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 and aldehyde 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 in whole cells would adversely affect the stereopurity 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; U.S. Patent Nos. 5,559,030; U.S. Patent No. 5,700,670 and U.S. Patent No. 5,891,685), reduction of dioxocarboxylic acids (*e.g.*, U.S. Patent No. 6,399,339), reduction of tert-butyl (S) chloro-5-hydroxy-3-oxohexanoate (*e.g.,* U.S. Patent No. 6,645,746 and WO 01/40450), reduction pyrrolotriazine-based compounds (*e.g.*, U.S. application No. 2006/0286646); reduction of substituted acetophenones (*e.g.*, U.S. Patent No. 6,800,477); and reduction of ketothiolanes (WO 2005/054491).

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

4. SUMMARY

[0007] The present disclosure provides ketoreductase polypeptides capable of reducing 5-((4*S*)-2-oxo-4-phenyl (1,3-oxazolidin-3-yl))-1-(4-fluorophenyl) pentane-1,5-dione ("the substrate") to (4*S*)-3-[(5*S*)-5-(4-fluorophenyl)-5-hydroxypentanoyl]-4-phenyl-1,3-oxazolidin-2-one ("the product"), polynucleotides encoding such polypeptides, and methods for using the polypeptides. The ketoreductase polypeptides are also capable of reducing 1-(4-fluorophenyl)-3(R)-[3-oxo-3-(4-fluorophenyl)propyl)]-4(S)-(4-hydroxyphenyl)-2-azetidinone, to the corresponding stereoisomeric alcohol 1-(4-fluorophenyl)- 3(R)-[3(S)-hydroxy-3(4-fluoropheny l)-propy1)]-4(S)-(4-hydroxypheny1)-2-azetidinone.

[0008] In one embodiment, the ketoreductase polypeptides described herein have an amino acid sequence that has one or more amino acid differences as compared to a reference amino acid sequence of a wild-type ketoreductase or an engineered ketoreductase that result in an improved property of the enzyme for the defined keto substrate. Generally, the engineered ketoreductase polypeptides have an improved property as compared to the naturally-occurring wild-type ketoreductase enzymes obtained

68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, and 126.

[0016] In some embodiments, an improved ketoreductase polypeptide comprises an amino acid sequence that corresponds to the sequence formulas of SEQ ID NO:161, 162 or 163 (or a region thereof, such as residues 90-211). SEQ ID NO:162 is based on the amino acid sequence of the *Lactobacillus kefir* ketoreductase of SEQ ID NO:4. The sequence formula of SEQ ID NO:161 is based on the amino acid sequence of the *Lactobacillus brevis* ketoreductase (SEQ ID NO:2). The sequence formula of SEQ ID NO:163 is based on the amino acid sequence of the *Lactobacillus minor* ketoreductase (SEQ ID NO:158). SEQ ID NO:161, 162 or 163 specify that residue X145 is a polar residue and residue X190 is cysteine. In one aspect, there is provided an engineered polypeptide having ketoreductase activity, wherein the polypeptide comprises the amino acid sequence of SEQ ID NO: 162, wherein the residue corresponding to X145 is an asparagine, glutamine, leucine, phenylalanine, or threonine residue.

[0017] In some embodiments, an improved ketoreductase polypeptide of the disclosure is based on the sequence formulas of SEQ ID NO:161, 162 or 163 and 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 the reference sequence of SEQ ID NO:128, 130, or 160, with the proviso that the ketoreductase amino acid sequence has at the residue corresponding to residue X145 a serine and at the amino acid residue corresponding to X190 a cysteine. In some embodiments, the ketoreductase polypeptides can have one or more amino acid residue differences as compared to SEQ ID NO:128, 130, or 160. These differences can be amino acid insertions, deletions, substitutions, or any combination of such changes. In some embodiments, the amino acid sequence differences can comprise non-conservative, conservative, as well as a combination of non-conservative and conservative amino acid substitutions. Various amino acid residue positions where such changes can be made are described herein.

[0018] In some embodiments, an improved ketoreductase polypeptide is based on the sequence formulas of SEQ ID NO:161, 162 or 163 and 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 90-211 of the reference sequence of SEQ ID NO:128, 130, or 160, with the proviso that the ketoreductase polypeptide amino acid sequence has at the residue corresponding to residue X145 a serine and at the amino acid residue corresponding to X190 a cysteine. In some embodiments, the amino acid sequence differences can comprise non-conservative, conservative, as well as a combination of non-conservative and conservative amino acid substitutions. Various amino acid residue positions where such changes can be made in the defined region are described herein.

[0019] In some embodiments, an improved ketoreductase 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 the amino acid sequence corresponding to SEQ ID NO: 8, 42, 44, 46, 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, and 126., wherein the improved ketoreductase polypeptide amino acid sequence includes any one set of the specified amino acid substitution combinations presented in Tables 3 and 4. In some embodiments, these ketoreductase polypeptides can have mutations at other amino acid residues.

[0020] 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. In some embodiments, the polynucleotides encode a ketoreductase polypeptides having at least the following amino acid sequence as compared to the amino acid sequence of SEQ ID NO:2, 4, or 158: the amino acid residue corresponding to X145 is a serine, and the amino acid residue corresponding to X190 is a cysteine. Exemplary polynucleotides include, but are not limited to, a polynucleotide sequence of SEQ ID NO: 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, and 125.

[0021] The present disclosure provides host cells comprising the polynucleotides and/or expression vectors described herein. The host cells may be *L. kefir* or *L. brevis*, or they may be a different organism, and as *E. coli*. The host cells can be used for the expression and isolation of the engineered ketoreductase enzymes described herein, or, alternatively, they can be used directly for the conversion of the substrate to the stereoisomeric product.

[0022] 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.

[0022a] In another aspect, there is provided a method for reducing a carbonyl group of a substrate compound to its corresponding alcohol group, the method comprising contacting the substrate compound with an engineered polypeptide described herein in the presence of a NADH or NADPH cofactor.

[0023] The ketoreductase enzymes described herein are capable of catalyzing the reduction reaction of the keto group in the compound of structural formula (**I**), 5-((4*S*)-2-oxo-4-phenyl (1,3-oxazolidin-3-yl))-1-(4-fluorophenyl) pentane-1,5-dione ("the substrate"),

WHAT IS CLAIMED IS:

1. An engineered polypeptide having ketoreductase activity, wherein the polypeptide comprises the amino acid sequence of SEQ ID NO: 162, wherein the residue corresponding to X145 is an asparagine, glutamine, leucine, phenylalanine, or threonine residue.

2. The engineered polypeptide of claim 1, wherein the residue corresponding to X190 is a cysteine, or proline residue.

3. The engineered polypeptide of claim 2, wherein the amino acid sequence further includes one or more of the following:
residue corresponding to X3 is aspartic acid, glutamic acid, serine, threonine, asparagine, or glutamine;
residue corresponding to X7 is glycine, methionine, alanine, valine, leucine, isoleucine, serine, threonine, asparagine, or glutamine;
residue corresponding to X17 is glycine, methionine, alanine, valine, leucine, isoleucine, serine, threonine, asparagine, or glutamine;
residue corresponding to X40 is histidine, lysine, arginine, serine, threonine, asparagine, or glutamine;
residue corresponding to X94 is glycine, methionine, alanine, valine, leucine, isoleucine, serine, threonine, asparagine, or glutamine;
residue corresponding to X96 is serine, threonine, asparagine, or glutamine;
residue corresponding to X108 arginine, lysine, serine, threonine, asparagine, glutamine, histidine;
residue corresponding to X117 is glycine, methionine, alanine, valine, leucine, isoleucine, serine, threonine, asparagine, or glutamine;
residue corresponding to X127 is lysine, arginine, serine, threonine, asparagine, or glutamine;
residue corresponding to X147 is glycine, methionine, alanine, valine, leucine, isoleucine, tyrosine, phenylalanine, or tryptophan;
residue corresponding to X152 is glycine, methionine, valine, leucine, or isoleucine, arginine, lysine, serine threonine, asparagine, or glutamine;
residue corresponding to X153 is glycine, threonine, glutamine, or valine;
residue corresponding to X157 is a serine, threonine, asparagine, or glutamine;
residue corresponding to X163 is a glycine, methionine, alanine, valine, leucine, or isoleucine;
residue corresponding to X176 is glycine, methionine, alanine, valine, leucine, or isoleucine;
residue corresponding to X194 is proline, arginine, lysine, serine, threonine, asparagine, or glutamine;
residue corresponding to X196 is leucine;
residue corresponding to X198 is aspartic acid, glutamic acid, arginine, lysine, serine, threonine, asparagine, glutamine, glycine, methionine, alanine, valine, leucine, or isoleucine;
residue corresponding to X199 is an aspartic acid, glutamic acid, glycine, methionine, alanine, valine, leucine, or isoleucine;
residue corresponding to X211 is a arginine or lysine;
residue corresponding to X223 is glycine, methionine, alanine, valine, leucine, or isoleucine;

residue corresponding to X226 is valine; or
residue corresponding to X250 is tryptophan;
wherein the amino acid sequence can optionally have one or more residue differences at other amino acid residues as compared to the reference sequence.

4. The recombinant polypeptide of claim 3, wherein the amino acid sequence includes the following:
residue corresponding to X94 is alanine or glycine;
residue corresponding to X96 is serine or valine;
residue corresponding to X145 is glutamine, leucine, or phenylalanine;
residue corresponding to X147 is leucine or methionine;
residue corresponding to X190 is proline;
residue corresponding to X196 is leucine, or valine; and
residue corresponding to X226 is isoleucine or valine;
wherein the amino acid sequence can optionally have one or more residue differences at other amino acid residues as compared to the reference sequence.

5. The engineered polypeptide of claim 2, wherein the amino acid sequence further includes one or more of the following:
residue corresponding to X3 is asparagine;
residue corresponding to X7 is serine;
residue corresponding to X17 is glutamine;
residue corresponding to X21 is phenylalanine;
residue corresponding to X25 is threonine;
residue corresponding to X29 is glycine or alanine;
residue corresponding to X40 is arginine;
residue corresponding to X42 is glycine;
residue corresponding to X53 is aspartic acid;
residue corresponding to X75 is arginine;
residue corresponding to X94 is asparagine, glycine, or serine;
residue corresponding to X95 is leucine or methionine;
residue corresponding to X96 is glutamine, asparagine, or threonine;
residue corresponding to X101 is glycine or asparagine;
residue corresponding to X105 is glycine;
residue corresponding to X108 is histidine or serine;
residue corresponding to X112 is aspartic acid;
residue corresponding to X113 is alanine;
residue corresponding to X117 is serine;
residue corresponding to X127 is arginine;
residue corresponding to X147 is leucine;
residue corresponding to X152 is methionine or lysine;
residue corresponding to X157 is threonine;
residue corresponding to X163 is isoleucine;
residue corresponding to X176 is valine;

residue corresponding to X194 is arginine or glutamine;
residue corresponding to X197 is valine or glutamic acid;
residue corresponding to X198 is glycine, glutamic acid, or lysine;
residue corresponding to X199 is aspartic acid;
residue corresponding to X200 is proline;
residue corresponding to X202 is glycine;
residue corresponding to X206 is glycine;
residue corresponding to X211 is arginine or lysine;
residue corresponding to X223 is valine; or
residue corresponding to X250 is isoleucine;
wherein the amino acid sequence can optionally have one or more residue differences at other amino acid residues as compared to the reference sequence.

6. The recombinant polypeptide of claim 3, wherein the amino acid sequence includes one or more of the following:
residue corresponding to X7 is serine;
residue corresponding to X108 is histidine or serine;
residue corresponding to X117 is serine;
residue corresponding to X152 is methionine or lysine; or
residue corresponding to X199 is aspartic acid;
wherein the amino acid sequence can optionally have one or more residue differences at other amino acid residues as compared to the reference sequence.

7. A method for reducing a carbonyl group of a substrate compound to its corresponding alcohol group, the method comprising contacting the substrate compound with an engineered polypeptide of claim 1 in the presence of a NADH or NADPH cofactor.