(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 August 2008 (28.08.2008)

PCT

(10) International Publication Number
WO 2008/103248 A1

(51) International Patent Classification:
C12N 9/04 (2006.01) *C12N 15/53* (2006.01)

(21) International Application Number: PCT/US2008/001683

(22) International Filing Date: 8 February 2008 (08.02.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/900,494 8 February 2007 (08.02.2007) US
60/900,430 8 February 2007 (08.02.2007) US

(71) Applicant *(for all designated States except US)*: **CODEXIS, INC.** [US/US]; 200 Penobscot Drive, Redwood City, CA 94063 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **CHING, Charlene** [US/US]; 4987 Kenson Drive, San Jose, CA 95124 (US). **GRUBER, John, M.** [US/US]; 505 Cypress Poing Drive, #177, Mountain View, CA 94043 (US). **HUISMAN, Gjalt, W.** [NL/US]; 2211 Howard Avenue, San Carlos, CA 94040 (US). **MUNDORRF, Emily** [US/US]; 2612 Belmont Canyon Road, Belmont, CA 94002 (US). **NEWMAN, Lisa, M.** [US/US]; 655 Jackson Street, San Jose, CA 95112 (US).

(74) Agents: OH, Euk, Charlie et al.; Dechert LLP, P.O.Box 390460, Mountain View, CA 94039 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
- *with international search report*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

WO 2008/103248 A1

(54) Title: KETOREDUCTASES AND USES THEREOF

(57) Abstract: The present disclosure provides engineered ketoreductase enzymes having improved properties as compared to a naturally occurring wild-type ketoreductase enzyme. Also provided are polynucleotides encoding the engineered ketoreductase enzymes, host cells capable of expressing the engineered ketoreductase enzymes, and methods of using the engineered ketoreductase enzymes to synthesize a variety of chiral compounds.

KETOREDUCTASES AND USES THEREOF

1. CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims benefit under 35 U.S.C. § 119(e) of application Serial No. 60/900,494, filed February 8, 2007 and application Serial No. 60/900,430, filed February 8, 2007, the contents of which are incorporated herein by reference.

2. REFERENCE TO SEQUENCE LISTING, TABLE OR COMPUTER PROGRAM

[0002] The "Sequence Listing" submitted concurrently herewith under 37 C.F.R. § 1.821(c) and the computer readable form (CRF) of the Sequence Listing submitted under 37 C.F.R. § 1.821(e) are incorporated by reference in its entirety. Two copies of the Sequence Listing, one on each of two compact discs labeled "Copy 1" and "Copy 2" are provided. Each electronic copy of the Sequence Listing was created on February 8, 2008 with a file size of 221 Kbytes. The file names are: Copy 1 – 376247-015WO.txt and Copy 2 - 376247-015WO.txt. The Sequence Listing incorporated herein by reference is the same as the CRF of the Sequence Listing being submitted concurrently herewith.

3. BACKGROUND

[0003] Enzymes belonging to the ketoreductase (KRED) or carbonyl reductase class (EC1.1.1.184) are useful for the synthesis of optically active alcohols from the corresponding prochiral ketone substrate. KREDs typically convert a ketone substrate to the corresponding alcohol product, but may also catalyze the reverse reaction, oxidation of an alcohol substrate to the corresponding ketone/aldehyde product. The reduction of ketones and the oxidation of alcohols by enzymes such as KRED requires a co-factor, most commonly reduced nicotinamide adenine dinucleotide (NADH) or reduced nicotinamide adenine dinucleotide phosphate (NADPH), and nicotinamide adenine dinucleotide (NAD) or nicotinamide adenine dinucleotide phosphate (NADP) for the oxidation reaction. NADH and NADPH serve as electron donors, while NAD and NADP serve as electron acceptors. It is frequently observed that ketoreductases and alcohol dehydrogenases accept either the phosphorylated or the non-phosphorylated co-factor (in its oxidized and reduced state), but not both.

[0004] KRED enzymes can be found in a wide range of bacteria and yeasts (for reviews: Kraus and Waldman, Enzyme catalysis in organic synthesis, Vols. 1 & 2,VCH Weinheim 1995; Faber, K., *Biotransformations in organic chemistry*, 4th Ed. Springer, Berlin Heidelberg New York. 2000; Hummel and Kula *Eur. J. Biochem.* 1989 184:1-13). Several KRED gene and enzyme sequences have been reported, *e.g.*, *Candida magnoliae* (Genbank Acc. No. JC7338; GI:11360538) *Candida parapsilosis* (Genbank Acc. No. BAA24528.1; GI:2815409), *Sporobolomyces salmonicolor* (Genbank Acc. No. AF160799; GI:6539734).

[0005] In order to circumvent many chemical synthetic procedures for the production of key compounds, ketoreductases are being increasingly employed for the enzymatic conversion of different keto substrates to chiral alcohol products. These applications can employ whole cells expressing the ketoreductase for biocatalytic ketone reductions, or purified enzymes in those instances where presence of multiple ketoreductases affects specificity and yield of the desired product. For in vitro applications, a co-factor (NADH or NADPH) regenerating enzyme such as glucose dehydrogenase (GDH), formate dehydrogenase etc. is used in conjunction with the ketoreductase. Examples using ketoreductases to generate useful chemical compounds include asymmetric reduction of 4-chloroacetoacetate esters (Zhou, *J. Am. Chem. Soc.* 1983 105:5925-5926; Santaniello, *J. Chem. Res.* (S) 1984:132-133; US Patent No. 5,559,030; US Patent No. 5,700,670 and US Patent No. 5,891,685); reduction of dioxocarboxylic acids (*e.g.*, US Patent No. 6,399,339); reduction of tert-butyl (S) chloro-5-hydroxy-3-oxohexanoate (*e.g.*, US Patent No. 6,645,746 and WO 01/40450); reduction pyrrolotriazine-based compounds (*e.g.*, US application No. 2006/0286646); reduction of substituted acetophenones (*e.g.*, US Patent No. 6,800,477); and reduction of hydroxythiolanes (WO 2005/054491).

[0006] It is desirable to identify other ketoreductase enzymes that can be used to carryout conversion of various keto substrates to its corresponding chiral alcohol products.

4. SUMMARY

[0007] The present disclosure provides engineered or recombinant ketoreductase polypeptides capable of reducing or converting the compound 1-[4-(4-fluoro2-methyl-IH-indol-5-yloxy)-5-methyl-pyrrolo[2, 1-f][1 ,2,4]triazin-6-yloxy]-propan-2-one to the corresponding product (*R*) 1-[4-(4-fluoro-2-methyl-IH-indol-5-yloxy)-5-methylpyrrolo[2, 1-f][1 ,2,4]triazin-6-yloxy]-propan-2-ol. The engineered or recombinant ketoreductases are also capable of reducing or converting acetophenone to (*R*)-1-phenylethanol. In the embodiments herein, the engineered ketoreductases have one or more improved properties in converting the substrates to the product as compared to the naturally-occurring wild-type ketoreductase enzymes of *Lactobacillus kefir* and *Lactobacillus brevis.*

[0008] In one aspect, the recombinant polypeptide capable of converting the 1-[4-(4-fluoro2-methyl-IH-indol-5-yloxy)-5-methyl-pyrrolo[2, 1-f][1 ,2,4]triazin-6-yloxy]-propan-2-one to the corresponding product (*R*)-1-[4-(4-fluoro-2-methyl-IH-indol-5-yloxy)-5-methylpyrrolo[2, 1-f][1 ,2,4]triazin-6-yloxy]-propan-2-ol comprises an amino acid sequence having (1) an aromatic amino acid or G at the amino acid residue corresponding to residue 94 of SEQ ID NO:2 or SEQ ID NO:4, and/or (2) an amino acid residue other than S and N at the amino acid residue corresponding to residue 96 of SEQ ID NO:2 or SEQ ID NO:4.

[0009] In some embodiments, the ketoreductase polypeptide comprises an amino acid sequence that is at least about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO:2 or SEQ ID NO:4 with the proviso that the residues corresponding to residue 94 is an aromatic amino acid residue or G. The engineered ketoreductase may optionally include one or more conservative substitutions at other residue positions within the amino acid sequence.

[0010] In some embodiments, the ketoreductase polypeptides comprise an amino acid sequence having at least 90% amino acid sequence identity to SEQ ID NO:2 or SEQ ID NO:4 and comprise an aromatic amino acid or G at the amino acid residue corresponding to residue 94 of SEQ ID NO:2 or SEQ ID NO:4.

[0011] In some embodiments, the ketoreductase polypeptide capable of converting the 1-[4-(4-fluoro2-methyl-IH-indol-5-yloxy)-5-methyl-pyrrolo[2, 1-f][1 ,2,4]triazin-6-yloxy]-propan-2-one to the corresponding product (*R*)-1-[4-(4-fluoro-2-methyl-IH-indol-5-yloxy)-5-methylpyrrolo[2, 1-f][1 ,2,4]triazin-6-yloxy]-propan-2-ol can comprise a region having an amino acid sequence that is at least about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a region or domain thereof of SEQ ID NO:2 or SEQ ID NO:4, such as residues 90-233, with the proviso that the amino acid residue corresponding to residue 94 is an aromatic amino acid residue or G. In some embodiments of these ketoreductase polypeptides, one or more of the remaining residues corresponding to residues 90-233 of SEQ ID NO:2 or SEQ ID NO:4 may have a conservative substitution.

[0012] The ketoreductase polypeptide capable of converting the compound 1-[4-(4-fluoro2-methyl-IH-indol-5-yloxy)-5-methyl-pyrrolo[2, 1-f][1 ,2,4]triazin-6-yloxy]-propan-2-one to the corresponding product (*R*) 1-[4-(4-fluoro-2-methyl-IH-indol-5-yloxy)-5-methylpyrrolo[2, 1-f][1 ,2,4]triazin-6-yloxy]-propan-2-ol can comprise, in addition to an aromatic amino acid or G at the amino acid residue corresponding to residue 94 of SEQ ID NO:2 or SEQ ID NO:4, one or more of the features selected from: residue 96 is any amino acid other than S/N; residue 153 is an aliphatic amino acid residue other than L; residue 199 is any amino acid residue other than L; residue 202 is G or an aliphatic amino acid residue other than A; and residue 206 is an aromatic amino acid residue.

[0013] In some embodiments, the ketoreductase polypeptide with the specified amino acid at residue 94 can comprise on or more of the following additional features selected from: residue 49 is a polar amino acid residue other than K; residue 53 is an acidic amino acid residue; residue 54 is a small or aliphatic amino acid residue other than T/P; residue 60 is an aliphatic amino acid residue other than V; residue 95 is an aliphatic amino acid other than V; residue 97 is a small amino acid or G; residue 109 is a basic amino acid residue other than K; residue 147 is an aliphatic amino acid residue; residue 165 is a hydroxyl or small amino acid residue; residue 197 is a small amino acid residue or G; residue 223 is an aliphatic amino acid residue other than L; and residue 233 is a small amino acid residue or G.

[0014] In another aspect, the recombinant ketoreductase polypeptide capable of converting the compound 1-[4-(4-fluoro2-methyl-IH-indol-5-yloxy)-5-methyl-pyrrolo[2, 1-f][1 ,2,4]triazin-6-yloxy]-propan-2-one to the corresponding product (*R*) 1-[4-(4-fluoro-2-methyl-IH-indol-5-yloxy)-5-methylpyrrolo[2, 1-f][1 ,2,4]triazin-6-yloxy]-propan-2-ol comprise an amino acid sequence having an amino acid other than S and N at the amino acid residue corresponding to residue 96 of SEQ ID NO:2 or SEQ ID NO:4. In some embodiments, these recombinant ketoreductase polypeptides can comprise an amino acid sequence with a G, F, Y, or I at the amino acid residue corresponding to residue 96 of SEQ ID NO:2 or SEQ ID NO:4.

[0015] In some embodiments, the ketoreductase polypeptide comprises an amino acid sequence that is at least about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO:2 or SEQ ID NO:4 with the proviso that the residue corresponding to residue 96 is an amino acid residue other than S and N. The engineered ketoreductase may optionally include one or more conservative substitutions at other residue positions within the amino acid sequence.

[0016] In some embodiments, these ketoreductase polypeptides can comprise an amino acid sequence that is at least about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO:2 or SEQ ID NO:4 with the proviso that the residue corresponding to residue 96 is G, F, Y, or I.

[0017] In some embodiments, the ketoreductase polypeptide capable of converting 1-[4-(4-fluoro2-methyl-IH-indol-5-yloxy)-5-methyl-pyrrolo[2, 1-f][1 ,2,4]triazin-6-yloxy]-propan-2-one to the corresponding product (*R*)-1-[4-(4-fluoro-2-methyl-IH-indol-5-yloxy)-5-methylpyrrolo[2, 1-f][1 ,2,4]triazin-6-yloxy]-propan-2-ol can comprise a region having an amino acid sequence that is at least about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a region or domain thereof of SEQ ID NO:2 or SEQ ID NO:4, such as residues 90-233, with the proviso that the residue corresponding to residue 96 is an amino acid residue other than S and N. In some embodiments of these ketoreductase polypeptides, one or more of the remaining residues corresponding to residues 90-233 of SEQ ID NO:2 or SEQ ID NO:4 may have a conservative substitution.

[0018] In some embodiments, the ketoreductase polypeptide comprises an amino acid sequence having at least 90% amino acid sequence identity to SEQ ID NO:2 or SEQ ID NO:4 and comprise a G, F, Y, or I at the amino acid residue corresponding to residue 96 of SEQ ID NO:2 or SEQ ID NO:4. These engineered ketoreductase may optionally include one or more conservative mutations at other residue positions within the polypeptide sequence.

[0019] In some embodiments, the ketoreductase polypeptide capable of converting the compound 1-[4-(4-fluoro2-methyl-IH-indol-5-yloxy)-5-methyl-pyrrolo[2, 1-f][1 ,2,4]triazin-6-yloxy]-propan-2-one to the corresponding product (*R*) 1-[4-(4-fluoro-2-methyl-IH-indol-5-yloxy)-5-methylpyrrolo[2, 1-f][1 ,2,4]triazin-6-yloxy]-propan-2-ol can comprise, in addition to G, F, Y, or I at the amino acid residue corresponding to residue 96 of SEQ ID NO:2 or

SEQ ID NO:4, one or more of the features selected from: residue 94 is an aromatic amino acid residue or G; residue 153 is an aliphatic amino acid residue other than L; residue 199 is any amino acid residue other than L; residue 202 is an aliphatic amino acid residue other than A; and residue 206 is an aromatic amino acid residue.

[0020] In some embodiments, the ketoreductase polypeptide with the specified amino acid residues at residue 96 can comprise one or more of the features selected from: residue 49 is a polar amino acid residue other than K; residue 53 is an acidic amino acid residue; residue 54 is a small or aliphatic amino acid residue other than T/P; residue 60 is an aliphatic amino acid residue other than V; residue 95 is an aliphatic amino acid other than V; residue 97 is a small amino acid or G; residue 109 is a basic amino acid residue other than K; residue 147 is an aliphatic amino acid residue; residue 165 is a hydroxyl or small amino acid residue; residue 197 is a small amino acid residue or G; residue 223 is an aliphatic amino acid residue other than L; and residue 233 is a small amino acid residue or G.

[0021] In some embodiments, the ketoreductase polypeptide capable of converting the compound 1-[4-(4-fluoro2-methyl-IH-indol-5-yloxy)-5-methyl-pyrrolo[2, 1-f][1 ,2,4]triazin-6-yloxy]-propan-2-one to the corresponding product (*R*) 1-[4-(4-fluoro-2-methyl-IH-indol-5-yloxy)-5-methylpyrrolo[2, 1-f][1 ,2,4]triazin-6-yloxy]-propan-2-ol is selected from SEQ ID NO: 6, 8, 10, 12, 14, 16, 18, 20, 22, 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, and 118.

[0022] As noted above, the recombinant ketoreductase polypeptides are also capable of reducing or converting acetophenone to the corresponding product (*R*)-1-phenylethanol. Additional polypeptides capable of converting acetophenone to the corresponding product (*R*)1-phenylethanol include recombinant ketoreductase polypeptides comprising an amino acid sequence having a G, I, C or an aromatic amino acid at the amino acid residue corresponding to residue 96 of SEQ ID NO:2 or SEQ ID NO:4.

[0023] In some embodiments, these ketoreductase polypeptides can comprise an amino acid sequence that is at least about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO:2 or SEQ ID NO:4 with the proviso that the residue corresponding to residue 96 is G, I, C or an aromatic amino acid. These engineered ketoreductase may optionally include one or more conservative mutations at other residue positions within the polypeptide sequence.

[0024] In some embodiments, the ketoreductase polypeptides capable of reducing or converting acetophenone to the corresponding product (*R*)-1-phenylethanol can comprise a region having an amino acid sequence that is at least about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a region or domain thereof of SEQ ID NO:2 or SEQ ID NO:4, such as residues 90-233, with the proviso that the residue corresponding to residue 96 is G, I, C or an aromatic amino acid. In some embodiments of these ketoreductase polypeptides, one or more of the remaining residues corresponding to residues 90-233 of SEQ ID NO:2 or SEQ ID NO:4 may have a conservative substitution.

[0025] In some embodiments, the ketoreductase polypeptide comprises an amino acid sequence having at least 90% amino acid sequence identity to SEQ ID NO:2 or SEQ ID NO:4 and comprise a G, I, C or an aromatic amino acid at the amino acid residue corresponding to residue 96 of SEQ ID NO:2 or SEQ ID NO:4.

[0026] In some embodiments, the ketoreductase polypeptide capable of reducing or converting acetophenone to the corresponding product (*R*)-1-phenylethanol can comprise, in addition to the G, I, C or an aromatic amino acid at the amino acid residue corresponding to residue 96, one or more of the features selected from: residue 94 is an aromatic amino acid residue or G; residue 153 is an aliphatic amino acid residue other than L; residue 199 is any amino acid residue other than L; residue 202 is an aliphatic amino acid residue other than A; and residue 206 is an aromatic amino acid residue.

[0027] In some embodiments, the ketoreductase polypeptides capable of reducing or converting acetophenone to the corresponding product (*R*)-1-phenylethanol can comprise additionally one or more of the features selected from: residue 49 is a polar amino acid residue other than K; residue 53 is an acidic amino acid residue; residue 54 is a small or aliphatic amino acid residue other than T/P; residue 60 is an aliphatic amino acid residue other than V; residue 95 is an aliphatic amino acid other than V; residue 97 is a small amino acid or G; residue 109 is a basic amino acid residue other than K; residue 147 is an aliphatic amino acid residue; residue 165 is a hydroxyl or small amino acid residue; residue 197 is a small amino acid residue or G; residue 223 is an aliphatic amino acid residue other than L; and residue 233 is a small amino acid residue or G.

[0028] In some embodiments, the ketoreductase polypeptide capable of reducing or converting acetophenone to the corresponding product (*R*)-1-phenylethanol is selected from SEQ ID NO: 6, 8, 10, 12, 14, 16, 18, 20, 22, 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, and 118. In some embodiments, the ketoreductase polypeptide capable of reducing or converting acetophenone to the corresponding product (*R*)-1-phenylethanol is selected from SEQ ID NO: 120, 122, 124, 126, 128, 130, 132, 134, 136, and 138.

[0029] In another aspect, the present disclosure provides polynucleotides encoding the engineered ketoreductases described herein or polynucleotides that hybridize to such polynucleotides under highly stringent conditions. The polynucleotide can include promoters and other regulatory elements useful for expression of the encoded engineered ketoreductase, and can utilize codons optimized for specific desired expression systems. Exemplary polynucleotides include SEQ ID NO: 5, 7, 9, 11, 13, 15, 17, 19, 21, 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, and 137.

[0030] In still another aspect, the present disclosure provides host cells comprising the polynucleotides and/or expression vectors described herein. The host cells may be *Lactobacillus kefir* or *Lactobacillus brevis*, or they may be a different organism. The host cells can be used for the expression and isolation of the engineered ketoreductase enzymes, or, alternatively, they can be used directly for the conversion of the keto substrate to the chiral alcohol product.

[0031] As will be appreciated by skilled artisans, the reduction reaction illustrated above generally requires a cofactor, which is normally NADH or NADPH, and can include a system for regenerating the cofactor, for example glucose and glucose dehydrogenase. In some embodiments employing purified engineered ketoreductase enzyme(s), such cofactors and optionally such cofactor regeneration systems, will typically be added to the reaction medium along with the substrate and the ketoreductase enzyme(s). Like the engineered ketoreductase enzyme, any enzyme(s) comprising the cofactor regeneration system can be supplied to the reaction mixture in the form of extracts or lysates of such cells, or as purified enzyme(s). In embodiments employing cell extracts or cell lysates, the cells used to generate the extracts or lysates can be engineered to express the enzyme(s) comprising the cofactor regeneration systems alone, or together with the engineered ketoreductase enzyme. In embodiments employing whole cells, the cells can be engineered to express the enzyme(s) comprising the cofactor regeneration systems and the engineered ketoreductase enzyme together.

[0032] Whether carrying out the method with whole cells, cell extracts or purified ketoreductase enzymes, a single ketoreductase enzyme may be used or, alternatively, mixtures of two or more ketoreductase enzymes may be used.

[0033] In various embodiments, the engineered enzymes can carry out the reduction or conversion reaction with a degree of enantioselectivity of ≥ 99%. Thus, the above reactions can be used as a standard reaction for assessing the activity of the engineered ketoreductase enzymes as compared to a reference ketoreductase, such as the ketoreductases of SEQ ID NO:2 or SEQ ID NO:4.

[0034] In some embodiments, because the engineered ketoreductase enzymes described herein are highly stereoselective, the resultant product of structural formula (II) ("Compound (II)") or structural formula (IV) ("Compound (IV)") can be recovered in substantially stereochemically pure form without the need to chirally separate it from the corresponding enantiomer.

5. BRIEF DESCRIPTION OF THE FIGURES

[0035] FIG.1 illustrates the role of ketoreductases in the conversion of a defined substrate of Compound (I) to the chiral alcohol product of Compound (II). The figure also shows use of a cofactor regenerating system involving glucose dehydrogenase (GDH) and glucose.

[0036] FIG. 2 illustrates the role of ketoreductases in the conversion of a defined substrate of Compound (III) to the chiral alcohol product of Compound (IV). The figure also shows use of a cofactor regenerating system involving glucose dehydrogenase (GDH) and glucose.

6. DETAILED DESCRIPTION

[0037] As used in this specification and the appended claims, the singular forms "a", "an" and "the" include plural referents unless the context clearly indicates otherwise. Thus, for example, reference to "a protein" includes more than one protein, and reference to "a compound" refers to more than one compound. In addition, the use of "or" means "and/or" unless stated otherwise. Similarly, "comprise," "comprises," "comprising" "include," "includes," and "including" are interchangeable and not intended to be limiting.

[0038] It is to be further understood that where descriptions of various embodiments use the term "comprising," those skilled in the art would understand that in some specific instances, an embodiment can be alternatively described using language "consisting essentially of" or "consisting of."

[0039] The section headings used herein are for organizational purposes only and not to be construed as limiting the subject matter described.

6.1 Definitions

[0040] As used herein, the following terms are intended to have the following meanings.

[0041] "Ketoreductase" and "KRED" are used interchangeably herein to refer to a polypeptide that is capable of reducing a keto group to its corresponding alcohol. More specifically, the ketoreductase polypeptides of the present disclosure are capable of stereoselectively reducing the compound of formula (I), *supra* to the alcohol product of formula (II), *supra* (see FIG. 1) and/or the compound of formula (III), *supra* to the alcohol product of formula (IV) (see FIG. 2). The polypeptide typically utilizes a cofactor reduced nicotinamide adenine dinucleotide (NADH) or reduced nicotinamide adenine dinucleotide phosphate (NADPH) as the reducing agent. Ketoreductases as used herein include naturally occurring (wild type) ketoreductases as well as non-naturally occurring engineered polypeptides generated by human manipulation (*i.e.*, recombinant polypeptides).

[0042] "Coding sequence" refers to that portion of a nucleic acid (*e.g.*, a gene) that encodes an amino acid sequence of a protein.

[0043] "Naturally-occurring" or "wild-type" refers to the form found in nature. For example, a naturally occurring or wild-type polypeptide or polynucleotide sequence is a sequence present in an organism that can be isolated from a source in nature and which has not been intentionally modified by human manipulation.

[0044] "Recombinant" when used with reference to, *e.g.*, a cell, nucleic acid, or polypeptide, refers to a material, or a material corresponding to the natural or native form of the material, that has been modified in a manner that would not otherwise exist in nature, or is identical thereto but produced or derived from synthetic materials and/or by manipulation using recombinant techniques. Non-limiting examples include, among others, recombinant cells expressing genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise expressed at a different level.

[0045] "Percentage of sequence identity" and "percentage homology" are used interchangeably herein to refer to comparisons among polynucleotides and polypeptides, and are determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage may be calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Alternatively, the percentage may be calculated by determining the number of positions at which either the identical nucleic acid base or amino acid residue occurs in both sequences or a nucleic acid base or amino acid residue is aligned with a gap to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Those of skill in the art appreciate that there are many established algorithms available to align two sequences. Optimal alignment of sequences for comparison can be conducted, *e.g.*, 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 Software Package), or by visual inspection (see generally, *Current Protocols in Molecular Biology*, F. M. 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 percent sequence identity and 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 analyses is publicly available through the National Center for Biotechnology Information website. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as, the neighborhood word score threshold (Altschul et al, supra). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff and Henikoff, 1989, *Proc Natl Acad Sci USA* 89:10915). Exemplary determination of sequence alignment and % sequence identity can employ the BESTFIT or GAP programs in the GCG Wisconsin Software package (Accelrys, Madison WI), using default parameters provided.

[0046] "Reference sequence" refers to a defined sequence used as a basis for a sequence comparison. A reference sequence may be a subset of a larger sequence, for example, a segment of a full-length gene or polypeptide sequence. Generally, a reference sequence is at least 20 nucleotide or amino acid residues in length, at least 25 residues in length, at least 50 residues in length, or the full length of the nucleic acid or polypeptide. Since two polynucleotides or polypeptides may each (I) comprise a sequence (*i.e.*, a portion of the complete sequence) that is similar between the two sequences, and (II) may further comprise a sequence that is divergent between the two sequences, sequence comparisons between two (or more) polynucleotides or polypeptide are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity.

[0047] "Comparison window" refers to a conceptual segment of at least about 20 contiguous nucleotide positions or amino acids residues wherein a sequence may be compared to a reference sequence of at least 20 contiguous nucleotides or amino acids and wherein the portion of the sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The comparison window can be longer than 20 contiguous residues, and includes, optionally 30, 40, 50, 100, or longer windows.

[0048] "Substantial identity" refers to a polynucleotide or polypeptide sequence that has at least 80 percent sequence identity, at least 85 percent identity and 89 to 95 percent sequence identity, more usually at least 98 percent sequence identity, at least 99 percent sequence identity or at least 99.5 percent or more sequence identity as compared to a reference sequence over a comparison window of at least 20 residue positions, frequently over a window of at least 30-50 residues, wherein the percentage of sequence identity is calculated by comparing the reference sequence to a sequence that includes deletions or additions which total 20 percent or less of the reference sequence over the window of comparison. In specific embodiments applied to polypeptides, the term "substantial identity" means that two polypeptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap weights, share at least 80 percent sequence identity, preferably at least 89 percent sequence identity, at least 95 percent sequence identity, at least 98% sequence identity, or at least 99 percent sequence identity. Preferably, residue positions which are not identical differ by conservative amino acid substitutions.

[0049] "Corresponding to", "reference to" or "relative to" when used in the context of the numbering of a given amino acid or polynucleotide sequence refers to the numbering of the residues of a specified reference sequence when the given amino acid or polynucleotide sequence is compared to the reference sequence. In other words, the residue number or residue position of a given polymer is designated with respect to the reference sequence rather than by the actual numerical position of the residue within the given amino acid or polynucleotide sequence. For example, a given amino acid sequence, such as that of an engineered ketoreductase, can be aligned to a reference sequence by introducing gaps to optimize residue matches between the two sequences. In these cases, although the gaps are present, the numbering of the residue in the given amino acid or polynucleotide sequence is made with respect to the reference sequence to which it has been aligned.

[0050] "Stereoselectivity" refers to the preferential formation in a chemical or enzymatic reaction of one stereoisomer over another. Stereoselectivity can be partial, where the formation of one stereoisomer is favored over the other, or it may be complete where only one stereoisomer is formed. When the stereoisomers are enantiomers, the stereoselectivity is referred to as enantioselectivity, the fraction (typically reported as a percentage) of one enantiomer in the sum of both. It is commonly reported in the art (typically as a percentage) as the enantiomeric excess calculated therefrom according to the formula [major enantiomer − minor enantiomer]/[major enantiomer + minor enantiomer]. Where the stereoisomers are diastereoisomers, the stereoselectivity is referred to as diastereoselectivity, the fraction (typically reported as a percentage) of one diastereomer in the sum with others.

[0051] "Highly stereoselective" refers to a ketoreductase polypeptide that is capable of converting or reducing the substrate to the corresponding product having the chemical formula (II) with at least about 85% stereomeric excess.

[0052] "Stereospecificity" refers to the preferential conversion in a chemical or enzymatic reaction of one stereoisomer over another. Stereospecificity can be partial, where the conversion of one stereoisomer is favored over the other, or it may be complete where only one stereoisomer is converted.

[0053] "Improved enzyme property" refers to a ketoreductase polypeptide that exhibits an improvement in any enzyme property as compared to a reference ketoreductase. For the engineered ketoreductase polypeptides described herein, the comparison is generally made to the wild-type ketoreductase enzyme, although in some embodiments, the reference ketoreductase can be another improved engineered ketoreductase. Enzyme properties for which improvement is desirable include, but are not limited to, enzymatic activity (which can be expressed in terms of percent conversion of the substrate), thermal stability, pH activity profile, cofactor requirements, refractoriness to inhibitors (*e.g.*, product inhibition), stereospecificity, and stereoselectivity (including enantioselectivity).

[0054] "Increased enzymatic activity" refers to an improved property of the engineered ketoreductase polypeptides, which can be represented by an increase in specific activity (*e.g.*, product produced/time/weight protein) or an increase in percent conversion of the substrate to the product (*e.g.*, percent conversion of starting amount of substrate to product in a specified time period using a specified amount of KRED) as compared to the reference ketoreductase enzyme. Exemplary methods to determine enzyme activity are provided in the Examples. Any property relating to enzyme activity may be affected, including the classical enzyme properties of K_m, V_{max} or k_{cat}, changes of which can lead to increased enzymatic activity. Improvements in enzyme activity can be from about 1.5 times the enzymatic activity of the corresponding wild-type ketoreductase enzyme, to as much as 2 times, 3 times, 4 times, 5 times, 10 times, 20 times, 25 times, 50 times, 75 times, 100 times, or more enzymatic activity than the naturally occurring ketoreductase or another engineered ketoreductase from which the ketoreductase polypeptides were derived. In specific embodiments, the engineered ketoreductase enzyme can exhibits improved enzymatic activity in the range of 1.5 to 50 times, 1.5 to 100 times greater than that of the parent ketoreductase enzyme. It is understood by the skilled artisan that the activity of any enzyme is diffusion limited such that the catalytic turnover rate cannot exceed the diffusion rate of the substrate, including any required cofactors. The theoretical maximum of the diffusion limit, or k_{cat}/K_m, is generally about 10^8 to 10^9 (M^{-1} s^{-1}). Hence, any improvements in the enzyme activity of the ketoreductase will have an upper limit related to the diffusion rate of the substrates acted on by the ketoreductase enzyme. Ketoreductase activity can be measured by any one of standard assays used for measuring ketoreductase, such as a decrease in absorbance or fluorescence of

NADPH due to its oxidation with the concomitant reduction of a ketone to an alcohol, or by product produced in a coupled assay. Comparisons of enzyme activities are made using a defined preparation of enzyme, a defined assay under a set condition, and one or more defined substrates, as further described in detail herein. Generally, when lysates are compared, the numbers of cells and the amount of protein assayed are determined as well as use of identical expression systems and identical host cells to minimize variations in amount of enzyme produced by the host cells and present in the lysates.

[0055] "Conversion" refers to the enzymatic conversion of the substrate to the corresponding product. "Percent conversion" refers to the percent of the substrate that is reduced to the product within a period of time under specified conditions. Thus, the "enzymatic activity" or "activity" of a ketoreductase polypeptide can be expressed as "percent conversion" of the substrate to the product.

[0056] "Thermostable" refers to a ketoreductase polypeptide that maintains similar activity (more than 60% to 80% or more, for example) after exposure to elevated temperatures (*e.g.*, 40-80°C) for a period of time (*e.g.*, 0.5-24 hrs) compared to the untreated enzyme.

[0057] "Solvent stable" refers to a ketoreductase polypeptide that maintains similar activity (more than, *e.g.*, 60% to 80%) after exposure to varying concentraions (*e.g.*, 5-99%) of a solvent(*e.g.*, isopropyl alcohol, tetrahydrofuran, 2-methyltetrahydrofuran, acetone, toluene, butylacetate, methyl tert-butylether, etc.) or solvent mixture, for a period of time (*e.g.*, 0.5-24 hrs) compared to the untreated enzyme.

[0058] "pH stable" refers to a ketoreductase polypeptide that maintains similar activity (more than, *e.g.*, 60% to 80%) after exposure to high or low pH (e.g., 4.5-6 or 8 to 12) for a period of time (*e.g.*, 0.5-24 hrs) compared to the untreated enzyme.

[0059] "Thermo- and solvent stable" refers to a ketoreductase polypeptide that are both thermostable and solvent stable.

[0060] "Derived from" as used herein in the context of engineered ketoreductase enzymes, identifies the originating ketoreductase enzyme, and/or the gene encoding such ketoreductase enzyme, upon which the engineering was based. For example, the engineered ketoreductase enzyme of SEQ ID NO: 10 was obtained by artificially evolving, over multiple generations the gene encoding the *Lactobacillus kefir* ketoreductase enzyme of SEQ ID NO:2. Thus, this engineered ketoreductase enzyme is "derived from" the wild-type ketoreductase of SEQ ID NO: 2.

[0061] "Hydrophilic Amino Acid or Residue" refers to an amino acid or residue having a side chain exhibiting a hydrophobicity of less than zero according to the normalized consensus hydrophobicity scale of Eisenberg *et al.*, 1984, *J. Mol. Biol.* 179:125-142. Genetically encoded hydrophilic amino acids include L-Thr (T), L-Ser (S), L-His (H), L-Glu (E), L-Asn (N), L-Gln (Q), L-Asp (D), L-Lys (K) and L-Arg (R).

[0062] "Acidic Amino Acid or Residue" refers to a hydrophilic amino acid or residue having a side chain exhibiting a pK value of less than about 6 when the amino acid is included in a peptide or polypeptide. Acidic amino acids typically have negatively charged side chains at physiological pH due to loss of a hydrogen ion. Genetically encoded acidic amino acids include L-Glu (E) and L-Asp (D).

[0063] "Basic Amino Acid or Residue" refers to a hydrophilic amino acid or residue having a side chain exhibiting a pK value of greater than about 6 when the amino acid is included in a peptide or polypeptide. Basic amino acids typically have positively charged side chains at physiological pH due to association with hydronium ion. Genetically encoded basic amino acids include L-Arg (R) and L-Lys (K).

[0064] "Polar Amino Acid or Residue" refers to a hydrophilic amino acid or residue having a side chain that is uncharged at physiological pH, but which has at least one bond in which the pair of electrons shared in common by two atoms is held more closely by one of the atoms. Genetically encoded polar amino acids include L-Asn (N), L-Gln (Q), L-Ser (S) and L-Thr (T).

[0065] "Hydrophobic Amino Acid or Residue" refers to an amino acid or residue having a side chain exhibiting a hydrophobicity of greater than zero according to the normalized consensus hydrophobicity scale of Eisenberg *et al.*, 1984, *J. Mol. Biol.* 179:125-142. Genetically encoded hydrophobic amino acids include L-Pro (P), L-Ile (I), L-Phe (F), L-Val (V), L-Leu (L), L-Trp (W), L-Met (M), L-Ala (A) and L-Tyr (Y).

[0066] "Aromatic Amino Acid or Residue" refers to a hydrophilic or hydrophobic amino acid or residue having a side chain that includes at least one aromatic or heteroaromatic ring. Genetically encoded aromatic amino acids include L-Phe (F), L-Tyr (Y) and L-Trp (W).

Although owing to the pKa of its heteroaromatic nitrogen atom L-His (H) it is sometimes classified as a basic residue, or as an aromatic residue as its side chain includes a heteroaromatic ring, herein histidine is classified as a hydrophilic residue or as a "constrained residue" (see below).

[0067] "Constrained amino acid or residue" refers to an amino acid or residue that has a constrained geometry. Herein, constrained residues include L-pro (P) and L-his (H). Histidine has a constrained geometry because it has a relatively small imidazole ring. Proline has a constrained geometry because it also has a five membered ring.

[0068] "Non-polar Amino Acid or Residue" refers to a hydrophobic amino acid or residue having a side chain that is uncharged at physiological pH and which has bonds in which the pair of electrons shared in common by two atoms is generally held equally by each of the two atoms (*i.e.*, the side chain is not polar). Genetically encoded non-polar amino acids include L-Gly (G), L-Leu (L), L-Val (V), L-Ile (I), L-Met (M) and L-Ala (A).

[0069] "Aliphatic Amino Acid or Residue" refers to a hydrophobic amino acid or residue having an aliphatic hydrocarbon side chain. Genetically encoded aliphatic amino acids include L-Ala (A), L-Val (V), L-Leu (L) and L-Ile (I).

[0070] "Cysteine". The amino acid L-Cys (C) is unusual in that it can form disulfide bridges with other L-Cys (C) amino acids or other sulfanyl- or sulfhydryl-containing amino acids. The "cysteine-like residues" include cysteine and other amino acids that contain sulfhydryl moieties that are available for formation of disulfide bridges. The ability of L-Cys (C) (and other amino acids with -SH containing side chains) to exist in a peptide in either the reduced free -SH or oxidized disulfide-bridged form affects whether L-Cys (C) contributes net hydrophobic or hydrophilic character to a peptide. While L-Cys (C) exhibits a hydrophobicity of 0.29 according to the normalized consensus scale of Eisenberg (Eisenberg *et al.*, 1984, *supra*), it is to be understood that for purposes of the present disclosure L-Cys (C) is categorized into its own unique group.

[0071] "Small Amino Acid or Residue" refers to an amino acid or residue having a side chain that is composed of a total three or fewer carbon and/or heteroatoms (excluding the α-carbon and hydrogens). The small amino acids or residues may be further categorized as aliphatic, non-polar, polar or acidic small amino acids or residues, in accordance with the above definitions. Genetically-encoded small amino acids include L-Ala (A), L-Val (V), L-Cys (C), L-Asn (N), L-Ser (S), L-Thr (T) and L-Asp (D).

[0072] "Hydroxyl-containing Amino Acid or Residue" refers to an amino acid containing a hydroxyl (-OH) moiety. Genetically-encoded hydroxyl-containing amino acids include L-Ser (S) L-Thr (T) and L-Tyr (Y).

[0073] "Conservative" amino acid substitutions or mutations refer to the interchangeability of residues having similar side chains, and thus typically involves substitution of the amino acid in the polypeptide with amino acids within the same or similar defined class of amino acids. However, as used herein, conservative mutations do not include substitutions from a hydrophilic to hydrophilic, hydrophobic to hydrophobic, hydroxyl-containing to hydroxyl-containing, or small to small residue, if the conservative mutation can instead be a substitution from an aliphatic to an aliphatic, non-polar to non-polar, polar to polar, acidic to acidic, basic to basic, aromatic to aromatic, or constrained to constrained residue. Further, as used herein, A, V, L, or I can be conservatively mutated to either another aliphatic residue or to another non-polar residue. The table below shows exemplary conservative substitutions.

Residue	**Possible Conservative Mutations**
A, L, V, I	Other aliphatic (A, L, V, I) Other non-polar (A, L, V, I, G, M)
G, M	Other non-polar (A, L, V, I, G, M)
D, E	Other acidic (D, E)
K, R	Other basic (K, R)
P, H	Other constrained (P, H)
N, Q, S, T	Other polar
Y, W, F	Other aromatic (Y, W, F)
C	None

[0074] "Non-conservative substitution" refers to substitution or mutation of an amino acid in the polypeptide with an amino acid with significantly differing side chain properties. Non-conservative substitutions may use amino acids between, rather than within, the defined groups listed above. In one embodiment, a non-conservative mutation affects (a) the structure of the peptide backbone in the area of the substitution (*e.g.*, proline for glycine) (b) the charge or hydrophobicity, or (c) the bulk of the side chain.

[0075] "Deletion" refers to modification to the polypeptide by removal of one or more amino acids to a reference polypeptide. Deletions can comprise removal of 1, 2, 3, 4, 5 or more amino acids, 10 or more amino acids, 15 or more amino acids, or 20 or more amino acids, up to 10% of the total number of amino acids, or up to 20% of the total number of amino acids making up the reference enzyme while retaining enzymatic activity and/or retaining the improved properties of an engineered ketoreductase enzyme. Deletions can be directed to the internal portions and/or terminal portions of the polypeptide. In various embodiments, the deletion can comprise a continuous segment or can be discontinuous.

[0076] "Insertion" refers to modification to the polypeptide by addition of one or more amino acids from the reference polypeptide. In some embodiments, the improved engineered ketoreductase enzymes comprise insertions of one or more amino acids to the naturally occurring ketoreductase polypeptide as well as insertions of one or more amino acids to other engineered ketoreductase polypeptides. Insertions can be in the internal portions of the polypeptide, or to the carboxy or amino terminus. Insertions as used herein include fusion proteins as is known in the art. The insertion can be a contiguous segment of amino acids or separated by one or more of the amino acids in the naturally occurring polypeptide.

[0077] "Different from" or "differs from" with respect to a designated reference sequence refers to difference of a given amino acid or polynucleotide sequence when aligned to the reference sequence. Generally, the differences can be determined when the two sequences are optimally aligned. Differences include insertions, deletions, or substitutions of amino acid residues in comparison to the reference sequence.

[0078] "Fragment" as used herein refers to a polypeptide that has an amino-terminal and/or carboxy-terminal deletion, but where the remaining amino acid sequence is identical to the corresponding positions in the sequence. Fragments can be at least 14 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 reference sequence, such as a wild-type (SEQ ID NO:2 or SEQ ID NO:4)or engineered ketoreductase polypeptide.

[0079] "Isolated polypeptide" refers to a polypeptide which is substantially separated from other contaminants that naturally accompany it, *e.g.*, protein, lipids, and polynucleotides. The term embraces polypeptides which have been removed or purified from their naturally-occurring environment or expression system (*e.g.*, host cell or *in vitro* synthesis). The improved ketoreductase enzymes may be present within a cell, present in the cellular medium, or prepared in various forms, such as lysates or isolated preparations. As such, in some embodiments, the improved ketoreductase enzyme can be an isolated polypeptide.

[0080] "Substantially pure polypeptide" refers to a composition in which the polypeptide species is the predominant species present (*i.e.*, on a molar or weight basis it is more abundant than any other individual macromolecular species in the composition), and is generally a substantially purified composition when the object species comprises at least about 50 percent of the macromolecular species present by mole or % weight. Generally, a substantially pure ketoreductase composition will comprise about 60 % or more, about 70% or more, about 80% or more, about 90% or more, about 95% or more, and about 98% or more of all macromolecular species by mole or % weight present in the composition. In some embodiments, the object species is purified to essential homogeneity (*i.e.*, contaminant species cannot be detected in the composition by conventional detection methods) wherein the composition consists essentially of a single macromolecular species. Solvent species, small molecules (<500 Daltons), and elemental ion species are not considered macromolecular species. In some embodiments, the isolated improved ketoreductases polypeptide is a substantially pure polypeptide composition.

[0081] "Stringent hybridization" is used herein to refer to conditions under which nucleic acid hybrids are stable. As known to those of skill in the art, the stability of hybrids is reflected in the melting temperature (T_m) of the hybrids. In general, the stability of a hybrid is a function of ion strength, temperature, G/C content, and the presence of chaotropic agents. The T_m values for polynucleotides can be calculated using known methods for predicting melting temperatures (see, *e.g.*, Baldino et al., *Methods Enzymology* 168:761-777; Bolton et al., 1962, *Proc. Natl. Acad. Sci. USA* 48:1390; Bresslauer et al., 1986, *Proc. Natl. Acad. Sci USA* 83:8893-8897; Freier et al., 1986, *Proc. Natl. Acad. Sci USA* 83:9373-9377; Kierzek et al., *Biochemistry* 25:7840-7846; Rychlik et al., 1990, *Nucleic Acids Res* 18:6409-6412 (erratum, 1991, *Nucleic Acids Res* 19:698); Sambrook et al., *supra*); Suggs et al., 1981, In *Developmental Biology Using Purified Genes* (Brown et al., eds.), pp. 683-693, Academic Press; and Wetmur, 1991, *Crit Rev Biochem Mol Biol* 26:227-259. All publications incorporate herein by reference). In some embodiments, the polynucleotide encodes the polypeptide disclosed herein and hybridizes under defined conditions, such as moderately stringent or highly stringent conditions, to the complement of a sequence encoding an engineered ketoreductase enzyme of the present disclosure.

[0082] "Hybridization stringency" relates to such washing conditions of nucleic acids. Generally, hybridization reactions are performed under conditions of lower stringency, followed by washes of varying but higher stringency. The term "moderately stringent hybridization" refers to conditions that permit target-polynucleotides to bind a complementary polynucleotide that has about 60% sequence identity, about 75% sequence identity, about 85% sequence identity; about 90% sequence identity, or with about 95% or greater sequence identity to the target-polynucleotide. Exemplary moderately stringent conditions are conditions equivalent to hybridization in 50% formamide, 5× Denhart's solution, 5×SSPE, 0.2% SDS at 42°C., followed by washing in 0.2×SSPE, 0.2% SDS, at 42°C.

[0083] "High stringency hybridization" refers generally to conditions that are about 10°C or less from the thermal melting temperature T_m as determined under the solution condition for a defined polynucleotide sequence. In some embodiments, a high stringency condition refers to conditions that permit hybridization of only those nucleic acid sequences that form stable hybrids in 0.018M NaCl at 65°C. (*i.e.*, if a hybrid is not stable in 0.018M NaCl at 65°C, it will not be stable under high stringency conditions, as contemplated herein). High stringency conditions can be provided, for example, by hybridization in conditions equivalent to 50% formamide, 5× Denhart's solution, 5×SSPE, 0.2% SDS at 42°C, followed by washing in 0.1×SSPE, and 0.1% SDS at 65°C. Other high stringency hybridization conditions, as well as moderately stringent conditions, are described in the references cited above.

[0084] "Heterologous" polynucleotide refers to any polynucleotide that is introduced into a host cell by laboratory techniques, and includes polynucleotides that are removed from a host cell, subjected to laboratory manipulation, and then reintroduced into a host cell.

[0085] "Codon optimized" refers to changes in the codons of the polynucleotide encoding a protein to those preferentially used in a particular organism such that the encoded protein is efficiently expressed in the organism of interest. Although the genetic code is degenerate in that most amino acids are represented by several codons, called "synonyms" or "synonymous" codons, it is well known that codon usage by particular organisms is nonrandom and biased towards particular codon triplets. This codon usage bias may be higher in reference to a given gene, genes of common function or ancestral origin, highly expressed proteins versus low copy number proteins, and the aggregate protein coding regions of an organism's genome. In some embodiments, the polynucleotides encoding the ketoreductases enzymes may be codon optimized for optimal production from the host organism selected for expression.

[0086] "Preferred, optimal, high codon usage bias codons" refers interchangeably to codons that are used at higher frequency in the protein coding regions than other codons that code for the same amino acid. The preferred codons may be determined in relation to codon usage in a single gene, a set of genes of common function or origin, highly expressed genes, the codon frequency in the aggregate protein coding regions of the whole organism, codon frequency in the aggregate protein coding regions of related organisms, or combinations thereof. Codons whose frequency increases with the level of gene expression are typically optimal codons for expression. A variety of methods are known for determining the codon frequency (*e.g.*, codon usage, relative synonymous codon usage) and codon preference in specific organisms, including multivariat analysis, for example, using cluster analysis or correspondence analysis, and the effective number of codons used in a gene (see GCG CodonPreference, Genetics Computer Group Wisconsin Package; CodonW, John Peden, University of Nottingham; McInerney, J. O, 1998, Bioinformatics 14:372-73; Stenico et al., 1994, Nucleic Acids Res. 222437-46; Wright, F., 1990, Gene 87:23-29). Codon usage tables are available for a growing list of organisms (see for example, Wada et al., 1992, *Nucleic Acids Res.* 20:2111-2118; Nakamura et al., 2000, *Nucl. Acids Res.* 28:292; Duret, et al., *supra*; Henaut and Danchin, "*Escherichia coli* and *Salmonella*," 1996, Neidhardt, et al. Eds., ASM Press, Washington D.C., p. 2047-2066. The data source for obtaining codon usage may rely on any available nucleotide sequence capable of coding for a protein. These data sets include nucleic acid sequences actually known to encode expressed proteins (*e.g.*, complete protein coding sequences-CDS), expressed sequence tags (ESTS), or predicted coding regions of genomic sequences (see for example, Mount, D., *Bioinformatics: Sequence and Genome Analysis*, Chapter 8, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 2001;

Uberbacher, E. C., 1996, *Methods Enzymol.* 266:259-281; Tiwari et al., 1997, *Comput. Appl. Biosci.* 13:263-270).

[0087] "Control sequence" is defined herein to include all components, which are necessary or advantageous for the expression of a polypeptide of the present disclosure. Each control sequence may be native or foreign to the nucleic acid sequence encoding the polypeptide. Such control sequences include, but are not limited to, a leader, polyadenylation sequence, propeptide sequence, promoter, signal peptide sequence, and transcription terminator. At a minimum, the control sequences include a promoter, and transcriptional and translational stop signals. The control sequences may be provided with linkers for the purpose of introducing specific restriction sites facilitating ligation of the control sequences with the coding region of the nucleic acid sequence encoding a polypeptide.

[0088] "Operably linked" is defined herein as a configuration in which a control sequence is appropriately placed at a position relative to the coding sequence of a polynucleotide sequence such that the control sequence directs or affects the expression of a polynucleotide and/or polypeptide encoded by the polynucleotide.

[0089] "Promoter sequence" is a nucleic acid sequence that is recognized by a host cell for expression of a polynucleotide, such as a polynucleotide containing the coding region. Generally, the promoter sequence contains transcriptional control sequences, which mediate expression of the polynucleotide. The promoter may be any nucleic acid sequence which shows transcriptional activity in the host cell of choice including mutant, truncated, and hybrid promoters, and may be obtained from genes encoding extracellular or intracellular polypeptides either homologous or heterologous to the host cell.

[0090] "Cofactor regeneration system" refers to a set of reactants that participate in a reaction that reduces the oxidized form of the cofactor (*e.g.*, $NADP^+$ to NADPH). Cofactors oxidized by the ketoreductase-catalyzed reduction of the keto substrate are regenerated in reduced form by the cofactor regeneration system. Cofactor regeneration systems comprise a stoichiometric reductant that is a source of reducing hydrogen equivalents and is capable of reducing the oxidized form of the cofactor. The cofactor regeneration system may further comprise a catalyst, for example an enzyme catalyst, that catalyzes the reduction of the oxidized form of the cofactor by the reductant. Cofactor regeneration systems to regenerate

NADH or NADPH from NAD^+ or $NADP^+$, respectively, are known in the art and may be used in the methods described herein.

6.2 Ketoreductase Enzymes

[0091] The present disclosure provides engineered or recombinant ketoreductase ("KRED") enzymes that are capable of reducing or converting a defined keto substrate to its corresponding alcohol product and having an improved property when compared with the naturally-occurring, wild-type ketoreductase enzyme obtained from *Lactobacillus kefir* or *Lactobacillus brevis*, or another reference ketoreductase enzyme, such as another engineered ketoreductase enzyme. In the embodiments herein, the recombinant ketoreductase polypeptides are capable of stereoselectively reducing or converting the substrate 1-[4-(4-fluoro-2-methyl-1H-indol-5-yloxy)-5-methyl-pyrrolo[2,1-f][1,2,4,]triazine-6-yloxy]-propan-2-one, as represented by the structure of formula (I), to the product (*R*)-1-[4-(4-fluoro-2-methyl-1H-indol-5-yloxy)-5-methyl-pyrrolo[2,1-f][1,2,4,]triazine-6-yloxy]-propan-2-ol, as represented by the structure of formula (II), as further discussed below. The ketoreductase have an improved property in converting the substrate of formula (I) to the product of formula (II) as compared to the naturally occurring ketoreductase enzymes of *Lactobacillus kefir* or *Lactobacillus bacillus*.

[0092] In some embodiments, the recombinant ketoreductase polypeptides are also capable of reducing or converting the substrate acetophenone, as represented by structural formula (III) to the chiral alcohol product (*R*) 1-phenylethanol, as represented by the structure of formula (IV), as further discussed below. Thus, the engineered ketoreductases can be compared to reference ketoreductases using the substrate 1-[4-(4-fluoro-2-methyl-1H-indol-5-yloxy)-5-methyl-pyrrolo[2,1-f][1,2,4,]triazine-6-yloxy]-propan-2-one or acetophenone, or both. In some embodiments, one reference substrate (*e.g.*, 1-[4-(4-fluoro-2-methyl-1H-indol-5-yloxy)-5-methyl-pyrrolo[2,1-f][1,2,4,]triazine-6-yloxy]-propan-2-one) can be substituted with the other reference substrate (*e.g.*, acetophenone) when comparing the activity of the engineered ketoreductases to the reference ketoreductases (*e.g.*, SEQ ID NO:2 or SEQ ID NO:4).

[0093] In the embodiments herein, the ketoreductase polypeptides comprise at least (1) an amino acid residue at position 94 which is an aromatic amino acid residue or G, and/or (2) an amino acid residue at position 96 which is an amino acid other than S/N in the corresponding residue position of the wild-type *L. kefir* or *L. brevis* sequences of SEQ ID NO:2 and 4, respectively. Thus, in various embodiments herein, the ketoreductases of the present disclosure comprise at least the following amino acid substitutions: (1) residue 94 is modified from A → an aromatic amino acid residue or G, and/or (2) residue 96 is modified from S/N → any amino acid other than S/N, for a polypeptide corresponding to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:4.

[0094] These non-naturally occurring ketoreductases can be generated by various well-known techniques, such as *in vitro* mutagenesis or directed evolution of the genetic material encoding the ketoreductase enzyme of *Lactobacillus kefir* or *Lactobacillus brevis*, and identifying polynucleotides that express engineered enzymes with a desired property. The capabilities of the KREDs can refer to their properties as single polypeptides or to their properties in the form of multimers, as they may be present in the wild-type enzyme.

[0095] Mutagenesis and directed evolution techniques useful for the purposes herein are amply described in the literature: Ling, et al., 1997, "Approaches to DNA mutagenesis: an overview," *Anal. Biochem.* 254(2):157-78; Dale et al., 1996, "Oligonucleotide-directed random mutagenesis using the phosphorothioate method," *Methods Mol. Biol.* 57:369-74; Smith, 1985, "In vitro mutagenesis," *Ann. Rev. Genet.* 19:423-462; Botstein et al., 1985, "Strategies and applications of in vitro mutagenesis," *Science* 229:1193-1201; Carter, 1986, "Site-directed mutagenesis," *Biochem. J.* 237:1-7; Kramer et al., 1984, "Point Mismatch Repair," *Cell*, 38:879-887; Wells et al., 1985, "Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites," *Gene* 34:315-323; Minshull et al., 1999, "Protein evolution by molecular breeding," *Curr Opin Chem Biol* 3:284-290; Christians et al., 1999, "Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling," *Nature Biotech* 17:259-264; Crameri et al., 1998, "DNA shuffling of a family of genes from diverse species accelerates directed evolution," *Nature* 391:288-291; Crameri et al., 1997, "Molecular evolution of an arsenate detoxification pathway by DNA shuffling," *Nature Biotech* 15:436-438; Zhang et al., 1997, "Directed evolution of an effective fructosidase from a galactosidase by DNA shuffling and screening," *Proc Natl Acad Sci USA* 94:45-4-4509; Crameri et al., 1996, "Improved green fluorescent protein by molecular evolution using DNA shuffling,' *Nature Biotech* 14:315-319; Stemmer, 1994, "Rapid evolution of a protein in vitro by DNA shuffling," *Nature* 370:389-391; Stemmer, 1994, "DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution," *Proc Natl Acad Sci USA* 91:10747-10751; WO 95/22625; WO 97/0078; WO 97/35966; WO 98/27230; WO 00/42651; WO 01/75767 and US Pat. 6,537,746. All publications are incorporated herein by reference.

[0096] The naturally occurring polynucleotide encoding the naturally occurring ketoreductase of *Lactobacillus kefir* and *Lactobacillus brevis* (also referred to as "alcohol dehydrogenase" or "ADH") can be obtained from the isolated polynucleotide known to encode the ketoreductase activity (*e.g.*, Genbank accession no. AAP94029 GI:33112056 for *Lactobacillus kefir* and Genbank accession no. CAD66648 GI:28400789 for *Lactobacillus brevis*). Alternatively, a polynucleotide encoding the naturally occurring ketoreductases can be synthesized by polynucleotide synthesis methodologies known in the art based on the reported polynucleotide sequence of the ketoreductase-encoding gene. In various embodiments, as further described below, the naturally occurring polynucleotide encoding the ketoreductase, as well as the engineered ketoreductases, can be codon optimized to the codons preferred by a specific host cell used for expression of the enzyme.

[0097] The parent or reference polynucleotide encoding the naturally occurring or wild type ketoreductase is subjected to mutagenic processes, for example random mutagenesis and recombination, to introduce mutations into the polynucleotide. The mutated polynucleotide is expressed and translated, thereby generating engineered ketoreductase enzymes with modifications to the polypeptide. As used herein, "modifications" include amino acid substitutions, deletions, and insertions. Any one or a combination of modifications can be introduced into the naturally occurring enzymatically active polypeptide to generate engineered enzymes, which are then screened by various methods to identify polypeptides, and corresponding polynucleotides, having a desired improvement in a specific enzyme property. A polynucleotide encoding an engineered ketoreductase with an improved property can be subjected to additional rounds of mutagenesis treatments to generate polypeptides with further improvements in the desired enzyme property. Enzyme properties for which improvement is desirable include, but are not limited to, enzymatic activity, thermal stability, pH activity profile, cofactor requirements, refractoriness to inhibitors (*e.g.*, product inhibition), sterospecificity, stereoselectivity, and solvent stability.

[0098] In some embodiments, the recombinant ketoreductase comprise engineered polypeptides derived from, and thus can be compared to, *Lactobacillus kefir* ketoreductase of SEQ ID NO:2 or *Lactobacillus brevis* ketoreductase of SEQ ID NO:4. In the descriptions here, the amino acid residue position is determined with respect to a reference polypeptide. For the reference sequences of SEQ ID NO:2 or SEQ ID NO:4, the numbering begins from the initiating methionine (M) residue (*i.e.*, M represents residue position 1), although it will be understood by the skilled artisan that this initiating methionine residue may be removed by biological processing machinery, such as in a host cell or *in vitro* translation system, to generate a mature protein lacking the initiating methionine residue. Where the amino acid residues at the same residue position differ between the two reference ketoreductases, the different residues are denoted by a "/" with the arrangement being "kefir residue/brevis residue". A substitution mutation, which is a replacement of an amino acid residue in a reference sequence, for example the wildtype ketoreductases of SEQ ID NO:2 and SEQ ID NO:4, with a different amino acid residue is denoted by the symbol "→".

[0099] The number of modifications to a reference polypeptide, such as the naturally occurring polypeptide of SEQ ID NO:2 or SEQ ID NO:4, that produces an improved ketoreductase property may comprise one or more amino acids, 2 or more amino acids, 3 or more amino acids, 4 or more amino acid, 5 or more amino acids, 6 or more amino acids, 8 or more amino acids, 10 or more amino acids, or 15 or more amino acids, 20 or more amino acids, up to 10% of the total number of amino acids, up to 20% of the total number of amino acids, or up to 30% of the total number of amino acids of the reference enzyme sequence. As such, the polypeptides of the present disclosure can differ from the reference polypeptide (e.g., SEQ ID NO:2 or SEQ ID NO:4) in 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14 16, 18, 20 or 25 or more amino acids, or up to 5% of the amino acids, up to 10% of the amino acids, up to 20% of the amino acids, or up to 30% of the total number of amino acids of the reference sequence.

[0100] In various embodiments, the modifications to the reference polypeptide to produce the improved enzyme property can comprise substitutions at 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14 16, 18, 20, 25, or more amino acids, or up to 5% of the amino acids, up to 10% of the amino acids, up to 20% of the amino acids, or up to 30% of the total number of amino acids of the reference sequence, such as SEQ ID NO:2 or SEQ ID NO:4. The substitutions for generating an improved ketoreductase can comprise conservative substitutions, non-conservative substitutions, as well as combinations of conservative and non-conservative substitutions.

[0101] In some embodiments, the ketoreductase polypeptides have increased enzyme activity in stereoselectively reducing or converting the substrate 1-[4-(4-fluoro-2-methyl-1H-indol-5-yloxy)-5-methyl-pyrrolo[2,1-f][1,2,4,]triazine-6-yloxy]-propan-2-one to the product (*R*)-1-[4-(4-fluoro-2-methyl-1H-indol-5-yloxy)-5-methyl-pyrrolo[2,1-f][1,2,4,]triazine-6-yloxy]-propan-2-ol as compared to the activity of a wild-type reference ketoreductases of SEQ ID NO:2 or SEQ ID NO:4. In some embodiments, the increased enzyme activity is at least 1.5 times or more the enzyme activity of the wild-type reference polypeptides. In some embodiments, the increased enzyme activity is at least 2.0 times or more enzyme activity, at least 3.0 times or more enzyme activity, at least 5 times or more enzyme activity, at least 10 times or more enzyme activity, at least 20 times or more enzyme activity, at least 25 times or more enzyme activity, at least 50 times or more enzyme activity, at least 75 times or more enzyme activity, at least 100 times or more enzyme activity as compared to the activity of reference polypeptides SEQ ID NO:2 or SEQ ID NO:4.

[0102] In some embodiments, the ketoreductase polypeptides have an increased conversion rate in stereoselectively reducing or converting 1-[4-(4-fluoro-2-methyl-1H-indol-5-yloxy)-5-methyl-pyrrolo[2,1-f][1,2,4,]triazine-6-yloxy]-propan-2-one to the product (*R*)-1-[4-(4-fluoro-2-methyl-1H-indol-5-yloxy)-5-methyl-pyrrolo[2,1-f][1,2,4,]triazine-6-yloxy]-propan-2-ol as compared to the conversion rate of a wild-type reference ketoreductase of SEQ ID NO:2 or SEQ ID NO:4 under a defined condition. In some embodiments, the engineered ketoreductases are characterized by a conversion of greater than 70%, of greater than 80%, of greater than 90%, of greater than 95%, of greater than 98% or of greater than 99% of the substrate under a defined condition. An exemplary defined condition is conversion in 24 hours of 10 g/L substrate with about 10 g/L of the KRED.

[0103] In some embodiments, the ketoreductase polypeptides have an increased enzyme activity in stereoselectively reducing or converting the substrate acetophenone to the product (*R*)-1-phenylethanol as compared to the activity of a wild-type reference ketoreductases of SEQ ID NO:2 or SEQ ID NO:4. In some embodiments, the increased enzyme activity is at least 1.5 times or more the enzyme activity of the wild-type reference polypeptides. In some embodiments, the increased enzyme activity is at least 2.0 times or more enzyme activity, at least 3.0 times or more enzyme activity, at least 5 times or more enzyme activity, at least 10 times or more enzyme activity, at least 20 times or more enzyme activity, at least 25 times or more enzyme activity, at least 50 times or more enzyme activity, at least 75 times or more enzyme activity, at least 100 times or more enzyme activity as compared to the activity of reference polypeptides SEQ ID NO:2 or SEQ ID NO:4.

[0104] In some embodiments, the ketoreductase polypeptides have an increased conversion rate in stereoselectively reducing or converting the substrate acetophenone to the product (*R*)-1-phenylethanol as compared to the activity of a wild-type reference ketoreductases of SEQ ID NO:2 or SEQ ID NO:4 under a defined condition. In some embodiments, the engineered ketoreductases are characterized by conversion of greater than 70%, of greater than 80%, of greater than 90%, of greater than 95%, of greater than 98% or of greater than 99% of the substrate under the defined condition. An exemplary defined condition is conversion in 24 hours of 10 g/L substrate with about 10 g/L of the KRED.

[0105] In various embodiments, the ketoreductase polypeptide is capable of stereoselectively reducing or converting the substrate 1-[4-(4-fluoro-2-methyl-1H-indol-5-yloxy)-5-methyl-pyrrolo[2,1-f][1,2,4,]triazine-6-yloxy]-propan-2-one to the product (*R*)-1-[4-(4-fluoro-2-methyl-1H-indol-5-yloxy)-5-methyl-pyrrolo[2,1-f][1,2,4,]triazine-6-yloxy]-propan-2-ol with at least 1.5 times the activity of SEQ ID NO:2 or SEQ ID NO:4, and comprises an amino acid sequence having (1) an aromatic amino acid or G at the corresponding residue position 94 of the reference sequence SEQ ID NO:2 or SEQ ID NO:4, or (2) an amino acid other than S/N at the corresponding residue position 96 of the reference sequence SEQ ID NO:2 or SEQ ID NO:4. In some embodiments, the amino acid residue at the corresponding residue position 96 is other than S and N. In some of these embodiments, the ketoreductase polypeptide can differ from the reference sequence in 2 or more amino acid residues, 3 or more amino acid residues, or 4 or more amino acid residues as compared to the reference sequence, as discussed below.

[0106] In some embodiments, the ketoreductase polypeptide capable of stereoselectively reducing or converting the substrate 1-[4-(4-fluoro-2-methyl-1H-indol-5-yloxy)-5-methyl-pyrrolo[2,1-f][1,2,4,]triazine-6-yloxy]-propan-2-one to the product (*R*)-1-[4-(4-fluoro-2-methyl-1H-indol-5-yloxy)-5-methyl-pyrrolo[2,1-f][1,2,4,]triazine-6-yloxy]-propan-2-ol comprises an amino acid sequence having an aromatic amino acid or G at the corresponding residue position 94 of the reference sequence SEQ ID NO:2 or SEQ ID NO:4. As such, in some embodiments, the ketoreductase polypeptide comprises a modification at residue 94 of A → an aromatic amino acid residue or G of the corresponding sequence of SEQ ID NO:2 or SEQ ID NO:4. In some of these embodiments, the ketoreductase polypeptide can differ from the reference sequence in 2 or more amino acid residues, 3 or more amino acid residues, or 4 or more amino acid residues as compared to the reference sequence, as discussed below.

[0107] In some embodiments, the ketoreductase polypeptide comprises an amino acid sequence having an amino acid F, W, H, or Y at the corresponding amino acid residue of residue 94 of SEQ ID NO:2 or SEQ ID NO:4. As such, in some embodiments, the ketoreductase polypeptide comprises a modification at residue 94 of A → F, W, H, or Y of the corresponding sequence of SEQ. ID NO:2 or SEQ ID NO:4. In some embodiments, the amino acid at residue 94 is F.

[0108] In some embodiments, the ketoreductase polypeptides capable of stereoselectively converting the substrate 1-[4-(4-fluoro-2-methyl-1H-indol-5-yloxy)-5-methyl-pyrrolo[2,1-f][1,2,4,]triazine-6-yloxy]-propan-2-one to the product (*R*)-1-[4-(4-fluoro-2-methyl-1H-indol-5-yloxy)-5-methyl-pyrrolo[2,1-f][1,2,4,]triazine-6-yloxy]-propan-2-ol with the specified amino acids at residue 94 can comprise modifications at other amino acid residues of the corresponding SEQ ID NO:2 or SEQ ID NO:4, including non-conservative or conservative substitutions. Non-conservative substitutions can be at amino acid residues corresponding to residue positions 53, 54, 96, 97, 147, 165, 153, 197, 199, 206, 223, and 233, as further discussed below. Additional substitutions, when present, can comprise one or more conservative substitutions at other amino acid residue positions.

[0109] Thus, in some embodiments, the recombinant ketoreductase polypeptides comprise an amino acid sequence which differs in 1 to 25 amino acid positions from SEQ ID NO: 2 with the proviso that the amino acid residue corresponding to residue 94 of SEQ ID NO:2 or SEQ ID NO:4 is aromatic amino acid or G. As such, some polypeptides comprise an amino sequence which differs from SEQ ID NO: 2 or SEQ ID NO:4 in 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 16, 18, 20, or 25 amino acid positions, for example, in 1-25 amino acid positions, in 1-20 amino acid positions, in 1-18 amino acid positions, in 1-16 amino acid positions, 1-14 amino acid positions, in 1-12 amino acid positions, in 1-11 amino acid positions, in 1-10 amino acid positions, in 1-9 amino acid positions, in 1-8 amino acid positions, in 1-7 amino acid positions, in 1-6 amino acid positions, in 1-5 amino acid positions, in 1-4 amino acid positions, in 1-3 amino acid positions, or in 1-2 amino acid positions.

[0110] In some embodiments, the ketoreductase polypeptides capable of stereoselectively reducing or converting the substrate 1-[4-(4-fluoro-2-methyl-1H-indol-5-yloxy)-5-methyl-pyrrolo[2,1-f][1,2,4,]triazine-6-yloxy]-propan-2-one to the product (*R*)-1-[4-(4-fluoro-2-methyl-1H-indol-5-yloxy)-5-methyl-pyrrolo[2,1-f][1,2,4,]triazine-6-yloxy]-propan-2-ol comprises an amino acid sequence with an aromatic amino acid or G at residue corresponding to residue 94 of SEQ ID NO:2 or SEQ ID NO:4 and one or more of the features selected from: (1) amino acid at residue corresponding to residue 96 is any amino acid other than S/N; (2) amino acid residue corresponding to residue 153 is an aliphatic amino acid other than L; (3) amino acid residue corresponding to residue 199 is any amino acid other than L; (4) amino acid residue corresponding to residue 202 is G or an aliphatic amino acid other than A; and (5) amino acid residue corresponding to residue 206 is an aromatic amino acid. Thus, in some embodiments, the ketoreductase polypeptides can comprise an amino acid sequence with modifications at residue 94 of A → an aromatic amino acid residue or G and one or more modifications selected from: 96 (S/N → any amino acid other than S/N); 153 (L → an aliphatic amino acid residue other than L); 199 (L → any amino acid residue other than L); 202 (A → G or an aliphatic amino acid residue other than A); and 206 (M → an aromatic amino acid residue) of the corresponding sequence of SEQ ID NO:2 or SEQ ID NO:4.

[0111] In some embodiments, the amino acid residues, and corresponding mutations, in the above residues are selected from: residue 153 is G or A; residue 199 is K, I, N, R, V, Q, or W; residue 202 is I, L, or G; and residue 206 is F.

[0112] In some embodiments, the ketoreductase polypeptides capable of stereoselectively reducing or converting the substrate 1-[4-(4-fluoro-2-methyl-1H-indol-5-yloxy)-5-methyl-pyrrolo[2,1-f][1,2,4,]triazine-6-yloxy]-propan-2-one to the product (*R*)-1-[4-(4-fluoro-2-methyl-1H-indol-5-yloxy)-5-methyl-pyrrolo[2,1-f][1,2,4,]triazine-6-yloxy]-propan-2-ol comprises an amino acid sequence with a G, F, Y, or I at the amino acid residue corresponding to residue 96 of SEQ ID NO:2 or SEQ ID NO:4. As such, in some embodiments, the ketoreductase polypeptides comprise a modification at residue 96 of S/N → G, F, Y, or I of the corresponding sequence of SEQ. ID NO:2 or SEQ ID NO:4. In some of these embodiments, the ketoreductase polypeptide can differ from the reference sequence in 2 or more amino acid residues, 3 or more amino acid residue, or 4 or more amino acid residues as compared to the reference sequence, as discussed below.

[0113] In some embodiments, the ketoreductase polypeptides are also capable of stereoselectively reducing or converting the substrate acetophenone to the product (*R*)-1-phenylethanol. In some of these embodiments, the ketoreductase polypeptide comprises an amino acid sequence having a G, I, C or an aromatic amino acid at the amino acid residue corresponding to residue 96 of SEQ ID NO:2 or SEQ ID NO:4. As such, in some embodiments, the ketoreductase polypeptides comprise a modification at residue 96 of S/N → G, I, C or an aromatic amino acid residue of the corresponding sequence of SEQ ID NO:2 or SEQ ID NO:4. In some of these embodiments, the ketoreductase polypeptide can differ from the reference sequence in 2 or more amino acid residues, 3 or more amino acid residue, or 4 or more amino acid residues as compared to the reference sequence, as discussed below.

[0114] In some embodiments, the ketoreductase polypeptides with the specified amino acids at residue 96 can comprise modifications at other amino acids residues of the corresponding SEQ ID NO:2 or SEQ ID NO:4, including non-conservative or conservative substitutions. Non-conservative substitutions can be at amino acid residues corresponding to residue positions 53, 54, 96, 97, 147, 165, 153, 197, 199, 206, 223, and 233, as further discussed below. Additional substitutions, when present, can comprise one or more conservative substitutions at other amino acid residue positions.

[0115] In some embodiments, ketoreductase polypeptide comprises an amino acid sequence which differs in 1 to 25 amino acid positions from SEQ ID NO: 2 with the proviso that the amino acid residue corresponding to residue 96 of SEQ ID NO:2 or SEQ ID NO:4 is a G, F, Y, or I. As such, some polypeptides comprise an amino sequence which differs from SEQ ID NO: 2 or SEQ ID NO:4 in 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 16, 18, 20, or 25 amino acid positions, for example, in 1-25 amino acid positions, in 1-20 amino acid positions, in 1-18 amino acid positions, in 1-16 amino acid positions, 1-14 amino acid positions, in 1-12 amino acid positions, in 1-11 amino acid positions, in 1-10 amino acid positions, in 1-9 amino acid positions, in 1-8 amino acid positions, in 1-7 amino acid positions, in 1-6 amino acid positions, in 1-5 amino acid positions, in 1-4 amino acid positions, in 1-3 amino acid positions, or in 1-2 amino acid positions.

[0116] In some embodiments, ketoreductase polypeptide comprises an amino acid sequence which differs in 1 to 25 amino acid positions from SEQ ID NO: 2 and comprises G, I, C or an aromatic amino acid at the residue corresponding to residue 96 of SEQ ID NO:2 or SEQ ID NO:4. As such, some polypeptides comprise an amino sequence which differs from SEQ ID NO: 2 or SEQ ID NO:4 in 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 16, 18, 20, or 25 amino acid positions, for example, in 1-25 amino acid positions, in 1-20 amino acid positions, in 1-18 amino acid positions, in 1-16 amino acid positions, 1-14 amino acid positions, in 1-12 amino acid positions, in 1-11 amino acid positions, in 1-10 amino acid positions, in 1-9 amino acid positions, in 1-8 amino acid positions, in 1-7 amino acid positions, in 1-6 amino acid positions, in 1-5 amino acid positions, in 1-4 amino acid positions, in 1-3 amino acid positions, or in 1-2 amino acid positions.

[0117] In some embodiments, the ketoreductase polypeptide comprises an amino acid sequence with the specified amino acid at residue 96 of SEQ ID NO:2 or SEQ ID NO:4 above, and one or more of the following features: (1) amino acid at residue corresponding to residue 94 is an aromatic amino acid or G; (2) amino acid residue corresponding to residue 153 is an aliphatic amino acid other than L; (3) amino acid residue corresponding to residue 199 is any amino acid other than L; (4) amino acid residue corresponding to residue 202 is G or an aliphatic amino acid other than A; and (5) amino acid residue corresponding to residue 206 is an aromatic amino acid.

[0118] Thus, in some embodiments, the ketoreductase polypeptide comprises an amino acid sequence with modifications at residue 96 of S/N → G, F, Y, or I and one or more modifications at residue: 94 (A → an aromatic amino acid residue or G); 153 (L → an aliphatic amino acid residue other than L); 199 (L → any amino acid residue other than L); 202 (A → G or an aliphatic amino acid residue other than A); and 206 (M → an aromatic amino acid residue) of the corresponding sequence of SEQ ID NO:2 or SEQ ID NO:4.

[0119] In some embodiments, the ketoreductase polypeptide comprises an amino acid sequence with modifications at residue 96 of S/N → G, I, C and one or more modifications at the following residues: 94 (A → an aromatic amino acid residue or G); 153 (L → an aliphatic amino acid residue other than L); 199 (L → any amino acid residue other than L); 202 (A →

G or an aliphatic amino acid residue other than A); and 206 (M → an aromatic amino acid residue) of the corresponding sequence of SEQ ID NO:2 or SEQ ID NO:4.

[0120] In some embodiments, the amino acid residues, and corresponding mutations, in the above residues can be selected from one or more of the following: residue 94 is F or G; residue 153 is G or A; residue 199 is K, I, N, R, V, Q, or W; residue 202 is I, L, or G; and residue 206 is F.

[0121] In some embodiments, additional mutations can be incorporated into all of the ketoreductase polypeptide embodiments above to enhance one or more properties of the polypeptide activity, such as, among others, enzyme activity, thermal stability, and/or solvent stability. Thus, in some embodiments, the ketoreductase polypeptides can comprise, in addition to all of embodiments above, one or more of the following features: (1) amino acid at residue corresponding to residue 49 is a polar amino acid residue other than K; (2) amino acid at residue corresponding to residue 53 is an acidic amino acid residue; (3) amino acid at residue corresponding to residue 54 is a small or aliphatic amino acid residue other than T/P; (4) amino acid at residue corresponding to residue 60 is an aliphatic amino acid residue other than V; (5) amino acid at residue corresponding to residue 95 is an aliphatic amino acid other than V; (6) amino acid at residue corresponding to residue 97 is a small amino acid or G; (7) amino acid at residue corresponding to residue 109 is a basic amino acid residue other than K; (8) amino acid at residue corresponding to residue 147 is an aliphatic amino acid residue; (9) amino acid at residue corresponding to residue 165 is a hydroxyl or small amino acid residue; (10) amino acid at residue corresponding to residue 197 is a small amino acid residue or G; (11) amino acid at residue corresponding to residue 223 is an aliphatic amino acid residue other than L; and (12) amino acid at residue corresponding to residue 233 is a small amino acid residue or G.

[0122] Thus, in these embodiments, the ketoreductase polypeptide can comprise an amino acid sequence with the specified modifications described for every embodiment of the ketoreductases above and one or more of the following modifications: 49 (K → a polar amino acid residue other than K); 53 (G/T → an acidic amino acid residue); 54 (T/P → a small or aliphatic amino acid residue other than T/P); 60 (V/F → an aliphatic amino acid residue other than V); 95 (V → an aliphatic amino acid other than V); 97 (K → a small amino acid or G); 109 (K → a basic amino acid residue other than K); 147 (F → an aliphatic amino acid residue); 165 (I → a hydroxyl or small amino acid residue); 197 (D → a small amino acid residue or G; 223 (I → an aliphatic amino acid residue other than L); and 233 (D/N → a small amino acid residue or G) of the corresponding sequence of SEQ ID NO:2 or SEQ ID NO:4.

[0123] In some embodiments, the ketoreductase polypeptides comprises an amino acid sequence that is at least about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO:2 or SEQ ID NO:4 with the proviso that the residues corresponding to (1) residue 94 is an aromatic amino acid residue or G, and/or (2) residue 96 is an amino acid residue other than S/N. In some embodiments, the amino acid residue at the corresponding residue position 96 is other than S and N.

[0124] In some embodiments, the ketoreductase polypeptides can comprise an amino acid sequence that is at least about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO:2 or SEQ ID NO:4 with the proviso that the residues corresponding to: (1) residue 94 is F, W, H, or Y; (2) residue 96 is G, F, Y, or I; or (3) residue 96 is G, I, C or an aromatic amino acid.

[0125] In some embodiments, the ketoreductase polypeptides of the invention can comprise a region having an amino acid sequence that is at least about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a region or domain thereof of SEQ ID NO:2 or SEQ ID NO:4, such as residues 90-233, with the proviso that the residues corresponding to (1) residue 94 is an aromatic amino acid residue or G, and/or (2) residue 96 is an amino acid residue other than S/N of the corresponding residue of SEQ ID NO:2 or SEQ ID NO:4. In some embodiments, the amino acid residue at the corresponding residue position 96 is other than S and N. In some embodiments of these ketoreductase polypeptides, one or more of the remaining residues corresponding to residues 90-233 of SEQ ID NO:2 or SEQ ID NO:4 may be conservatively mutated.

[0126] In some embodiments, the ketoreductase polypeptides can comprise a region having an amino acid sequence that is at least about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a region or domain thereof of SEQ ID NO:2 or SEQ ID NO:4, such as residues 90-233, with the proviso that the residues corresponding to: (1) residue 94 is F, W, H, or Y; (2) residue 96 is G, F, Y, or I; or (3) residue

96 is G, I, C or an aromatic amino acid of the corresponding residue of SEQ ID NO:2 or SEQ ID NO:4. In some embodiments of these ketoreductase polypeptides, one or more of the remaining residues corresponding to residues 90-233 of SEQ ID NO:2 or SEQ ID NO:4 may be conservatively mutated.

[0127] In some embodiments, the engineered ketoreductase polypeptide is selected from the amino acid sequences recited in Table 1. Table 1 ranks the enzyme activity measured for the reduction or conversion of 1-[4-(4-fluoro-2-methyl-1H-indol-5-yloxy)-5-methyl-pyrrolo[2,1-f][1,2,4,]triazine-6-yloxy]-propan-2-one to the product (*R*)-1-[4-(4-fluoro-2-methyl-1H-indol-5-yloxy)-5-methyl-pyrrolo[2,1-f][1,2,4,]triazine-6-yloxy]-propan-2-ol using the listed engineered ketoreductase polypeptides compared to the activity measured using ADH-LK (SEQ ID NO:2).

[0128] Table 1:

Nucleic Acid SEQ ID NO.	Polypeptide SEQ ID NO.	Mutations from ADH-LK (SEQ ID NO:2)	Activity Ranking (see legend)
5	6	L17Q; A94F; S96Y; F147L; L153G; L199I	+++
7	8	G53D; S96G	+
9	10	S96G	+
11	12	S96G; K109R	+
13	14	S96Y	+
15	16	A94F; S96P	+
17	18	S96F	+
19	20	T54A; A94F; S96G; K109R; F147L; L153A; D233G	+++
21	22	V43A; V60A; A94G; F147L	+++
23	24	V43A; F74L; S96I; F147L; L153G; L199K; I223V	+++
25	26	V43A; F74L; S96G; F147L; L153A; D197G; L199I; A202L	++

Nucleic Acid SEQ ID NO.	Polypeptide SEQ ID NO.	Mutations from ADH-LK (SEQ ID NO:2)	Activity Ranking (see legend)
27	28	V43A; F74L; A94G; S96G; F147L; L153A; D197G; L199I; A202L; I223V	+++
29	30	V43A; F74L; A94G; S96Y; F147L; L153A; K211R; I223V	++
31	32	V43A; F74L; A94G; F147L; L199K; A202I	++
33	34	V43A; F74L; A94G; F147L; L199I; A202I; M206F; I223V	++
35	36	V43A; F74L; A94F; S96Y; L153A; I165T	++
37	38	V43A; A94G; F147L	+++
39	40	V43A; A94G; E106G; F147L; A202L	++
41	42	V43A; A94F; V95I; S96G; F147L; L153G; D197G; D233G	+++
43	44	V43A; A94F; S96D; F147L; L153G; D197G; D233G	+++
45	46	V43A; A94F; S96G; F147L; L153G; D197G; L199Q; D233G	+++
47	48	V43A; A94F;S96G; F147L; L153G; D197G; L199R; D233G	+++
49	50	V43A; A94F; S96G; F147L; L153G; D197G; D233G	++
51	52	V43A; A94F; S96G; K97G; F147L; L153G; D197G; M205T; D233G	+++
53	54	V43A; A94F; S96Y; L153G; D197G; L199N; A202I; I223V	++
55	56	V43A; T54A; A94G; S96G; F147L; I223V	+++
57	58	F74L; A94G; S96G; F147L; L153A; A202I; M206F	++

Nucleic Acid SEQ ID NO.	Polypeptide SEQ ID NO.	Mutations from ADH-LK (SEQ ID NO:2)	Activity Ranking (see legend)
59	60	F74L; A94G; S96G; F147L; L153A; D197G; L199K	++
61	62	F74L; A94G; F147L; L199K; A202I	+++
63	64	F74L; A94G; F147L; D197G; I223V	++
65	66	A94G; F147L; L199I; A202I	++
67	68	A94G; F147L; L199I; A202L; M206F; K211R	++
69	70	A94F; V95I; S96Y; F147L; L153G	+++
7`	72	A94F; V95L; S96Y; F147L; L153G	+++
73	74	A94F; S96G; F147L; L153A; L199V; D233G	+++
75	76	A94F; S96Y; L124Q; F147L; L153G	+++
77	78	A94F; S96Y; F147L; L153G	++
79	80	A94F; S96Y; K109R; F147L; L153A; D233G	+++
81	82	A94F; S96F; K109R; F147L; L153A; D233G	+++
83	84	T77A; A94F; S96Y; F147L; L153G; L199W	+++
85	86	T54A; A94F; S96G; F147L; L153A; D233G	+++
87	88	T54A; A94F; S96G; F147L; L153A; L199V; D233G	+++
89	90	T54A; A94F; S96Y; F147L; L153A; D233G	+++
91	92	T54A; A94F; S96Y; F147L; L153G; D233G	+++
93	94	T54A; A94F; S96Y; K109R; F147L; L153A; L199I; D233G;	+++
95	96	T54A; A94F; S96Y; K109R; F147L; L153A; L199R	+++
97	98	T54A; A94F; S96Y; K109R; F147L; L153A; D233G	+++

Nucleic Acid SEQ ID NO.	Polypeptide SEQ ID NO.	Mutations from ADH-LK (SEQ ID NO:2)	Activity Ranking (see legend)
99	100	T54A; A94F; K109R; F147L; L153A; D233G	+++
101	102	T54A; A94F; K109R; F147L; L153A; L199K; N221D; D233G	+++
103	104	T54A; A94F; S96F; K109R; F147L; L153A; D233G	+++
105	106	K49R; A94G; F147L; D197G; A202G	++
107	108	V43A; F74L; A94F; S96Y; L153G	++
109	110	T54A; A94F; S96G; K109R; F147L; L153A; L199K; D233G	+++
111	112	I11V; A94F	+
113	114	A94G	+

[1] Activity Ranking: + Up to 20-fold improved in activity compared to ADH-LK. ++ From 21- to 80-fold improved in activity compared to ADH-LK. +++ More than 80-fold improved in activity compared to ADH-LK.

[0129] In some embodiments, the engineered ketoreductase enzymes are selected from the amino acid sequences recited in Table 2. Table 2 ranks the enzyme activity measured for the reduction or conversion of acetophenone to (*R*)-1-phenylethanol using the listed engineered ketoreductase polypeptides derived from ADH-LK compared to the activity measured using ADH-LK (SEQ ID NO:2).

[0130] Table 2

Nucleic Acid SEQ ID NO.	Polypeptide SEQ ID NO.	Mutations from ADH-LK (SEQ ID NO:2) or ADH-LB (SEQ ID NO:4)	Activity Ranking[1] (see legend)
115	116	ADH-LB: A96G	
117	118	ADH-LB: N96F	
119	120	ADH-LK: S96F	+++
121	122	ADH-LK: S96G; F147L; L199N	+++

123	124	ADH-LK: S96F; F147L	+++
125	126	ADH-LK: S96Y; F147L	+++
127	128	ADH-LK: S96I	+++
129	130	ADH-LK: A94H	++
131	132	ADH-LK: S96C	++
133	134	ADH-LK: S96W	++
135	136	ADH-LK: S96I; F147L	+++
137	138	ADH-LK: A94S; F147L	+++

[1]Activity Ranking: Acetophenone conversion to (*R*) 1-phenylethanol in the method of Example 16: + <70% conversion; ++ 70-90% conversion; +++ >90% conversion.

[0131] In some embodiments, the ketoreductase polypeptide comprises an amino acid sequence that is at least about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO:2 or SEQ ID NO:4, wherein the amino acid sequence comprises any one of the set of mutations contained in any one of the polypeptide sequences listed in Table 1 or Table 2. In some embodiments, the ketoreductase polypeptide comprises any one of the set of the amino acid residues at the specified residue positions of Table I or Table 2 and which differs from the corresponding sequence from about 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-14, 1-16, 1-18, 1-20, or 1-25 amino acid positions. In some embodiments, the ketoreductase polypeptide comprises any one of the set of mutations listed in Table 1 or Table 2, and additionally from about 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-14, 1-16, 1-18, 1-20, or 1-25 conservative substitutions at other residues. Thus, in some embodiments, the ketoreductase polypeptide can differ in 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 16, 18, 20, or 25 amino acid positions as compared to the amino acid sequences in Table 1 or Table 2.

[0132] In some embodiments, the ketoreductase polypeptide can comprise a region having an amino acid sequence that is at least about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a region or domain thereof, such as residues 90-233, of SEQ ID NO: 6, 8, 10, 12, 14, 16, 18, 20, 22, 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, or 118. In some embodiments, the ketoreductase polypeptides can comprise a region having an amino acid sequence that is at least about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a region or domain thereof, such as residues 90-233 of SEQ ID NO: 120, 122, 124, 126, 128, 130, 132, 134, 136, or 138.

[0133] In some embodiments, the ketoreductase polypeptide is selected from SEQ ID NOS: 6, 8, 10, 12, 14, 16, 18, 20, 22, 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, and 118. In some embodiments, the ketoreductase polypeptide is selected from SEQ ID NOS: 120, 122, 124, 126, 128, 130, 132, 134, 136, and 138.

[0134] In some embodiments, where the ketoreductase polypeptide comprises an amino acid sequence with an aromatic acid or G at the residue corresponding to residue 94 of SEQ ID NO:2 or SEQ ID NO:4, the ketoreductase polypeptide can be selected from the amino acid sequences of SEQ ID NOS: 6, 16, 20, 22, 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, and 128.

[0135] In some embodiments, where the ketoreductase polypeptide comprises an amino acid sequence with an amino acid other than S/N at the residue corresponding to residue 96 of SEQ ID NO:2 or SEQ ID NO:4, the ketoreductase polypeptide can be selected from the amino acid sequences of SEQ ID NOS: 6, 8, 10, 12, 14, 16, 18, 20, 24, 26, 28, 30, 36, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 104, 108, 110, 118, 120, 122, 124, 126, 128, 132, 134, and 136.

[0136] In some embodiments, segments of the ketoreductase polypeptides can be deleted to generate polypeptide fragments. The term "fragment" as used herein refers to a polypeptide that has an amino-terminal and/or carboxy-terminal deletion, but where the remaining amino acid sequence is identical to the corresponding positions in the sequence. Fragments can be at least 14 amino acids long, at least 20 amino acids long, at least 50 amino acids long or longer. In some embodiments, the fragments are up to 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% of the full-length recombinant ketoreductase polypeptide above, including fragments of the ketoreductase polypeptides of SEQ ID NO: 6, 8, 10, 12, 14, 16, 18, 20, 22, 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, and 118. In some embodiments, the fragments are up to 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% of the full-length recombinant ketoreductase polypeptide above, including fragments of the ketoreductase polypeptides of SEQ ID NO: 120, 122, 124, 126, 128, 130, 132, 134, 136, and 138.

[0137] The improved ketoreductase enzymes may be present within a cell, present in the cellular medium, or prepared in various forms, such as lysates or isolated preparations. As such, in some embodiments, the improved ketoreductase enzyme can be an isolated polypeptide. The term "isolated polypeptide" refers to a polypeptide which is substantially separated from other contaminants that naturally accompany it, *e.g.*, protein, lipids, and polynucleotides. The term embraces polypeptides which have been removed or purified from their naturally-occurring environment or expression system (*e.g.*, host cell or *in vitro* synthesis).

[0138] In some embodiments, the isolated improved ketoreductases polypeptide is a substantially pure polypeptide composition. The term "substantially pure polypeptide" refers to a composition in which the polypeptide species is the predominant species present (*i.e.*, on a molar or weight basis it is more abundant than any other individual macromolecular species in the composition), and is generally a substantially purified composition when the object species comprises at least about 50 percent of the macromolecular species present by mole or % weight. Generally, a substantially pure ketoreductase composition will comprise about 60 % or more, about 70% or more, about 80% or more, about 90% or more, about 95% or more, and about 98% or more of all macromolecular species by mole or % weight present in the composition. In some embodiments, the object species is purified to essential homogeneity (*i.e.*, contaminant species cannot be detected in the composition by conventional detection methods) wherein the composition consists essentially of a single macromolecular species. Solvent species, small molecules (<500 Daltons), and elemental ion species are not considered macromolecular species.

6.3 Polynucleotides Encoding Engineered Ketoreductases

[0139] In another aspect, the present disclosure provides polynucleotides encoding the engineered ketoreductase polypeptides. The polynucleotides may be operatively linked to one or more heterologous regulatory or control sequences that control gene expression to create a recombinant polynucleotide capable of expressing the polypeptide. Expression constructs containing a heterologous polynucleotide encoding the engineered ketoreductase can be introduced into appropriate host cells to express the corresponding ketoreductase.

[0140] Because of the knowledge of the codons corresponding to the various amino acids, availability of a polypeptide sequence provides a description of all the polynucleotides capable of encoding the subject polypeptide. The degeneracy of the genetic code, where the same amino acids are encoded by alternative or synonymous codons allows an extremely large number of nucleic acids to be made, all of which encode the improved ketoreductase enzymes disclosed herein. Thus, having identified a particular amino acid sequence, those skilled in the art could make any number of different nucleic acids by simply modifying the sequence of one or more codons in a way which does not change the amino acid sequence of the protein. In this regard, the present disclosure specifically contemplates each and every possible variation of polynucleotides that could be made by selecting combinations based on the possible codon choices, and all such variations are to be considered specifically disclosed for any polypeptide disclosed herein, including the amino acid sequences presented in Table 1 and Table 2. As such, the polynucleotides of the present disclosure include any and all possible polynucleotide sequences that encode the ketoreductase polypeptide of SEQ ID NO: 6, 8, 10, 12, 14, 16, 18, 20, 22, 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, or 138.

[0141] In some embodiments, the polynucleotides encoding the ketoreductases enzymes may be codon optimized for optimal production from the host organism selected for expression. For example, preferred codons used in bacteria are used to express the gene in bacteria; preferred codons used in yeast are used for expression in yeast; and preferred codons used in mammals are used for expression in mammalian cells. By way of example, the polynucleotide of SEQ ID NO: 1 has been codon optimized for expression in *E. coli*, but otherwise encodes the naturally occurring ketoreductase of *Lactobacillus kefir*.

[0142] In some embodiments, all codons need not be replaced to optimize the codon usage of the ketoreductases since the natural sequence will comprise preferred codons and because use of preferred codons may not be required for all amino acid residues. Consequently, codon optimized polynucleotides encoding the ketoreductase enzymes may contain preferred codons at about 40%, 50%, 60%, 70%, 80%, or greater than 90% of codon positions of the full length coding region.

[0143] In some embodiments, the polynucleotides encoding the engineered ketoreductases are selected from SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 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, and 117. In some embodiments, the polynucleotides encoding the engineered ketoreductases are selected from SEQ ID NOS: 119, 121, 123, 125, 127, 129, 131, 133, 135, and 137.

[0144] These polynucleotides encode the corresponding polypeptides represented by the amino acid sequences listed in Table 1 and Table 2, which were derived by subjecting the *E. coli* codon optimized *Lactobacillus kefir* gene to directed gene evolution techniques described herein.

[0145] In some embodiments, the polynucleotides comprise polynucleotides that encode the polypeptides described herein but have about 80% or more sequence identity, about 85% or more sequence identity, about 90% or more sequence identity, about 91% or more sequence identity, about 92% or more sequence identity, about 93% or more sequence identity, about 94% or more sequence identity, about 95% or more sequence identity, about 96% or more sequence identity, about 97% or more sequence identity, about 98% or more sequence identity, or about 99% or more sequence identity at the nucleotide level to a reference polynucleotide encoding an engineered ketoreductase. In some embodiments, the reference polynucleotide is selected from polynucleotide sequences of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 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, and 117. In some embodiments, the reference polynucleotide is selected from sequences of SEQ ID NOS: 119, 121, 123, 125, 127, 129, 131, 133, 135, and 137.

[0146] In some embodiments, the polynucleotide encodes the polypeptide disclosed herein and hybridizes under defined conditions, such as moderately stringent or highly stringent conditions, to the complement of a sequence encoding an engineered ketoreductase enzyme of the present disclosure. As such, in some embodiments, the polynucleotides encoding the ketoreductase polypeptides comprises a polynucleotide that hybridizes under highly stringent conditions to a polynucleotide selected from SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 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, and 117. In some embodiments, the polynucleotides encoding the ketoreductase polypeptides comprises a polynucleotide that hybridizes under highly stringent conditions to a polynucleotide selected from SEQ ID NOS: 119, 121, 123, 125, 127, 129, 131, 133, 135, and 137.

[0147] An isolated polynucleotide encoding an improved ketoreductase polypeptide may be manipulated in a variety of ways to provide for expression of the polypeptide. Manipulation of the isolated polynucleotide prior to its insertion into a vector may be desirable or necessary depending on the expression vector. The techniques for modifying polynucleotides and nucleic acid sequences utilizing recombinant DNA methods are well known in the art. Guidance is provided in Sambrook et al., 2001, *Molecular Cloning: A Laboratory Manual*, 3rd Ed., Cold Spring Harbor Laboratory Press; and *Current Protocols in Molecular Biology*, Ausubel. F. ed., Greene Pub. Associates, 1998, updates to 2006.

[0148] In some embodiments, the control sequence may be an appropriate promoter sequence, which can be obtained from genes encoding extracellular or intracellular polypeptides either homologous or heterologous to the host cell. For bacterial host cells, suitable promoters for directing transcription of the nucleic acid constructs of the present disclosure, include the promoters obtained from the *E. coli* lac operon, *Streptomyces coelicolor* agarase gene (dagA), *Bacillus subtilis* levansucrase gene (sacB), *Bacillus licheniformis* alpha-amylase gene (amyL), *Bacillus stearothermophilus* maltogenic amylase gene (amyM), *Bacillus amyloliquefaciens* alpha-amylase gene (amyQ), *Bacillus licheniformis* penicillinase gene (penP), *Bacillus subtilis* xylA and xylB genes, and prokaryotic beta-lactamase gene (Villa-Kamaroff et al., 1978, *Proc. Natl Acad. Sci. USA* 75: 3727-3731), as well as the *tac* promoter (DeBoer et al., 1983, *Proc. Natl Acad. Sci. USA* 80: 21-25). Further promoters are described in "Useful proteins from recombinant bacteria" in *Scientific American*, 1980, 242:74-94; and in Sambrook et al., *supra*.

[0149] For filamentous fungal host cells, suitable promoters for directing the transcription of the nucleic acid constructs of the present disclosure include promoters obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *Aspergillus niger* neutral alpha-amylase, *Aspergillus niger* acid stable alpha-amylase, *Aspergillus niger* or *Aspergillus awamori* glucoamylase (glaA), *Rhizomucor miehei* lipase, *Aspergillus oryzae* alkaline protease, *Aspergillus oryzae* triose phosphate isomerase, *Aspergillus nidulans* acetamidase, and *Fusarium oxysporum* trypsin-like protease (WO 96/00787), as well as the NA2-tpi promoter (a hybrid of the promoters from the genes for *Aspergillus niger* neutral alpha-amylase and *Aspergillus oryzae* triose phosphate isomerase), and mutant, truncated, and hybrid promoters thereof.

[0150] In a yeast host, useful promoters can be from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* galactokinase (GAL1), *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP), and *Saccharomyces cerevisiae* 3-phosphoglycerate kinase. Other useful promoters for yeast host cells are described by Romanos et al., 1992, *Yeast* 8:423-488.

[0151] In some embodiments, the control sequence may also be a suitable transcription terminator sequence, a sequence recognized by a host cell to terminate transcription. The terminator sequence is operably linked to the 3' terminus of the nucleic acid sequence encoding the polypeptide. Any terminator which is functional in the host cell of choice may be used in the present invention.

[0152] For example, exemplary transcription terminators for filamentous fungal host cells can be obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Aspergillus niger* alpha-glucosidase, and *Fusarium oxysporum* trypsin-like protease.

[0153] Exemplary terminators for yeast host cells can be obtained from the genes for *Saccharomyces cerevisiae* enolase, *Saccharomyces cerevisiae* cytochrome C (CYC1), and *Saccharomyces cerevisiae* glyceraldehyde-3-phosphate dehydrogenase. Other useful terminators for yeast host cells are described by Romanos et al., 1992, *supra*.

[0154] In some embodiments, the control sequence may also be a suitable leader sequence, a nontranslated region of an mRNA that is important for translation by the host cell. The leader sequence is operably linked to the 5' terminus of the nucleic acid sequence encoding the polypeptide. Any leader sequence that is functional in the host cell of choice may be used. Exemplary leaders for filamentous fungal host cells are obtained from the genes for *Aspergillus oryzae* TAKA amylase and *Aspergillus nidulans* triose phosphate isomerase. Suitable leaders for yeast host cells are obtained from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* 3-phosphoglycerate kinase, *Saccharomyces cerevisiae* alpha-factor, and *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP).

[0155] In some embodiments, the control sequence may also be a polyadenylation sequence, a sequence operably linked to the 3' terminus of the nucleic acid sequence and which, when transcribed, is recognized by the host cell as a signal to add polyadenosine residues to transcribed mRNA. Any polyadenylation sequence which is functional in the host cell of choice may be used in the present invention. Exemplary polyadenylation sequences for filamentous fungal host cells can be from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Fusarium oxysporum* trypsin-like protease, and *Aspergillus niger* alpha-glucosidase. Useful polyadenylation sequences for yeast host cells are described by Guo and Sherman, 1995, *Mol Cell Bio* 15:5983-5990.

[0156] In some embodiments, the control sequence may also be a signal peptide coding region that codes for an amino acid sequence linked to the amino terminus of a polypeptide and directs the encoded polypeptide into the cell's secretory pathway. The 5' end of the coding sequence of the nucleic acid sequence may inherently contain a signal peptide coding region naturally linked in translation reading frame with the segment of the coding region that encodes the secreted polypeptide. Alternatively, the 5' end of the coding sequence may contain a signal peptide coding region that is foreign to the coding sequence. The foreign signal peptide coding region may be required where the coding sequence does not naturally contain a signal peptide coding region.

[0157] Alternatively, the foreign signal peptide coding region may simply replace the natural signal peptide coding region in order to enhance secretion of the polypeptide. However, any signal peptide coding region which directs the expressed polypeptide into the secretory pathway of a host cell of choice may be used in the present invention.

[0158] Effective signal peptide coding regions for bacterial host cells are the signal peptide coding regions obtained from the genes for *Bacillus* NCIB 11837 maltogenic amylase, *Bacillus stearothermophilus* alpha-amylase, *Bacillus licheniformis* subtilisin, *Bacillus licheniformis* beta-lactamase, *Bacillus stearothermophilus* neutral proteases (nprT, nprS, nprM), and *Bacillus subtilis* prsA. Further signal peptides are described by Simonen and Palva, 1993, *Microbiol Rev* 57: 109-137.

[0159] Effective signal peptide coding regions for filamentous fungal host cells can be the signal peptide coding regions obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* neutral amylase, *Aspergillus niger* glucoamylase, *Rhizomucor miehei* aspartic proteinase, *Humicola insolens* cellulase, and *Humicola lanuginosa* lipase.

[0160] Useful signal peptides for yeast host cells can be from the genes for *Saccharomyces cerevisiae* alpha-factor and *Saccharomyces cerevisiae* invertase. Other useful signal peptide coding regions are described by Romanos et al., 1992, *supra.*

[0161] In some embodiments, the control sequence may also be a propeptide coding region that codes for an amino acid sequence positioned at the amino terminus of a polypeptide. The resultant polypeptide is known as a proenzyme or propolypeptide (or a zymogen in some cases). A propolypeptide is generally inactive and can be converted to a mature active polypeptide by catalytic or autocatalytic cleavage of the propeptide from the propolypeptide. The propeptide coding region may be obtained from the genes for *Bacillus subtilis* alkaline protease (aprE), *Bacillus subtilis* neutral protease (nprT), *Saccharomyces cerevisiae* alpha-factor, *Rhizomucor miehei* aspartic proteinase, and *Myceliophthora thermophila* lactase (WO 95/33836).

[0162] Where both signal peptide and propeptide regions are present at the amino terminus of a polypeptide, the propeptide region is positioned next to the amino terminus of a polypeptide and the signal peptide region is positioned next to the amino terminus of the propeptide region.

[0163] It may also be desirable to add regulatory sequences, which allow the regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory systems are those which cause the expression of the gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. In prokaryotic host cells, suitable regulatory sequences include the lac, tac, and trp operator systems. In yeast host cells, suitable regulatory systems include, as examples, the ADH2 system or GAL1 system. In filamentous fungi, suitable regulatory sequences include the TAKA alpha-amylase promoter, *Aspergillus niger* glucoamylase promoter, and *Aspergillus oryzae* glucoamylase promoter.

[0164] Other examples of regulatory or control sequences are those which allow for gene amplification. In eukaryotic systems, these include the dihydrofolate reductase gene, which is amplified in the presence of methotrexate, and the metallothionein genes, which are amplified with heavy metals. In these cases, the nucleic acid sequence encoding the KRED polypeptide of the present invention would be operably linked with the control sequence.

[0165] Thus, in another aspect, the present disclosure is also directed to a recombinant expression vector comprising a polynucleotide encoding an engineered ketoreductase polypeptide or a variant thereof, and one or more expression regulating regions such as a promoter and a terminator, a replication origin, etc., depending on the type of hosts into which they are to be introduced. The various nucleic acid and control sequences described above may be joined together to produce a recombinant expression vector which may include one or more convenient restriction sites to allow for insertion or substitution of the nucleic acid sequence encoding the polypeptide at such sites. Alternatively, the nucleic acid sequence of the present disclosure may be expressed by inserting the nucleic acid sequence or a nucleic acid construct comprising the sequence into an appropriate vector for expression. In creating the expression vector, the coding sequence is located in the vector so that the coding sequence is operably linked with the appropriate control sequences for expression.

[0166] The recombinant expression vector may be any vector (*e.g.*, a plasmid or virus), which can be conveniently subjected to recombinant DNA procedures and can bring about the expression of the polynucleotide sequence. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which the vector is to be introduced. The vectors may be linear or closed circular plasmids.

[0167] The expression vector may be an autonomously replicating vector, *i.e.*, a vector that exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, *e.g.*, a plasmid, an extrachromosomal element, a minichromosome, or an artificial chromosome. The vector may contain any means for assuring self-replication. Alternatively, the vector may be one which, when introduced into the host cell, is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. Furthermore, a single vector or plasmid or two or more vectors or plasmids which together contain the total DNA to be introduced into the genome of the host cell, or a transposon may be used.

[0168] The expression vector of the present invention preferably contains one or more selectable markers, which permit easy selection of transformed cells. A selectable marker is a gene the product of which provides for biocide or viral resistance, resistance to heavy metals, prototrophy to auxotrophs, and the like. Examples of bacterial selectable markers are the *dal* genes from *Bacillus subtilis* or *Bacillus licheniformis*, or markers, which confer antibiotic resistance such as ampicillin, kanamycin, chloramphenicol (Example 1) or tetracycline resistance. Suitable markers for yeast host cells are ADE2, HIS3, LEU2, LYS2, MET3, TRP1, and URA3.

[0169] Selectable markers for use in a filamentous fungal host cell include, but are not limited to, amdS (acetamidase), argB (ornithine carbamoyltransferase), bar (phosphinothricin acetyltransferase), hph (hygromycin phosphotransferase), niaD (nitrate reductase), pyrG (orotidine-5'-phosphate decarboxylase), sC (sulfate adenyltransferase), and trpC (anthranilate synthase), as well as equivalents thereof. Embodiments for use in an Aspergillus cell include the amdS and pyrG genes of *Aspergillus nidulans* or *Aspergillus oryzae* and the bar gene of *Streptomyces hygroscopicus*.

[0170] The expression vectors of the present invention preferably contain an element(s) that permits integration of the vector into the host cell's genome or autonomous replication of the vector in the cell independent of the genome. For integration into the host cell genome, the vector may rely on the nucleic acid sequence encoding the polypeptide or any other element of the vector for integration of the vector into the genome by homologous or nonhomologous recombination.

[0171] Alternatively, the expression vector may contain additional nucleic acid sequences for directing integration by homologous recombination into the genome of the host cell. The additional nucleic acid sequences enable the vector to be integrated into the host cell genome at a precise location(s) in the chromosome(s). To increase the likelihood of integration at a precise location, the integrational elements should preferably contain a sufficient number of nucleic acids, such as 100 to 10,000 base pairs, preferably 400 to 10,000 base pairs, and most preferably 800 to 10,000 base pairs, which are highly homologous with the corresponding target sequence to enhance the probability of homologous recombination. The integrational elements may be any sequence that is homologous with the target sequence in the genome of the host cell. Furthermore, the integrational elements may be non-encoding or encoding nucleic acid sequences. On the other hand, the vector may be integrated into the genome of the host cell by non-homologous recombination.

[0172] For autonomous replication, the vector may further comprise an origin of replication enabling the vector to replicate autonomously in the host cell in question. Examples of bacterial origins of replication are P15A ori (as shown in the plasmid of FIG. 5) or the origins of replication of plasmids pBR322, pUC19, pACYCl77 (which plasmid has the P15A ori), or pACYC184 permitting replication in *E. coli*, and pUB110, pE194, pTA1060, or pAM.beta.1 permitting replication in Bacillus. Examples of origins of replication for use in a yeast host cell are the 2 micron origin of replication, ARS1, ARS4, the combination of ARS1 and CEN3, and the combination of ARS4 and CEN6. The origin of replication may be one having a mutation which makes it's functioning temperature-sensitive in the host cell (see, *e.g.*, Ehrlich, 1978, *Proc Natl Acad Sci. USA* 75:1433).

[0173] More than one copy of a nucleic acid sequence of the present invention may be inserted into the host cell to increase production of the gene product. An increase in the copy number of the nucleic acid sequence can be obtained by integrating at least one additional copy of the sequence into the host cell genome or by including an amplifiable selectable marker gene with the nucleic acid sequence where cells containing amplified copies of the selectable marker gene, and thereby additional copies of the nucleic acid sequence, can be selected for by cultivating the cells in the presence of the appropriate selectable agent.

[0174] Many of the expression vectors for use in the present invention are commercially available. Suitable commercial expression vectors include p3xFLAGTM$^{\text{TM}}$ expression vectors from Sigma-Aldrich Chemicals, St. Louis MO., which includes a CMV promoter and hGH polyadenylation site for expression in mammalian host cells and a pBR322 origin of replication and ampicillin resistance markers for amplification in *E. coli.* Other suitable expression vectors are pBluescriptII SK(-)® and pBK-CMV, which are commercially available from Stratagene, LaJolla CA, and plasmids which are derived from pBR322 (Gibco BRL), pUC (Gibco BRL), pREP4, pCEP4 (Invitrogen) or pPoly (Lathe et al., 1987, *Gene* 57:193-201).

6.4 Host Cells for Expression of Ketoreductase Polypeptides

[0175] In another aspect, the present disclosure provides a host cell comprising a polynucleotide encoding an improved ketoreductase polypeptide of the present disclosure, the polynucleotide being operatively linked to one or more control sequences for expression of the ketoreductase enzyme in the host cell. Host cells for use in expressing the ketoreductase polypeptides encoded by the expression vectors of the present invention are well known in the art and include but are not limited to, bacterial cells, such as *E. coli*, *Lactobacillus kefir*, *Lactobacillus brevis*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells (*e.g.*, *Saccharomyces cerevisiae* or *Pichia pastoris* (ATCC Accession No. 201178)); insect cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as CHO, COS, BHK, 293, and Bowes melanoma cells; and plant cells. Appropriate culture mediums and growth conditions for the above-described host cells are well known in the art.

[0176] Polynucleotides for expression of the ketoreductase may be introduced into cells by various methods known in the art. Techniques include among others, electroporation, biolistic particle bombardment, liposome mediated transfection, calcium chloride transfection, and protoplast fusion. Various methods for introducing polynucleotides into cells will be apparent to the skilled artisan.

[0177] An exemplary host cell is *Escherichia coli* W3110. The expression vector was created by operatively linking a polynucleotide encoding an improved ketoreductase into the plasmid pCK110900 operatively linked to the *lac* promoter under control of the *lac*I repressor. The expression vector also contained the P15a origin of replication and the chloramphenicol resistance gene. Cells containing the subject polynucleotide in *Escherichia coli* W3110 were isolated by subjecting the cells to chloramphenicol selection.

6.5 Methods of Generating Engineered Ketoreductase Polypeptides.

[0178] To make the improved ketoreductase polynucleotides and polypeptides of the present disclosure, the naturally-occurring ketoreductase enzyme that catalyzes the reduction reaction is obtained from *Lactobacillus kefir* or *Lactobacillus brevis*. In some embodiments, the parent polynucleotide sequence is codon optimized to enhance expression of the ketoreductase in a specified host cell. As an illustration, the parental polynucleotide sequence encoding the wild-type ketoreductase polypeptide of *Lactobacillus kefir* was constructed from oligonucleotides prepared based upon the known polypeptide sequence of *Lactobacillus kefir* KRED sequence available in Genbank database (Genbank accession no. AAP94029 GI:33112056). The parental polynucleotide sequence, designated as SEQ ID NO: 1, was codon optimized for expression in *E. coli* and the codon-optimized polynucleotide cloned into an expression vector, placing the expression of the ketoreductase gene under the control of the lac promoter and lacI repressor gene. Clones expressing the active ketoreductase in *E. coli* were identified and the genes sequenced to confirm their identity. The sequence designated (SEQ ID NO: 1) was the parent sequence utilized as the starting point for all experiments and library construction of engineered ketoreductases evolved from the *Lactobacillus kefir* ketoreductase.

[0179] The engineered ketoreductases can be obtained by subjecting the polynucleotide encoding the naturally occurring ketoreductase to mutagenesis and/or directed evolution methods, as discussed above. An exemplary directed evolution technique is mutagenesis and/or DNA shuffling as described in Stemmer, 1994, *Proc Natl Acad Sci USA* 91:10747-10751; WO 95/22625; WO 97/0078; WO 97/35966; WO 98/27230; WO 00/42651; WO 01/75767 and US Pat. 6,537,746. Other directed evolution procedures that can be used include, among others, staggered extension process (StEP), in vitro recombination (Zhao et al., 1998, *Nat. Biotechnol.* 16:258–261), mutagenic PCR (Caldwell et al., 1994, *PCR Methods Appl.* 3:S136-S140), and cassette mutagenesis (Black et al., 1996, *Proc Natl Acad Sci USA* 93:3525-3529).

[0180] The clones obtained following mutagenesis treatment are screened for engineered ketoreductases having a desired improved enzyme property. Measuring enzyme activity from the expression libraries can be performed using the standard biochemistry technique of monitoring the rate of decrease (via a decrease in absorbance or fluorescence) of NADH or

NADPH concentration, as it is converted into NAD$^+$ or NADP$^+$. In this reaction, the NADH or NADPH is consumed (oxidized) by the ketoreductase as the ketoreductase reduces a ketone substrate to the corresponding hydroxyl group. The rate of decrease of NADH or NADPH concentration, as measured by the decrease in absorbance or fluorescence, per unit time indicates the relative (enzymatic) activity of the ketoreductase polypeptide in a fixed amount of the lysate (or a lyophilized powder made therefrom). Where the improved enzyme property desired is thermal stability, enzyme activity may be measured after subjecting the enzyme preparations to a defined temperature and measuring the amount of enzyme activity remaining after heat treatments. Clones containing a polynucleotide encoding a ketoreductase are then isolated, sequenced to identify the nucleotide sequence changes (if any), and used to express the enzyme in a host cell.

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

[0182] Engineered ketoreductase enzymes expressed in a host cell can be recovered from the cells and or the culture medium using any one or more of the well known techniques for protein purification, including, among others, lysozyme treatment, sonication, filtration, salting-out, ultra-centrifugation, and chromatography. Suitable solutions for lysing and the high efficiency extraction of proteins from bacteria, such as *E. coli*, are commercially available under the trade name CelLytic B^{TM} from Sigma-Aldrich of St. Louis MO.

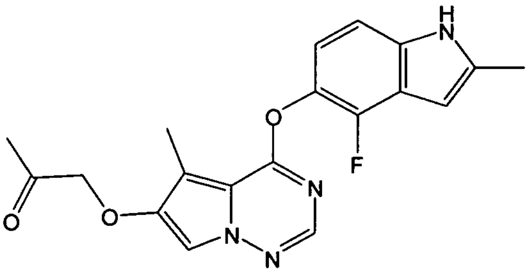
[0183] Chromatographic techniques for isolation of the ketoreductase polypeptide include, among others, reverse phase chromatography high performance liquid chromatography, ion exchange chromatography, gel electrophoresis, and affinity chromatography. Conditions for purifying a particular enzyme will depend, in part, on factors such as net charge, hydrophobicity, hydrophilicity, molecular weight, molecular shape, etc., and will be apparent to those having skill in the art.

[0184] In some embodiments, affinity techniques may be used to isolate the improved ketoreductase enzymes. For affinity chromatography purification, any antibody which specifically binds the ketoreductase polypeptide may be used. For the production of antibodies, various host animals, including but not limited to rabbits, mice, rats, etc., may be immunized by injection with a compound. The compound may be attached to a suitable carrier, such as BSA, by means of a side chain functional group or linkers attached to a side chain functional group. Various adjuvants may be used to increase the immunological response, depending on the host species, including but not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, and potentially useful human adjuvants such as BCG (bacilli Calmette Guerin) and Corynebacterium parvum.

6.6 Methods of Using the Engineered Ketoreductase Enzymes and Compounds Prepared Therewith

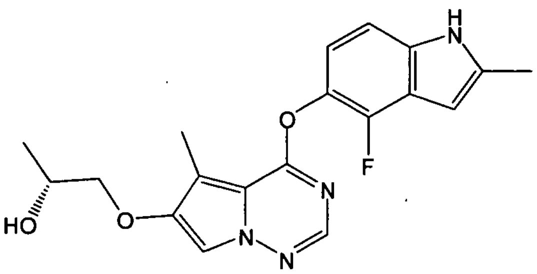
[0185] The ketoreductase enzymes described herein can catalyze the reduction of the substrate 1-[4-(4-fluoro-2-methyl-1H-indol-5-yloxy)-5-methyl-pyrrolo[2,1-f][1,2,4,]triazine-6-yloxy]-propan-2-one, as represented by the structure of formula (I):

(I)



to the chiral alcohol product (*R*)-1-[4-(4-fluoro-2-methyl-1H-indol-5-yloxy)-5-methyl-pyrrolo[2,1-f][1,2,4,]triazine-6-yloxy]-propan-2-ol, as represented by structure of formula (II)

(II)

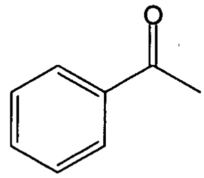


[0186] The above compounds useful in the synthesis of protein kinase inhibitors, and are encompassed within a class of compound described in US application No. US 2006/0004007, US 2006/0128709; US 2006/0257400; US20060264438; and US20060286646 (all publications incorporated herein by reference). Thus, other compounds of the class with the requisite keto group can be used as substrates for the ketoreductases disclosed herein.

[0187] In various embodiments, the methods can comprise contacting or mixing the compound of formula (I) with a ketoreductase of the present disclosure under reaction conditions suitable for conversion of the substrate to the compound of structural formula (II). Exemplary reactions conditions are described in the Examples. Other suitable reactions conditions will be apparent to the skilled artisan. Exemplary ketoreductase polypeptides that can be used in the method include, but are not limited to, SEQ ID NO: 6, 8, 10, 12, 14, 16, 18, 20, 22, 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, and 118.

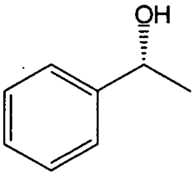
[0188] In some embodiments, the ketoreductase polypeptides described herein are capable of converting the substrate acetophenone, as represented by structural formula (III):

(III)



to the corresponding chiral alcohol product, (*R*)-1-phenylethanol, as represented by structural formula (IV):

(IV)



[0189] In various embodiments, the method can comprise contacting or mixing the compound of formula (III) with a ketoreductase of the present disclosure under reaction conditions suitable for conversion of the substrate to the compound of structural formula (IV). Exemplary reactions conditions are described in the Examples (*see*, *e.g.*, Example 15). Other suitable reactions conditions will be apparent to the skilled artisan. Exemplary ketoreductase polypeptides that can be used in the method include, but are not limited to, SEQ ID NO: 6, 8, 10, 12, 14, 16, 18, 20, 22, 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, or 138. In some embodiments, the ketoreductase polypeptides that can be used for reduction or conversion of acetophenone to (*R*)1-phenylethanol is selected from the polypeptides of SEQ ID NO: 120, 122, 124, 126, 128, 130, 132, 134, 136, or 138.

[0190] In some embodiments, the resultant product is enriched in a particular stereoisomer, i.e., compound (II) or compound (IV). As used herein, a compound is "enriched" in a particular stereoisomer when that stereoisomer is present in excess over any other stereoisomer present in the compound. A compound that is enriched in a particular stereoisomer will typically comprise at least about 60% or more, 70% or more, 80% or more, 90% or more, 95% or more, 96% or more, 97% or more, 98% or more, 99% or more, or 99.5% or more, of the specified stereoisomer. The amount of enrichment of a particular stereoisomer can be confirmed using conventional analytical methods routinely used by those of skill in the art.

[0191] In some embodiments, the amount of undesired stereoisomers may be less than 10%, for example, less than 9%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, less than 1% or even less than 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.2%, or 0.1%. Stereoisomerically enriched compounds that contain at least about 99.5% or more of the desired stereoisomer are referred to herein as "substantially pure" stereoisomers. In some embodiments, compounds that are substantially pure in a specified stereoisomer contain greater than 99.0%, 99.2%, 99.4%, 99.6%, 99.8 %, or even greater purity. Stereoisomerically enriched compounds that contain $\geq$ 99.9% of the desired stereoisomer are referred to herein as "pure" stereoisomers.

[0192] As is known by those of skill in the art, ketoreductase-catalyzed reduction reactions typically require a cofactor. Reduction reactions catalyzed by the engineered ketoreductase enzymes described herein also typically require a cofactor, although many embodiments of the engineered ketoreductases require far less cofactor than reactions catalyzed with wild-type ketoreductase enzymes. As used herein, the term "cofactor" refers to a non-protein compound that operates in combination with a ketoreductase enzyme. Cofactors suitable for use with the engineered ketoreductase enzymes described herein include, but are not limited to, $NADP^+$ (nicotinamide adenine dinucleotide phosphate), NADPH (the reduced form of $NADP^+$), NAD^+ (nicotinamide adenine dinucleotide) and NADH (the reduced form of NAD^+). Generally, the reduced form of the cofactor is added to the reaction mixture. The reduced NAD(P)H form can be optionally regenerated from the oxidized $NAD(P)^+$ form using a cofactor regeneration system.

[0193] Suitable exemplary cofactor regeneration systems that may be employed include, but are not limited to, glucose and glucose dehydrogenase, formate and formate dehydrogenase, glucose-6-phosphate and glucose-6-phosphate dehydrogenase, a secondary (*e.g.*, isopropanol) alcohol and secondary alcohol dehydrogenase, phosphite and phosphite dehydrogenase, molecular hydrogen and hydrogenase, and the like. These systems may be used in combination with either $NADP^+$/NADPH or NAD^+/NADH as the cofactor. Electrochemical regeneration using hydrogenase may also be used as a cofactor regeneration system. *See, e.g.*, US Pat. Nos. 5,538,867 and 6,495,023, both of which are incorporated herein by reference. Chemical cofactor regeneration systems comprising a metal catalyst and a reducing agent (for example, molecular hydrogen or formate) are also suitable. *See, e.g.*, PCT publication WO 2000/053731, which is incorporated herein by reference.

[0194] The terms "glucose dehydrogenase" and "GDH" are used interchangeably herein to refer to an NAD^+ or $NADP^+$-dependent enzyme that catalyzes the conversion of D-glucose and NAD^+ or $NADP^+$ to gluconic acid and NADH or NADPH, respectively. Equation (1), below, describes the glucose dehydrogenase-catalyzed reduction of NAD^+ or $NADP^+$ by glucose.

$$(1) \quad \text{glucose} + NAD(P)^+ + H_2O \xrightarrow{\text{GDH}} \text{gluconic acid} + NAD(P)H + H^+$$

[0195] Glucose dehydrogenases that are suitable for use in the practice of the methods described herein include both naturally occurring glucose dehydrogenases, as well as non-naturally occurring glucose dehydrogenases. Naturally occurring glucose dehydrogenase encoding genes have been reported in the literature. For example, the *Bacillus subtilis* 61297 GDH gene was expressed in *E. coli* and was reported to exhibit the same physicochemical properties as the enzyme produced in its native host (Vasantha et al., 1983, *Proc. Natl. Acad. Sci. USA* 80:785). The gene sequence of the *B. subtilis* GDH gene, which corresponds to Genbank Acc. No. M12276, was reported by Lampel et al., 1986, *J. Bacteriol.* 166:238-243, and in corrected form by Yamane et al., 1996, *Microbiology* 142:3047-3056 as Genbank Acc. No. D50453. Naturally occurring GDH genes also include those that encode the GDH from *B. cereus* ATCC 14579 (*Nature*, 2003, 423:87-91; Genbank Acc. No. AE017013) and *B. megaterium* (*Eur. J. Biochem.*, 1988, 174:485-490, Genbank Acc. No. X12370; *J. Ferment. Bioeng.*, 1990, 70:363-369, Genbank Acc. No. GI216270). Glucose dehydrogenases from *Bacillus sp.* are provided in PCT publication WO 2005/018579 as SEQ ID NOS: 10 and 12 (encoded by polynucleotide sequences corresponding to SEQ ID NOS: 9 and 11, respectively, of the PCT publication), the disclosure of which is incorporated herein by reference.

[0196] Non-naturally occurring glucose dehydrogenases may be generated using known methods, such as, for example, mutagenesis, directed evolution, and the like. GDH enzymes having suitable activity, whether naturally occurring or non-naturally occurring, may be readily identified using the assay described in Example 4 of PCT publication WO 2005/018579, the disclosure of which is incorporated herein by reference. Exemplary non-naturally occurring glucose dehydrogenases are provided in PCT publication WO 2005/018579 as SEQ ID NOS: 62, 64, 66, 68, 122, 124, and 126. The polynucleotide sequences that encode them are provided in PCT publication WO 2005/018579 as SEQ ID NOS: 61, 63, 65, 67, 121, 123, and 125, respectively. All of these sequences are incorporated herein by reference. Additional non-naturally occurring glucose dehydrogenases that are suitable for use in the ketoreductase-catalyzed reduction reactions disclosed herein are provided in US application publication Nos. 2005/0095619 and 2005/0153417, the disclosures of which are incorporated herein by reference.

[0197] Glucose dehydrogenases employed in the ketoreductase-catalyzed reduction reactions described herein may exhibit an activity of at least about 10 μmol/min/mg and sometimes at least about 10^2 μmol/min/mg or about 10^3 μmol/min/mg, up to about 10^4 μmol/min/mg or higher in the assay described in Example 4 of PCT publication WO 2005/018579.

[0198] The ketoreductase-catalyzed reduction reactions described herein are generally carried out in a solvent. Suitable solvents include water, organic solvents (*e.g.*, ethyl acetate, butyl acetate, 1-octanol, heptane, octane, methyl t-butyl ether (MTBE), toluene, and the like), and ionic liquids (*e.g.*, 1-ethyl 4-methylimidazolium tetrafluoroborate, 1-butyl-3-methylimidazolium tetrafluoroborate, 1-butyl-3-methylimidazolium hexafluorophosphate, and the like). In some embodiments, aqueous solvents, including water and aqueous co-solvent systems, are used.

[0199] Exemplary aqueous co-solvent systems have water and one or more organic solvent. In general, an organic solvent component of an aqueous co-solvent system is selected such that it does not completely inactivate the ketoreductase enzyme. Appropriate co-solvent systems can be readily identified by measuring the enzymatic activity of the specified engineered ketoreductase enzyme with a defined substrate of interest in the candidate solvent system, utilizing an enzyme activity assay, such as those described herein.

[0200] The organic solvent component of an aqueous co-solvent system may be miscible with the aqueous component, providing a single liquid phase, or may be partly miscible or immiscible with the aqueous component, providing two liquid phases. Generally, when an aqueous co-solvent system is employed, it is selected to be biphasic, with water dispersed in an organic solvent, or vice-versa. Generally, when an aqueous co-solvent system is utilized, it is desirable to select an organic solvent that can be readily separated from the aqueous phase. In general, the ratio of water to organic solvent in the co-solvent system is typically in the range of from about 90:10 to about 10:90 (v/v) organic solvent to water, and between 80:20 and 20:80 (v/v) organic solvent to water. The co-solvent system may be pre-formed prior to addition to the reaction mixture, or it may be formed *in situ* in the reaction vessel.

[0201] The aqueous solvent (water or aqueous co-solvent system) may be pH-buffered or unbuffered. Generally, the reduction can be carried out at a pH of about 10 or below, usually in the range of from about 5 to about 10. In some embodiments, the reduction is carried out at a pH of about 9 or below, usually in the range of from about 5 to about 9. In some embodiments, the reduction is carried out at a pH of about 8 or below, often in the range of from about 5 to about 8, and usually in the range of from about 6 to about 8. The reduction may also be carried out at a pH of about 7.8 or below, or 7.5 or below. Alternatively, the reduction may be carried out a neutral pH, *i.e.*, about 7.

[0202] During the course of the reduction reactions, the pH of the reaction mixture may change. The pH of the reaction mixture may be maintained at a desired pH or within a desired pH range by the addition of an acid or a base during the course of the reaction. Alternatively, the pH may be controlled by using an aqueous solvent that comprises a buffer. Suitable buffers to maintain desired pH ranges are known in the art and include, for example, phosphate buffer, triethanolamine buffer, and the like. Combinations of buffering and acid or base addition may also be used.

[0203] When the glucose/glucose dehydrogenase cofactor regeneration system is employed, the co-production of gluconic acid (pKa=3.6), as represented in equation (3) causes the pH of the reaction mixture to drop if the resulting aqueous gluconic acid is not otherwise neutralized. The pH of the reaction mixture may be maintained at the desired level by standard buffering techniques, wherein the buffer neutralizes the gluconic acid up to the buffering capacity provided, or by the addition of a base concurrent with the course of the conversion. Combinations of buffering and base addition may also be used. Suitable buffers to maintain desired pH ranges are described above. Suitable bases for neutralization of gluconic acid are organic bases, for example amines, alkoxides and the like, and inorganic bases, for example, hydroxide salts (*e.g.*, NaOH), carbonate salts (*e.g.*, $NaHCO_3$), bicarbonate salts (*e.g.*, K_2CO_3), basic phosphate salts (*e.g.*, K_2HPO_4, Na_3PO_4), and the like. The addition of a base concurrent with the course of the conversion may be done manually while monitoring the reaction mixture pH or, more conveniently, by using an automatic titrator as a pH stat. A combination of partial buffering capacity and base addition can also be used for process control.

[0204] When base addition is employed to neutralize gluconic acid released during a ketoreductase-catalyzed reduction reaction, the progress of the conversion may be monitored by the amount of base added to maintain the pH. Typically, bases added to unbuffered or partially buffered reaction mixtures over the course of the reduction are added in aqueous solutions.

[0205] In some embodiments, the co-factor regenerating system can comprises a formate dehydrogenase. The terms "formate dehydrogenase" and "FDH" are used interchangeably herein to refer to an NAD^+ or $NADP^+$-dependent enzyme that catalyzes the conversion of formate and NAD^+ or $NADP^+$ to carbon dioxide and NADH or NADPH, respectively. Formate dehydrogenases that are suitable for use as cofactor regenerating systems in the ketoreductase-catalyzed reduction reactions described herein include both naturally occurring formate dehydrogenases, as well as non-naturally occurring formate dehydrogenases. Formate dehydrogenases include those corresponding to SEQ ID NOS: 70 (*Pseudomonas sp.*) and 72 (*Candida boidinii*) of PCT publication WO 2005/018579, which are encoded by polynucleotide sequences corresponding to SEQ ID NOS: 69 and 71, respectively, of PCT publication 2005/018579, the disclosures of which are incorporated herein by reference. Formate dehydrogenases employed in the methods described herein, whether naturally occurring or non-naturally occurring, may exhibit an activity of at least about 1 μmol/min/mg, sometimes at least about 10 μmol/min/mg, or at least about 10^2 μmol/min/mg, up to about 10^3 μmol/min/mg or higher, and can be readily screened for activity in the assay described in Example 4 of PCT publication WO 2005/018579.

[0206] As used herein, the term "formate" refers to formate anion (HCO_2^-), formic acid (HCO_2H), and mixtures thereof. Formate may be provided in the form of a salt, typically an alkali or ammonium salt (for example, HCO_2Na, $KHCO_2NH_4$, and the like), in the form of formic acid, typically aqueous formic acid, or mixtures thereof. Formic acid is a moderate acid. In aqueous solutions within several pH units of its pKa (pKa=3.7 in water) formate is present as both HCO_2^- and HCO_2H in equilibrium concentrations. At pH values above about pH 4, formate is predominantly present as HCO_2^-. When formate is provided as formic acid, the reaction mixture is typically buffered or made less acidic by adding a base to provide the desired pH, typically of about pH 5 or above. Suitable bases for neutralization of formic acid include, but are not limited to, organic bases, for example amines, alkoxides and the like, and inorganic bases, for example, hydroxide salts (*e.g.*, NaOH), carbonate salts (*e.g.*, $NaHCO_3$), bicarbonate salts (*e.g.*, K_2CO_3), basic phosphate salts (*e.g.*, K_2HPO_4, Na_3PO_4), and the like.

[0207] For pH values above about pH 5, at which formate is predominantly present as HCO_2^-, Equation (2) below, describes the formate dehydrogenase-catalyzed reduction of NAD^+ or $NADP^+$ by formate.

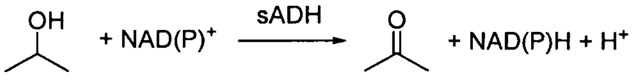
$$(2) \quad HCO_2^- + NAD(P)^+ \xrightarrow{FDH} CO_2 + NAD(P)H$$

[0208] When formate and formate dehydrogenase are employed as the cofactor regeneration system, the pH of the reaction mixture may be maintained at the desired level by standard buffering techniques, wherein the buffer releases protons up to the buffering capacity provided, or by the addition of an acid concurrent with the course of the conversion. Suitable acids to add during the course of the reaction to maintain the pH include organic acids, for example carboxylic acids, sulfonic acids, phosphonic acids, and the like, mineral acids, for example hydrohalic acids (such as hydrochloric acid), sulfuric acid, phosphoric acid, and the like, acidic salts, for example dihydrogenphosphate salts (*e.g.*, KH_2PO_4), bisulfate salts (*e.g.*, $NaHSO_4$) and the like. Some embodiments utilize formic acid, whereby both the formate concentration and the pH of the solution are maintained.

[0209] When acid addition is employed to maintain the pH during a reduction reaction using the formate/formate dehydrogenase cofactor regeneration system, the progress of the conversion may be monitored by the amount of acid added to maintain the pH. Typically, acids added to unbuffered or partially buffered reaction mixtures over the course of conversion are added in aqueous solutions.

[0210] The terms "secondary alcohol dehydrogenase" and "sADH" are used interchangeably herein to refer to an NAD^+ or $NADP^+$-dependent enzyme that catalyzes the conversion of a secondary alcohol and NAD^+ or $NADP^+$ to a ketone and NADH or NADPH, respectively. Equation (3), below, describes the reduction of NAD^+ or $NADP^+$ by a secondary alcohol, illustrated by isopropanol.

(3)



[0211] Secondary alcohol dehydrogenases that are suitable for use as cofactor regenerating systems in the ketoreductase-catalyzed reduction reactions described herein include both naturally occurring secondary alcohol dehydrogenases , as well as non-naturally occurring secondary alcohol dehydrogenases. Naturally occurring secondary alcohol dehydrogenases include known alcohol dehydrogenases from, *Thermoanerobium brockii*, *Rhodococcus etythropolis*, *Lactobacillus kefir*, and *Lactobacillus brevis,* and non-naturally occurring secondary alcohol dehydrogenases include engineered alcohol dehdyrogenases derived therefrom. Secondary alcohol dehydrogenases employed in the methods described herein, whether naturally occurring or non-naturally occurring, may exhibit an activity of at least about 1 μmol/min/mg, sometimes at least about 10 μmol/min/mg, or at least about 10^2 μmol/min/mg, up to about 10^3 μmol/min/mg or higher.

[0212] Suitable secondary alcohols include lower secondary alkanols and aryl-alkyl carbinols. Examples of lower secondary alcohols include isopropanol, 2-butanol, 3-methyl-2-butanol, 2-pentanol, 3-pentanol, 3,3-dimethyl-2-butanol, and the like. In one embodiment the secondary alcohol is isopropanol. Suitable aryl-akyl carbinols include unsubstituted and substituted 1-arylethanols.

[0213] When a secondary alcohol and secondary alcohol dehydrogenase are employed as the cofactor regeneration system, the resulting NAD^+ or $NADP^+$ is reduced by the coupled oxidation of the secondary alcohol to the ketone by the secondary alcohol dehydrogenase. Some engineered ketoreductases also have activity to dehydrogenate a secondary alcohol reductant. In some embodiments using secondary alcohol as reductant, the engineered ketoreductase and the secondary alcohol dehydrogenase are the same enzyme.

[0214] In carrying out embodiments of the ketoreductase-catalyzed reduction reactions described herein employing a cofactor regeneration system, either the oxidized or reduced form of the cofactor may be provided initially. As described above, the cofactor regeneration system converts oxidized cofactor to its reduced form, which is then utilized in the reduction of the ketoreductase substrate.

[0215] In some embodiments, cofactor regeneration systems are not used. For reduction reactions carried out without the use of a cofactor regenerating systems, the cofactor is added to the reaction mixture in reduced form.

[0216] In some embodiments, when the process is carried out using whole cells of the host organism, the whole cell may natively provide the cofactor. Alternatively or in combination, the cell may natively or recombinantly provide the glucose dehydrogenase.

[0217] In carrying out the enantioselective reduction reactions described herein, the engineered ketoreductase enzyme, and any enzymes comprising the optional cofactor regeneration system, may be added to the reaction mixture in the form of the purified enzymes, whole cells transformed with gene(s) encoding the enzymes, and/or cell extracts and/or lysates of such cells. The gene(s) encoding the engineered ketoreductase enzyme and the optional cofactor regeneration enzymes can be transformed into host cells separately or together into the same host cell. For example, in some embodiments one set of host cells can be transformed with gene(s) encoding the engineered ketoreductase enzyme and another set can be transformed with gene(s) encoding the cofactor regeneration enzymes. Both sets of transformed cells can be utilized together in the reaction mixture in the form of whole cells, or in the form of lysates or extracts derived therefrom. In other embodiments, a host cell can be transformed with gene(s) encoding both the engineered ketoreductase enzyme and the cofactor regeneration enzymes.

[0218] Whole cells transformed with gene(s) encoding the engineered ketoreductase enzyme and/or the optional cofactor regeneration enzymes, or cell extracts and/or lysates thereof, may be employed in a variety of different forms, including solid (*e.g.*, lyophilized, spray-dried, and the like) or semisolid (*e.g.*, a crude paste).

[0219] The cell extracts or cell lysates may be partially purified by precipitation (ammonium sulfate, polyethyleneimine, heat treatment or the like, followed by a desalting procedure prior to lyophilization (*e.g.*, ultrafiltration, dialysis, and the like). Any of the cell preparations may be stabilized by crosslinking using known crosslinking agents, such as, for example, glutaraldehyde or immobilization to a solid phase (*e.g.*, Eupergit C, and the like).

[0220] The solid reactants (*e.g.*, enzyme, salts, etc.) may be provided to the reaction in a variety of different forms, including powder (*e.g.*, lyophilized, spray dried, and the like), solution, emulsion, suspension, and the like. The reactants can be readily lyophilized or spray dried using methods and equipment that are known to those having ordinary skill in the art. For example, the protein solution can be frozen at -80°C in small aliquots, then added to a prechilled lyophilization chamber, followed by the application of a vacuum. After the removal of water from the samples, the temperature is typically raised to 4°C for two hours before release of the vacuum and retrieval of the lyophilized samples.

[0221] The quantities of reactants used in the reduction reaction will generally vary depending on the quantities of product desired, and concomitantly the amount of ketoreductase substrate employed. The following guidelines can be used to determine the amounts of ketoreductase, cofactor, and optional cofactor regeneration system to use. Generally, keto substrates can be employed at a concentration of about 20 to 300 grams/liter using from about 50 mg to about 5 g of ketoreductase and about 10 mg to about 150 mg of cofactor. Those having ordinary skill in the art will readily understand how to vary these quantities to tailor them to the desired level of productivity and scale of production. Appropriate quantities of optional cofactor regeneration system may be readily determined by routine experimentation based on the amount of cofactor and/or ketoreductase utilized. In general, the reductant (*e.g.*, glucose, formate, isopropanol, etc.) is utilized at levels above the equimolar level of ketoreductase substrate to achieve essentially complete or near complete conversion of the ketoreductase substrate.

[0222] The order of addition of reactants is not critical. The reactants may be added together at the same time to a solvent (*e.g.*, monophasic solvent, biphasic aqueous co-solvent system, and the like), or alternatively, some of the reactants may be added separately, and some together at different time points. For example, the cofactor regeneration system, cofactor, ketoreductase, and ketoreductase substrate may be added first to the solvent.

[0223] For improved mixing efficiency when an aqueous co-solvent system is used, the cofactor regeneration system, ketoreductase, and cofactor may be added and mixed into the aqueous phase first. The organic phase may then be added and mixed in, followed by addition of the ketoreductase substrate. Alternatively, the ketoreductase substrate may be premixed in the organic phase, prior to addition to the aqueous phase

[0224] Suitable conditions for carrying out the ketoreductase-catalyzed reduction reactions described herein include a wide variety of conditions which can be readily optimized by routine experimentation that includes, but is not limited to, contacting the engineered ketoreductase enzyme and substrate at an experimental pH and temperature and detecting product, for example, using the methods described in the Examples provided herein.

[0225] The ketoreductase catalyzed reduction is typically carried out at a temperature in the range of from about 15°C to about 75°C. For some embodiments, the reaction is carried out at a temperature in the range of from about 20°C to about 55°C. In still other embodiments, it is carried out at a temperature in the range of from about 20°C to about 45°C. The reaction may also be carried out under ambient conditions.

[0226] The reduction reaction is generally allowed to proceed until essentially complete, or near complete, reduction of substrate is obtained. Reduction of substrate to product can be monitored using known methods by detecting substrate and/or product. Suitable methods include gas chromatography, HPLC, and the like. Conversion yields of the alcohol reduction product generated in the reaction mixture are generally greater than about 50%, may also be greater than about 60%, may also be greater than about 70%, may also be greater than about 80%, may also be greater than 90%, and are often greater than about 97%.

7. EXAMPLES

[0227] Various features and embodiments of the disclosure are illustrated in the following representative examples, which are intended to be illustrative, and not limiting.

[0228] In the following description, wherever glucose dehydrogenase (GDH) is used, it is GDH CDX901, obtainable from Julich Chiral Solutions, Jülich, Germany.

7.1 Example 1: Wild-type Ketoreductase Gene Acquisition and Construction of Expression Vectors.

[0229] Ketoreductase (KRED) encoding genes were designed for expression in *E. coli* based on the reported amino acid sequence of the ketoreductase and a codon optimization algorithm as described in Example 1 of US provisional application Serial No. 60/848,950, incorporated herein by reference. Genes were synthesized using oligonucleotides composed of 42 nucleotides and cloned into expression vector pCK110900 (depicted as FIG. 3 in United States Patent Application Publication 20060195947) under the control of a *lac* promoter. The expression vector also contained the P15a origin of replication and the chloramphenicol resistance gene. Resulting plasmids were transformed into *E. coli* W3110 using standard methods. Codon optimized genes and the encoding polypeptides *adh-LB* gene (Genbank

Acc. No.: GI:28400789) and *adh-LK* gene (Genbank Acc. No.: AAP94029.1; GI:33112056) can be found as SEQ ID NO:1 and 3. The activity of the wild-type ketoreductases was confirmed as described in US provisional application Serial No. 60/848,950. Other ketoreductases used herein, including codon optimized genes and the encoding polypeptides for *Ydl124wp* (Genbank Acc. No.:NP_010159.1; GI:6320079), *adh-LB* gene (Genbank Acc. No.: 1NXQ_A; GI:30749782), *adh-RE* gene (Genbank Acc. No.: AAN73270.1; GI:34776951), *Yprlp* gene (Genbank Acc. No.: NP_010656.1; GI:6320576), *Gre2p* gene (Genbank Acc. No.: NP_014490.1; GI:6324421) are also disclosed in US provisional application Serial No. 60/848,950, which is incorporated herein by reference.

[0230] Polynucleotides encoding engineered ketoreductases of the present disclosure were likewise cloned into vector pCK110900 for expression in *E. coli* W3110.

7.2 Example 2: LC/MS/MS Assay for Substrate Specificity and Conversion.

[0231] A single microbial colony of *E. coli* containing a plasmid with the ketoreductase gene of interest was inoculated into 50 ml Luria Bertani broth containing 30 μg/ml chloramphenicol and 1% glucose. Cells were grown overnight (at least 16 hrs) in an incubator at 30°C with shaking at 250 rpm. The culture was diluted into 250 ml Terrific Broth (12 g/L bacto-tryptone, 24g/L yeast extract, 4 ml/L glycerol, 65mM potassium phosphate, pH 7.0, 1mM $MgSO_4$, 30 μg/ml chloramphenicol) in 1 liter flask) to an optical density at 600 nm (OD600) of 0.2 and allowed to grow at 30°C. Expression of the ketoreductase gene was induced with 1mM IPTG when the OD600 of the culture is 0.6 to 0.8 and incubated overnight (at least 16 hrs). Cells were harvested by centrifugation (5000rpm, 15 min, 4°C) and the supernatant discarded. The cell pellet was resuspended with an equal volume of cold (4°C) 100mM triethanolamine (chloride) buffer, pH 7.0 (including 2mM $MgSO_4$ in the case of ADH-LK and ADH-LB and engineered ketoreductases derived therefrom), and harvested by centrifugation as above. The washed cells were resuspended in two volumes of the cold triethanolamine (chloride) buffer and passed through a French Press twice at 12000psi while maintained at 4°C. Cell debris was removed by centrifugation (9000 rpm, 45 min., 4°C). The clear lysate supernatant was collected and stored at -20°C. Lyophilization of frozen clear lysate provided a dry powder of crude ketoreductase enzyme.

[0232] The activity of the wild-type ketoreductases was confirmed as described US provisional application Serial No. 60/848,950. To a solution of 1 mL 100mM (sodium) phosphate buffer, pH 7.5, were added 10 mg ketoreductase powder, 50 mg NAD(P)H, 100 μL isopropanol and 10 mg 4'-chloroacetophenone or unsubstituted acetophenone. The reaction mixture was stirred at room temperature for 16 hours, then extracted with 1 mL MTBE. A sample of the MTBE phase was analyzed by chiral HPLC for the conversion of the 4'-chloroacetophenone and the enantiomeric composition of the product, 1-(4'-chlorophenyl)ethanol.

7.3 Example 3: Production of Ketoreductases; Fermentation Procedure.

[0233] In an aerated agitated 15L fermenter, 6.0L of growth medium containing 0.88 g/L ammonium sulfate, 0.98 g/L of sodium citrate; 12.5 g/L of dipotassium hydrogen phosphate trihydrate, 6.25g/L of potassium dihydrogen phosphate, 6.2 g/L of Tastone-154 yeast extract, 0.083 g/L ferric ammonium citrate, and 8.3 ml/L of a trace element solution containing 2 g/L of calcium chloride dihydrate, 2.2 g/L of zinc sulfate septahydrate, 0.5 g/L manganese sulfate monohydrate, 1 g/L cuprous sulfate heptahydrate, 0.1 g/L ammonium molybdate tetrahydrate and 0.02 g/L sodium tetraborate decahydrate was brought to a temperature of 30°C. The fermenter was inoculated with a late exponential culture of E. coli W3110, containing a plasmid with the ketoreductase gene of interest, grown in a shake flask as described in Example 3 to a starting OD600 of 0.5 to 2.0. The fermenter was agitated at 500-1500 rpm and air was supplied to the fermentation vessel at 1.0-15.0 L/min to maintain dissolved oxygen level of 30% saturation or greater. The pH of the culture was controlled at 7.0 by addition of 20% v/v ammonium hydroxide. Growth of the culture was maintained by the addition of a feed solution containing 500g/L cerelose, 12 g/L ammonium chloride and 10.4 g/L magnesium sulfate heptahydrate. After the culture reached an OD600 of 50, the expression of ketoreductase was induced by the addition of isopropyl-b-D-thiogalactoside (IPTG) to a final concentration of 1mM. The culture was grown for another 14 hours. The culture was then chilled to 4°C and maintained at 4°C until harvested. Cells were harvested by centrifugation at 5000G for 40 minutes in a Sorval RC12BP centrifuge at 4°C. Harvested cells were used directly in the following downstream recovery process or were stored at 4°C until such use.

[0234] The cell pellet was resuspended in 2 volumes of 100 mM triethanolamine (chloride) buffer, pH 6.8, at 4°C to each volume of wet cell paste. The intracellular ketoreductase was released from the cells by passing the suspension through a homogenizer fitted with a twostage homogenizing valve assembly using a pressure of 12000 psig. The cell homogenate was cooled to 4°C immediately after disruption. A solution of 10% w/v polyethyleneimine, pH 7.2, was added to the lysate to a final concentration of 0.5% w/v and stirred for 30 minutes. The resulting suspension was clarified by centrifugation at 5000G in a standard laboratory centrifuge for 30 minutes. The clear supernatant was decanted and concentrated ten fold using a cellulose ultrafiltration membrane with a molecular weight cut off of 30Kd. The final concentrate was dispensed into shallow containers, frozen at -20°C and lyophilized to powder. The ketoreductase powder was stored at -80°C.

7.4 Example 4 Analytical methods to determine conversion and enantiomeric excess of Compound (II).

[0235] Reversed Phase HPLC assay for conversion. The following HPLC method was used to analyze the reduction of the compound of formula (I) to the compound of formula (II) in high throughput:

Instrument:	Agilent 1100 series HP
Method name:	1549-ISO
Column type:	Eclipse XDB C18
Column size:	2.1 x 50mm
Packing size:	3.5μm C18 Zorbax XDB
Run time	3 minutes
Mobile phase	50% acetylnitrile 50% 0.25% acetic acid
Flow Rate:	0.6 ml/min, Temperature: Ambient
Detection:	UV 250 nm
Elution time:	alcohol: 0.9 minute ketone: 1.2 minute.

[0236] Chiral HPLC for determination of enantiomeric excess. The following HPLC method was used to separate and analyze the (S) and (R) enantiomers of the compound of formula (II):

Instrument:	Agilent 1100 series HP

Column:	Chiralpak AD-H, 250 × 4.6 mm,
Solvent:	Isocratic, 80% A (Heptane) and 20% B (Heptane-Isopropanol 50:50)
Flow Rate:	1 ml/min, Temperature: Ambient
Detection:	UV 220 nm
Run Time:	45 min

[0237] Retention times were as follows 29.1 min for (I), 35.4 min for R-(II), and 38.7 min for S-(II). The R- and S- alcohols showed baseline separation with 1:1 area ratio.

7.5 Example 5: Evaluation of Wild-Type Ketoreductases for Activity to Reduce Compound (I) using Glucose and Glucose Dehydrogenase for Cofactor Regeneration:

[0238] Reaction mixtures containing 30 mg/L KRED, 12 mg/ml 1-[4-(4-fluoro-2-methyl-1*H*-indol-5-yloxy)-5-methyl-pyrrolo[2,1-*f*][1,2,4]triazin-6-yloxy]propan-2-one (I) (obtainable according to scheme 1 and Example 1 in WO06130657, incorporated herein by reference), 1.5 mg/mL GDH CDX901 (Julich Chiral Solutions, Jülich, Germany), 1mM NADP$^+$, 66 mg/mL glucose, 200 mM NaH_2PO_4/Na_2HPO_4 (pH 7), 100mM triethanolamine/chloride buffer, pH 7.0, 1 mM $MgSO_4$ in 1ml reaction volume were incubated with stirring at room temperature for 20 hr, then extracted with 3 mL ethyl acetate and analyzed by the methods as described in Example 4 or according to Example 2 of WO06130657. The results are shown in Table 3.

Table 3: Activity of wild-type KREDs on Compound (I)

Enzyme	Activity	Enantioselectivity
GREc	+	S
YDLC	+	S and R
YPrR	+	S
ADH-LK	+	R
RhoC	+	S

+ indicates that activity was observed.

[0239] The wild-type ketoreductase, ADH-LK, provides R-(II) with >99.9% enantioselectivity (>99.9% e.e.).

[0240] The Example illustrates the evaluation of wild-type ketoreductases for activity and enantioselectivity when used in combination with a cofactor regeneration system (glucose and glucose dehydrogenase). Wild-type ketoreductase, ADH-LK, provides the desired (R)-enantiomer.

7.6 Example 6: Evaluation of Wild-Type Ketoreductases for Activity to Reduce Compound (I) using Isopropyl alcohol for Cofactor Regeneration:

[0241] Reaction mixtures containing 15 mg/ml KRED; 2 mg/ml (I) in DMF or in THF (added by 10x-dilution from a 20 mg/ml stock solution), 2 mg/mL NADP+, 0.4-0.5 ml IPA, 0.4 ml 100mM triethanolamine/chloride buffer, pH 7.0, 1mM $MgSO_4$ in 1ml reaction volume were incubated with stirring at room temperature for 8 hr (when DMF was used) or 16 hr (when THF was used). Samples were analyzed by the methods of Examples 4 and 5.

[0242] Under these reaction conditions ADH-LK from *L. kefiri* provided 21% conversion in DMF and 61% conversion in THF, while ADH-LB from *L. brevis* provided 70% conversion in DMF and 26% conversion in THF

[0243] The Example illustrates the evaluation of wild-type *Lactobacillus* ketoreductases for activity and enantioselectivity when used in combination with IPA for cofactor regeneration. Wild-type ketoreductase, ADH-LK, provides the desired (R)-enantiomer in high e.e. in the presence of 10% THF.

7.7 Example 7: High Throughput HPLC Assay for Ketoreductase Activity on Compound (I) using glucose/glucose dehydrogenase for co-factor recycling.

[0244] Library colonies were picked using a Q-bot® robotic colony picker (Genetix USA, Inc., Beaverton, OR) into 96-well shallow well microtiter plates containing 180 µL Luria Bertani broth (LB), 1% glucose and 30 µg/mL chloramphenicol (CAM). Cells were grown overnight at 37°C with shaking at 250 rpm. 10 µL of this culture was then transferred into 96-deep well plates containing 390 µL Terrific broth (TB) and 30 µg/mL CAM. After incubation of deep-well plates at 30°C with shaking at 250 rpm for 2.5 to 3 hours (OD_{600} 0.6-0.8), recombinant gene expression by the cell cultures was induced by isopropyl thiogalactoside (IPTG) to a final concentration of 1 mM. The plates were then incubated at 30°C with shaking at 250 rpm for overnight.

[0245] Cells were pelleted via centrifugation, resuspended in 300 µL lysis buffer and lysed by shaking at room temperature for at least 2 hours. The lysis buffer contained 100 mM triethanolamine (chloride) buffer, pH 7.0-7.2, 1 mg/mL lysozyme and 750 µg/mL polymixin B sulfate.

[0246] Ketoreductase activity was measured by transferring measured quantities of the lysis mixtures into the wells of microtiter plates containing 205 µL an assay mixture containing 1.7mg/mL GDH CDX901, 0.7 mg/ml NADP+, 66.7 mg/mL glucose, 200mM NaH_2PO_4/Na_2HPO_4 (pH 7), 100mM triethanolamine/chloride buffer, pH 7.0, 1mM $MgSO_4$. Reactions were initiated by addition of 25 µL 2 mg/ml (I) in DMF (final concentration of (I) is 0.2 mg/ml) and incubated at 25°C for 1 hr. 500 µL acetonitrile was added per well, mixed well after which 200 µL was transferred to a Solvinert (Millipore, MA) filter plate. The filtrate was collected in a Nunc round bottom plate by centrifugation of the solvenert filter plate at 200 rpm for 1 minute. The sample plate was sealed to prevent evaporation of the solvent and analyzed by HPLC according to Example 4.

7.8 Example 8: High Throughput HPLC Assay for Ketoreductase Activity on Compound (I) using IPA for co-factor recycling.

[0247] Cell pellets were prepared according to Example 8, resuspended in 150 µL 200mM triethanolamine/chloride buffer, pH 7.0 with 1mM $MgSO_4$. 150 µL isopropylalcohol was added to the resuspended cells and after sealing the plates, the cells were lysed by shaking at room temperature on an orbital shaker for at least 120 minutes.

[0248] Ketoreductase activity was measured by transferring measured quantities of the lysis mixtures into the wells of a deep-well (2 ml) microtiter plates containing 175 µL an assay mixture consisting of 100 mM triethanolamine/chloride buffer, pH 7.0, 25% isopropyl alcohol (IPA), 2% acetone, 1mM $MgSO_4$ and 0.7 mg/ml NADP+, as well as 50 µL of 5 mg/ml (I) in 50%THF/50%IPA.

[0249] Reactions were initiated by addition of 25 µL lysate (diluted with equal volume of IPA if necessary), heat sealed, and incubated in a shaking incubator at 25 to 50°C for 18 hr. At the end of the reaction 500 µL ethyl acetate was added to each well, the plate resealed and shaken vigorously for at least 5 minutes on an orbital shaker. Plates were centrifuged for 20 to 30 sec. at 4000 rpm. 140 μL acetonitrile and 70 μL of the ethylacetate reaction mix were transferred to a Solvinert filter plate (Millipore, MA) and the filtrates were collected in a Nunc round bottom plate by centrifugation of the solvenert filter plate at 200 rpm for 3 minutes. The sample plate was sealed to prevent evaporation of the solvent and analyzed by HPLC according to Example 4.

7.9 Example 9: High Throughput Fluorescence Prescreen for Ketoreductases active on Isopropanol.

[0250] Cells were grown, harvested and lysed according to Example 8.

[0251] In 96-well black microtiter plates 20 μL of sample (diluted in 100 mM triethanolamine/chloride buffer, pH 7.0, 1mM $MgSO_4$ if necessary) was added to 180 μL of an assay mixture consisting of 100 mM triethanolamine/chloride buffer, pH 7.0, 2% isopropyl alcohol (IPA), 1mM $MgSO_4$ and reaction progress measured by following the reduction in fluorescence of NADP upon conversion to NADPH at 445 nm after excitation at 330 nm in a Flexstation (Molecular Devices, USA).

7.10 Example 10: High Throughput Screen for Acetone Stable Ketoreductases

[0252] Cell pellets were prepared according to Example 8, resuspended in 150 μL 200mM triethanolamine/chloride buffer, pH 7.0 with 1mM $MgSO_4$. 150 μL of a mix containing 76% IPA, 20% THF and 4% acetone was added to the resuspended cells and after sealing the plates, the cells were lysed by shaking at room temperature on an orbital shaker for 18 hrs.

[0253] Ketoreductase activity was measured by transferring measured quantities of the lysis mixtures into the wells of a deep-well (2 ml) microtiter plates containing 175 μL an assay mixture consisting of 80 mM triethanolamine/chloride buffer, pH 7.0, 26.2 to 37.1% isopropyl alcohol (IPA), 1.8% acetone, 1mM $MgSO_4$ and 0.7 mg/ml NADP+, as well as 50 μL of 5 mg/ml (I) in 50%THF/50%IPA.

[0254] Reactions were initiated by addition of 25 μL lysate (diluted with equal volume of IPA if necessary), heat sealed, and incubated in a shaking incubator at 25 to 50°C for 18 hr. At the end of the reaction 500 μL ethyl acetate was added to each well, the plate resealed and shaken vigorously for at least 5 minutes on an orbital shaker. Plates were centrifuged for 20 to 30 sec. at 4000 rpm (3220 x g). 140 μL acetonitrile and 70 μL of the ethylacetate reaction mix were transferred to a Solvinert filter plate (Millipore, MA) and the filtrates were collected in a Nunc round bottom plate by centrifugation of the solvenert filter plate at 200 rpm for 3 minutes. The sample plate was sealed to prevent evaporation of the solvent and analyzed by HPLC according to Example 4.

7.11 Example 11: Improved Activity of Engineered Ketoreductases Derived from Wild-type ADH-LK for the Reduction of Compound (I) to (*R*)-2 using Isopropyl alcohol for Cofactor Regeneration:

[0255] Reaction mixtures containing 15 mg/ml KRED; 2 mg/ml (I) in THF, 2 mg/mL NADP+, 0.4 ml IPA, 0.5 ml 100mM triethanolamine/chloride buffer, pH 7.0, 1mM $MgSO_4$ in 1ml reaction volume were incubated with stirring at room temperature for 16 hr. Samples were analyzed by the methods of Examples 4 and 5.

[0256] Under these reaction conditions ADH-LK gave 61% conversion, ADH-LB gave 26% conversion in THF, a ADH-LK variant with SEQ ID NO:114 gave 100% conversion and a ADH-LK variant with SEQ ID NO:18 gave 93% conversion.

[0257] When tested under similar reaction conditions but with 3 mg/ml KRED and 1 mg/ml (I) in THF, conversions with ADH-LK, SEQ ID NO:114 and SEQ ID NO:18 after a 2 hr reaction time were 11, 100 and 23% respectively. Introduction of the mutation A94G from SEQ ID NO:114 into ADH-LB provided SEQ ID NO:116. Ketoreductases with SEQ ID NO:114 and SEQ ID NO:116 have similar activity in 10% DMF.

[0258] This Example illustrates, by comparisons of ketoreductase amounts, reaction times, and conversions, that engineered ketoreductases derived from the wild-type ketoreductase ADH-LK provide improved activity compared to ketoreductase ADH-LK.

7.12 Example 12: Improved Conversion of Compound (I) by Engineered Ketoreductases Derived from ADH-LK.

[0259] The conversion by engineered ADH-LK polypeptides was determined by incubating 1 g/L of various KRED under conditions as described in Example 7.

[0260] Table 4 gives the SEQ ID NO. corresponding to the ketoreductase powder, the number of amino acid mutations from the wild-type ADH-LK and the conversion of Compound (I) to (R)-2 in the 20 minute reaction.

Table 4: Improved activity of engineered ADH-LK variants		
Ketoreductases SEQ ID NO.	Mutations from ADH-LK	Conversion
SEQ ID NO:2	-	+
SEQ ID NO:8	2	++++
SEQ ID NO:10	1	++++
SEQ ID NO:12	2	++++
SEQ ID NO:14	1	+++
SEQ ID NO:16	2	+++
SEQ ID NO:18	1	++++
SEQ ID NO:112	2	++++
SEQ ID NO:114	1	++++

+: <20% conversion; ++: 25-75% conversion; +++: >75% conversion.

7.13 Example 13: Improved Conversion and Tolerance to Acetone of Engineered Ketoreductases Derived from ADH-LK

[0261] The conversion and tolerance to acetone of engineered ADH-LK variants was determined by incubating 1 g/L of various KRED with 0.5 g/L (I) in a mixture of 10% THF, 40% IPA, 50% 100 mM triethanolamine-chloride, 1 mM $MgSO_4$ pH 8.0, 0.7 mg/ml $NADP^+$ in the absence or presence of 2% acetone for 20 min at room temperature followed by determination of conversion according to Example 4.

[0262] Table 5 gives the SEQ ID NO. corresponding to the ketoreductase powder, the number of amino acid mutations from the wild-type ADH-LK and the conversion of (I) to (R)-2 in the 20 minute reaction.

Table 5: Tolerance of engineered ADH-LK variant towards acetone.			
Ketoreductases SEQ ID NO.	Mutations from ADH-LK	Conversion without acetone	Conversion with acetone

SEQ ID NO:114	1	76%	12%
SEQ ID NO:50	7	98%	62%
SEQ ID NO:78	4	99.10%	96.0%
SEQ ID NO:110	8	98.6%	98.2%

7.14 Example 14: Preparation of R-isomer of Compound (II).

[0263] To a 20 ml sample vial (21mm OD) containing a magnetic stir bar (4x12mm) was added 3.5 mg of NADP$^+$ (mono sodium salt from Oriental Yeast, Japan), 25 mg of KRED with SEQ ID NO:78 and 500 mg of (I). A mixture of 1.5 ml 2-methylTHF (Aldrich, USA), 1.0 ml isopropyl alcohol, and 2.5 ml 100 mM triethanolamine/chloride (pH 8), 1mM $MgSO_4$, was added to the solids and the resulting three-phase mixture was stirred and heated (oil bath) at 40°C for 24 hr. A sample was taken from the stirred mixture for analysis according to Example 4 or according to Example 2 of WO06130657 at which point (I) was completely converted to R-(II) of >99.9% e.e.

7.15 Example 15: High Throughput Chiral GC Assay for Acetophenone Reduction using IPA for co-factor recycling:

[0264] Chiral GC analysis:

Instrument:	Astec Chiraldex B-DP column (30m x 0.25 mm)
Temperature:	110°C
Inlet temperature:	250°C
Split:	1:100
Pressure:	15 psi Helium
Detector:	FID, 250°C
Retention time:	Ketone: 6.6 minutes
	(*R*)-alcohol: 9.1 minutes
	(*S*)-alcohol: 9.5 minutes

7.16 Example 16: Improved Activity of Engineered Ketoreductases Derived from Wild-type Kefir for the Reduction of acetophenone to (*R*)-1-Phenylethanol using Isopropyl alcohol for Cofactor Regeneration:

[0265] Cell lysates were prepared by picking colonies using a Q-bot® robotic colony picker (Genetix USA, Inc., Beaverton, OR) into 96-well shallow well microtiter plates containing 180 µL Luria Bertani broth (LB), 1% glucose and 30 µg/mL chloramphenicol (CAM). Cells were grown overnight at 37°C with shaking at 250 rpm. 10 µL of this culture was then transferred into 96-deep well plates containing 390 µL Terrific broth (TB) and 30 µg/mL CAM. After incubation of deep-well plates at 30°C with shaking at 250 rpm for 2.5 to 3 hours (OD_{600} 0.6-0.8), recombinant gene expression by the cell cultures was induced by isopropyl thiogalactoside (IPTG) to a final concentration of 1 mM. The plates were then incubated at 30°C with shaking at 250 rpm for overnight.

[0266] Cells were pelleted via centrifugation, resuspended in 300 µL lysis buffer and lysed by shaking at room temperature for at least 2 hours. The lysis buffer contained 100 mM triethanolamine (chloride) buffer, pH 7.0-7.2, 1 mg/mL lysozyme and 750 µg/mL polymixin B sulfate. To 100 µL cell lysate was added in 50 µL 100 mM Triethanolamine-HCl buffer, pH 7.0, containing 0.5 mM NADP sodium salt, 300 µL isopropanol, and 50 µL 50 g/L acetophenone in tetrahydrofuran (THF) and after sealing the plates, the cells were incubated by shaking at room temperature at 850 rpm on an orbital shaker for 4 hrs.

[0267] The product of the reaction (1-phenylethanol) was extracted by adding 1 mL of ethyl acetate to each sample, and after sealing the microtiter plate, shaking at 850 rpm at room temperature for 10 minutes. Plates were centrifuged for 2 minutes at 4,000 rpm in a plate centrifuge (3220 x g) at 4°C and 200 µL of the organic phase from each well was transferred to a shallow well plate and the plates sealed prior to chiral GC analysis.

[0268] Table 6 gives the SEQ ID NO. corresponding to the ketoreductase, the number of amino acid mutations from the wild-type KEFc (ADH-LK) and the conversion of acetophenone to (*R*)-1-phenylethanol.

Table 6: Conversion of acetophenone.

Ketoreductases SEQ ID NO.	Number of amino acid mutations from ADH-LK	Conversion
ADH-LK	-	+
SEQ ID NO:16	2	++
SEQ ID NO:66	4	++
SEQ ID NO:120	1	+++
SEQ ID NO:122	3	+++
SEQ ID NO:124	2	+++
SEQ ID NO:126	2	+++
SEQ ID NO:128	1	+++
SEQ ID NO:130	1	++
SEQ ID NO:132	1	++
SEQ ID NO:134	1	++
SEQ ID NO:136	2	+++
SEQ ID NO:138	2	+++

+: <70% conversion; ++: 70-90% conversion; +++: >90% conversion.

[0269] All publications, patents, patent applications and other documents cited in this application are hereby incorporated by reference in their entireties for all purposes to the same extent as if each individual publication, patent, patent application or other document were individually indicated to be incorporated by reference for all purposes.

[0270] While various specific embodiments have been illustrated and described, it will be appreciated that various changes can be made without departing from the spirit and scope of the invention(s).

WHAT IS CLAIMED IS:

1. A recombinant ketoreductase polypeptide capable of converting the compound 1-[4-(4-fluoro2-methyl-IH-indol-5-yloxy)-5-methyl-pyrrolo[2, 1-f][1 ,2,4]triazin-6-yloxy]-propan-2-one to 1-[4-(4-fluoro-2-methyl-IH-indol-5-yloxy)-5-methylpyrrolo[2, 1-f][1 ,2,4]triazin-6-yloxy]-propan-2-ol, with at least 1.5 times the activity of the polypeptide of SEQ ID NO:2, the polypeptide comprising an amino acid sequence with at least 90% sequence identity to SEQ ID NO:2 or SEQ ID NO:4 and having (a) an aromatic amino acid or G at the amino acid residue corresponding to residue 94 of SEQ ID NO:2 or SEQ ID NO:4, or (b) an amino acid other than S and N at the amino acid residue corresponding to residue 96 of SEQ ID NO:2 or SEQ ID NO:4.

2. The recombinant polypeptide of claim 1 in which the amino acid sequence comprises an aromatic amino acid or G at the amino acid residue corresponding to residue 94 of SEQ ID NO:2 or SEQ ID NO:4.

3. The recombinant polypeptide of claim 2 in which the residue 94 is F, W, H, or Y.

4. The recombinant polypeptide of claim 2 in which the amino acid sequence further comprises one or more features selected from: residue 96 is any amino acid other than S/N; residue 153 is an aliphatic amino acid residue other than L; residue 199 is any amino acid residue other than L; residue 202 is G or an aliphatic amino acid residue other than A; and residue 206 is an aromatic amino acid residue.

5. The recombinant polypeptide of claim 4 in which the amino acid sequence has one or more of features selected from: residue 153 is G or A; residue 199 is K, I, N, R, V, Q, or W; residue 202 is I, L, or G; and residue 206 is F.

6. The recombinant polypeptide of any one of claims 1 to 5 in which the amino acid sequence further comprises one or more of the features selected from: residue 49 is a polar amino acid residue other than K; residue 53 is an acidic amino acid residue; residue 54 is a small or aliphatic amino acid residue other than T/P; residue 60 is an aliphatic amino acid

residue other than V; residue 95 is an aliphatic amino acid other than V; residue 97 is a small amino acid or G; residue 109 is a basic amino acid residue other than K; residue 147 is an aliphatic amino acid residue; residue 165 is a hydroxyl or small amino acid residue; residue 197 is a small amino acid residue or G; residue 223 is an aliphatic amino acid residue other than L; and residue 233 is a small amino acid residue or G.

7. The recombinant polypeptide of claim 6 in which the amino acid sequence comprises one or more features selected from: residue 49 is R; residue 53 is D; residue 54 is A; residue 60 is A; residue 95 is L; residue 97 is G; residue 109 is R; residue 147 is L; residue 165 is T; residue 197 is G; residue 223 is V; and residue 233 is G.

8. The recombinant polypeptide of claim 7 in which the amino acid sequence has one or more features selected from: residue 54 is A; residue 109 is R; residue 147 is L; and residue 233 is G.

9. The recombinant polypeptide of claim 1 in which the amino acid sequence comprises an amino acid other than S and N at the amino acid residue corresponding to residue 96 of SEQ ID NO:2 or SEQ ID NO:4.

10. The recombinant polypeptide of claim 9 in which the amino acid sequence has a G, F, Y, or I at the amino acid residue corresponding to residue 96 of SEQ ID NO:2 or SEQ ID NO:4.

11. The recombinant polypeptide of claim 10 in which the amino acid sequence further comprises one or more of the following features selected from: residue 94 is an aromatic amino acid residue or G; residue 153 is an aliphatic amino acid residue other than L; residue 199 is any amino acid residue other than L; residue 202 is an aliphatic amino acid residue other than A; or residue 206 is an aromatic amino acid residue.

12. The recombinant polypeptide of claim 11 in which the amino acid sequence comprises one or more of the following features selected from: residue 94 is F or G; residue

153 is G or A; residue 199 is K, I, N, R, V, Q, or W; residue 202 is I, L, or G; and residue 206 is F.

13. The recombinant polypeptide of any one of claims 9 to 12 in which the amino acid sequence further comprises one or more of the following features selected from: residue 49 is a polar amino acid residue other than K; residue 53 is an acidic amino acid residue; residue 54 is a small or aliphatic amino acid residue other than T/P; residue 60 is an aliphatic amino acid residue other than V; residue 95 is an aliphatic amino acid other than V; residue 97 is a small amino acid or G; residue 109 is a basic amino acid residue other than K; residue 147 is an aliphatic amino acid residue; residue 165 is a hydroxyl or small amino acid residue; residue 197 is a small amino acid residue or G; residue 223 is an aliphatic amino acid residue other than L; and residue 233 is a small amino acid residue or G.

14. The recombinant polypeptide of claim 13 in which the amino acid sequence comprises on or more features selected from: residue 49 is R; residue 53 is D; residue 54 is A; residue 60 is A; residue 95 is L; residue 97 is G; residue 109 is R; residue 147 is L; residue 165 is T; residue 197 is G; residue 223 is V; and residue 233 is G.

15. The recombinant polypeptide of claim 14 in which the amino acid sequence has one or more of the features selected from: residue 54 is A; residue 109 is R; residue 147 is L; and residue 233 is G.

16. The recombinant polypeptide of claim 9 in which the amino acid sequence comprises a G, I, C or an aromatic amino acid at the amino acid residue corresponding to residue 96 of SEQ ID NO:2 or SEQ ID NO:4.

17. The recombinant polypeptide of claim 16 in which the amino acid sequence further comprises one or more of the features selected from: residue 94 is an aromatic amino acid residue or G; residue 153 is an aliphatic amino acid residue other than L; residue 199 is any amino acid residue other than L; residue 202 is an aliphatic amino acid residue other than A; and residue 206 is an aromatic amino acid residue.

18. The recombinant polypeptide of claim 17 in which the amino acid sequence comprises one or more of the features selected from: residue 94 is F or G; residue 153 is G or A; residue 199 is K, I, N, R, V, Q, or W; residue 202 is I, L, or G; and residue 206 is F.

19. The recombinant polypeptide of any one of claims 16 to 18 in which the amino acid sequence further comprises one or more of the features selected from: residue 49 is a polar amino acid residue other than K; residue 53 is an acidic amino acid residue; residue 54 is a small or aliphatic amino acid residue other than T/P; residue 60 is an aliphatic amino acid residue other than V; residue 95 is an aliphatic amino acid other than V; residue 97 is a small amino acid or G; residue 109 is a basic amino acid residue other than K; residue 147 is an aliphatic amino acid residue; residue 165 is a hydroxyl or small amino acid residue; residue 197 is a small amino acid residue or G; residue 223 is an aliphatic amino acid residue other than L; and residue 233 is a small amino acid residue or G.

20. The recombinant polypeptide of claim 19 in which the amino acid sequence comprises one or more of the features selected from: residue 49 is R; residue 53 is D; residue 54 is A; residue 60 is A; residue 95 is L; residue 97 is G; residue 109 is R; residue 147 is L; residue 165 is T; residue 197 is G; residue 223 is V; and residue 233 is G.

21. The recombinant polypeptide of claim 20 in which the amino acid sequence has one or more of the features selected from: residue 54 is A; residue 109 is R; residue 147 is L; and residue 233 is G.

22. The recombinant polypeptide of claim 1 which comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 6, 8, 10, 12, 14, 16, 18, 20, 22, 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, and 138.

23. The recombinant polypeptide of any one of claims 1 to 21 in which the recombinant polypeptide is an isolated polypeptide.

24. The recombinant polypeptide of claim 23 in which the isolated polypeptide is a substantially pure polypeptide.

25. A polynucleotide encoding the recombinant polypeptide of any one of claims 1 to 22.

26. The polynucleotide of claim 25 selected from the group consisting of SEQ ID NO: 5, 7, 9, 11, 13, 15, 17, 19, 21, 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, and 137.

27. The polynucleotide of claim 25 in which the polynucleotide hybridizes under high stringency condition to a polynucleotide selected from the group consisting of SEQ ID NO: 5, 7, 9, 11, 13, 15, 17, 19, 21, 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, and 137.

28. An expression vector comprising the polynucleotide of any one of claims 25 and 27 operably linked to control sequences suitable for directing expression in a host cell.

29. The expression vector of claim 28 in which the control sequence is a promoter.

30. The expression vector of claim 29 in which the promoter comprises an *E. coli.* promoter.

31. The expression vector of claim 29 in which the control sequence is a secretion signal.

32. A host cell comprising the expression vector of any one of claim 28 to 31.

33. The host cell of claim 32 which is homologous with the cell type of the wild-type ketoreductase enzyme from which the engineered ketoreductase enzyme was derived.

34. The host cell of claim 32 which is heterologous with the cell type of the wild-type ketoreductase enzyme from which the engineered ketoreductase enzyme was derived.

35. The host cell of claim 32 in which the cell is *E. coli.*

36. The host cell of claim 32 or 35 in which the codons comprising the expression vector have been optimized for expression in said host cell.

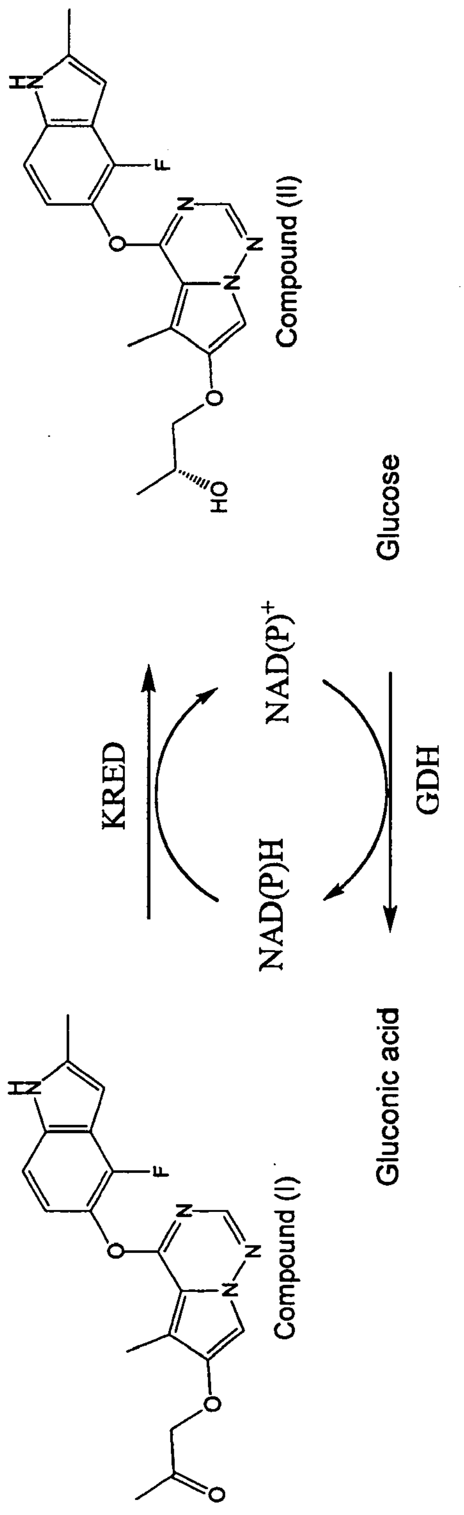

Compound (I)
Compound (II)
KRED
GDH
NAD(P)H
NAD(P)+
Glucose
Gluconic acid
F
HN
O
N
HO

FIG. 1

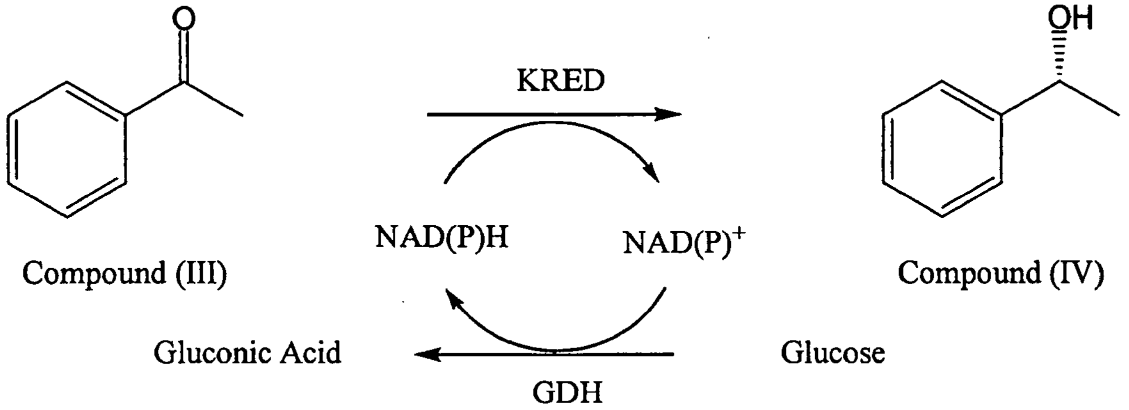

O
Compound (III)
KRED
NAD(P)H
NAD(P)+
GDH
Gluconic Acid
Glucose
OH
Compound (IV)

FIG. 2

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2008/001683

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N9/04 C12N15/53

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, Sequence Search, WPI Data, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2005/017135 A (CODEXIS INC [US]) 24 February 2005 (2005-02-24) the whole document	1-36
A	US 2003/054520 A1 (BOMMANUS BETTINA [US] ET AL) 20 March 2003 (2003-03-20) the whole document	1-36
A	WO 2007/012428 A (IEP GMBH [DE]; TSCHENTSCHER ANKE [DE]; GUPTA ANTJE [DE]; BOBKOVA MARIA) 1 February 2007 (2007-02-01) page 5, paragraph 4	1-36

-/--

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search 16 July 2008	Date of mailing of the international search report 31/07/2008
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Cupido, Marinus

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	-& DATABASE EPO Proteins [Online] 19 April 2007 (2007-04-19), "Sequence 4 from Patent WO2007012428." XP002488479 retrieved from EBI accession no. EPOP:CS539287 Database accession no. CS539287	
T	ANN M. THAYER: "Competitors want to get a piece of Lipitor" CHEMICAL & ENGINEERING NEWS, vol. 84, no. 33, 14 August 2006 (2006-08-14), pages 26-27, XP002488478	1-36

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2008/001683

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2005017135 A	24-02-2005	CA 2533838 A1	24-02-2005
		EP 1654354 A1	10-05-2006
		JP 2007502124 T	08-02-2007
US 2003054520 A1	20-03-2003	DE 10037101 A1	07-02-2002
		EP 1176203 A1	30-01-2002
		JP 2002330781 A	19-11-2002
WO 2007012428 A	01-02-2007	AT 502395 A4	15-03-2007
		AU 2006274252 A1	01-02-2007
		CA 2625834 A1	01-02-2007
		EP 1907556 A1	09-04-2008
		KR 20080036617 A	28-04-2008