6-33), Block 015, 333 North Century Avenue, Ningbo, Zhejiang 315000, Kina**
**YANG, Zhuhong, (6-21)-(6-33), Block 015, 333 North Century Avenue, Ningbo, Zhejiang 315000, Kina**

(74) Fuldmægtig i Danmark: **Zacco Denmark A/S, Arne Jacobsens Allé 15, 2300 København S, Danmark**

(54) Benævnelse: **Manipulerede aldolase-polypeptider og deres anvendelser i syntese af beta-hydroxy-alpha-aminosyrer**

(56) Fremdragne publikationer:
**EP-A2- 0 460 883**
**WO-A1-2018/219108**
**CA-A1- 2 648 697**
**CN-A- 104 673 814**
**JP-A- 2006 121 983**
**VORHOLT, J.: "L-threonine aldolase", , 10 May 2017 (2017-05-10), XP055769610, Retrieved from the Internet: URL:https://www.uniprot.org/uniprot/A0A0Q8 I770.txt?version=6 [retrieved on 2021-01-27]**
**DUAN; J. ET AL.: "L-threonine aldolase", , 10 May 2017 (2017-05-10), XP055769608, Retrieved from the Internet: URL:https://www.uniprot.org/uniprot/K9NI78 .txt?version=19 [retrieved on 2021-01-27]**
**DE GROOT, N.N.: "L-threonine aldolase", , 10 May 2017 (2017-05-10), XP055769602, Retrieved from the Internet: URL:https://www.uniprot.org/uniprot/A0A1H2 NQ51.txt?version=5 [retrieved on 2021-01-27]**

**Fortsættes ...**

MISONO, H. ET AL.: "L-threonine aldolase", , 10 May 2017 (2017-05-10), XP055768431, Retrieved from the Internet: URL:https://www.uniprot.org/uniprot/Q59IT3 .txt?version=60 [retrieved on 2021-01-25]
DÜCKERS, N. ET AL.: "Threonine aldolases - screening, properties and applications in the synthesis of non-proteinogenic beta-hydroxy-alpha-amino acids", APPLIED MICROBIOLOGY AND BIOTECHNOLOGY, vol. 88, no. 2, 4 August 2010 (2010-08-04) , pages 409-424, XP019841749,
DATABASE GenBank 10 August 2005 (2005-08-10), "phenylserine aldolase [Pseudomonas putida", XP055556916, Database accession no. BAD91544.1
GWON, H.J. et al.: "Diastereoselective synthesis of L-threo-3, 4-dihydroxyphenylserine by low-specific L-threonine aldolase mutants", Biotechnol. Lett., vol. 32, no. 1, 17 September 2009 (2009-09-17), pages 143-149, XP019766769,
FESKO, K.: "Threonine aldolases: perspectives in engineering and screening the enzymes with enhanced substrate and stereo specificities", Appl. Microbiol. Biotechnol., vol. 100, no. 6, 26 January 2016 (2016-01-26), pages 2579-2590, XP035631555,
LEE, S.J. et al.: "High-throughput screening methods for selecting 1-threonine aldolases with improved activity", Journal of Molecular Catalysis B: Enzymatic, vol. 26, no. 3-6, 31 December 2003 (2003-12-31), pages 265-272, XP002992768,
STEINREIBER, J. ET AL.: "Threonine aldolases-an emerging tool for organic synthesis", TETRAHEDRON, vol. 63, no. 4, 16 December 2006 (2006-12-16), pages 918-926, XP005805698, DOI: 10.1016/J.TET.2006.11.035

# DESCRIPTION

**Technical field**

**[0001]** The present invention relates to the field of biotechnology, in particular to engineered aldolase polypeptides and their application in industrial biocatalysis.

**Background technique**

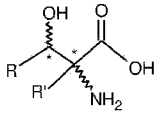
**[0002]** β-hydroxy-α-amino acids are a kind of important amino acids which are critical intermediates to construct a variety of natural products and drugs. They have many important biological activities. The chemical structure of these compounds generally has two chiral centers with multiple stereoisomers (Scheme 1). (2S, 3R)-2-amino-3-hydroxy-3- (4-nitrophenyl) propanoic acid (a compound of formula A2 as shown in Scheme 2), is a β-hydroxy-α-amino acid and is available as an intermediate in the synthesis of D-(-)-threo-2-amino-1-(4-nitrophenyl)-1,3-propanediol.

**[0003]** At present, there are two main routes in the synthesis of D-(-)-threo-2-amino-1-(4-nitrophenyl)-1,3-propanediol (or Levoamine): p-nitrobenzaldehyde or p-nitro-acetophenone is used as the starting material respectively. Using p-nitrobenzaldehyde as the starting material requires less number of substrates and is a relatively simpler process. However, due to the presence of multiple isomeric structures, this route requires excessive p-nitrobenzaldehyde; it also requires calcium borohydride as reducing agent which is very expensive. Therefore, the current industrial process is mainly the second route. P-nitro-acetophenone is subject to bromination, amination, acylation, formaldehyde condensation, reduction and hydrolysis to produce racemic product which is then resoluted by tartaric acid to obtain D-(-)-threo-2-amino-1-(4-nitrophenyl)-1,3-propanediol. This process uses a lot of irritating organic solvents such as acetic anhydride, benzyl chloride, acetic acid, bromine, giving terrible environmental impact and also making it very difficult to manage the production site. And this synthesis process consists of many steps including protection, deprotection, resolution, etc, which is tedious and makes overall yield low.

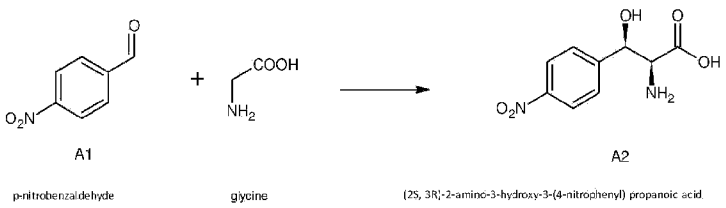
**[0004]** It has been reported that aldolase can condense aldehydes and amino acids to form β-hydroxy-α-amino acids. The condition of this enzymatic reaction is mild and its pollution is little, but the stereoselectivity of wild-type aldolases s is not good enough to meet industrial application request. The present invention provides a series of engineered polypeptides which high stereoselectivity. These engineered polypeptides were developed through directed evolution towards the selection of (2S, 3R) -2-amino-3-hydroxy-3-(4-nitrophenyl) propanoic acid. According to the enzymatic process disclosed by this invention, p-nitrobenzaldehyde and glycine are directly and selectively condensed into (2S, 3R)-2-amino-3-hydroxy-3-(4-nitrophenyl) propanoic acid, thus determining two chiral centers in one step. (2S, 3R)-2-amino-3-hydroxy-3-(4-nitrophenyl) propanoic acid can be simply esterified and reduced to give D-(-)-threo-2-amino-1-(4-nitrophenyl)-1,3-propanediol. This entire process is simple and easy to operate, and does not require chemical resolution, effectively reducing production cost. At the same time, the condition of enzyme-catalyzed reaction is mild without requiring strong acid, alkali, high temperature or high pressure environment, alleviating the requirements of production equipment, effectively guaranteeing the quality of products, enabling economical and environmental-friendly production.

**BRIEF DESCRIPTION OF THE SCHEMES**

**[0005]**




**Scheme 1** depicts β-hydroxy-α-amino acids (* refers to chiral center).




A1 A2

p-nitrobenzaldehyde glycine (2S, 3R)-2-amino-3-hydroxy-3-(4-nitrophenyl) propanoic acid

**Scheme 2** depicts the synthesis of A2 by an asymmetric reaction catalyzed by an engineered aldolase polypeptide of the present invention.

**Content of the invention**

**1. Overview**

**[0006]** The present invention provides engineered polypeptides with high stereoselectivity, high catalytic activity and good stability, which can synthesize β-hydroxy-α-amino acids asymmetrically, and in particular asymmetrically synthesize (2S, 3R)-2-amino-3-hydroxy-3-(4-nitrophenyl) propanoic acid. The present invention also provides gene sequences of engineered polypeptides, recombinant expression vectors containing the genes, engineered strains and efficient methods for the production thereof, as well as reaction processes for the asymmetric synthesis of β-hydroxy-α-amino acids using engineered polypeptides.

**[0007]** In the first aspect, the present invention provides engineered aldolase polypeptides comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, and 230. with improved catalytic properties. These engineered polypeptides are derived from directed evolution of a wild-type aldolase which is less stereoselective toward the product, through substitutions, insertions, or deletions of a number of amino acid residues. The wild-type aldolase is from *Pseudomonas putida* which consists of 357 amino acids and has the sequence shown in SEQ ID No: 2. The wild-type aldolase showed low stereoselectivity for the product. As measured by the inventors, in the reaction of converting p-nitrobenzaldehyde (i.e. A1) with glycine to produce (2S, 3R)-2-amino-3-hydroxy-3-(4-nitrophenyl) propanoic acid (i.e. A2) in Scheme 2 using SEQ ID No: 2, the diastereomeric excess number (i.e. de) for A2 is ≤ 40%. Said wild-type aldolase of *Pseudomonas putida* has been described in the database entry MISONO, H. ET AL.: "L-threonine aldolase" (UniProt accession number Q59IT3,10.05.2017, retrieved from the internet: URL:https://www.uniprot.org/uniprot/Q59IT3.txt?version=60) and the publications DÜCKERS, N. ET AL.: "Threonine aldolases - screening, properties and applications in the synthesis of non-proteinogenic beta-hydroxy-alpha-amino acids" (APPLIED MICROBIOLOGY AND BIOTECHNOLOGY, vol. 88, no. 2, 04.08.2010, pages 409-424) and STEINREIBER, J. ET AL.: "Threonine aldolases-an emerging tool for organic synthesis" (TETRAHEDRON, vol. 63, no. 4, 16.12.2006, pages 918-926).

**[0008]** According to the invention, engineered aldolase polypeptides of the present invention are capable of converting A1 and glycine to A2 at a stereoselectivity greater than that of SEQ ID No: 2. Under the indicated reaction conditions, the engineered aldolase polypeptides of the present invention are capable of producing at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93% 94%, 95%, 96%, 97%, 98% or 99% or more diastereomeric excess for A2. In some embodiments, the reaction conditions include 20% organic solvent (including but not limited to dimethyl sulfoxide, ethanol or methanol) and temperature of about 30 °C and a pH of about 6.0.

**[0009]** According to the invention, the engineered aldolase polypeptides are capable of converting A1 and glycine to A2 at a higher stereoselectivity than the polypeptide of SEQ ID NO: 2 under the indicated reaction conditions. The engineered aldolase polypeptides comprise an amino acid sequence selected from the group consisting of SEQ ID NOs: 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, and 230.

**[0010]** The identity between two amino acid sequences or two nucleotide sequences can be obtained by commonly used algorithms in the art and can be calculated according to default parameters by using NCBI Blastp and Blastn software, or by using the Clustal W algorithm (Nucleic Acid Research, 22 (22): 4673-4680, 1994). For example, using the Clustal W algorithm, the amino acid sequence identity of SEQ ID NO: 2 to SEQ ID NO: 184 is 93.3%.

**[0011]** In some examples, engineered aldolase polypeptides can comprise an amino acid sequence that differs in one or more residues compared to the sequence of SEQ ID NO: 2 in the residue position: X16, X17, X19, X26, X32, X33, X37, X38, X39, X41, X42, X43, X44, X45, X46, X47, X48, X49, X91, X92, X118, X132, X134, X154, X164, X168, X176, X182, X185, X189,

X191, X216, X217, X218, X227, X234, X237, X244, X247, X262, X282, X284, X285, X288, X291, X292, X293, X294, X295, X302, X305, X316, X318, X319, X320, X324, X352; for example, engineered aldolase polypeptides comprise an amino acid sequence comprising at least one of the following features (these features are substitutions of amino acid residues with the reference sequence of SEQ ID NO: 2): D16E, N17G, N17E, A19W, A19N, A26V, A26L, H32V, S33N, A37T, A37M, A37K, G38D, G38P, G38S, G38E, G38A, P39L, P39A, G41Y, T42M, D43P, D43Y, E44D, L45I, T46H, A47H, Q48L, V49S, P91H, P91S, P91L, P91K, P91N, A92W, P118R, P118I, P118G, R132S, K134Q, V154G, V154S, V154A, V154F, V154R, E164R, D168N, G176P, S182A, A185T, V189S, L191H, V216C, L217W, A218C, A218S, T227P, S234R, R237T, S244I, S244V, M247Y, M247H, L262I, E282R, E282Y, E282K, L284K, L284F, L284A, L284V, G285P, G285S, G285K, E288I, E288T, G291F, G291V, G291W, G291Y, G292K, G292V, T293P, E294K, E294M, A295G, A295Q, L302M, A305T, A305P, G316K, G316S, G316V, G316R, Y318G, Y318L, H319Y, H319V, D320K, D320E, P324L, D352Y, D352Q, D352A; Or, in addition to the abovementioned differences, engineered aldolase polypeptides comprise insertions or 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 of amino acid residues.

**[0012]** In another aspect, this invention provides polynucleotide sequences encoding engineered aldolase polypeptides. In some embodiments, a polynucleotide can be part of an expression vector having one or more control sequences for the expression of an engineered aldolase polypeptide. In some embodiments, polynucleotides can comprise sequences corresponding to SEQ ID No: 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, and 229.

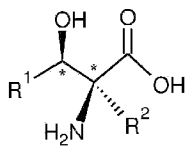
**[0013]** As known to people skilled in the art, due to the degeneracy of the nucleotide codons, the polynucleotide sequences encoding amino acid sequences SEQ ID No: 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, and 230 are not limited to SEQ ID No: 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, and 229.

**[0014]** The polynucleotide sequences of the engineered polypeptides of the present invention may also be any other polynucleotide sequences encoding amino acid sequences SEQ ID No: 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, and 230.

**[0015]** In another aspect, this disclosure provides polynucleotides comprising sequences encoding engineered aldolase polypeptides, expression vectors and host cells capable of expressing engineered aldolase polypeptides. In some embodiments, the host cell can be bacterial host cell, such as E. coli. The host cell can be used to express and isolate the engineered aldolase described herein, or alternatively be directly used in the reaction for conversion of substrates to products.

**[0016]** In some embodiments, the engineered aldolase in the form of whole cell, crude extract, isolated enzyme, or purified enzyme can be used alone or in an immobilized form, such as immobilization on a resin.

**[0017]** The present disclosure also provides the process of the asymmetric synthesis of β-hydroxy-α-amino acid compounds using the herein disclosed engineered aldolase polypeptides, the resulting β-hydroxy-α-amino acid products having the structure shown in Formula (I) :



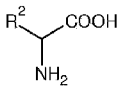

(I) ,

**[0018]** The β-hydroxy-α-amino acid products of formula (I) have the indicated stereochemical configuration shown at the chiral center marked with an*; the β-hydroxy-α-amino acid products of formula (I) are in a diastereoisomeric excess over the other isomers, where

$R^1$ is optionally substituted or unsubstituted aryl or heteroaryl, or optionally substituted or unsubstituted $C_1$-$C_8$ hydrocarbyl;

$R^2$ is -H, -$CH_2OH$, -$CH_2SH$, -$CH_2SCH_3$, or optionally substituted or unsubstituted $C_1$-$C_4$ hydrocarbyl, the process comprising that, under suitable reaction conditions of reacting the aldehyde substrate and the amino acid substrate to obtain β-hydroxy-α-amino acid products, the aldehyde substrate of formula (II) and the amino acid substrate of formula (III):

(II)

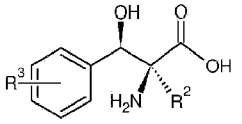


(III)

were contacted with the aldolase polypeptides, wherein the aldolase polypeptides are engineered aldolase polypeptides described herein. In some examples, the engineered aldolase polypeptides have at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93 %, 94%, 95%, 96%, 97%, 98%, 99% or more of sequence identity to SEQ ID NO: 2 and are capable of condensing the aldehyde substrate of formula (II) and the amino acid substrate of formula (III) to obtain the β-hydroxy-α-amino acid products of formula (I) at higher conversion or higher stereoselectivity compared to SEQ ID NO: 2.

**[0019]** According to the invention, the β-hydroxy-α-amino acid product of formula (I) is present in diastereomeric excess of at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93% , 94%, 95%, 96%, 97%, 98%, or 99% or greater.

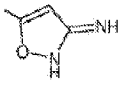
**[0020]** In some embodiments of this process, the β-hydroxy-α-amino acid products of formula (I) are:



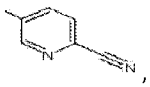
,

wherein $R^3$ is $C_1$-$C_4$ hydrocarbyl, -H, halogen such as -F, -Cl, -Br and -I, -$NO_2$, -NO, -$SO_2R'$ or -SOR', -SR', -NR' -C(O)NR', -$SO_2NH_2$ or -$SONH_2$, -CN, $CF_3$, wherein each R' is independently selected from -H or ($C_1$-$C_4$) hydrocarbyl;

$R^3$ may also be

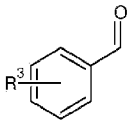


or



,

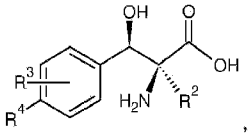
$R^2$ is -H, -$CH_3$, -$CH_2CH_3$, -CH ($CH_3$)$_2$, -$CH_2OH$, -$CH_2SH$ or -$CH_2SCH_3$, and the aldehyde substrate of formula (II) is:



.

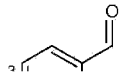
**[0021]** In some embodiments, $R^3$ is in the para position of the phenyl ring. In some embodiments, $R^3$ is in the meta position of the phenyl ring. In some embodiments, $R^3$ is ortho to the phenyl ring. In some embodiments, $R^3$ is both para and meta to the phenyl ring. In some embodiments, $R^3$ is both para and ortho to the phenyl ring. In some embodiments, $R^3$ is both meta and ortho to the phenyl ring.

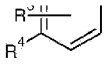
**[0022]** In some embodiments of this process, the β-hydroxy-α-amino acid product of formula (I) is:



,

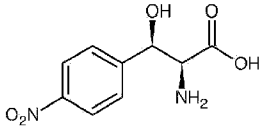
wherein $R^4$ is defined same as $R^3$ above, $R^3$ and $R^2$ are as defined above, and the aldehyde substrate of formula (II) is:





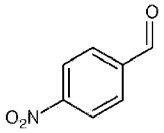
**[0023]** In some embodiments, $R^3$ is in the meta position of the phenyl ring. In some embodiments, $R^3$ is ortho to the phenyl ring.

**[0024]** In some embodiments, the engineered aldolase polypeptides can be used in the production process of diastereomeric excess of the compound of formula A2, (2S, 3R)-2-amino-3-hydroxy-3-(4-nitrophenyl) propanoic acid:



A2

**[0025]** In these embodiments, the production process comprises that, under suitable reaction conditions for converting compound of formula A1 to compound of formula A2, in a suitable organic solvent, in the presence of glycine, the compound of formula A1:



A1

were contacted with the engineered aldolase polypeptides disclosed herein. According to the invention, the compound of formula (I) or the compound of formula A2 is present in a diastereomeric excess of at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93% 94%, 95%, 96%, 97%, 98% or 99% or more. Specific embodiments of engineered aldolase polypeptides for use in this method are further provided in the detailed description. An engineered aldolase polypeptide that is used in the above process comprises one or more sequences selected from the amino acid sequences corresponding to SEQ ID NOs: 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, and 230.

**[0026]** In another aspect, this disclosure provides a process of producing (2S, 3R) -2-amino-3-hydroxy-3-(4-nitrophenyl) propanoic acid using an engineered aldolase polypeptide disclosed herein. In some embodiments, the process comprises that, under suitable reaction conditions for converting p-nitrobenzaldehyde to (2S, 3R)-2-amino-3-hydroxy-3-(4-nitrophenyl) propanoic acid, in the presence of glycine, the p-nitrobenzaldehyde is contacted with the engineered aldolase polypeptides described herein.

**[0027]** Any of the processes for the preparation of a compound of formula (I) or a compound of formula A2 using an engineered polypeptide as disclosed herein can be performed under a range of suitable reaction conditions, which including, but not limited to, amino donor, pH, temperature, buffer, solvent system, substrate loading, polypeptide loading, cofactor loading, pressure and reaction time range. For example, preparing a compound of formula (I) or a compound of formula A2 may be performed, wherein suitable reaction conditions include: (a) about 10 g/L to 200 g/L of a substrate compound (e.g. compound (II) or A1); (b) about 0.5 g/L to 10 g/L of engineered polypeptide; (c) about 30 g/L to 300 g/L of glycine loading; (d) about 0.1mM-5mM PLP cofactor; (e) from 0% (v/v) to about 60% (v/v) of organic solvent, including but not limited to, dimethylsulfoxide (DMSO), Dimethylformamide (DMF), isopropyl acetate, methanol, ethanol, propanol or isopropanol (IPA); (F) a pH of about 4.0 to about 8.0; and (g) a temperature of about 10 °C to about 60 °C.

**[0028]** According to the invention, the process is capable of forming the product (2S, 3R)-2-amino-3-hydroxy-3-(4-nitrophenyl) propanoic acid in a diastereomeric excess of at least 60%.

**[0029]** In some embodiments, the process is capable of forming the product (2S, 3R)-2-amino-3-hydroxy-3-(4-nitrophenyl) propanoic acid in a diastereomeric excess of at least 70%.

**[0030]** In some embodiments, the process is capable of forming the product (2S, 3R)-2-amino-3-hydroxy-3-(4-nitrophenyl) propanoic acid in a diastereomeric excess of at least 80%.

**[0031]** In some embodiments, the process is capable of forming the product (2S, 3R)-2-amino-3-hydroxy-3-(4-nitrophenyl) propanoic acid in a diastereomeric excess of at least 85%.

**[0032]** In some embodiments, the process is capable of forming the product (2S, 3R)-2-amino-3-hydroxy-3-(4-nitrophenyl) propanoic acid in a diastereomeric excess of at least 90%.

**[0033]** In some embodiments, the process is capable of forming the product (2S, 3R) -2-amino-3-hydroxy-3-(4-nitrophenyl)propanoic acid in a diastereomeric excess of at least 95%.

**[0034]** In some embodiments, the process is capable of forming the product (2S, 3R) -2-amino-3-hydroxy-3-(4-nitrophenyl)propanoic acid in a diastereomeric excess of at least 99%.

**2. Details**

**2.1 Definition**

**[0035]** Unless expressly defined otherwise, technical and scientific terms used in this disclosure have the meanings that are commonly understood by people skilled in the art.

**[0036]** "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, myristoylation, ubiquitination, etc.). This definition includes D-amino acids and L-amino acids, as well as mixtures of D-amino acids and L-amino acids.

**[0037]** "Engineered aldolase", "engineered aldolase polypeptide", "aldolase polypeptide", "improved aldolase polypeptide", and "engineered polypeptide" are used interchangeably herein.

**[0038]** "Polynucleotide" and "nucleic acid" are used interchangeably herein.

**[0039]** "Cofactor" as used herein refers to a non-protein compound that operates in conjunction with an enzyme in a catalytic reaction. As used herein, "cofactor" is intended to encompass the vitamin B6 family compounds PLP, PN, PL, PM, PNP and PMP, which are sometimes also referred to as coenzymes.

**[0040]** "Pyridoxal phosphate", "PLP", "pyridoxal 5'-phosphate", "PYP" and "P5P" are used interchangeably herein to refer to compounds that act as coenzyme in aldolase reactions.

**[0041]** "Coding sequence" refers to that portion of a nucleic acid (e.g., a gene) that encodes an amino acid sequence of a protein.

**[0042]** "Naturally occurring" or "wild-type" refers to the form found in nature. For example, a naturally-occurring or wild-type polypeptide or polynucleotide sequence is a sequence that is present in an organism that can be isolated from sources in nature and which has not been intentionally modified by manual procedures.

**[0043]** "Recombinant" or "engineered" or "non-naturally occurring" when used with reference to, for example, a cell, nucleic acid or polypeptide, refers to a material or material corresponding to the native or native form of the material, that has been modified in a manner that would not otherwise exist in nature, or is identical thereto but produced or derived from synthetic material and/or by manipulation using recombinant techniques.

**[0044]** "Sequence identity" and "homology" are used interchangeably herein to refer to comparisons between polynucleotide sequences or polypeptide sequences ("sequence identity" and "homology" are generally expressed as a percentage), and are determined by comparing two optimally aligned sequences over a comparison window, where 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 can be calculated by determining the number of positions at which either the identical nucleic acid base or amino acid residue occurs in both sequences or a nucleic acid base or amino acid residue is aligned with a gap to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage

of sequence identity. Those skilled in the art will appreciate that there are many established algorithms available to align two sequences. The optimal alignment of sequences for comparison can be conducted, for example, by the local homology algorithm of Smith and Waterman, 1981, Adv. Appl. Math. 2: 482, by the Homology alignment algorithm of Needleman and Wunsch, 1970, J. Mol. Biol. 48: 443, by the search for similarity method of Pearson and Lipman, 1988, Proc. Natl. Acad. Sci. USA 85: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the GCG Wisconsin Package) or by visual inspection (see generally, Current Protocols in Molecular Biology, FM Ausubel et al. eds., Current Protocols, a Joint Venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (1995 Supplement) (Ausubel)). Examples of algorithms that are suitable for determining the percent sequence identity and percent sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al., 1990, J. Mol. Biol. 215: 403-410 and Altschul et al., 1977, Nucleic Acids Res. 3389-3402, respectively. Software for performing BLAST analysis is publicly available through the National Center for Biotechnology Information website. The algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold scores T when aligned with a word of the same length in the database sequence. T is referred to as, the neighborhood word score threshold (Altschul et al., Supra). These initial neighborhood word hits serve as seeds for initiating searches to find longer HSPs that contain them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. For nucleotide sequences, the cumulative scores are calculated using the parameters M (reward score for matched pair of residues; always> 0) and N (penalty score for mismatched residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. The extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quality X from its maximum achieved value; the cumulative score goes 0 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, the expected value (E) of 10, M = 5, N = -4, and a comparison of both strands as a default value. For amino acid sequences, the BLASTP program uses as defaults the wordlength (W) of 3, the expected value (E) of 10 and the BLOSUM62 scoring matrix (see Henikoff and Henikoff, 1989, Proc Natl Acad Sci USA 89: 10915). Exemplary determination of sequence alignments and % sequence identity can employ the BESTFIT or GAP programs in the GCG Wisconsin Software package (Accelrys, Madison WI), using the default parameters provided.

**[0045]** "Reference sequence" refers to a defined sequence that is used as a basis for sequence comparison. The reference sequence may be a subset of a larger sequence, for example, a full-length gene or a fragment of a polypeptide sequence. In general, a reference sequence is at least 20 nucleotides or amino acid residues in length, at least 25 residues long, at least 50 residues in length, or the full length of the nucleic acid or polypeptide. Because two polynucleotides or polypeptides may each (1) comprise a sequence (i.e., a portion of the complete sequence) that is similar between two sequences, and (2) may further comprise sequences that is divergent between the two sequences, sequence comparisons between two (or more) polynucleotides or polypeptides are typically performed by comparing the sequences of the two polynucleotides or polypeptides over a "comparison window" to identify and compare local regions of sequence similarity. For example, a "reference sequence" is not intended to be limited to a wild-type sequence, and may comprise engineered or altered sequences. For example, "a reference sequence with leucine at the residue corresponding to X39 based on SEQ ID NO: 2" refers to a reference sequence wherein the corresponding residue at position X39 in SEQ ID NO: 2 which is proline, has been altered to leucine.

**[0046]** A "comparison window" refers to a conceptual segment of at least about 20 contiguous nucleotide positions or amino acid residues, wherein the sequence may be compared to a reference sequence of at least 20 contiguous nucleotides or amino acids and wherein the portions of the sequence in the comparison window may comprise additions or deletions (i.e., gaps) of 20% or less as compared to a 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 optionally include 30, 40, 50, 100 or more residues.

**[0047]** In the context of the numbering for a given amino acid or polynucleotide sequence, "corresponding to," "reference to" or "relative to" refers to the numbering of the residues of a specified reference 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 sequence 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 an engineered aldolase can be aligned to a reference sequence by introducing gaps to optimize residue matches between the two sequences. In these cases, although there are gaps, the numbering of the residue in a given amino acid or polynucleotide sequence is made with respect to the reference sequence to which they have been aligned.

**[0048]** "Amino acid difference" or "residue difference" refers to the difference in amino acid residues at a position of a polypeptide sequence relative to the amino acid residue at a corresponding position in the reference sequence. The positions

of amino acid differences are generally referred to herein as "Xn", where n refers to the corresponding position in the reference sequence on which the residue differences are based. For example, "a residue difference at position X39 as compared to SEQ ID NO: 2" refers to the difference in amino acid residues at the polypeptide position corresponding to position 39 of SEQ ID NO: 2. Thus, if the reference polypeptide of SEQ ID NO: 2 has a proline at position 39, then "a residue difference at position X39 as compared to SEQ ID NO: 2" refers to an amino acid substitution of any residue other than proline at the position of the polypeptide corresponding to position 39 of SEQ ID NO: 2. In most of the examples herein, the specific amino acid residue difference at the position is indicated as "XnY", wherein "Xn" specified to the corresponding position as described above, and "Y" is the single letter identifier of the amino acid found in the engineered polypeptide (i.e., a different residue than in the reference polypeptide). In some examples (e.g., in Table 1), the present disclosure also provides specific amino acid differences denoted by the conventional notation "AnB", where A is a single letter identifier of a residue in the reference sequence, "n" is the number of residue position in the reference sequence, and B is the single letter identifier for the residue substitution in the sequence of the engineered polypeptide. In some examples, an engineered polypeptide of this disclosure may comprise one or more amino acid residue differences relative to a reference sequence, which is indicated by a list of specific positions at which residue differences are present relative to a reference sequence. In some embodiments, more than one amino acid residue can be used in a specific residue position of an engineered polypeptide, the various amino acid residues that can be used are separated by a "/" (e.g., X39L/X39A).

**[0049]** "Deletion" refers to the modification of a polypeptide by removing one or more amino acids from a reference polypeptide. Deletions can include the removal of one or more amino acids, two or more amino acids, five or more amino acids, ten or more amino acids, fifteen or more amino acids, or twenty or more amino acids, up to 10% of the total number of amino acids of the enzyme, or up to 20% of the total number of amino acids making up the reference enzyme while retaining the enzymatic activity of the engineered aldolase and/or retaining the improved properties of the engineered aldolase. Deletion may involve the internal portion and/or the terminal portion of the polypeptide. In various examples, deletions may include a contiguous segment or may be discontinuous.

**[0050]** "Insertion" refers to the modification of a polypeptide by adding one or more amino acids from a reference polypeptide. In some examples, the improved engineered aldolase comprises insertions of one or more amino acids to a naturally-occurring aldolase polypeptide as well as insertions of one or more amino acids to other engineered aldolase polypeptides. It can be inserted in the internal portions of the polypeptide or inserted to the carboxyl or amino terminus. As used herein, insertions include fusion proteins known in the art. The insertion can be a contiguous segment of amino acids or separated by one or more amino acids in naturally-occurring or engineered polypeptides.

**[0051]** "Fragment" as used herein refers to a polypeptide having an amino terminal and/or carboxyl terminal deletion, but where the remaining amino acid sequence is identical to the corresponding position in the sequence. Fragments may be at least 10 amino acids long, at least 20 amino acids long, at least 50 amino acids long or longer, and up to 70%, 80%, 90%, 95%, 98% and 99% of the full-length aldolase polypeptide.

**[0052]** An "isolated polypeptide" refers to a polypeptide that is substantially separated from other substances with which it is naturally associated, such as proteins, lipids, and polynucleotides. The term comprises polypeptides that have been removed or purified from their naturally occurring environment or expression system (e.g., in host cells or in vitro synthesis). Engineered aldolase polypeptides may be present in the cell, in the cell culture medium, or prepared in various forms, such as lysates or isolated preparations. As such, in some embodiments, the engineered aldolase polypeptide may be an isolated polypeptide.

**[0053]** "Chiral center" refers to a carbon atom connecting four different groups.

**[0054]** "Stereoselectivity" refers to the preferential formation of one stereoisomer over the other in a chemical or enzymatic reaction. Stereoselectivity can be partial, with the formation of one stereoisomer is favored over the other; or it may be complete where only one stereoisomer is formed. When the stereoisomers are enantiomers, the stereoselectivity is referred to as enantioselectivity. It is often reported as "enantiomeric excess" (ee for short). When the stereoisomers are diastereomers, the stereoselectivity is referred to as diastereoselectivity. It is often reported as " diastereomeric excess" (de for short). The fraction, typically a percentage, is generally reported in the art as optionally reported as the diastereomeric excess (i.e., de) derived therefrom according to the following formula: [major diastereomer - minor diastereomer] / [major diastereomer + minor diastereomer]. In some instances, only two diastereomers were detected in the product formed by the engineered aldolase polypeptides of the present disclosure: (2S, 3R)-2-amino-3-hydroxy-3-(4-nitrophenyl) propanoic acid (i.e., A2) and (2S, 3S) -2-amino-3-hydroxy-3- (4-nitrophenyl) propanoic acid (i.e., A3), the de value for A2 in the product is calculated as follows : [A2 - A3]/[A2 + A3].

**[0055]** "Stereoisomers," "stereoisomeric forms," and similar expressions are used interchangeably herein to refer to all isomers resulting from a difference in orientation of atoms in their space only. It includes enantiomers and compounds that have more

than one chiral center and are not mirror images of one another (i.e., diastereomers).

**[0056]** "Improved enzyme properties" refers to an enzyme property that is better or more desirable for a specific purpose as compared to a reference aldolase such as a wild-type aldolase or another improved engineered aldolase. Improved enzyme properties are exhibited by engineered aldolase polypeptides in this disclosure. Enzyme properties that are expected to be improved include, but are not limited to, enzyme activity (which can be expressed as a percentage of substrate conversion), thermal stability, solvent stability, pH activity characteristics, cofactor requirements, tolerance to inhibitors (e.g., substrate or product inhibition), stereospecificity and stereoselectivity (including enantioselectivity or diastereoselectivity).

**[0057]** "Conversion" refers to the enzymatic transformation of a substrate to the corresponding product. "Percent conversion" or "conversion" refers to the percentage of substrate that is converted to product within a period of time under the specified conditions. Thus, "enzymatic activity" or "activity" of an aldolase polypeptide can be expressed as the "percent conversion" of the substrate to the product.

**[0058]** "Thermostable" means that an aldolase polypeptide that retains similar activity (e.g., greater than 50%) after being exposed to an elevated temperature (e.g., 30-80 °C) for a period of time (0.5-24h).

**[0059]** "Solvent-stable" refers to an aldolase polypeptide that maintains similar activity (for example more than 50% to 80%) after exposure to varying solvent (ethanol, isopropanol, dimethylsulfoxide (DMSO), tetrahydrofuran, 2- Methyltetrahydrofuran, acetone, toluene, butyl acetate, methyl tert-butyl ether, etc.) for a period of time (e.g., 0.5-24 hours).

**[0060]** "Suitable reaction conditions" refer to those conditions (e.g., enzyme loading, substrate loading, cofactor loading, temperature, pH, buffer, co-solvent, etc.) in the biocatalytic reaction system, under which the aldolase polypeptide of the present disclosure can convert a substrate to a desired product compound. Exemplary "suitable reaction conditions" are provided in the present disclosure and illustrated by examples.

**[0061]** "Hydrocarbyl" refers to a straight or branched hydrocarbon group. The number of subscripts following the symbol "C" specifies the number of carbon atoms that a particular group may contain. For example, "$C_1$-$C_8$" refers to a straight or branched chain hydrocarbyl group having 1 to 8 carbon atoms. Hydrocarbyl groups may optionally be substituted with one or more substituent groups. "Aryl" means a monovalent aromatic hydrocarbon radical of 6 to about 20 carbon atoms. "Heteroaryl" and "heteroaromatic" refer to an aryl group in which one or more of the carbon atoms of the parent aromatic ring system is/are replaced by a heteroatom (0, N, or S). "Substituted", when used to modify a specified group or radical, means that one or more hydrogen atoms of the specified group or radical are each replaced, independently of one another, by identical or different substituents. "Substituted hydrocarbyl, aryl, or heteroaryl" refers to a hydrocarbyl, aryl, or heteroaryl group in which one or more hydrogen atoms are replaced by other substituents. "Optional" or "optionally" means that the described event or circumstance may or may not occur; for example, "optionally substituted aryl" refers to an aryl group that may or may not be substituted. This description includes both substituted aryl groups and unsubstituted aryl groups.

**[0062]** As used herein, "compound" refers to any compound encompassed by the structural formulas and/or chemical names indicated with the compounds disclosed herein. Compounds may be identified by their chemical structure and/or chemical name. When the chemical structure and chemical name conflict, the chemical structure determines the identity of the compound. Unless specifically stated or indicated otherwise, the chemical structures described herein encompass all possible isomeric forms of the described compounds.

**2.2 Engineered aldolase**

**[0063]** Table 1 below illustrates the engineered aldolase polypeptides developed by the present invention. Each row gives the polynucleotide sequence number and amino acid sequence number of a particular engineered aldolase polypeptide, as well as the residue difference compared to SEQ ID No: 2. The level of activity or stereoselectivity of each exemplified engineered aldolase polypeptide is indicated as "+", with the specific meanings given in Table 2.

Table 1

| Polynucleotide SEQ ID No | Amino acid SEQ ID No | Residue difference relative to SEQ ID NO:2 | Activity or Stereoselectivity |
|---|---|---|---|
| 1 | 2 | - | |
| 3 | 4 | N17G; | + |

| Polynucleotide SEQ ID No | Amino acid SEQ ID No | Residue difference relative to SEQ ID NO:2 | Activity or Stereoselectivity |
|---|---|---|---|
| 5 | 6 | N17E; | + |
| 7 | 8 | P39L; | + |
| 9 | 10 | G41Y; | + |
| 11 | 12 | D43P; | + |
| 13 | 14 | D43Y; | + |
| 15 | 16 | D16E;E44D;P91H;R132S;S244I;M247Y;G316K;Y318G;H319Y;D320K; | + |
| 17 | 18 | D16E;L45I;P91H;R132S;S244I;M247Y;G316K;Y318G;H319Y;D320K; | + |
| 19 | 20 | D16E;A47H;P91H;R132S;S244I;M247Y;G316K;Y318G;H319Y;D320K; | + |
| 21 | 22 | D16E;G38D;P91H;R132S;G176P;T227P;S244I;M247Y;G285P;G316K;Y318G;H319Y; D320K; | + |
| 23 | 24 | D16E;L451;P91S;R132S;S2441;M247Y;G316K;Y318G;H319Y;D320K; | ++ |
| 25 | 26 | D16E;L45I;P91H;R132S;S244I;M247Y;A305T;G316K;Y318G;H319Y;D320K; | ++ |
| 27 | 28 | D16E;L45I;P91H;R132S;S244I;M247Y;G316S;Y318G;H319Y;D320K; | ++ |
| 29 | 30 | D16E;L45I;P91H;R132S;S244I;M247Y;G316K;Y318G;H319V;D320K; | ++ |
| 31 | 32 | D16E;L451;Q48L;P91H;R132S;D168N;S2441;M247Y;G316K;Y318G;H319Y;D320K; | ++ |
| 33 | 34 | D16E;L451;V49S;P91H;R132S;S2441;M247Y;E294K;G316K;Y318G;H319Y;D320K; | ++ |
| 35 | 36 | D16E;L451;P91H;R132S;K134Q;S2441;M247Y;G316K;Y318G;H319Y;D320K; | ++ |
| 37 | 38 | D16E;A26V;L45I;P91H;R132S;S244I;M247Y;G316K;Y318G;H319Y;D320K; | ++ |
| 39 | 40 | D16E;A26L;L45I;P91H;R132S;E164R;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320K; | ++ |
| 41 | 42 | D16E;A26L;E44D;L45I;P91H;R132S;E164R;S244I;M247H;G316K;Y318G;H319Y;D320K; | ++ |
| 43 | 44 | D16E;T42M;L45I;P91H;R132S;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++ |
| 45 | 46 | D16E;E44D;L45I;P91H;R132S;E164R;S244I;M247Y;L302M;G316K;Y318G;H319Y;D320E; | ++ |
| 47 | 48 | D16E;A26L;E44D;L45I;P91H;R132S;E164R;S244I;M247Y;L302M;G316K;Y318G;H319Y;D320K; | ++ |
| 49 | 50 | D16E;A26L;L45I;P91H;R132S;E164R;S244I;M247H;L302M;G316K;Y318G;H319Y;D320E; | ++ |
| 51 | 52 | D16E;A26L;L45I;P91H;R132S;E164R;S244I;M247Y;G316K;Y318G;H319Y;D320K; | ++ |
| 53 | 54 | D16E;A26L;L45I;P91H;R132S;S244I;M247Y;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++ |
| 55 | 56 | D16E;A26L;L45I;P91H;R132S;E164R;S244I;M247H;L302M;G316K;Y318G;H319Y;D320K; | ++ |

| Polynucleotide SEQ ID No | Amino acid SEQ ID No | Residue difference relative to SEQ ID NO:2 | Activity or Stereoselectivity |
|---|---|---|---|
| 57 | 58 | D16E;A26L;T42M;E44D;L45I;P91H;R132S;E164R;S244I;M247Y;L302M;G316K;Y318G;H319Y;D320E; | ++ |
| 59 | 60 | D16E;A26L;T42M;E44D;L45I;P91H;R132S;E164R;S244I;M247H;G316K;Y318G;H319Y;D320K; | ++ |
| 61 | 62 | D16E;A26L;T42M;E44D;L45I;P91H;R132S;E164R;S244I;M247Y;L302M;G316K;Y318G;H319Y;D320K; | ++ |
| 63 | 64 | D16E;A26L;E44D;L45I;P91H;R132S;S244I;M247H;L302M;G316K;Y318G;H319Y;D320K; | ++ |
| 65 | 66 | D16E;L45I;P91H;R132S;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++ |
| 67 | 68 | D16E;A26L;L45I;P91H;R132S;E164R;S244I;M247Y;L302M;G316K;Y318G;H319Y;D320K;D352Y; | ++ |
| 69 | 70 | D16E;T42M;L45I;P91H;R132S;E164R;S244I;M247H;L302M;G316K;Y318G;H319Y;D320E; | ++ |
| 71 | 72 | D16E;A26L;L45I;T46H;P91H;R132S;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++ |
| 73 | 74 | D16E;A26L;E44D;L45I;P91H;R132S;S244I;M247Y;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++ |
| 75 | 76 | D16E;A26L;L45I;P91H;R132S;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++ |
| 77 | 78 | D16E;T42M;E44D;L45I;P91H;R132S;E164R;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320K; | ++ |
| 79 | 80 | D16E;E44D;L45I;P91H;R132S;E164R;S244I;M247H;L302M;G316K;Y318G;H319Y;D320E; | ++ |
| 81 | 82 | D16E;A26L;T42M;E44D;L45I;P91H;R132S;E164R;S182A;R237T;S244I;M247Y;L302M;G316K;Y318G;H319Y;D320K; | ++ |
| 83 | 84 | D16E;A26L;E44D;L45I;P91H;R132S;E164R;S244I;M247H;L302M;G316K;Y318L;H319Y;D320E; | ++ |
| 85 | 86 | D16E;G38P;L45I;P91H;R132S;S244I;M247Y;G316K;Y318G;H319Y;D320K; | ++ |
| 87 | 88 | D16E;A26L;T42M;E44D;L45I;P91H;R132S;E164R;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |
| 89 | 90 | D16E:A26L:T42M:L45I:P91H:R132S:E164R:S244I:M247H:L302M:A305P:G316K:Y3 | +++ |

| Polynucleotide SEQ ID No | Amino acid SEQ ID No | Residue difference relative to SEQ ID NO:2 | Activity or Stereoselectivity |
|---|---|---|---|
| | | 18G;H319Y;D320E; | |
| 91 | 92 | D16E;A26L;T42M;L45I;P91L;R132S;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |
| 93 | 94 | D16E;T42M;L45I;P91K;R132S;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |
| 95 | 96 | D16E;A19W;T42M;L45I;P91H;R132S;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |
| 97 | 98 | D16E;A19N;T42M;L45I;P91H;R132S;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |
| 99 | 100 | D16E;H32V;T42M;L45I;P91H;R132S;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |
| 101 | 102 | D16E;S33N;T42M;L45I;P91H;R132S;S244I;M247H;L302M;A305P;G316K;Y318G;H3 | +++ |
| | | 19Y;D320E; | |
| 103 | 104 | D16E;T42M;L45I;P91H;P118R;R132S;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |
| 105 | 106 | D16E;T42M;L45I;P91H;R132S;V154G;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |
| 107 | 108 | D16E;T42M;L45I;P91H;R132S;A218C;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |
| 109 | 110 | D16E;T42M;L45I;P91H;R132S;L217W;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |
| 111 | 112 | D16E;T42M;L45I;P91H;R132S;V216C;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |
| 113 | 114 | D16E;T42M;L45I;P91H;R132S;S244I;M247H;G292K;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |
| 115 | 116 | D16E;T42M;L45I;P91H;R132S;S244I;M247H;E294M;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |
| 117 | 118 | D16E;T42M;L45I;P91H;R132S;S244I;M247H;G291F;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |

| Polynucleotide SEQ ID No | Amino acid SEQ ID No | Residue difference relative to SEQ ID NO:2 | Activity or Stereoselectivity |
|---|---|---|---|
| 119 | 120 | D16E;T42M;L45I;P91H;R132S;S244I;M247H;G291V;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |
| 121 | 122 | D16E;T42M;L45I;P91H;R132S;L191H;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |
| 123 | 124 | D16E;G38S;T42M;L45I;P91H;R132S;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |
| 125 | 126 | D16E;T42M;L45I;P91H;R132S;S244I;M247H;E288I;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |
| 127 | 128 | D16E;T42M;L45I;P91H;R132S;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E;D352Q; | +++ |
| 129 | 130 | D16E;G38E;T42M;L45I;P91H;R132S;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |
| 131 | 132 | D16E;T42M;L45I;P91H;R132S;S244I;M247H;G285S;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |
| 133 | 134 | D16E;T42M;L45I;P91H;R132S;S244I;M247H;G285K;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |
| 135 | 136 | D16E;T42M;L45I;P91H;R132S;S244I;M247H;L284K;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |
| 137 | 138 | D16E;T42M;L45I;P91H;R132S;S244I;M247H;L284F;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |
| 139 | 140 | D16E;T42M;L45I;P91H;R132S;S244I;M247H;E282R;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |
| 141 | 142 | D16E;T42M;L45I;P91H;R132S;V189S;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |
| 143 | 144 | D16E;T42M;L45I;P91H;R132S;S244I;M247H;A295G;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++ |
| 145 | 146 | D16E;T42M;L45I;P91H;R132S;S244I;M247H;L302M;A305P;G316V;Y318G;H319Y;D320E; | +++ |
| 147 | 148 | D16E;T42M;L45I;P91H;R132S;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E;P324L; | +++ |

| Polynucleotide SEQ ID No | Amino acid SEQ ID No | Residue difference relative to SEQ ID NO:2 | Activity or Stereoselectivity |
|---|---|---|---|
| 149 | 150 | D16E;L45I;P91H;R132S;A185T;S244I;M247Y;G316K;Y318G;H319Y;D320K; | +++ |
| 151 | 152 | D16E;A19N;T42M;L45I;P91H;R132S;V154S;S244I;M247H;G291W;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++++ |
| 153 | 154 | D16E;A19N;S33N;T42M;L45I;P91H;R132S;V154S;V216C;L217W;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++++ |
| 155 | 156 | D16E;A19N;S33N;T42M;L45I;P91L;R132S;V154A;V216C;L217W;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++++ |
| 157 | 158 | D16E;A19N;S33N;T42M;L45I;P91H;R132S;V154S;A218C;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++++ |
| 159 | 160 | D16E;A19N;H32V;S33N;T42M;L45I;P91H;R132S;V154S;L217W;A218C;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++++ |
| 161 | 162 | D16E;A19N;H32V;S33N;T42M;L45I;P91H;R132S;V154S;V216C;L217W;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++++ |
| 163 | 164 | D16E;A19N;A26L;T42M;L45I;P91H;R132S;V216C;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++++ |
| 165 | 166 | D16E;A19N;H32V;S33N;T42M;L45I;P91H;R132S;V154A;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++++ |
| 167 | 168 | D16E;A19N;A26L;H32V;S33N;T42M;L45I;P91L;R132S;V154S;V216C;L217W;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++++ |
| 169 | 170 | D16E;A19N;S33N;G38E;T42M;L45I;P91H;R132S;S244I;M247H;G291W;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++++ |
| 171 | 172 | D16E;A19N;H32V;T42M;L45I;P91H;R132S;S244I;M247H;G291W;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++++ |
| 173 | 174 | D16E;A19N;T42M;L45I;P91H;R132S;V154S;V216C;L217W;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++++ |
| 175 | 176 | D16E;A19N;H32V;G38A;T42M;L45I;P91H;R132S;S244I;M247H;G291Y;E294M;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++++ |
| 177 | 178 | D16E;A19N;H32V;G38E;T42M;L45I;P91H;R132S;S244I;M247H;L262I;G291Y;E294M;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++++ |
| 179 | 180 | D16E;A19N;S33N;G38S;T42M;L45I;P91H;R132S;S244I;M247H;G291Y;E294M;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++++ |

| Polynucleotide SEQ ID No | Amino acid SEQ ID No | Residue difference relative to SEQ ID NO:2 | Activity or Stereoselectivity |
|---|---|---|---|
| | | | |
| 181 | 182 | D16E;A19N;S33N;G38A;T42M;L45I;P91H;R132S;S244I;M247H;E294M;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++++ |
| 183 | 184 | D16E;A19N;H32V;G38S;T42M;L45I;P91H;R132S;S244I;M247H;L262I;G291W;G292V;T293P;E294K;A295Q;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++++ |
| 185 | 186 | D16E;A19N;A37T;G38A;P39A;T42M;L45I;P91H;R132S;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++++ |
| 187 | 188 | D16E;A19N;A37M;G38P;P39A;T42M;L45I;P91H;R132S;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++++ |
| 189 | 190 | D16E;A19N;A37K;G38A;P39A;T42M;L45I;P91H;R132S;S244I;M247H;L302M;A305P;G316K;Y318G;H319Y;D320E; | ++++ |
| 191 | 192 | D16E;A19N;G38S;T42M;L45I;P91H;R132S;A218S;S244I;M247H;E288T;G291W;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++++ |
| 193 | 194 | D16E;A19N;G38S;T42M;L45I;P91H;R132S;A218C;S244I;M247H;L262I;G291W;G292K;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++++ |
| 195 | 196 | D16E;A19N;T42M;L45I;P91H;R132S;V154F;S244I;M247H;L302M;A305P;G316R;Y318G;H319Y;D320E;D352Q; | +++++ |
| 197 | 198 | D16E;A19N;T42M;L45I;P91H;R132S;V154R;S234R;S244V;M247H;L302M;A305P;G316V;Y318G;H319Y;D320E;D352A; | +++++ |
| 199 | 200 | D16E;A19N;T42M;L45I;P91H;R132S;V154A;S234R;S244V;M247H;L302M;A305P;G316V;Y318G;H319Y;D320E;D352Q; | +++++ |
| 201 | 202 | D16E;A19N;T42M;L45I;P91H;R132S;V154R;S234R;S244I;M247H;L302M;A305P;G316V;Y318G;H319Y;D320E;D352Q; | +++++ |
| 203 | 204 | D16E;A19N;H32V;G38E;T42M;L45I;P91H;R132S;S244I;M247H;L262I;E294M;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++++ |
| 205 | 206 | D16E;A19N;H32V;G38S;P39A;T42M;L45I;P91H;R132S;S244I;M247H;L262I;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++++ |
| 207 | 208 | D16E;A19N;T42M;L45I;P91H;A92W;R132S;S244I;M247H;E282R;L284A;G285K;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++++ |
| 209 | 210 | D16E;A19N;T42M;L45I;P91N;A92W;P118I;R132S;S244I;M247H;E282Y;L284A;G285K;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++++ |

| Polynucleotide SEQ ID No | Amino acid SEQ ID No | Residue difference relative to SEQ ID NO:2 | Activity or Stereoselectivity |
|---|---|---|---|
| | | | |
| 211 | 212 | D16E;A19N;T42M;L45I;P91N;P118I;R132S;S244I;M247H;E282K;L284A;G285P;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++++ |
| 213 | 214 | D16E;A19N;T42M;L45I;P91N;P118R;R132S;S244I;M247H;E282Y;L284A;G285P;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++++ |
| 215 | 216 | D16E;A19N;T42M;L45I;P91H;A92W;R132S;S244I;M247H;L284A;G285K;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++++ |
| 217 | 218 | D16E;A19N;T42M;L45I;P91H;R132S;S244I;M247H;E282R;L284A;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++++ |
| 219 | 220 | D16E;A19N;T42M;L45I;P91N;A92W;P118R;R132S;S244I;M247H;E282R;L284A;G285K;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++++ |
| 221 | 222 | D16E;A19N;T42M;L45I;P91N;A92W;R132S;S244I;M247H;E282Y;G285K;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++++ |
| 223 | 224 | D16E;A19N;T42M;L45I;P91H;A92W;R132S;S244I;M247H;E282K;L284A;G285P;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++++ |
| 225 | 226 | D16E;A19N;T42M;L45I;P91H;A92W;P118G;R132S;S244I;M247H;E282Y;L284V;G285P;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++++ |
| 227 | 228 | D16E;A19N;T42M;L45I;P91H;A92W;R132S;S244I;M247H;E282Y;L284A;G285P;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++++ |
| 229 | 230 | D16E;A19N;G38S;T42M;L45I;P91H;R132S;A218S;S244I;M247H;L262I;G291W;G292K;L302M;A305P;G316K;Y318G;H319Y;D320E; | +++++ |

Table 2

| Activity or Stereoselectivity | Description | Reaction condition |
|---|---|---|
| + | Conversion of substrate A1≥ 10%, de for A2 in product≥ 50%, reaction time ≤ 8 hours | Loading of enzyme powder 4g/L, loading of substrate A1 40g/L, loading of glycine 178g/L, 50μM PLP, 25%(v/v) Ethanol, 30°C |
| ++ | Conversion of substrate A1≥ 50%, de for A2 in product≥ 60%, reaction time ≤ 8 hours | Loading of enzyme powder 4g/L, loading of substrate A1 40g/L, loading of glycine 178g/L, 50μM PLP, 40%(v/v) Ethanol, 35°C |
| +++ | Conversion of substrate A1≥ 70%, de for A2 in product≥ 70%, reaction time ≤ 8 hours | Loading of enzyme powder 4g/L, loading of substrate A1 40g/L, loading of glycine 178g/L, 50μM PLP, 40%(v/v) Ethanol, 40°C |
| ++++ | Conversion of substrate A1≥ 80%, de for A2 in product≥ 80%, reaction time ≤ 8 hours | Loading of enzyme powder 6g/L, loading of substrate A1 100g/L, loading of glycine 178g/L, 50μM PLP, 40%(v/v) Ethanol, 45°C |
| +++++ | Conversion of substrate A1≥ 80%, de for A2 in product≥ 90%, reaction time | Loading of enzyme powder 9g/L, loading of substrate A1 200g/L, loading of glycine 178g/L, |

| Activity or Stereoselectivity | Description | Reaction condition |
|---|---|---|
| | ≤ 8 hours | 50μM PLP, 40%(v/v) Ethanol, 45°C |

**[0064]** The amino acid sequences listed in Table 1 (i.e., even sequence identifiers of SEQ ID NO: 2 to 230) each contain 357 amino acid residues. SEQ ID NO: 232, 234, or 236 has a different number of deletion or substitution of amino acid residues as compared to SEQ ID No: 2. The engineered aldolase polypeptides represented by SEQ ID NO: 232, 234, 236 exhibit higher stereoselectivity and/or activity than SEQ ID No: 2 under the reaction conditions of +, ++, +++, ++++ or +++++ as shown in Table 2.

**2.3 Polynucleotides, control sequences, expression vectors and host cells that can be used to produce engineered aldolase polypeptides**

**[0065]** In another aspect, this disclosure provides polynucleotides encoding engineered polypeptides having aldolase activity described herein. The polynucleotides can be linked to one or more heterologous regulatory sequences that control gene expression to produce recombinant polynucleotides that are capable of expressing the engineered polypeptides. Expression constructs comprising a heterologous polynucleotide encoding an engineered aldolase may be introduced into a suitable host cell to express the corresponding engineered aldolase polypeptide.

**[0066]** As apparent to one skilled in the art, the availability of protein sequences and knowledge of codons corresponding to a variety of amino acids provide an illustration of all possible polynucleotides that encode the protein sequence of interest. The degeneracy of the genetic code, in which the same amino acid is encoded by selectable or synonymous codons, allows for the production of an extremely large number of polynucleotides, all of which encode the engineered aldolase polypeptides disclosed herein. Thus, upon determination of a particular amino acid sequence, one skilled in the art can generate any number of different polynucleotides by merely modifying one or more codons in a manner that does not alter the amino acid sequence of the protein. In this regard, this disclosure specifically contemplates each and every possible alteration of a polynucleotide that can be made by selecting a combination based on possible codon selections, for any of the polypeptides disclosed herein, comprising those amino acid sequences of exemplary engineered polypeptides listed in Table 1 , and any of the polypeptides disclosed as even sequence identifiers of SEQ ID NOS: 4 to 236 in the Sequence Listing.

**[0067]** In various embodiments, the codons are preferably selected to accommodate the host cell in which the recombinant protein is produced. For example, codons preferred for bacteria are used to express genes in bacteria; codons preferred for yeast are used to express genes in yeast; and codons preferred for mammals are used for gene expression in mammalian cells.

**[0068]** For example, the polynucleotides encode polypeptides comprising amino acid sequences that are 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 that is an even sequence identifier of SEQ ID NO: 4-236. Wherein the polypeptides have aldolase activity and one or more of the improved properties described herein, for example, the ability to convert compound A1 to compound A2 with increased stereoselectivity compared to the polypeptide of SEQ ID NO: 2.

**[0069]** For example, the polynucleotides encode engineered aldolase polypeptides comprising amino acid sequences having a percentage of identity described above and having one or more amino acid residue differences as compared to SEQ ID NO: 2. In some examples, the present disclosure provides engineered polypeptides having aldolase activity, wherein the engineered polypeptides comprise a combination that has at least 80% sequence identity to the reference sequence of SEQ ID NO: 2 with residue differences that is selected from the following positions: X16, X17, X19, X26, X32, X33, X37, X38, X39, X41, X42, X43, X44, X45, X46, X47, X48, X49, X91, X92, X118, X132, X134, X154, X164, X168, X176, X182, X185, X189, X191, X216, X217, X218, X227, X234, X237, X244, X247, X262, X282, X284, X285, X288, X291, X292, X293, X294, X295, X302, X305, X316, X318, X319, X320, X324, X352.

**[0070]** In some examples, the polynucleotides encoding the engineered aldolase polypeptides comprises sequences having odd sequence identifier of SEQ ID NO: 3-235.

**[0071]** In some examples, the polynucleotides encode polypeptides as described herein; but at the nucleotide level, the polynucleotides have about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98 %, 99% or more sequence identity to reference polynucleotides encoding engineered aldolase

polypeptides as described herein. In some examples, the reference polynucleotides are selected from the sequences having the odd sequence identifiers of SEQ ID NOs: 3 - 235.

**[0072]** The isolated polynucleotides encoding engineered aldolase polypeptides can be manipulated to enable the expression of the engineered polypeptides in a variety of ways, which comprises further modification of the sequences by codon optimization to improve expression, insertion into suitable expression elements with or without additional control sequences, and transformation into a host cell suitable for expression and production of the engineered polypeptides.

**[0073]** Depending on the expression vector, manipulation of the isolated polynucleotide prior to insertion of the isolated polynucleotide into the vector may be desirable or necessary. Techniques for modifying polynucleotides and nucleic acid sequences using recombinant DNA methods are well known in the art. Guidance is provided below: Sambrook et al., 2001, Molecular Cloning: A Laboratory Manual, Third Edition, Cold Spring Harbor Laboratory Press; and Current Protocols in Molecular Biology, Ausubel. F. Eds., Greene Pub. Associates, 1998, 2010 Year update.

**[0074]** In another aspect, this invention also relates to recombinant expression vectors, depending on the type of host they are to be introduced into, including a polynucleotide encoding an engineered aldolase polypeptide or variant thereof, and one or more expression regulatory regions, such as promoters and terminators, origin of replication and the like. Alternatively, the nucleic acid sequence of the present disclosure can be expressed by inserting the nucleic acid sequence or the nucleic acid construct comprising the sequence into an appropriate expression vector. In generating the expression vector, the coding sequence is located in the vector such that the coding sequence is linked to a suitable control sequence for expression.

**[0075]** The recombinant expression vector can be any vector (e.g., a plasmid or virus) that can be conveniently used in recombinant DNA procedures and can result in the expression of a polynucleotide sequence. The choice of vector will generally depend on the compatibility of the vector with the host cell to be introduced into. The vector can be linear or closed circular plasmid. The expression vector may be an autonomously replicating vector, i.e., a vector that exists as an extrachromosomal entity whose replication is independent of chromosomal replication such as plasmids, extrachromosomal elements, minichromosomes, or artificial chromosomes. The vector may contain any tools for ensuring self-copying. Alternatively, the vector may be a vector that, when introduced into a host cell, integrates into the genome and replicates with the chromosome into which it is integrated. Moreover, a single vector or plasmid or two or more vectors or plasmids that together comprise the total DNA to be introduced into the genome of the host cell may be used.

**[0076]** Many expression vectors useful to the embodiments of the present disclosure are commercially available. An exemplary expression vector can be prepared by inserting a polynucleotide encoding an engineered aldolase polypeptide to plasmid pACYC-Duet-1 (Novagen).

**[0077]** In another aspect, this invention provides host cells comprising polynucleotides encoding engineered aldolase polypeptides of the present disclosure. The polynucleotide is linked to one or more control sequences for expression of aldolase polypeptides in a host cell. Host cells for expression of polypeptides encoded by the expression vectors of the present disclosure are well known in the art, including, but not limited to, bacterial cells such as E. coli, Arthrobacter KNK168, Streptomyces, and Salmonella typhimurium cells; fungal cells such as yeast cells (e.g., Saccharomyces cerevisiae or Pichia pastoris); insect cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as CHO, COS, BHK, 293 and Bowes melanoma cells; and plant cells. An exemplary host cell is E. coli BL21 (DE3). The above host cells may be wild-type or may be engineered cells through genomic edition, such as knockout of the wild-type aldolase gene carried in the host cell's genome. Suitable media and growth conditions for the above host cells are well known in the art.

**[0078]** Polynucleotides used to express engineered aldolases can be introduced into cells by a variety of methods known in the art. Techniques comprise, among others, electroporation, bio-particle bombardment, liposome-mediated transfection, calcium chloride transfection, and protoplast fusion. Different methods of introducing polynucleotides into cells are obvious to those skilled in the art.

**2.4 Process of producing an engineered aldolase polypeptide**

**[0079]** Engineered aldolase can be obtained by subjecting a polynucleotide encoding an aldolase to mutagenesis and/or directed evolution. An exemplary directional evolution technique can be found in "Biocatalysis for the Pharmaceutical Industry: Discovery, Development, and Manufacturing" (2009 John Wiley & Sons Asia (Pte) Ltd. ISBN: 978-0-470-82314-9).

**[0080]** When the sequence of an engineered polypeptide is known, the encoding polynucleotide may be prepared by standard solid-phase methods according to known synthetic methods. In some examples, fragments of up to about 100 bases can be

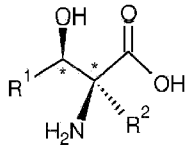
synthesized separately and then ligated (e.g., by enzymatic or chemical ligation methods or polymerase-mediated methods) to form any desired contiguous sequence. For example, the polynucleotides and oligonucleotides of the present disclosure can be prepared by chemical synthesis using, for example, the classic phosphoramidite methods described by Beaucage et al., 1981, Tet Lett 22: 1859-69, or Matthes et al. People, 1984, EMBO J. 3: 801-05, as typically practiced in automated synthesis methods. According to the phosphoramidite method, oligonucleotides are synthesized, purified, annealed, ligated, and cloned into a suitable vector, for example, in an automated DNA synthesizer. In addition, essentially any nucleic acid is available from any of a variety of commercial sources.

**[0081]** In some embodiments, the present invention also provides a process for preparing or producing an engineered aldolase polypeptide that is capable of converting Compound A1 to Compound A2 under suitable reaction conditions, wherein the process comprises culturing a host cell capable of expressing a polynucleotide encoding an engineered polypeptide under culture conditions suitable for the expression of the polypeptide. In some embodiments, the process of preparing a polypeptide further comprises isolating the polypeptide. Engineered polypeptides may be expressed in suitable cells and isolated (or recovered) from the host cell and/or culture medium using any one or more of the well-known techniques for protein purification, the techniques for protein purification include, among others, lysozyme treatment, sonication, filtration, salting out, ultracentrifugation and chromatography.

**2.5 Methods of using an engineered aldolase and compounds prepared therewith**

**[0082]** In another aspect, the engineered aldolase polypeptides described herein can asymmetrically condense aldehyde substrates and amino acid substrates. The present disclosure also provides process of preparing a wide range of compounds (I) or structural analogs thereof using an engineered aldolase polypeptide disclosed herein.

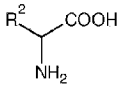
**[0083]** According to the invention, engineered aldolase polypeptides according to the invention are used in a process of preparing a compound of structural formula (I):



(I)

the β-hydroxy-α-amino acid product of formula (I) has the indicated stereochemical configuration at the chiral center marked with an *; the β-hydroxy-α-amino acid product of formula (I) is in diastereomeric excess over the other isomers, where $R^1$ is optionally substituted or unsubstituted aryl or heteroaryl, or optionally substituted or unsubstituted $C_1$-$C_8$ alkyl; $R^2$ is -H, -$CH_2OH$, -$CH_2SH$, -$CH_2SCH_3$, or optionally substituted or unsubstituted $C_1$-$C_4$ hydrocarbyl. The process herein comprises that, under reaction conditions suitable for converting the aldehyde substrate and the amino acid substrate to β-hydroxy-α-amino acid product, the aldehyde substrate of formula (II) and the amino acid substrate of formula (III)

(II)



(III)

are contacted with an aldolase polypeptide, wherein the aldolase polypeptide is an engineered aldolase polypeptide of the invention.

**[0084]** In some examples, the engineered aldolase polypeptides have at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more of sequence identity with SEQ ID NO:2, and are capable of condensing the aldehyde substrate of formula (II) and the amino acid substrate of formula (III) to form β-Hydroxy-a-amino acid product of formula (I) with a higher conversion and/or higher stereoselectivity than SEQ ID NO: 2.

**[0085]** According to the invention, the β-hydroxy-α-amino acid product of formula (I) is present in a diastereomeric excess of at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater.

**[0086]** As noted above, aldolase polypeptides useful in the process of the present disclosure may be characterized according to the ability of condensation of p-nitrobenzaldehyde and glycine to (2S, 3R)-2-amino-3-hydroxy-3-(4-nitrophenyl) propanoic acid. Thus, in any of the embodiments of the process disclosed herein, the process is carried out, wherein the engineered aldolase polypeptides are capable of condensing p-nitrobenzaldehyde and glycine to (2S, 3R)-2-amino-3-hydroxy-3-(4-nitrophenyl) propanoic acid with a higher conversion and/or higher stereoselectivity than SEQ ID NO: 2 , and comprise an

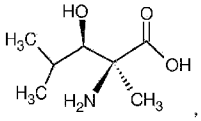
amino acid sequence selected from the group consisting of SEQ ID NOs: 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, and 230.

**[0087]** In some embodiments of the above process, $R^1$ is optionally substituted or unsubstituted $C_1$-$C_8$ alkyl. In some embodiments, $R^1$ is optionally substituted or unsubstituted phenyl. In some embodiments, $R^1$ is optionally substituted or unsubstituted pyridyl. In some embodiments, $R^1$ is optionally substituted or unsubstituted aryl or heteroaryl. In some embodiments, $R^1$ is optionally substituted or unsubstituted phenyl, and substitution occurs at either (ortho, meta or para) of the phenyl ring or any two of the substitutions occurring simultaneously on the phenyl ring, the substituents are selected from the group consisting of$C_1$-$C_4$ hydrocarbyl, halogen (e.g., -F, -Cl, -Br and -I), -$NO_2$, -NO, -$SO_2$R' or -SOR', -SR', -NR'R', -OR', -$CO_2$R' or -COR', -C(O)NR', -$SO_2NH_2$ or -$SONH_2$, -CN, $CF_3$, wherein each R' is independently selected from -H or ($C_1$-$C_4$) alkyl. In some embodiments, ($C_1$-$C_4$) alkyl is a halogen-substituted hydrocarbon.

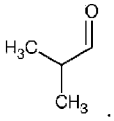
**[0088]** In some embodiments of the above process, $R^2$ is -H, -$CH_3$, -$CH_2CH_3$, -CH ($CH_3$) $_2$, -$CH_2OH$, -$CH_2SH$, or -$CH_2SCH_3$.

**[0089]** According to the invention, the β-hydroxy-α-amino acid product of formula (I) is present in a diastereomeric excess of at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater.

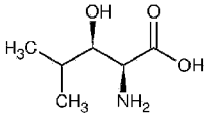
**[0090]** In some embodiments of this process, the β-hydroxy-α-amino acid product of formula (I) is (2S, 3R) - (+) - 2-amino-3-hydroxy-2,4-methylpentanoic acid:



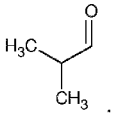
and the amino acid substrate of formula (III) is L-alanine, the aldehyde substrate of formula (II) is isobutyraldehyde:



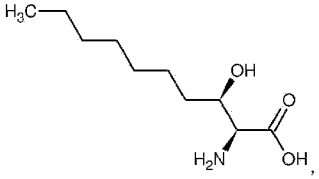
**[0091]** In some embodiments of this process, the β-hydroxy-α-amino acid product of formula (I) is (2S, 3R) - (+) - 2-amino-3-hydroxy-4-methylpentanoic acid:



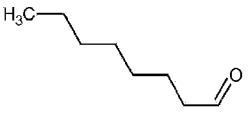
and the amino acid substrate of formula (III) is glycine, the aldehyde substrate of formula (II) is isobutyraldehyde:



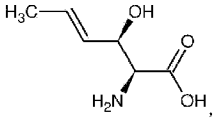
**[0092]** In some embodiments of this process, the β-hydroxy-α-amino acid product of formula (I) is (2S, 3R) -2-amino-3-hydroxydecanoic acid:



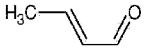
and the amino acid substrate of formula (III) is glycine, the aldehyde substrate of formula (II) is n-octanal:



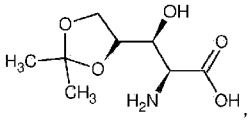
**[0093]** In some embodiments of this process, the β-hydroxy-α-amino acid product of formula (I) is (2S, 3R, 4E) -2-amino-3-hydroxy-4-hexenoic acid:



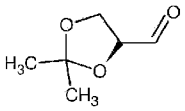
and the amino acid substrate of formula (III) is glycine, the aldehyde substrate of formula (II) is crotonaldehyde:



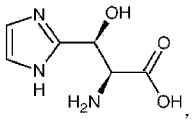
**[0094]** In some embodiments of this process, the β-hydroxy-α-amino acid product of formula (I) is (2S, 3S) -2- amino-3- [(4S) 3-dioxolan-4-yl] -3-hydroxypropanoic acid:



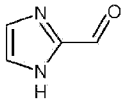
and the amino acid substrate of formula (III) is glycine, the aldehyde substrate of formula (II) is (4S) -2,2-dimethyl-1,3-dioxolane-4-carbaldehyde:



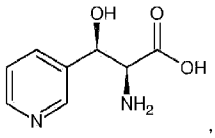
**[0095]** In some embodiments of this process, the β-hydroxy-α-amino acid product of formula (I) is (2S, 3S) -2-amino-3-hydroxy-3- (1H- imidazol-2-yl) propanoic acid:



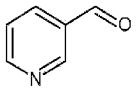
and the amino acid substrate of formula (III) is glycine, the aldehyde substrate of formula (II) is 1H-imidazole-2-carboxaldehyde:



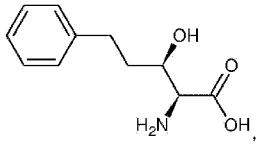
**[0096]** In some embodiments of this process, the β-hydroxy-α-amino acid product of formula (I) is (2S, 3R) -2-amino-3-hydroxy-3- (pyridin-3-yl) propanoic acid:



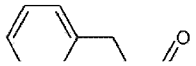
and the amino acid substrate of formula (III) is glycine, the aldehyde substrate of formula (II) is pyridine carboxaldehyde:



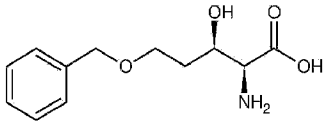
**[0097]** In some embodiments of this process, the β-hydroxy-α-amino acid product of formula (I) is (2S, 3R) -2-amino-3-hydroxy-5-phenylpentanoic acid:



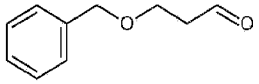
and the amino acid substrate of formula (III) is glycine, the aldehyde substrate of formula (II) is 3-phenylpropionaldehyde:



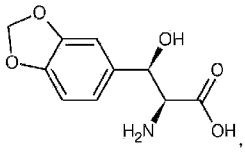
**[0098]** In some embodiments of this process, the β-hydroxy-α-amino acid product of formula (I) is (2S, 3R) -2-amino-5- (benzyloxy) -3-hydroxyvaleric acid:



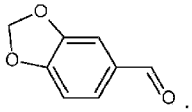
and the amino acid substrate of formula (III) is glycine, the aldehyde substrate of formula (II) is 3- (benzyloxy) propanal:



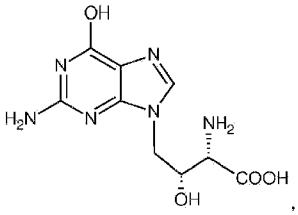
**[0099]** In some embodiments of this process, the β-hydroxy-α-amino acid product of formula (I) is (2S, 3R) -2-amino-3- (1,3-1,3-benzodioxol-5-yl) -3-hydroxypropanoic acid:



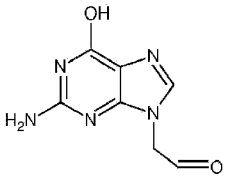
and the amino acid substrate of formula (III) is glycine, the aldehyde substrate of formula (II) is methylenedioxybenzene-5-carbaldehyde:



**[0100]** In some embodiments of this process, the β-hydroxy-α-amino acid product of formula (I) is (2S, 3R) -2-amino- 4- (2-amino-6-hydroxy-9H-purin-9- Yl) -3-hydroxybutyrate:

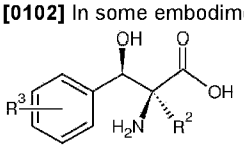


and the amino acid substrate of formula (III) is glycine, the aldehyde substrate of formula (II) is (2-amino-6-hydroxy-9H-purin-9-yl) acetaldehyde:



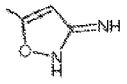
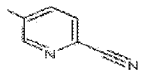
**[0101]** According to the invention, the β-hydroxy-α-amino acid product of Formula (I) produced in the above process is present in a diastereomeric excess of at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more.

**[0102]** In some embodiments, the β-hydroxy-α-amino acid product of structural formula (I) is:



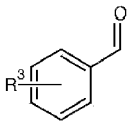
wherein $R^3$ is $C_1$-$C_4$ hydrocarbyl, -H, halogen(such as -F, -Cl, -Br and -I), $-NO_2$, -NO, $-SO_2R'$ or -SOR', -SR', -NR'R', -OR', $-CO_2R'$ or -COR', -C(O)NR', $-SO_2NH_2$ or $-SONH_2$, -CN, $CF_3$, wherein each R' is independently selected from -H or ( $C_1$-$C_4$) hydrocarbyl;

$R^3$ may also be

or

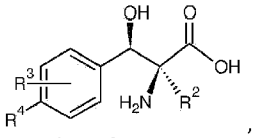
$R^2$ is -H, $-CH_3$, $-CH_2CH_3$, $-CH(CH_3)_2$, $-CH_2OH$, $-CH_2SH$ or $-CH_2SCH_3$, the aldehyde substrate of formula (II) is:



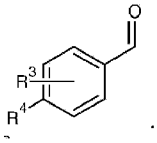
**[0103]** In some embodiments, $R^3$ is in the para position of the phenyl ring. In some embodiments, $R^3$ is in the meta position of the phenyl ring. In some embodiments, $R^3$ is ortho to the phenyl ring. In some embodiments, $R^3$ is both para and meta to the phenyl ring. In some embodiments, $R^3$ is both para and ortho to the phenyl ring. In some embodiments, $R^3$ is both meta and ortho to the phenyl ring.

**[0104]** According to the invention, the β-hydroxy-α-amino acid product of Formula (I) produced in the above process is present in a diastereomeric excess of at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% or more.

**[0105]** In some embodiments of this process, the β-hydroxy-α-amino acid product of formula (I) is:

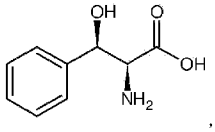


**[0106]** Wherein $R^4$ is $R^3$ as defined above, $R^3$ and $R^2$ are as defined above, the aldehyde substrate of formula (II) is:

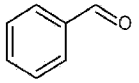


**[0107]** In some embodiments, $R^3$ is in the meta position of the phenyl ring. In some embodiments, $R^3$ is ortho to the phenyl ring. According to the invention, the β-hydroxy-α-amino acid product of Formula (I) produced in the above process is present in a diastereomeric excess of at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more.

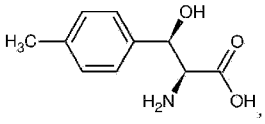
**[0108]** In some embodiments of this process, the β-hydroxy-α-amino acid product of formula (I) is (2S, 3R) -2-amino-3-hydroxy-3-benzoic acid:



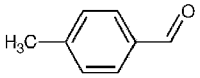
and the amino acid substrate of formula (III) is glycine, the aldehyde substrate of formula (II) is benzaldehyde:



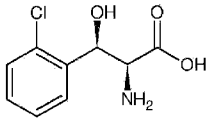
**[0109]** In some embodiments of this process, the β-hydroxy-α-amino acid product of formula (I) is (2S, 3R) -2- amino-3-hydroxy-3-(4-methylphenyl) propanoic acid :



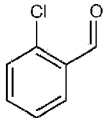
and the amino acid substrate of formula (III) is glycine, the aldehyde substrate of formula (II) is 4-methylbenzaldehyde:



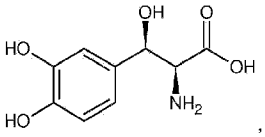
**[0110]** In some embodiments of this process, the β-hydroxy-α-amino acid product of structural formula (I) is (2S, 3R) -2-amino-3- (2- chlorophenyl) -3-hydroxypropanoic acid:



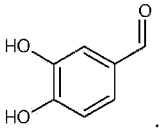
and the amino acid substrate of formula (III) is glycine, the aldehyde substrate of formula (II) is 2-chlorobenzaldehyde:



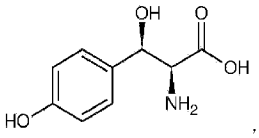
**[0111]** In some embodiments of this process, the β-hydroxy-α-amino acid product of formula (I) is (2S, 3R) -2-amino-3- (3,4-dihydroxybenzene) -3-hydroxypropanoic acid:



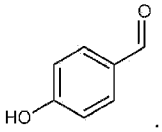
and the amino acid substrate of formula (III) is glycine and the aldehyde substrate of formula (II) is 3,4-dihydroxybenzaldehyde:



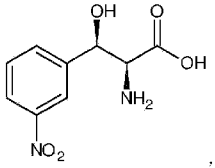
**[0112]** In some embodiments of this process, the β-hydroxy-α-amino acid product of formula (I) is (2S, 3R) -2-amino-3-hydroxy-3- (4- hydroxyphenyl) propanoic acid:



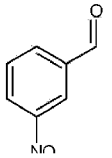
and the amino acid substrate of formula (III) is glycine, the aldehyde substrate of formula (II) is 4-hydroxybenzaldehyde:



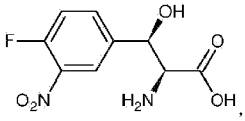
**[0113]** In some embodiments of this process, the β-hydroxy-α-amino acid product of formula (I) is (2S, 3R) -2-amino-3-hydroxy-3- (3-nitrophenyl) propanoic acid:



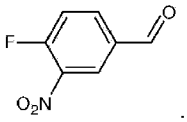
and the amino acid substrate of formula (III) is glycine, the aldehyde substrate of formula (II) is 3-nitrobenzaldehyde:



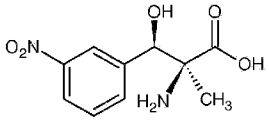
**[0114]** In some embodiments of this process, the β-hydroxy-α-amino acid product of formula (I) is (2S, 3R) -2-amino-3- (4-fluoro-3-nitrophenyl) -3-hydroxypropanoic acid:



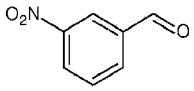
and the amino acid substrate of formula (III) is glycine and the aldehyde substrate of formula (II) is 4-fluoro-3-nitrobenzaldehyde:



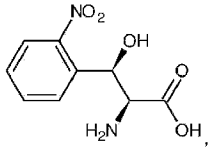
**[0115]** In some embodiments of this process, the β-hydroxy-α-amino acid product of formula (I) is (2S, 3R) -2-amino-3-hydroxy-2-methyl-3-(3-nitrophenyl) propanoic acid:



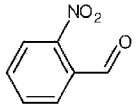
and the amino acid substrate of formula (III) is L-alanine, the aldehyde substrate of formula (II) is 3-nitrobenzaldehyde:



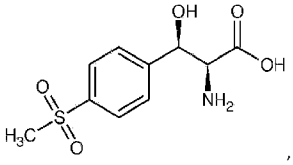
**[0116]** In some embodiments of this process, the β-hydroxy-α-amino acid product of formula (I) is (2S, 3R) -2-amino-3-hydroxy-3- (2-nitrophenyl) propanoic acid:



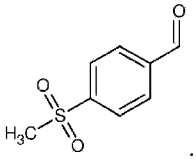
and the amino acid substrate of formula (III) is glycine, the aldehyde substrate of formula (II) is 2-nitrobenzaldehyde:



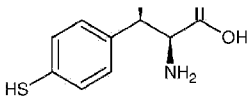
**[0117]** In some embodiments of this process, the β-hydroxy-α-amino acid product of formula (I) is (2S, 3R) -3- [p-(methylsulfonyl) phenyl] 3-hydroxy-2-amino-propanoic acid:



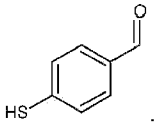
and the amino acid substrate of formula (III) is glycine, the aldehyde substrate of formula (II) is p-methyl sulfone benzaldehyde:



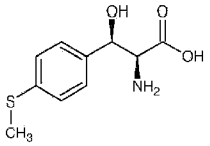
**[0118]** In some embodiments of this process, the β-hydroxy-α-amino acid product of formula (I) is (2S, 3R) -2-amino-3-hydroxy-3- (4- mercaptophenyl) propanoic acid:



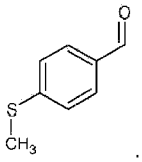
and the amino acid substrate of formula (III) is glycine, the aldehyde substrate of formula (II) is 4-mercaptobenzaldehyde:



**[0119]** In some embodiments of this process, the β-hydroxy-α-amino acid product of formula (I) is (2S, 3R) -2-amino-3-hydroxy-3- (4-mercaptomethylbenzene) propanoic acid:

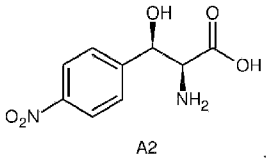


and the amino acid substrate of formula (III) is glycine, the aldehyde substrate of formula (II) is 4-mercaptomethylberzaldehyde: According to the invention,



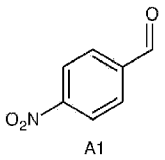
the β-hydroxy-α-amino acid product of Formula (I) produced in the above process is present in a diastereomeric excess of at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more .

**[0120]** According to the invention, the improved engineered aldolase polypeptide of the invention is used in the preparation of a diastereomeric excess of the compound of formula A2 (2S, 3R)-2-amino-3-hydroxy-3-(4-nitrophenyl) propanoic acid:



A2

**[0121]** In these embodiments, the process comprises that, in a suitable organic solvent, in the presence of glycine, under reaction conditions suitable for converting the compound of formula A1 to a compound of formula A2, the compound of formula A1 is contacted with the engineered aldolase polypeptides disclosed herein.



A1

**[0122]** According to the invention, the compound of Formula (I) or the compound of Formula A2 is present in a diastereomeric excess of at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93% 94%, 95%, 96%, 97%, 98%, 99% or greater. Engineered aldolase polypeptides that can be used in the above process comprise amino acid sequences selected from SEQ ID NO: 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, and 230.

**[0123]** As described herein and exemplified in the examples, the present disclosure contemplates a range of suitable reaction conditions that may be used in the process herein, including but not limited to pH, temperature, buffers, solvent systems, substrate loadings, mixtures of product diastereomers, polypeptide loading, cofactor loading, pressure, and reaction time. Additional suitable reaction conditions for performing a method of biocatalytically converting substrate compounds to a product compound using engineered aldolase polypeptides described herein can be readily optimized by routine experimentation, which including but not limited to that the engineered aldolase polypeptide is contacted with substrate compounds under

experimental reaction conditions of varying concentration, pH, temperature, solvent conditions, and the product compound is detected, for example, using the methods described in the Examples provided herein.

**[0124]** As described above, engineered polypeptides having aldolase activity for use in the process of the present disclosure generally comprises amino acid sequences that have at least 80%, 85 %, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% sequence identity to the reference amino acid sequence selected from any one of the even numbered sequences of SEQ ID NO: 4 to 236.

**[0125]** The substrate compounds in the reaction mixture can be varied, taking into consideration of, for example, the amount of the desired product compound, the effect of the substrate concentration on the enzyme activity, the stability of the enzyme under the reaction conditions, and the percent conversion of substrate to product. In some examples of the process, the suitable reaction conditions include at least about 0.5 to about 400 g/L, about 1 to about 400 g/L, about 5 to about 400 g/L, about 10 to about 400 g/L, or about 50 to about 400 g/L of loading of substrate (II) or substrate A1. In some examples, suitable reaction conditions include 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 100 g/L, at least about 150 g/L, at least about 200 g/L, at least about 250 g/L, at least about 300 g/L, at least about 350 g/L, at least about 400 g/L or even more of loading of substrate (II) or substrate A1. The values for the substrate loading provided herein are based on the molecular weight of compound (II) or A1, however it is also contemplated that the equivalent molar amounts of various hydrates and salts of compound (II) or A1 may also be used in the process.

**[0126]** In the process described herein, the engineered aldolase polypeptides use an amino acid and an aldehyde compound to form a product compound. In some embodiments, the amino acids in the reaction conditions include compounds selected from glycine, D,L-alanine, D,L-serine, D,L-cysteine, D,L-leucine, D,L-isoleucine, D,L-methionine, D,L-threonine or D,L-valine. In some embodiments, the amino acid is glycine. In some embodiments, suitable reaction conditions include amino acids present in a loading of at least about 1 times of the molar loading of substrate (II). In some embodiments, glycine is present at a loading of 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 times of the molar loading of substrate (II).

**[0127]** Suitable reaction conditions for the process generally also include the presence of a cofactor in the reaction mixture. Because the engineered aldolases typically use members of the vitamin B6 family, the reaction conditions may include one or more compounds 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 examples, suitable reaction conditions may include a cofactor selected from the group consisting of 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 examples, the cofactor is PLP. Accordingly, in some examples, suitable reaction conditions may include 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 examples, the reaction conditions include 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 a PLP concentration of about 0.2 g/L or less.

**[0128]** In some examples 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 examples of the process (e.g., using partially purified, or purified aldolase), the process may further include the step of adding cofactor to the enzymatic reaction mixture. In some examples, cofactor is added either at the beginning of the reaction and/or additional cofactor is added during the reaction.

**[0129]** In the embodiments of the reaction, the reaction conditions may include a suitable pH. As noted above, the desired pH or desired pH range can be maintained by using an acid or base, a suitable buffer, or a combination of buffer and added acid or base. The pH of the reaction mixture can be controlled before and/or during the reaction. In some embodiments, suitable reaction conditions include a solution pH of about 4 to about 8, a pH of about 5 to about 7, a pH of about 6 to about 7. In some embodiments, the reaction conditions include a solution pH of about 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5 or 8.

**[0130]** In embodiments of the processes herein, suitable temperatures can be used for the reaction conditions, taking into consideration of, for example, the increase in reaction rate at higher temperatures, the activity of the enzyme for sufficient duration of the reaction. Accordingly, in some embodiments, suitable reaction conditions include a temperature of about 10°C to about 60 °C, about 25°C to about 50 °C, about 25°C to about 40 °C, or about 25°C to about 30 °C. In some embodiments, suitable reaction temperatures include a temperature of about 25 °C, 30 °C, 35 °C, 40 °C, 45 °C, 50 °C, 55 °C, or 60 °C. In some embodiments, the temperature during the enzymatic reaction can be maintained at a certain temperature throughout the reaction. In some embodiments, the temperature during the enzymatic reaction may be adjusted over a temperature profile during the course of the reaction.

**[0131]** The processes of using the engineered aldolases are generally carried out in a solvent. Suitable solvents include water,

aqueous buffer solutions, organic solvents, and/or co-solvent systems, which generally include aqueous solvents and organic solvents. The aqueous solutions (water or aqueous co-solvent systems) can be pH-buffered or unbuffered. In some examples, the processes of using an engineered aldolase polypeptide are generally carried out in an aqueous co-solvent system comprising an organic solvent (e.g., methanol, ethanol, propanol, isopropanol (IPA), dimethyl sulfoxide (DMSO), dimethylformamide (DMF), isopropyl acetate, ethyl acetate, butyl acetate, 1-octanol, heptane, octane, methyl tert-butyl ether (MTBE) , Toluene, etc.), ionic liquids (for example, 1-ethyl 4-methylimidazolium tetrafluoroborate, 1-butyl-3-methylimidazolium tetrafluoroborate, 1-butyl-3-methylimidazolium hexafluorophosphate, and the like). The organic solvent component of the aqueous co-solvent system may be miscible with the aqueous component, providing a single liquid phase, or may be partially miscible or immiscible with the aqueous component, providing two liquid phases. Exemplary aqueous co-solvent system comprises water and one or more organic solvents. In general, the organic solvent component of the aqueous co-solvent system is selected such that it does not completely inactivate the aldolase. Suitable co-solvent system can be readily identified by measuring the enzymatic activity of a particular engineered aldolase with a defined substrate of interest in the candidate solvent system, utilizing enzymatic activity assays, such as those described herein. In some examples of the process, suitable reaction conditions include an aqueous co-solvent comprising ethanol at a concentration of about 1% to about 100% (v/v), about 1% to about 60% (v/v), about 2% to about 60% (v/v), about 5% to about 60% (v/v), from about 10% to about 60% (v/v), from about 10% to about 50% (v/v) ,or about 10% to about 40% (v/v). In some examples of the process, suitable reaction conditions include an aqueous co-solvent comprising ethanol at a concentration of at least about 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55 %, or 60% (v/v).

**[0132]** Suitable reaction conditions can include a combination of reaction parameters that provide for the biocatalytic conversion of the substrate compounds to its corresponding product compound. Accordingly, in some examples of the process, the combination of reaction parameters comprises: (a) substrate A1 loading of about 10 g/L to about 200 g/L; (b) glycine loading is about 3 to 10 times the molar amount of substrate A1; (C) engineered polypeptide concentration of about 0.5 g/L to 10 g/L; (d) PLP cofactor concentration of about 0.1 mM to 10 mM; (e) DMSO or ethanol concentration of about 20% (v/v) to about 60% (v/v); (f) pH of about 4.0 to 8.0; and(g) temperature of about 10°C to 60°C.

**[0133]** Exemplary reaction conditions include the assay conditions provided in Table 2 and Example 3.

**[0134]** In carrying out the reaction described herein, the engineered aldolase polypeptide may be added to the reaction mixture in the partially purified or purified forms, whole cells transformed with the gene encoding the engineered aldolase polypeptide, and/or as cell extracts and/or lysates of such cells. Whole cells transformed with the gene encoding the engineered aldolase or cell extracts, lysates thereof, and isolated enzymes can be used in a wide variety of different forms, including solids (e.g., lyophilized, spray dried, or the like) or semisolid (e.g., a crude paste). The cell extract or cell lysate may be partially purified by precipitation (e.g., ammonium sulfate, polyethyleneimine, heat treatment or the like), followed by desalting procedures (e.g., ultrafiltration, dialysis, and the like) prior to lyophilization. Any of the enzyme preparations can be stabilized by crosslinking using known crosslinking agents, such as glutaraldehyde, or immobilization to a solid phase material (such as a resin).

**[0135]** In some embodiments of the reactions described herein, the reaction is performed under suitable reaction conditions described herein, wherein the engineered aldolase polypeptide is immobilized to a solid support. Solid supports useful for immobilizing the engineered aldolase enzyme for carrying out the reaction include but are not limited to beads or resins such as polymethacrylates with epoxy functional groups, polymethacrylates with amino epoxy functional groups, polymethacrylates, styrene/DVB copolymer or polymethacrylates 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.

**[0136]** In some examples, wherein an engineered polypeptide is expressed in the form of a secreted polypeptide, a culture medium containing the secreted polypeptide can be used in the process herein.

**[0137]** In some examples, the solid reactants (e.g., enzymes, salts, etc.) can be provided to the reaction in a variety of different forms, including powders (e.g., lyophilized, spray dried, etc.), solutions, emulsions, suspensions and the like. The reactants can be readily lyophilized or spray-dried using methods and instrumentation known to one skilled in the art. For example, the protein solution can be frozen at -80 °C in small aliquots, and then added to the pre-chilled lyophilization chamber, followed by the application of a vacuum.

**[0138]** In some examples, the order of addition of reactants is not critical. The reactants may be added together to the solvent at the same time (e.g., monophasic solvent, a biphasic aqueous co-solvent system, etc.), or alternatively, some reactants may be added separately, and some may be added together at different time points. For example, the cofactor, aldolase, and substrates may be added first to the solvent. For improved mixing efficiency when using aqueous co-solvent systems, aldolase

and cofactors may be added and mixed into the aqueous phase first. The organic phase can then be added and mixed in, followed by addition of the substrates. Alternatively, the substrates can be premixed in the organic phase prior to addition to the aqueous phase.

**[0139]** Different features and embodiments of the present disclosure are exemplified in the following representative examples, which are intended to be illustrative.

**3. Examples**

**[0140]** The following examples further illustrate the present invention.

**[0141]** In the following examples, experimental methods with conditions not specified, were conducted at the commonly used conditions or according to the supplier's' suggestion.

**Example 1: Gene Cloning and Construction of Expression Vectors**

**[0142]** The amino acid sequence of the wild-type aldolase from *Pseudomonas putida* can be retrieved from NCBI, and the corresponding nucleic acids were then synthesized by a vendor using conventional techniques in the art and cloned into the expression vector pACYC-Duet-1. The recombinant expression plasmid was transformed into E. coli BL21 (DE3) competent cells under the conditions of 42 °C and thermal shock for 90 seconds. The transformation solution was plated on LB agar plates containing chloramphenicol which was then incubated overnight at 37 °C. Recombinant transformants were obtained.

**Example 2: Recombinant expression of aldolase polypeptides**

**[0143]** The resulting transformant such as recombinant E. coli BL21 (DE3) from example 1 was inoculated into LB medium containing chloramphenicol (peptone 10 g/L, yeast extract 5 g/L, NaCl 10 g/L, pH 7.0) which was then cultured in a shaking incubator at 30 °C, 250rpm overnight. The overnight culture was subcultured into a 1 L flask containing 250 mL of TB medium (tryptone 12 g/L, yeast extract 24 g/L, glycerol 4 mL/L, PBS) at 30 °C, 250rpm in a shaking incubator. When the $OD_{600}$ of subculture broth reached 0.6 ~ 0.8, IPTG was added to induce the expression of recombinant aldolase at a final concentration of 0.1 mmol/L. After expression overnight, the culture was centrifuged to get resting cells. The pelleted resting cells were suspended in a pH 7.4 buffer, and then sonicated in an ice bath to get cell lysate., The supernatant of cell lysate was collected by centrifugation as a crude enzyme solution of the recombinant aldolase, and the supernatant was further freeze-dried using a lyophilizer to obtain crude enzyme powder.

**[0144]** According to the recombinant expression process using shaking flasks as mentioned above, a miniaturized expression process in 96-well plate was performed by proportionally reducing the scale. The crude enzyme solution was obtained through chemical lysis rather than ultrasonication.

**Example 3: Reaction conditions and analytical methods for measuring activity and stereoselectivity of aldolase polypeptides**

**[0145]** p-nitrobenzaldehyde was added at a final concentration of 7.5 g/L in a 96-well plate, where p-nitrobenzaldehyde was dissolved in ethanol (EtOH) prior to its addition. The final concentration of ethanol in the system was 40 % (v/v), while glycine was added at 10 times the molar amount of p-nitrobenzaldehyde (i.e., 37.4/L), and pyridoxal phosphate (PLP) was added at the final concentration of 0.05mmol/L, and finally the crude enzyme solution was added. The total volume of the reaction was 200 µl. After the reaction was run for 4 hours, the reaction was quenched with 50% acetonitrile to inactivate aldolase polypeptides. The quenched reaction was centrifuged and resulting supernatant was diluted and then subjected to HPLC analysis to determine the substrate conversion and the de value for product A2.

**[0146]** Enzymatic reaction was scaled up to 5 mL of total reaction volume on the basis of the above 96-well microplate reaction. The loading of p-nitrobenzaldehyde was 40 g/L, the loading of glycine was 199.2 g/L, the final concentration of PLP was 0.05 mmol/L, concentration of ethanol in the system was 30 % (v/v) and crude enzyme powder loading was 4g/L.

**[0147]** The analytical method for the determination of the conversion and the de value of the product was as follows: the reaction solution was centrifuged and the supernatant was diluted with 50% acetonitrile to a product concentration of less than 1 g/L. 10 µl of this diluted sample was injected into an Agilent 1260 HPLC to analyze the conversion. The column was Phenomenex Chirex 3126 (D) -penicillamine 150 * 4.6 mm, mobile phase was 3 mM copper sulfate: methanol = 90: 10, at a flow rate of 1 mL per minute, at a column temperature of 50 °C, and the detection wavelength was 235 nm. The retention time of (2S, 3R)-2-amino-3-hydroxy-3-(4-nitrophenyl) propanoic acid was 22.53 min; the retention time of (2R, 3R) -2-amino-3-hydroxy-3-(4- nitrophenyl) propanoic acid was 24.17 minutes; the retention time of p-nitrobenzaldehyde was 28.65 minutes; the retention time of (2S, 3S) -2-amino-3-hydroxy-3-(4-nitrophenyl) propanoic acid and (2R, 3S) -2-amino-3-hydroxy-3- (4-nitrophenyl) propanoic acid was 64.19 minutes. The total analysis time was 70 minutes.

**Example 4: Construction of aldolase mutant library**

**[0148]** Quikchange kit (supplier: Agilent) was preferably used here. The sequence design of the mutagenesis primers was performed according to the instructions of the kit. The PCR system consisted of 10 µl of 5x Buffer, 1 µl of 10 mM dNTP, 1 µl of plasmid DNA template (50 ng/ µl), 0.75 µl (10 uM) each of the upstream and downstream primers, 0.5 µl of high fidelity enzyme and 36 µl of ddH2O, The PCR primer has a NNK codon at the mutation position.

**[0149]** PCR amplification steps: (1) 98 °C pre-denaturation 3min; (2) 98 °C denaturation 10s; (3) annealing and extension 3min at 72 °C; steps of (2) ~ (3) repeated 25 times; (5) extension 10min at 72 °C; (6) cooling to 4 °C, 2µl of DpnI was added to the PCR product and the plasmid template was eliminated by overnight digestion at 37 °C. The digested PCR product was_transformed into E. coli BL21 (DE3) competent cells and plated on LB agar plates containing chloramphenicol to obtain a site-saturation mutagenesis library.

**Example 5: High-throughput screening of aldolase mutant libraries**

**[0150]** Mutant colonies were picked from the LB agar plates, inoculated into 200 µl of LB medium (containing chloramphenicol) in a 96-well shallow plate and cultured overnight at 30 °C. 20µl of the above culture was used to inoculate 400 µl of TB medium (including chloramphenicol) in a deep-well plate. When $OD_{600}$ of deep-well culture reached 0.6 ~ 0.8, and IPTG was added to induce expression at a final concentration of 1 mM, and the expression undertook at 30 °C overnight. Once the overnight expression was done, the culture was centrifuged at 4000 rpm for 10 minutes to obtain cell pellets to which 200 µl of a chemical lysis reagent (1 g/L lysozyme, 0.5 g/L PMBS) was added to break the cells. Then cell lysate was centrifuged at 4000rpm for 10 minutes, and subsequently 60 µl of supernatant per well were transferred into a deep well plate containing the reaction solution described in Example 3. The reaction was shaken at 30~50 °C for desired time, and finally quenched with 50% acetonitrile. Samples were taken for analysis.

**Example 6: Fermentation process for the expression of engineered aldolase**

**[0151]** A single microbial colony of E. coli containing a plasmid bearing the target aldolase gene was inoculated into a 50 mL LB broth containing 30 µg/mL chloramphenicol (5.0 g/L Yeast Extract, 10 g/L Tryptone, 10 g/L sodium chloride). Cells were incubated overnight (at least 16 hours) with shaking at 250 rpm in a 30 °C shaker. When the $OD_{600}$ of the culture reached 1.6 to 2.2, the culture was used to inoculate medium in fermentor.

**[0152]** A 5 L fermentor containing 2.0 L of growth medium was sterilized in a 121 °C autoclave for 30 minutes. The fermentor was inoculated with the abovementioned culture. Temperature of fermentor was maintained at 37 °C. The growth medium in fermentor was agitated at 200-800 rpm and air was supplied to the fermentation vessel at 2-8 L/min to maintain the dissolved oxygen level at 30% saturation or greater. The culture was maintained at pH 7.0 by addition of 25-28% v/v ammonium hydroxide. Cell growth was maintained by feeding a feed solution containing 500 g/L of dextrose glucose monohydrate, 12 g/L ammonium chloride, and 5 g/L magnesium sulfate heptahydrate. After the $OD_{600}$ of culture reached 25 ± 5, the temperature of fermentor was decreased and maintained at 30 °C, and the expression of aldolase polypeptides was induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. Fermentation process then continued for additional 18 hours. After the fermentation process was complete, cells were harvested using a Thermo Multifuge X3R centrifuge at 8000 rpm for 10 minutes at 4 °C. Harvested cells were used directly in the downstream recovery process or stored frozen at -20 °C.

**[0153]** 6 g of cell pellet was resuspended in 30 mL of 100 mM potassium phosphate buffer containing 250 μM pyridoxal 5'-phosphate (PLP), pH 7.5 at 4 °C. The cells were then homogenized into cell lysate using a homogenizer. The cell lysate was clarified using a Thermo Multifuge X3R centrifuge at 8000 rpm for 10 minutes at 4 °C. The clarified supernatant was dispensed into a shallow container, frozen at -20 °C and lyophilized to a enzyme powder. The aldolase enzyme powder was stored frozen at -20 °C. Example 7: Asymmetric synthesis of (2S, 3R)-2-amino-3-hydroxy-3-(4-nitrophenyl) propanoic acid from aldehydes and amino acids catalyzed by aldolase polypeptides Taking a total volume of 1.0 L as an example, the following items were added to the reaction vessel: 178 g of glycine, 30 g of p-nitrobenzaldehyde, 942 mL of a 25% (v/v) aqueous ethanol solution, 4 g of enzyme powder of SEQ ID NO: 6, 5 mL of PLP stock solution (10 mM). The reaction temperature was set at 30 °C and the stirring speed was 400 rpm. After 8 hours of reaction, the total conversion of the substrate was ≥ 20% and de ≥ 95% for the product A2. Supernatant was obtained by filtration of the reaction, and the supernatant was concentrated to precipitate a solid crude product. The crude solid was washed with 300 mL of pure water for 30 minutes by stirring at 25 °C. Filtration was applied, and the filter cake was vacuum dried to obtain pure product (2S, 3R)-2-amino-3-hydroxy-3-(4-nitrophenyl) propanoic acid (chemical purity 99.5%, de ≥ 99%).

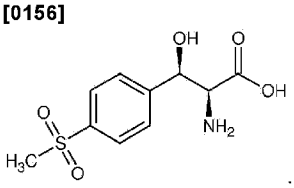
**Example 8: Preparation of D-(-)-threo-2-amino-l-(4-nitrophenyl)-l,3-propanediol from (2S, 3R)-2-amino-3-hydroxy-3-(4-nitrophenyl) propanoic acid.**

**[0154]** 1000mL of anhydrous methanol was added to a reaction vessel with ice bath, and it was stirred for 1 hour. The temperature of the reaction vessel was maintained at 5 °C, and within 1 hour, 128 mL of thionyl chloride was slowly added dropwise into the reaction vessel. After the addition of thionyl chloride was completed, the reaction mixture was stirred in an ice bath for 1 hour. Then 100 g of (2S, 3R)-2-amino-3-hydroxy-3-(4-nitrophenyl) propanoic acid was added into the reaction, followed by raising reaction temperature to 25 °C and stirring for 3 hours. Then the reaction temperature was slowly raised to 65 °C (reflux). Reflux reaction was carried out for about 24 hours. After etherification was complete, $SO_2$, HCl, and methanol were removed by depressurization at 40 °C until no liquid flew out. Then 1000mL ice water were added to cool down the reaction. At the same time, KOH solution was added dropwise to adjust the pH of reaction to about 8.0, and the reaction was stirred for 1 hour. Finally, the reaction was filtered, and the filtered cake was washed with water. 80 g of a white solid substance was obtained after drying the filter cake which was the ester product.

**[0155]** In order to obtain D-(-)-threo-2-amino-1-(4-nitrophenyl)-1,3-propanediol by reducing the ester product, 1200mL THF, 75g ester product were added into a reaction vessel, and it was stirred for 30min. Then 12g NaBH4 slowly added into the reaction which was stirred for 1 hour and then heated to reflux (50-55 °C). Subsequently, 160mL methanol were slowly added dropwise into the reaction within 30min, followed by stirring for 3 hours to finish the reaction. Concentrated hydrochloric acid were used to adjust the pH of finished reaction to ≤ 2, and it was stirred overnight. The reaction was filtered, and THF and methanol were removed from the filtrate under reduced pressure. 500mL of pure water were then added to the filtrate, and KOH were added to adjust pH ≥ 10. The filtrate was kept at 4 °C for crystallization to occur. The crystallized substance was recovered by filtration, and filter cake was dried to get about 55g D-(-)-threo-2-amino-1-(4-nitrophenyl)-1,3-propanediol.

**Example 9 Asymmetric synthesis of (2S, 3R)-3-[p-(Methylsulfonyl) phenyl]-3-hydroxy-2-amino-propanoic acid catalyzed by aldolase polypeptides**

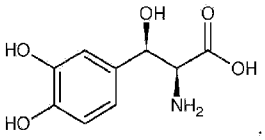
**[0156]**




**[0157]** Taking a total volume of 1.0 L as an example, the following items were added to the reaction vessel: 178 g of glycine, 40 g of p-methylsulfonylbenzaldehyde, 958 mL of a 40% (v/v) aqueous ethanol solution, 4 g of the enzyme powder of SEQ ID NO: 18, 5 mL of PLP stock solution (10 mM). The reaction temperature was set at 30 °C and the stirring speed was 400 rpm. After 6 hours of reaction, the conversion of p-methylsulfonylbenzaldehyde was ≥ 40%. The de for product (2S, 3R) -3- [p-(methylsulfonyl) phenyl] -hydroxy-2-amino-propanoic acid was ≥ 90%.

**Example 10 Asymmetric synthesis of (2S, 3R)-2-amino-3-(3,4-dihydroxybenzene)-3-hydroxypropanoic acid catalyzed by aldolase polypeptides**

**[0158]**



**[0159]** Taking a total reaction volume of 1.0 L for example, the following items were added to the reaction vessel: 55 g of glycine, 10 g of 3,4-dihydroxybenzaldehyde, 960 mL of deionized water, 10 g of enzyme powder of SEQ ID NO: 44, 5 mL of PLP stock solution (10 mM). The reaction temperature was set at 30 °C, stirring speed was 400rpm. After 2 hours of reaction, the total conversion of the substrate 3,4-dihydroxybenzaldehyde was ≥ 40%.

# REFERENCES CITED IN THE DESCRIPTION

Cited references

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

**Non-patent literature cited in the description**


- **MISONO, H. et al.**L-threonine aldolase, [0007]
- **DÜCKERS, N. et al.**Threonine aldolases - screening, properties and applications in the synthesis of non-proteinogenic beta-hydroxy-alpha-amino acidsAPPLIED MICROBIOLOGY AND BIOTECHNOLOGY, 2010, vol. 88, 2409-424 [0007]
- **STEINREIBER, J et al.**Threonine aldolases-an emerging tool for organic synthesisTETRAHEDRON, 2006, vol. 63, 4918-926 [0007]
- Nucleic Acid Research, 1994, vol. 22, 224673-4680 [0010]
- **SMITH AND WATERMAN**Adv. Appl. Math, 1981, vol. 2, 482- [0044]
- **NEEDLEMANWUNSCH**J. Mol. Biol, 1970, vol. 48, 443- [0044]
- **PEARSONLIPMAN**Proc. Natl. Acad. Sci. USA, 1988, vol. 85, 2444- [0044]
- Current Protocols in Molecular Biology**FM AUSUBEL ET AL.**Current ProtocolsGreene Publishing Associates, Inc. and John Wiley & Sons, Inc.19950000 [0044]
- **ALTSCHUL et al.**J. Mol. Biol., 1990, vol. 215, 403-410 [0044]
- **ALTSCHUL et al.**Nucleic Acids Res., 1977, 3389-3402 [0044]
- **HENIKOFF AND HENIKOFF**Proc Natl Acad Sci USA, 1989, vol. 89, 10915- [0044]
- **SAMBROOK et al.**Molecular Cloning: A Laboratory ManualCold Spring Harbor Laboratory Press20010000 [0073]
- **AUSUBEL. F**Current Protocols in Molecular BiologyGreene Pub. Associates19980000 [0073]
- Biocatalysis for the Pharmaceutical Industry: Discovery, Development, and ManufacturingJohn Wiley & Sons Asia (Pte) Ltd20090000 [0079]
- **BEAUCAGE et al.**Tet Lett, 1981, vol. 22, 1859-69 [0080]
- **MATTHES et al.**PeopleEMBO J, 1984, vol. 3, 801-05 [0080]

## Patentkrav

**1.** Manipuleret aldolase-polypeptid omfattende en aminosyresekvens udvalgt fra gruppen bestående af SEQ ID NOs: 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228 og 230.

**2.** Polypeptid immobiliseret på et fast materiale ved hjælp af kemisk binding eller en fysisk adsorptionsfremgangsmåde, hvor polypeptidet er udvalgt blandt aldolase-polypeptiderne ifølge krav 1.

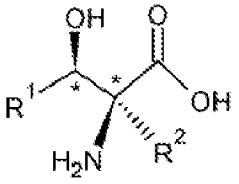
**3.** Polynukleotid, der koder for polypeptidet ifølge krav 1 eller 2; eventuelt hvor polynukleotidsekvensen er SEQ ID NO: 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227 og 229.

**4.** Ekspressionsvektor omfattende polynukleotidet ifølge krav 3, som eventuelt omfatter et plasmid, et cosmid, en bakteriofag eller en virusvektor.

**5.** Værtscelle omfattende ekspressionsvektoren ifølge krav 4, hvor værtscellen fortrinsvis er E. coli.

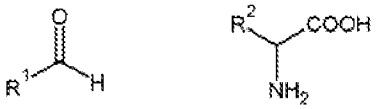
**6.** Fremgangsmåde til at fremstille et aldolase-polypeptid, der omfatter trinnene med at dyrke værtscellen ifølge krav 5 og opnå et aldolase-polypeptid fra kulturen.

**7.** Fremgangsmåde til at fremstille en β-hydroxy-α-aminosyre med formel (I):



(I)

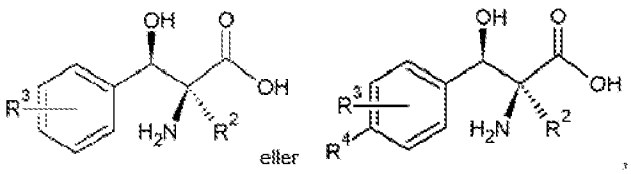
hvor β-hydroxy-α-aminosyren med formel (I) har den angivne stereokemiske konfiguration ved det chirale center markeret med en *;
$R^1$ eventuelt er substitueret eller usubstitueret aryl eller heteroaryl, eller eventuelt substitueret eller usubstitueret $C_1$-$C_8$-hydrocarbyl;
$R^2$ er -H, $-CH_2OH$,$-CH_2SH$,$-CH_2SCH_3$, eller eventuelt substitueret eller usubstitueret $C_1$-$C_4$-hydrocarbyl;
hvilken fremgangsmåde omfatter at bringe et aldehydsubstrat med formel (II) og et aminosyresubstrat med formel (III) i kontakt



(II) (III)

med det manipulerede aldolase-polypeptid ifølge krav 1 eller 2, under egnede reaktionsbetingelser til omdannelse af aldehydsubstratet for at producere β-hydroxy-α-aminosyren,
hvor β-hydroxy-α-aminosyren med formel (I) opnås i diastereomert overskud.

**8.** Fremgangsmåde ifølge krav 7, hvor β-hydroxy-α-aminosyren med formel (I) er:

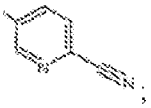


eller

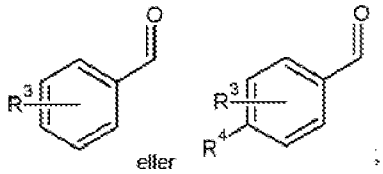
hvor $R^2$ er -H, $-CH_3$,$-CH_2CH_3$,$-CH(CH_3)_2$,$-CH_2OH$,$-CH_2SH$ eller $-CH_2SCH_3$;
$R^3$ er $C_1$-$C_4$-hydrocarbyl, -H, halogen (såsom -F, -Cl, -Br eller -I), $-NO_2$, -NO, -

SO2R' eller -SOR', -SR', -NR'R', -OR', -$CO_2$R' eller -COR', -C(O)NR' , -$SO_2NH_2$ eller -$SONH_2$, -CN, -$CF_3$, hvor hvert R' er uafhængigt udvalgt blandt -H eller ($C_1$-$C_4$)-hydrocarbyl; $R^3$ kan også være

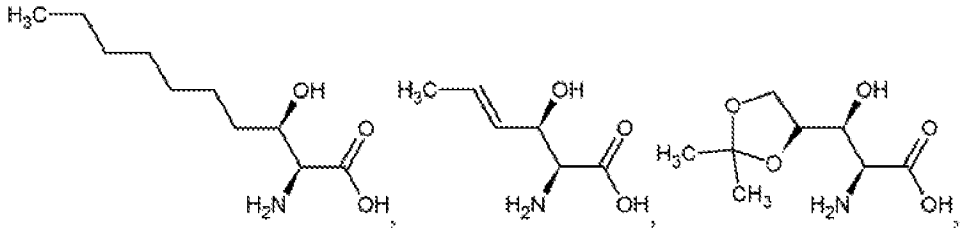
eller

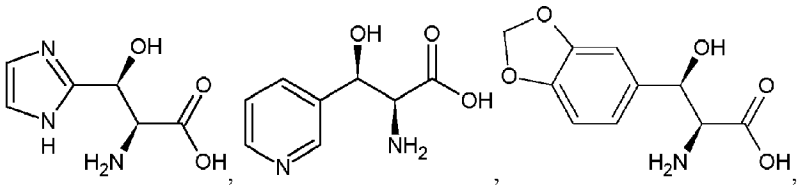


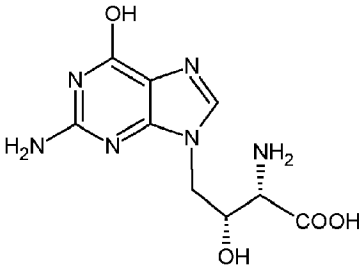
$R^4$ er $C_1$-$C_4$-hydrocarbyl, -H, halogen (såsom -F, -Cl, -Br eller -I), -$NO_2$, -NO, -$SO_2$R eller -SOR, -SR' , -NR'R ', -OR', -$CO_2$R' eller -COR', -C(O)NR', -$SO_2NH_2$ eller -$SONH_2$, -CN, -$CF_3$, hvor hvert R' er uafhængigt udvalgt blandt -H eller ($C_1$-$C_4$)-hydrocarbyl; og aldehydsubstratet med formel (II) er:

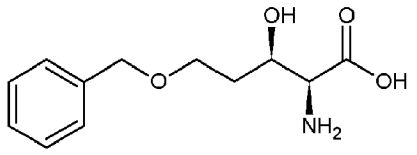


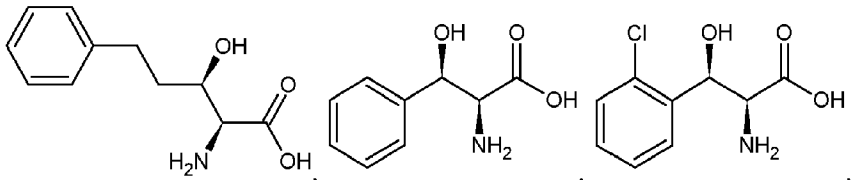

eventuelt hvor $R^3$ er i parapositionen af phenylringen, eller $R^3$ er i metapositionen af phenylringen, eller $R^3$ er i orthopositionen af phenylringen; eller $R^3$ er i parapositionen af phenylringpositionen og metapositionen, eller $R^3$ både er ved parapositionen og orthopositionen af phenylringen, eller R3 både er ved meta- og orthoposition på phenylringen; og/eller
eventuelt hvor β-hydroxy-α-aminosyren med formel (I) er:

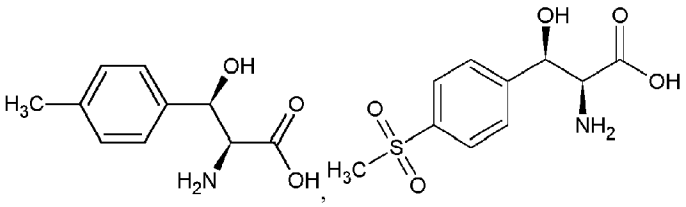


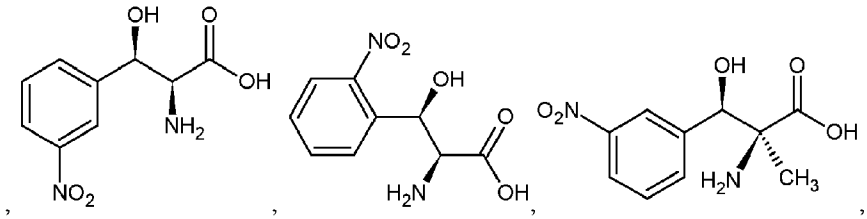


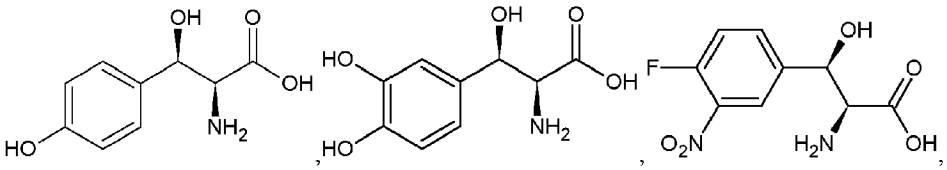


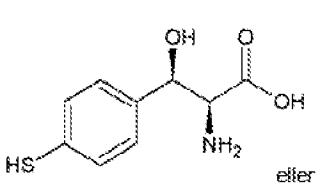




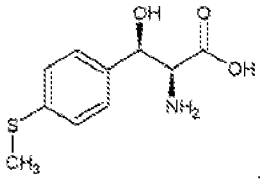




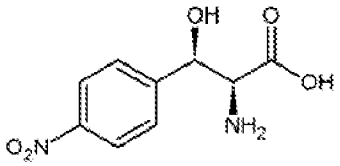





eller

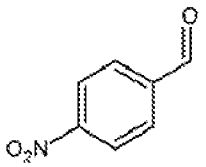



**9.** Fremgangsmåde til fremstilling af en forbindelse med formel A2 (2S, 3R)-2-amino-3-hydroxy-3-(4-nitrophenyl)-propionsyre:




A2

hvilken fremgangsmåde omfatter at bringe p-nitrobenzaldehyd med formel A1




A1

i kontakt med et manipuleret aldolase-polypeptid ifølge krav 1 eller 2, under nærvær af glycin, i et egnet opløsningsmiddel, under egnede reaktionsbetingelser for at omdanne forbindelsen med formel A1 til forbindelsen med formel A2.

**10.** Fremgangsmåde ifølge et af kravene 7-9, hvor β-hydroxy-α-aminosyreproduktet er til stede i diastereomert overskud på mindst 60 %, 70 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % eller mere; og/eller

hvor reaktionen udføres i et opløsningsmiddel omfattende vand, methanol, ethanol, propanol, isopropanol, isopropylacetat, dimethylsulfoxid (DMSO) eller dimethylformamid (DMF); og/eller

hvor reaktionsbetingelserne indbefatter en temperatur på 10°C til 60 °C; og/eller

hvor reaktionsbetingelserne indbefatter pH-værdi 4,0 til pH-værdi 8,0; og/eller

hvor aldehydsubstratet er til stede ved en belastning på 5 g/L til 400 g/L.