Title: ALCOHOL DEHYDROGENASES (ADH) USEFUL FOR FERMENTIVE PRODUCTION OF LOWER ALKYL ALCOHOLS

Abstract: The invention relates to suitable candidate ADH enzymes for production of lower alkyl alcohols including isobutanol. The invention also relates to recombinant host cells that comprise such ADH enzymes and methods for producing lower alkyl alcohols in the same.
w o 2011/090753 A 2 1 I 1 1 I I I 1 I lllll I ll l l I I lll l ll III I I I I I

Publi h d:
— without international search report and to be republished upon receipt of that report (Rule 48.2(g)) — with sequence listing part of description (Rule 5.2(a))
ALCOHOL DEHYDROGENASES (ADH) USEFUL FOR FERMENTIVE PRODUCTION OF LOWER ALKYL ALCOHOLS

BACKGROUND OF THE INVENTION

Field of the Invention

[0001] The invention relates to the fields of industrial microbiology and alcohol production. Specifically, the invention relates suitable alcohol dehydrogenases for the production of lower alkyl alcohols via an engineered pathway in microorganisms. More specifically, the invention relates to suitable alcohol dehydrogenases for the production of butanol, particularly isobutanol, via an engineered pathway in microorganisms.

Background Art

[0002] Butanol is an important industrial chemical, useful as a fuel additive, as a feedstock chemical in the plastics industry, and as a food grade extractant in the food and flavor industry. Each year 10 to 12 billion pounds of butanol are produced by petrochemical means and the need for this commodity chemical will likely increase in the future.


[0004] Isobutanol is produced biologically as a by-product of yeast fermentation. It is a component of "fusel oil" that forms as a result of the incomplete metabolism of amino acids by this group of fungi. Isobutanol is specifically produced from catabolism of L-valine. After the amine group of L-valine is harvested as a nitrogen source, the resulting a-keto acid is decarboxylated and reduced to isobutanol by enzymes of the so-called Ehrlich pathway (Dickinson et al., J. Biol. Chem. 275:25752-25756, 1998). Yields of fusel oil and/or its components achieved during beverage fermentation are typically low. For example, the concentration of isobutanol produced in beer fermentation is reported to
be less than 16 parts per million (Garcia et al., Process Biochemistry 2P:303-309, 1994). Addition of exogenous L-valine to the fermentation mixture increases the yield of isobutanol, as described by Dickinson et al, *supra*, wherein it is reported that a yield of isobutanol of 3 g/L is obtained by providing L-valine at a concentration of 20 g/L in the fermentation mixture. In addition, production of n-propanol, isobutanol and isoamylalcohol has been shown by calcium alginate immobilized cells of *Zymomonas mobilis*. A 10% glucose-containing medium supplemented with either L-Leu, L-Ile, L-Val, α-ketoisocaproic acid (a-KCA), α-ketobutyric acid (a-KBA) or α-ketoisovaleric acid (α-KVA) was used (Oaxaca, et al., *Acta Biotechnol.* 77:523-532, 1991). α-KCA increased isobutanol levels. The amino acids also gave corresponding alcohols, but to a lesser degree than the keto acids. An increase in the yield of C3-C5 alcohols from carbohydrates was shown when amino acids leucine, isoleucine, and/or valine were added to the growth medium as the nitrogen source (PCT Publ. No. WO 2005/040392).

Whereas the methods described above indicate the potential of isobutanol production via biological means, these methods are cost prohibitive for industrial scale isobutanol production.

For an efficient biosynthetic process, an optimal enzyme is required at the last step to rapidly convert isobutyraldehyde to isobutanol. Furthermore, an accumulation of isobutyraldehyde in the production host normally leads to undesirable cellular toxicity.

Alcohol dehydrogenases (ADHs) are a family of proteins comprising a large group of enzymes that catalyze the interconversion of aldehydes and alcohols (de Smidt et al, *FEMS Yeast Res.*, 5:967-978, 2008), with varying specificities for different alcohols and aldehydes. There is a need to identify suitable ADH enzymes to catalyze the formation of product alcohols in recombinant microorganisms. There is also a need to identify a suitable ADH enzyme that would catalyze the formation of isobutanol at a high rate, with specific affinity for isobutyraldehyde as the substrate and in the presence of high levels of isobutanol.

**BRIEF SUMMARY OF THE INVENTION**

One aspect of the invention is directed to a recombinant microbial host cell comprising a heterologous polynucleotide that encodes a polypeptide wherein the polypeptide has alcohol dehydrogenase activity. In embodiments, the recombinant
microbial host cell further comprises a biosynthetic pathway for the production of a lower alkyl alcohol, wherein the biosynthetic pathway comprises a substrate to product conversion catalyzed by a polypeptide with alcohol dehydrogenase activity. In embodiments, the polypeptide has alcohol dehydrogenase activity and one or more of the following characteristics: (a) the $K_M$ value for a lower alkyl aldehyde is lower for the polypeptide relative to a control polypeptide having the amino acid sequence of SEQ ID NO: 26; (b) the $\frac{3}{4}$ value for a lower alkyl alcohol for the polypeptide is higher relative to a control polypeptide having the amino acid sequence of SEQ ID NO: 26; and (c) the $k_{cat}/K_M$ value for a lower alkyl aldehyde for the polypeptide is higher relative to a control polypeptide having the amino acid sequence of SEQ ID NO: 26. In embodiments, the polypeptide having alcohol dehydrogenase activity has two or more of the above-listed characteristics. In embodiments, the polypeptide preferentially uses NADH as a cofactor. In embodiments, the polypeptide having alcohol dehydrogenase activity has three of the above-listed characteristics. In embodiments, the biosynthetic pathway for production of a lower alkyl alcohol is a butanol, propanol, isopropanol, or ethanol biosynthetic pathway. In one embodiment, the biosynthetic pathway for production of a lower alkyl alcohol is a butanol biosynthetic pathway.

Accordingly, one aspect of the invention is a recombinant microbial host cell comprising: a biosynthetic pathway for production of a lower alkyl alcohol, the biosynthetic pathway comprising a substrate to product conversion catalyzed by a polypeptide with alcohol dehydrogenase activity and one or more, two or more, or all of the following characteristics: (a) the $K_M$ value for isobutyraldehyde is lower for said polypeptide relative to a control polypeptide having the amino acid sequence of SEQ ID NO: 26; (b) the $\frac{3}{4}$ value for isobutanol for said polypeptide is higher relative to a control polypeptide having the amino acid sequence of SEQ ID NO: 26; and (c) the $k_{cat}/K_M$ value isobutyraldehyde for said polypeptide is higher relative to a control polypeptide having the amino acid sequence of SEQ ID NO: 26. In embodiments, the biosynthetic pathway for production of a lower alkyl alcohol is a butanol, propanol, isopropanol, or ethanol biosynthetic pathway. In embodiments, the polypeptide with alcohol dehydrogenase activity has at least 90% identity to the amino acid sequence of SEQ ID NO: 21, 22, 23, 24, 25, 31, 32, 34, 35, 36, 37, or 38. In embodiments, the polypeptide with alcohol dehydrogenase activity has the amino acid sequence of SEQ ID NO: 31. In embodiments,
the polypeptide with alcohol dehydrogenase activity is encoded by a polynucleotide having at least 90% identity to a nucleotide sequence of SEQ ID NO: 1, 2, 3, 4, 5, 6, 11, 12, 14, 15, 16, or 17. In embodiments, polypeptide having alcohol dehydrogenase activity catalyzes the conversion of isobutyraldehyde to isobutanol in the presence of isobutanol at a concentration of at least about 10 g/L, at least about 15 g/L, or at least about 20 g/L.

[0010] In embodiments, the biosynthetic pathway for production of a lower alkyl alcohol is an isobutanol biosynthetic pathway comprising heterologous polynucleotides encoding polypeptides that catalyze substrate to product conversions for each step of the following steps: (a) pyruvate to acetolactate; (b) acetolactate to 2,3-dihydroxyisovalerate; (c) 2,3-dihydroxyisovalerate to α-ketoisovalerate; (d) α-ketoisovalerate to isobutyraldehyde; and (e) isobutyraldehyde to isobutanol; and wherein said microbial host cell produces isobutanol. In embodiments, (a) the polypeptide that catalyzes a substrate to product conversion of pyruvate to acetolactate is acetolactate synthase having the EC number 2.2.1.6; (b) the polypeptide that catalyzes a substrate to product conversion of acetolactate to 2,3-dihydroxyisovalerate is acetoxyhydroxy acid isomeroreductase having the EC number 1.1.186; (c) the polypeptide that catalyzes a substrate to product conversion of 2,3-dihydroxyisovalerate to alpha-ketoisovalerate is acetoxyhydroxy acid dehydratase having the EC number 4.2.1.9; and (d) the polypeptide that catalyzes a substrate to product conversion of alpha-ketoisovalerate to isobutyraldehyde is branched-chain alpha-keto acid decarboxylase having the EC number 4.1.1.72. In embodiments, the biosynthetic pathway for production of a lower alkyl alcohol is an isobutanol biosynthetic pathway comprising heterologous polynucleotides encoding polypeptides that catalyze substrate to product conversions for each step of the following steps: (a) pyruvate to acetolactate; (b) acetolactate to 2,3-dihydroxyisovalerate; (c) 2,3-dihydroxyisovalerate to α-ketoisovalerate; (d) α-ketoisovalerate to isobutyryl-CoA; (e) isobutyryl-CoA to isobutyraldehyde; and (f) isobutyraldehyde to isobutanol; and wherein said microbial host cell produces isobutanol. In embodiments, the biosynthetic pathway for production of a lower alkyl alcohol is an isobutanol biosynthetic pathway comprising heterologous polynucleotides encoding polypeptides that catalyze substrate to product conversions for each step of the following steps: (a) pyruvate to acetolactate; (b) acetolactate to 2,3-dihydroxyisovalerate; (c) 2,3-dihydroxyisovalerate to α-ketoisovalerate; (d) α-
ketoisovalerate to valine; (e) valine to isobutylamine; (e) isobutylamine to isobutyraldehyde; and (f) isobutyraldehyde to isobutanol; and wherein said microbial host cell produces isobutanol.

[0011] Also provided herein are recombinant microbial host cells comprising a biosynthetic pathway for the production of a lower alkyl alcohol and a heterologous polynucleotide encoding a polypeptide with alcohol dehydrogenase activity having at least 85% identity to the amino acid sequence of SEQ ID NO: 21, 22, 23, 24, 25, 31, 32, 34, 35, 36, 37, or 38. In embodiments, the biosynthetic pathway for the production of a lower alkyl alcohol is a 2-butanol biosynthetic pathway comprising heterologous polynucleotides encoding polypeptides that catalyze substrate to product conversions for each of the following steps: (a) pyruvate to alpha-acetolactate; (b) alpha-acetolactate to acetoin; (c) acetoin to 2,3-butanediol; (d) 2,3-butanediol to 2-butanone; and (e) 2-butanone to 2-butanol; and wherein said microbial host cell produces 2-butanol. In embodiments, (a) the polypeptide that catalyzes a substrate to product conversion of pyruvate to acetolactate is acetolactate synthase having the EC number 2.2.1.6; (b) the polypeptide that catalyzes a substrate to product conversion of acetolactate to acetoin is acetolactate decarboxylase having the EC number 4.1.1.5; (c) the polypeptide that catalyzes a substrate to product conversion of acetoin to 2,3-butanediol is butanediol dehydrogenase having the EC number 1.1.1.76 or EC number 1.1.1.4; (d) the polypeptide that catalyzes a substrate to product conversion of butanediol to 2-butanone is butanediol dehydratase having the EC number 4.2.1.28; and (e) the polypeptide that catalyzes a substrate to product conversion of 2-butanone to 2-butanol is 2-butanol dehydrogenase having the EC number 1.1.1.1. In embodiments, the polypeptide having alcohol dehydrogenase activity comprises an amino acid sequence with at least 95% identity to the amino acid sequence of SEQ ID NO: 21, 22, 23, 24, 25, 27, 31, 32, 34, 35, 36, 37, or 38. In embodiments, the polypeptide having alcohol dehydrogenase activity comprises an amino acid sequence with at least 95% identity to the amino acid sequence of SEQ ID NO: 31.

[0012] In embodiments, the biosynthetic pathway for the production of a lower alkyl alcohol is a 1-butanol biosynthetic pathway comprises heterologous polynucleotides encoding polypeptides that catalyze substrate to product conversions for each of the following steps: (a) acetyl-CoA to acetoacetyl-CoA; (b) acetoacetyl-CoA to 3-
hydroxybutyryl-CoA; (c) 3-hydroxybutyryl-CoA to crotonyl-CoA; (d) crotonyl-CoA to butyryl-CoA; (e) butyryl-CoA to butyraldehyde; and (f) butyraldehyde to 1-butanol; and wherein said microbial host cell produces 1-butanol. In embodiments, (a) the polypeptide that catalyzes a substrate to product conversion of acetyl-CoA to acetoacetyl-CoA is acetyl-CoA acetyltransferase having the EC number 2.3.1.9 or 2.3.1.16; (b) the polypeptide that catalyzes a substrate to product conversion of acetoacetyl-CoA to 3-hydroxybutyryl-CoA is 3-hydroxybutyryl-CoA dehydrogenase having the EC number 1.1.1.35, 1.1.1.30, 1.1.1.157, or 1.1.1.36; (c) the polypeptide that catalyzes a substrate to product conversion of 3-hydroxybutyryl-CoA to crotonyl-CoA is crotonase having the EC number 4.2.1.17 or 4.2.1.55; (d) the polypeptide that catalyzes a substrate to product conversion of crotonyl-CoA to butyryl-CoA is butyryl-CoA dehydrogenase having the EC number 1.3.1.44 or 1.3.1.38; (e) the polypeptide that catalyzes a substrate to product conversion of butyryl-CoA to butyrylaldehyde is butyraldehyde dehydrogenase having the EC number 1.2.1.57; and (f) the polypeptide that catalyzes a substrate to product conversion of butyrylaldehyde to 1-butanol is 1-butanol dehydrogenase. In embodiments, the polypeptide having alcohol dehydrogenase activity comprises an amino acid sequence with at least 95% identity to the amino acid sequence of SEQ ID NO: 21, 22, 23, 24, 25, 27, 31, 32, 34, 35, 36, 37, or 38. In embodiments, the polypeptide having alcohol dehydrogenase activity comprises an amino acid sequence with at least 95% identity to the amino acid sequence of SEQ ID NO: 31.

[0013] In embodiments, the recombinant microbial host cell is selected from the group consisting of: bacteria, cyanobacteria, filamentous fungi and yeasts. In embodiments, the host cell is a bacterial or cyanobacterial cell. In embodiments, the genus of the host cells is selected from the group consisting of: Salmonella, Arthrobacter, Bacillus, Brevibacterium, Clostridium, Corymb acterium, Gluconobacter, Nocardia, Pseudomonas, Rhodococcus, Streptomyces, Zymomonas, Escherichia, Lactobacillus, Enterococcus, Alcaligenes, Klebsiella, Serratia, Shigella, Alcaligenes, Erwinia, Paenibacillus, and Xanthomonas. In embodiments, the genus of the host cells provided herein is selected from the group consisting of: Saccharomyces, Pichia, Hansenula, Yarrowia, Aspergillus, Kluyveromyces, Pachysolen, Rhodotorula, Zygosaccharomyces, Galactomyces, Schizosaccharomyces, Torulaspora, Debayomyces, Williopsis, Dekkera, Kloeckera, Metschnikowia, Issatchenkia, and Candida.
Another aspect of the present invention is a method for producing isobutanol comprising: (a) providing a recombinant microbial host cell comprising an isobutanol biosynthetic pathway, the pathway comprising a heterologous polypeptide which catalyzes the substrate to product conversion of isobutyraldehyde to isobutanol wherein the polypeptide has at least 90% identity to the amino acid sequence of SEQ ID NO: 21, 22, 23, 24, 25, 27, 31, 32, 34, 35, 36, 37, or 38; and (b) contacting the host cell of (a) with a carbon substrate under conditions whereby isobutanol is produced. In embodiments, the heterologous polypeptide which catalyzes the substrate to product conversion of isobutyraldehyde to isobutanol has at least 90% identity to the amino acid sequence of SEQ ID NO: 31. Another aspect is a method for producing 2-butanol comprising: (a) providing a recombinant microbial host cell comprising a 2-butanol biosynthetic pathway, the pathway comprising a heterologous polypeptide having at least 90% identity to the amino acid sequence of SEQ ID NO: 21, 22, 23, 24, 25, 27, 31, 32, 34, 35, 36, 37, or 38; and (b) contacting the host cell of (a) with a carbon substrate under conditions whereby 2-butanol is produced. In embodiments, the heterologous polypeptide has at least 90% identity to the amino acid sequence of SEQ ID NO: 31. Another aspect is a method for producing 1-butanol comprising: (a) providing a recombinant microbial host cell comprising a 1-butanol biosynthetic pathway, the pathway comprising a heterologous polypeptide having at least 90% identity to the amino acid sequence of SEQ ID NO: 21, 22, 23, 24, 25, 27, 31, 32, 34, 35, 36, 37, or 38; and (b) contacting the host cell of (a) with a carbon substrate under conditions whereby 1-butanol is produced. In embodiments, the heterologous polypeptide has at least 90% identity to the amino acid sequence of SEQ ID NO: 31.

Also provided herein are methods for the production of a lower alkyl alcohol comprising: (a) providing a recombinant host cell provided herein; (b) contacting said host cell with a fermentable carbon substrate in a fermentation medium under conditions whereby the lower alkyl alcohol is produced; and (c) recovering said lower alkyl alcohol. In embodiments, said fermentable carbon substrate is selected from the group consisting of: monosaccharides, oligosaccharides, and polysaccharides. In embodiments, monosaccharide is selected from the group consisting: glucose, galactose, mannose, rhamnose, xylose, and fructose. In embodiments, said oligosaccharide is selected from the group consisting of: sucrose, maltose, and lactose. In embodiments, polysaccharide is
selected from the group consisting of: starch, cellulose, and maltodextrin. In embodiments, the conditions are anaerobic, aerobic, or microaerobic. In embodiments, said lower alkyl alcohol is produced at a titer of at least about 10 g/L, at least about 15 g/L, or at least about 20 g/L. In embodiments, said lower alkyl alcohol is selected from the group consisting of: butanol, isobutanol, propanol, isopropanol, and ethanol.

[0016] In embodiments, isobutanol is produced. In embodiments, the method for producing isobutanol comprises: (a) providing a recombinant host cell comprising a heterologous polypeptide which catalyzes the substrate to product conversion of isobutyraldehyde to isobutanol and which has one or more of the following characteristics: (i) the $K_M$ value of a lower alkyl aldehyde is lower for the polypeptide relative to a control polypeptide having the amino acid sequence of SEQ ID NO: 26; (ii) the $K_i$ value for a lower alkyl aldehyde for the polypeptide is higher relative to control polypeptide having the amino acid sequence of SEQ ID NO: 26; (iii) the $k_{cat}/K_M$ value for a lower alkyl aldehyde for the polypeptide is higher relative to a control polypeptide having the amino acid sequence of SEQ ID NO: 26; and (b) contacting the host cell of (a) with a carbon substrate under conditions whereby isobutanol is produced.

[0017] In embodiments, 1-butanol is produced. In embodiments, the method for producing 1-butanol comprises: (a) providing a recombinant microbial host cell comprising a heterologous polypeptide which catalyzes the substrate to product conversion of butyraldehyde to 1-butanol and which has one or more of the following characteristics: (i) the $K_M$ value for a lower alkyl aldehyde is lower for the polypeptide relative to a control polypeptide having the amino acid sequence of SEQ ID NO: 26; (ii) the $\frac{1}{3}$ value for a lower alkyl alcohol for the polypeptide is higher relative to a control polypeptide having the amino acid sequence of SEQ ID NO: 26; and (iii) the $k_{cat}/K_M$ value for a lower alkyl aldehyde for the polypeptide is higher relative to a control polypeptide having the amino acid sequence of SEQ ID NO: 26; and (b) contacting the host cell of (a) with a carbon substrate under conditions whereby 1-butanol is produced.

[0018] Also provided herein are methods for screening candidate polypeptides having alcohol dehydrogenase activity, said method comprising: a) providing a candidate polypeptide and a cofactor selected from the group consisting of NADH and NADPH; b) monitoring a change in $A_{340nm}$ over time in the presence or absence of a lower alkyl alcohol for the candidate polypeptide; and c) selecting those candidate polypeptides
where the change in $A_{340 \text{ nm}}$ is a decrease, and the decrease is faster in the absence of the lower alkyl alcohol with respect to the decrease in the presence of the lower alkyl alcohol. In embodiments, the methods further comprise (d) providing a control polypeptide having the amino acid sequence of either SEQ ID NO: 21 or 26 and NADH; (e) monitoring a change in $A_{340 \text{ nm}}$ over time in the presence or absence of a lower alkyl alcohol for the control polypeptide; (f) comparing the changes observed in (e) with the changes observed in (b); and (g) selecting those candidate polypeptides where the decrease in $A_{340 \text{ nm}}$ in the absence of the lower alkyl alcohol is faster than the decrease observed for the control polypeptide. In embodiments, the methods further comprise (d) providing a control polypeptide having the amino acid sequence of either SEQ ID NO: 21 or 26 and NADH; (e) monitoring a change in $A_{340 \text{ nm}}$ over time in the presence or absence of a lower alkyl alcohol for the control polypeptide; (f) comparing the changes observed in (e) with the changes observed in (b); and (g) selecting those candidate polypeptides where the decrease in $A_{340 \text{ nm}}$ in the presence of the lower alkyl alcohol is faster than the decrease observed for the control polypeptide.

Also provided herein is use of an alcohol dehydrogenase having at least about 80% identity to an amino acid sequence of SEQ ID NO: 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 in a microbial host cell to catalyze the conversion of isobutyraldehyde to isobutanol; wherein said host cell comprises an isobutanol biosynthetic pathway.

**BRIEF DESCRIPTION OF THE DRAWINGS/FIGURES AND SEQUENCES**

Figure 1 shows the results of semi-physiological time-course assays showing isobutyraldehyde reduction by NAD(P)H, catalyzed by ADH candidate enzymes in the presence and absence of isobutanol. Enzymatic activity is measured by following changes in absorbance at 340 nm. In each panel, $A_{340 \text{ nm}}$ of NADH or NADPH alone, in the presence of all other reactants except the enzyme, was used as a control. Panel A shows the change in absorbance at 340 nm over time for *Achromobacter xylosoxidans* SadB. Panel B shows the change in absorbance at 340 nm over time for horse liver ADH. Panel C shows the change in absorbance at 340 nm over time for *Saccharomyces cerevisiae* ADH6. Panel D shows the change in absorbance at 340 nm over time for *Saccharomyces cerevisiae* ADH7. Panel E shows the change in absorbance at 340 nm
over time for *Beijierickia indica* ADH. Panel F shows the change in absorbance at 340 nm over time for *Clostridium beijerinckii* ADH. Panel G shows the change in absorbance at 340 nm over time for *Rattus norvegicus* ADH. Panel H shows the change in absorbance at 340 nm over time for *Therm. sp. ATN1* ADH.

Figure 2 shows the results of semi-physiological time-course assays comparing the level of isobutanol inhibition observed with horse liver ADH and *Achromobacter xylosoxidans* SadB in the same figure. The assays are as described for Figure 1.

Figure 3 is an alignment of the polypeptide sequences of *Pseudomonas putida* formaldehyde dehydrogenase (IkoLA) (SEQ ID NO: 79), horse liver ADH (2ohxA) (SEQ ID NO: 21), *Clostridium beijerinckii* ADH (IpedA) (SEQ ID NO: 29), *Pyrococcus horikoshii* L-theronine 3-dehydrogenase (2d8aA) (SEQ ID NO: 80), and *Achromobacter xylosoxidans* SadB (SEQ ID NO: 26).

Figure 4 is a phylogenetic tree of oxidoreductase enzymes obtained as hits from (i) a protein BLAST search for similar sequences in *Saccharomyces cerevisiae*, *E. coli*, *Homo sapiens*, *C. elegans*, *Drosophila melanogaster*, and *Arabidopsis thaliana*, and (ii) a protein BLAST search of Protein Data Bank (PDB) for similar sequences using horse liver ADH and *Achromobacter xylosoxidans* SadB as queries.

Figure 5 is a phylogenetic tree of oxidoreductase enzyme sequences more closely related in sequence to *Achromobacter xylosoxidans* SadB among hits from a protein BLAST search of nonredundant protein sequence database (nr) at NCBI using *Achromobacter xylosoxidans* SadB as query.

Figure 6 is an illustration of example pyruvate to isobutanol biosynthetic pathways.

Figure 7 shows the Michaelis-Menten plots describing the properties of the enzymes pertaining to isobutyraldehyde reduction. Figure 7A shows results of assays to determine the $\frac{3}{4}$ for isobutanol for ADH6 and Figure 7B shows results of assays to determine the $\frac{3}{4}$ for isobutanol for BiADH.

Figure 8A shows the results of semi-physiological time-course assays, which were as described for Figure 1. Panel A shows the change in absorbance at 340 nm over time for the ADH from *Phenylobacterium zucineum*. Panel B shows the change in absorbance at 340 nm over time for *Methylocella silvestris* BL2. Panel C shows the change in absorbance at 340 nm over time for *Acinetobacter baumannii* AYE.
Figure 9 depicts the pdcl::ilvD::FBA-alsS::trxl A locus. The alsS gene integration in the pdcl-trxl intergenic region is considered a "scarless" insertion since vector, marker gene and loxP sequences are lost.

The following sequences provided in the accompanying sequence listing, filed electronically herewith and incorporated herein by reference, conform with 37 C.F.R. 1.821-1.825 ("Requirements for Patent Applications Containing Nucleotide Sequences and/or Amino Acid Sequence Disclosures - the Sequence Rules") and are consistent with World Intellectual Property Organization (WIPO) Standard ST.25 (2009) and the sequence listing requirements of the EPO and PCT (Rules 5.2 and 49.5(a-bis), and Section 208 and Annex C of the Administrative Instructions). The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

SEQ ID NOs: 1 and 7-20 are codon-optimized polynucleotide sequences.

SEQ ID NOs: 2 and 3 are polynucleotide sequences from Saccharomyces cerevisiae.

SEQ ID NOs: 4 and 5 are polynucleotide sequences from Clostridium acetobutylicum.

SEQ ID NO: 6 is a polynucleotide sequence from Achromobacter xylosoxidans.

SEQ ID NOs: 21-40 and 79-80 are polypeptide sequences.

SEQ ID NOs: 41-50 and 52-57 and 59-74 and 77-78 are primers.

SEQ ID NO: 51 is the sequence of the pRS423::TEF(M4)-xpkl+EN01-eutD plasmid.

SEQ ID NO: 58 is the sequence of the pUC19-URA3::pdcl ::TEF(M4)-xpkl ::kan plasmid.

SEQ ID NO: 75 is the sequence of the pLH468 plasmid.

SEQ ID NO: 76 is the BiADH coding region (codon optimized for yeast) plus 5'homology to GPM promoter and 3'homology to ADH1 terminator.

SEQ ID NO: 81 is the sequence of the pRS426::GPD-xpkl+ADH-eutD plasmid.

**DETAILED DESCRIPTION OF THE INVENTION**

The stated problems are solved as described herein by devising and using a suitable screening strategy for evaluating various candidate ADH enzymes. The
screening strategy can be used to identify ADH enzymes having desirable characteristics. These identified ADH enzymes can be used to enhance the biological production of lower alkyl alcohols, such as isobutanol. Also provided are recombinant host cells that express the identified desirable ADH enzymes and provided methods for producing lower alkyl alcohols using the same.

[0042] The present invention describes a method for screening large numbers of alcohol dehydrogenase (ADH) enzymes for their ability to rapidly convert isobutyraldehyde to isobutanol in the presence of high concentrations of isobutanol. Also described in the present invention is a new ADH that is present in the bacterium Beijerinckia indica subspecies indica ATCC 9039. The Beijerinckia indica ADH enzyme can be used in the production of isobutanol from isobutyraldehyde in a recombinant microorganism having an isobutyraldehyde source.

[0043] The present invention meets a number of commercial and industrial needs. Butanol is an important industrial commodity chemical with a variety of applications, where its potential as a fuel or fuel additive is particularly significant. Although only a four-carbon alcohol, butanol has an energy content similar to that of gasoline and can be blended with any fossil fuel. Butanol is favored as a fuel or fuel additive as it yields only C0_{2} and little or no SO_{2} or NO_{2} when burned in the standard internal combustion engine. Additionally butanol is less corrosive than ethanol, the most preferred fuel additive to date.

[0044] In addition to its utility as a biofuel or fuel additive, butanol has the potential of impacting hydrogen distribution problems in the emerging fuel cell industry. Fuel cells today are plagued by safety concerns associated with hydrogen transport and distribution. Butanol can be easily reformed for its hydrogen content and can be distributed through existing gas stations in the purity required for either fuel cells or vehicles.

[0045] The present invention produces butanol from plant derived carbon sources, avoiding the negative environmental impact associated with standard petrochemical processes for butanol production. In one embodiment, the present invention provides a method for the selection and identification of ADH enzymes that increase the flux in the last reaction of the isobutanol biosynthesis pathway; the conversion of isobutyraldehyde to isobutanol. In one embodiment, the present invention provides a method for the selection and identification of ADH enzymes that increase the flux in the last reaction of
the 1-butanol biosynthesis pathway; the conversion of butyrylaldehyde to 1-butanol. In one embodiment, the present invention provides a method for the selection and identification of ADH enzymes that increase the flux in the last reaction of the 2-butanol biosynthesis pathway; the conversion of 2-butanone to 2-butanol. Particularly useful ADH enzymes are those that are better able to increase the flux in the isobutyraldehyde to isobutanol conversion reaction when compared to known control ADH enzymes. The present invention also provides for recombinant host cells expressing such identified ADH enzymes and methods for using the same.

[0046] The following definitions and abbreviations are to be used for the interpretation of the claims and the specification.

[0047] The term "invention" or "present invention" as used herein is meant to apply generally to all embodiments of the invention as described in the claims as presented or as later amended and supplemented, or in the specification.

[0048] The term "isobutanol biosynthetic pathway" refers to the enzymatic pathway to produce isobutanol from pyruvate.

[0049] The term "1-butanol biosynthetic pathway" refers to the enzymatic pathway to produce 1-butanol from pyruvate.

[0050] The term "2-butanol biosynthetic pathway" refers to the enzymatic pathway to produce 2-butanol from acetyl-CoA.

[0051] The term "NADH consumption assay" refers to an enzyme assay for the determination of the specific activity of the alcohol dehydrogenase enzyme, which is measured as a stoichiometric disappearance of NADH, a cofactor for the enzyme reaction, as described in Racker, *J Biol. Chem.*, 154:313-319 (1950).

[0052] "ADH" is the abbreviation for the enzyme alcohol dehydrogenase.

[0053] The terms "isobutyraldehyde dehydrogenase," "secondary alcohol dehydrogenase," "butanol dehydrogenase," "branched-chain alcohol dehydrogenase," and "alcohol dehydrogenase" will be used interchangeably and refer the enzyme having the EC number, EC 1.1.1.1 (Enzyme Nomenclature 1992, Academic Press, San Diego). Preferred branched-chain alcohol dehydrogenases are known by the EC number 1.1.1.265, but may also be classified under other alcohol dehydrogenases (specifically, EC 1.1.1.1 or 1.1.1.2). These enzymes utilize NADH (reduced nicotinamide adenine dinucleotide) and/or NADPH as an electron donor.
As used herein, "heterologous" refers to a polynucleotide, gene or polypeptide not normally found in the host organism but that is introduced or is otherwise modified. "Heterologous polynucleotide" includes a native coding region from the host organism, or portion thereof, that is reintroduced or otherwise modified in the host organism in a form that is different from the corresponding native polynucleotide as well as a coding region from a different organism, or portion thereof. "Heterologous gene" includes a native coding region, or portion thereof, that is reintroduced or is otherwise modified from the source organism in a form that is different from the corresponding native gene as well as a coding region from a different organism. For example, a heterologous gene may include a native coding region that is a portion of a chimeric gene including non-native regulatory regions that is reintroduced into the native host. "Heterologous polypeptide" includes a native polypeptide that is reintroduced or otherwise modified in the host organism in a form that is different from the corresponding native polypeptide as well as a polypeptide from another organism.

The term "carbon substrate" or "fermentable carbon substrate" refers to a carbon source capable of being metabolized by host organisms of the present invention. Non-limited examples of carbon sources that can be used in the invention include monosaccharides, oligosaccharides, polysaccharides, and one-carbon substrates or mixtures thereof.

The terms "k_{cat}" and "K_M" and "K_i" are known to those skilled in the art and are described in Enzyme Structure and Mechanism, 2nd ed. (Ferst, W.H. Freeman: NY, 1985; pp 98-120). The term "k_{cat}" often called the "turnover number," is defined as the maximum number of substrate molecules converted to product molecules per active site per unit time, or the number of times the enzyme turns over per unit time. k_{cat} = V_{max} / [E], where [E] is the enzyme concentration (Ferst, supra).

The term "catalytic efficiency" is defined as the k_{cat}/K_M of an enzyme. "Catalytic efficiency" is used to quantitate the specificity of an enzyme for a substrate.

The term "specific activity" means enzyme units/mg protein where an enzyme unit is defined as moles of product formed/minute under specified conditions of temperature, pH, [S], etc.
The terms "slow," "slower," "faster," or "fast" when used in reference to an enzyme activity relates to the turnover number of the enzyme as compared with a standard.

The term "control polypeptide" refers to a known polypeptide having known alcohol dehydrogenase activity. Non-limiting examples of control polypeptides suitable for use in the invention include Achromobacter xylosidans SadB and horse liver ADH.

The term "NAD(P)H" is used to refer to either NADH or NADPH.

The term "lower alkyl alcohol" refers to any straight-chain or branched, saturated or unsaturated, alcohol molecule with 1-10 carbon atoms.

The term "lower alkyl aldehyde" refers to any straight-chain or branched, saturated or unsaturated, aldehyde molecule with 1-10 carbon atoms.

The term "butanol" as used herein refers to 1-butanol, 2-butanol, isobutanol, or mixtures thereof.

The term "biosynthetic pathway for production of a lower alkyl alcohol" as used herein refers to an enzyme pathway to produce lower alkyl alcohols. For example, isobutanol biosynthetic pathways are disclosed in U.S. Patent Application Publication No. 2007/0092957, which is incorporated by reference herein.

As used herein, the term "yield" refers to the amount of product per amount of carbon source in g/g. The yield may be exemplified for glucose as the carbon source. It is understood unless otherwise noted that yield is expressed as a percentage of the theoretical yield. In reference to a microorganism or metabolic pathway, "theoretical yield" is defined as the maximum amount of product that can be generated per total amount of substrate as dictated by the stoichiometry of the metabolic pathway used to make the product. For example, the theoretical yield for one typical conversion of glucose to isopropanol is 0.33 µg. As such, a yield of isopropanol from glucose of 29.7 µg would be expressed as 90% of theoretical or 90% theoretical yield. It is understood that while in the present disclosure the yield is exemplified for glucose as a carbon source, the invention can be applied to other carbon sources and the yield may vary depending on the carbon source used. One skilled in the art can calculate yields on various carbon sources. The term "NADH" means reduced nicotinamide adenine dinucleotide.

The term "NADPH" means reduced nicotinamide adenine dinucleotide phosphate.

The term "NAD(P)H" is used to refer to either NADH or NADPH.
Polypeptides and Polynucleotides for Use in the Invention

[0068] The ADH enzymes used in the invention comprise polypeptides and fragments thereof. As used herein, term "polypeptide" is intended to encompass a singular "polypeptide" as well as plural "polypeptides," and refers to a molecule composed of monomers (amino acids) linearly linked by amide bonds (also known as peptide bonds). The term "polypeptide" refers to any chain or chains of two or more amino acids, and does not refer to a specific length of the product. Thus, peptides, dipeptides, tripeptides, oligopeptides, "protein," "amino acid chain," or any other term used to refer to a chain or chains of two or more amino acids, are included within the definition of "polypeptide," and the term "polypeptide" may be used instead of, or interchangeably with any of these terms. The term "polypeptide" is also intended to refer to the products of post-expression modifications of the polypeptide, including without limitation glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, or modification by non-naturally occurring amino acids.

[0069] A polypeptide of the invention may be of a size of about 10 or more, 20 or more, 25 or more, 50 or more, 75 or more, 100 or more, 200 or more, 500 or more, 1,000 or more, or 2,000 or more amino acids. Polypeptides may have a defined three-dimensional structure, although they do not necessarily have such structure. Polypeptides with a defined three-dimensional structure are referred to as folded, and polypeptides which do not possess a defined three-dimensional structure, but rather can adopt a large number of different conformations, and are referred to as unfolded.

[0070] Also included as polypeptides of the present invention are derivatives, analogs, or variants of the foregoing polypeptides, and any combination thereof. The terms "active variant," "active fragment," "active derivative," and "analog" refer to polypeptides of the present invention and include any polypeptides that are capable of catalyzing the reduction of a lower alkyl aldehyde. Variants of polypeptides of the present invention include polypeptides with altered amino acid sequences due to amino acid substitutions, deletions, and/or insertions. Variants may occur naturally or be non-naturally occurring. Non-naturally occurring variants may be produced using art-known mutagenesis techniques. Variant polypeptides may comprise conservative or non-conservative amino acid substitutions, deletions and/or additions. Derivatives of polypeptides of the present
invention, are polypeptides which have been altered so as to exhibit additional features not found on the native polypeptide. Examples include fusion proteins. Variant polypeptides may also be referred to herein as "polypeptide analogs." As used herein a "derivative" of a polypeptide refers to a subject polypeptide having one or more residues chemically derivatized by reaction of a functional side group. Also included as "derivatives" are those peptides which contain one or more naturally occurring amino acid derivatives of the twenty standard amino acids. For example, 4-hydroxyproline may be substituted for proline; 5-hydroxylysine may be substituted for lysine; 3-methylhistidine may be substituted for histidine; homoserine may be substituted for serine; and ornithine may be substituted for lysine.

[0071] A "fragment" is a unique portion of an ADH enzyme which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous amino acid residues. A fragment may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 100 or 200 amino acids of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

[0072] Alternatively, recombinant variants encoding these same or similar polypeptides can be synthesized or selected by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce various restriction sites, may be introduced to optimize cloning into a plasmid or viral vector or expression in a host cell system. Mutations in the polynucleotide sequence may be reflected in the polypeptide or domains of other peptides added to the polypeptide to modify the properties of any part of the polypeptide, to change characteristics such as the \(K_M\) for a lower alkyl aldehyde, the \(K_M\) for a lower alkyl alcohol, the \(\frac{3}{4}\) for a lower alkyl alcohol, etc.

[0073] Preferably, amino acid "substitutions" are the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, \(i.e.,\)
conservative amino acid replacements. "Conservative" amino acid substitutions may be
made on the basis of similarity in polarity, charge, solubility, hydrophobicity,
hydrophilicity, and/or the amphipathic nature of the residues involved. For example,
nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline,
phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine,
serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic)
amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino
acids include aspartic acid and glutamic acid. "Insertions" or "deletions" are preferably in
the range of about 1 to about 20 amino acids, more preferably 1 to 10 amino acids. The
variation allowed may be experimentally determined by systematically making insertions,
deletions, or substitutions of amino acids in a polypeptide molecule using recombinant
DNA techniques and assaying the resulting recombinant variants for activity.

By a polypeptide having an amino acid or polypeptide sequence at least, for
example, 95% "identical" to a query amino acid sequence of the present invention, it is
intended that the amino acid sequence of the subject polypeptide is identical to the query
sequence except that the subject polypeptide sequence may include up to five amino acid
alterations per each 100 amino acids of the query amino acid sequence. In other words, to
obtain a polypeptide having an amino acid sequence at least 95% identical to a query
amino acid sequence, up to 5% of the amino acid residues in the subject sequence may be
inserted, deleted, or substituted with another amino acid. These alterations of the
reference sequence may occur at the amino or carboxy terminal positions of the reference
amino acid sequence or anywhere between those terminal positions, interspersed either
individually among residues in the reference sequence or in one or more contiguous
groups within the references sequence.

As a practical matter, whether any particular polypeptide is at least 80%, 85%,
90%, 95%, 96%, 97%, 98%, or 99% identical to a reference polypeptide can be
determined conventionally using known computer programs. A preferred method for
determining the best overall match between a query sequence (a sequence of the present
invention) and a subject sequence, also referred to as a global sequence alignment, can be
determined using the FASTDB computer program based on the algorithm of Brutlag et
al., Comp. Appl. Biosci. 5:237-245 (1990). In a sequence alignment the query and subject
sequences are either both nucleotide sequences or both amino acid sequences. The result
of said global sequence alignment is in percent identity. Preferred parameters used in a
FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1,
Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window
Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the
length of the subject amino acid sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence due to N- or C-terminal
deletions, not because of internal deletions, a manual correction must be made to the
results. This is because the FASTDB program does not account for N- and C-terminal
truncations of the subject sequence when calculating global percent identity. For subject
sequences truncated at the N- and C-termini, relative to the query sequence, the percent
identity is corrected by calculating the number of residues of the query sequence that are
N- and C-terminal of the subject sequence, which are not matched/aligned with a
 corresponding subject residue, as a percent of the total bases of the query sequence.
Whether a residue is matched/aligned is determined by results of the FASTDB sequence
alignment. This percentage is then subtracted from the percent identity, calculated by the
above FASTDB program using the specified parameters, to arrive at a final percent
identity score. This final percent identity score is what is used for the purposes of the
present invention. Only residues to the N- and C-termini of the subject sequence, which
are not matched/aligned with the query sequence, are considered for the purposes of
manually adjusting the percent identity score. That is, only query residue positions
outside the farthest N- and C-terminal residues of the subject sequence.

For example, a 90 amino acid residue subject sequence is aligned with a 100
residue query sequence to determine percent identity. The deletion occurs at the N-
terminus of the subject sequence and therefore, the FASTDB alignment does not show a
matching/alignment of the first 10 residues at the N-terminus. The 10 unpaired residues
represent 10% of the sequence (number of residues at the N- and C-termini not
matched/total number of residues in the query sequence) so 10% is subtracted from the
percent identity score calculated by the FASTDB program. If the remaining 90 residues
were perfectly matched the final percent identity would be 90%. In another example, a
90 residue subject sequence is compared with a 100 residue query sequence. This time
the deletions are internal deletions so there are no residues at the N- or C-termini of the
subject sequence which are not matched/aligned with the query. In this case, the percent
identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to be made for the purposes of the present invention.

Polypeptides useful in the invention include those that are at least about 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the sequences set forth in Table 5, including active variants, fragments, or derivatives thereof. The invention also encompasses polypeptides comprising amino acid sequences of Table 5 with conservative amino acid substitutions.

In one embodiment of the invention, polypeptides having alcohol dehydrogenase activity to be expressed in the recombinant host cells of the invention have amino acid sequences that are at least about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, and SEQ ID NO: 40. In another embodiment of the invention, a polypeptide having alcohol dehydrogenase activity to be expressed in the recombinant host cells of the invention has an amino acid sequence selected from the group consisting of: SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, and SEQ ID NO: 40, or an active variant, fragment or derivative thereof. In one embodiment, polypeptides having alcohol dehydrogenase activity are encoded by polynucleotides that have been codon-optimized for expression in a specific host cell.

In one embodiment of the invention, polypeptides having alcohol dehydrogenase activity to be expressed in the recombinant host cells of the invention comprise a amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to SEQ ID
In another embodiment, the polypeptide comprises the amino acid sequence of SEQ ID NO: 22 or an active variant, fragment or derivative thereof.

In one embodiment of the invention, polypeptides having alcohol dehydrogenase activity to be expressed in the recombinant host cells of the invention comprise a amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to SEQ ID NO: 23. In another embodiment, the polypeptide comprises the amino acid sequence of SEQ ID NO: 23 or an active variant, fragment or derivative thereof.

In one embodiment of the invention, polypeptides having alcohol dehydrogenase activity to be expressed in the recombinant host cells of the invention comprise a amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to SEQ ID NO: 31. In another embodiment, the polypeptide comprises the amino acid sequence of SEQ ID NO: 31 or an active variant, fragment or derivative thereof.

In one embodiment of the invention, polypeptides having alcohol dehydrogenase activity to be expressed in the recombinant host cells of the invention comprise a amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to SEQ ID NO: 29. In another embodiment, the polypeptide comprises the amino acid sequence of SEQ ID NO: 29 or an active variant, fragment or derivative thereof.

ADH enzymes suitable for use in the present invention and fragments thereof are can be encoded by polynucleotides. The term "polynucleotide" is intended to encompass a singular nucleic acid as well as plural nucleic acids, and refers to an isolated nucleic acid molecule or construct, e.g., messenger RNA (mRNA), virally-derived RNA, or plasmid DNA (pDNA). A polynucleotide may comprise a conventional phosphodiester bond or a non-conventional bond (e.g., an amide bond, such as found in peptide nucleic acids (PNA)). The term "nucleic acid" refers to any one or more nucleic acid segments, e.g., DNA or RNA fragments, present in a polynucleotide. Polynucleotides according to the present invention further include such molecules produced synthetically. Polynucleotides of the invention may be native to the host cell or heterologous. In addition, a polynucleotide or a nucleic acid may be or may include a regulatory element such as a promoter, ribosome binding site, or a transcription terminator.
As used herein, a "coding region" or "ORF" is a portion of nucleic acid which consists of codons translated into amino acids. Although a "stop codon" (TAG, TGA, or TAA) is not translated into an amino acid, it may be considered to be part of a coding region, if present, but any flanking sequences, for example promoters, ribosome binding sites, transcriptional terminators, introns, 5' and 3' non-translated regions, and the like, are not part of a coding region.

The term "promoter" refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental or physiological conditions. Promoters which cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters." It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

In certain embodiments, the polynucleotide or nucleic acid is DNA. In the case of DNA, a polynucleotide comprising a nucleic acid, which encodes a polypeptide normally may include a promoter and/or other transcription or translation control elements operably associated with one or more coding regions. An operable association is when a coding region for a gene product, e.g., a polypeptide, is associated with one or more regulatory sequences in such a way as to place expression of the gene product under the influence or control of the regulatory sequence(s). Two DNA fragments (such as a polypeptide coding region and a promoter associated therewith) are "operably associated" if induction of promoter function results in the transcription of mRNA encoding the desired gene product and if the nature of the linkage between the two DNA fragments does not interfere with the ability of the expression regulatory sequences to direct the expression of the gene product or interfere with the ability of the DNA template to be transcribed. Thus, a promoter region would be operably associated with a nucleic acid encoding a polypeptide if the promoter was capable of affecting transcription of that
nucleic acid. Other transcription control elements, besides a promoter, for example enhancers, operators, repressors, and transcription termination signals, can be operably associated with the polynucleotide. Suitable promoters and other transcription control regions are disclosed herein.

A variety of translation control elements are known to those of ordinary skill in the art. These include, but are not limited to ribosome binding sites, translation initiation and termination codons, and elements derived from viral systems (particularly an internal ribosome entry site, or IRES, also referred to as a CITE sequence).

In other embodiments, a polynucleotide of the present invention is RNA, for example, in the form of messenger RNA (mRNA). RNA of the present invention may be single stranded or double stranded.

Polynucleotide and nucleic acid coding regions of the present invention may be associated with additional coding regions which encode secretory or signal peptides, which direct the secretion of a polypeptide encoded by a polynucleotide of the present invention.

As used herein, the term "transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "recombinant" or "transformed" organisms.

The term "expression," as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide.

The terms "plasmid," "vector," and "cassette" refer to an extra chromosomal element often carrying genes which are not part of the central metabolism of the cell, and usually in the form of circular double-stranded DNA fragments. Such elements may be autonomously replicating sequences, genome integrating sequences, phage or nucleotide sequences, linear or circular, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of introducing a promoter fragment and DNA sequence for a selected gene product along with appropriate 3' untranslated sequence into a cell. "Transformation cassette" refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that facilitates transformation of a
particular host cell. "Expression cassette" refers to a specific vector containing a foreign
gene and having elements in addition to the foreign gene that allow for enhanced expression of that gene in a foreign host.

The term "artificial" refers to a synthetic, or non-host cell derived composition, e.g., a chemically-synthesized oligonucleotide.

By a nucleic acid or polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence of the present invention, it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence.

As a practical matter, whether any particular nucleic acid molecule or polypeptide is at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to a nucleotide sequence or polypeptide sequence of the present invention can be determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al., Comp. Appl. Biosci. 6:237-245 (1990). In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's to Ts. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identity are: Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty-30, Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the subject nucleotide sequences, whichever is shorter.

If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences
truncated at the 5' or 3' ends, relative to the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only bases outside the 5' and 3' bases of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score.

For example, a 90 base subject sequence is aligned to a 100 base query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignment of the first 10 bases at 5' end. The 10 unpaired bases represent 10% of the sequence (number of bases at the 5' and 3' ends not matched/total number of bases in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 bases were perfectly matched the final percent identity would be 90%. In another example, a 90 base subject sequence is compared with a 100 base query sequence. This time the deletions are internal deletions so that there are no bases on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only bases 5' and 3' of the subject sequence which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to be made for the purposes of the present invention.

Polynucleotides useful in the invention include those that are at least about 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the nucleotide sequences set forth in Table 4, below, including variants, fragments or derivatives thereof that encode polypeptides with active alcohol dehydrogenase activity.

The terms "active variant," "active fragment," "active derivative," and "analog" refer to polynucleotides of the present invention and include any polynucleotides that encode polypeptides capable of catalyzing the reduction of a lower alkyl aldehyde.
Variants of polynucleotides of the present invention include polynucleotides with altered nucleotide sequences due to base pair substitutions, deletions, and/or insertions. Variants may occur naturally or be non-naturally occurring. Non-naturally occurring variants may be produced using art-known mutagenesis techniques. Derivatives of polynucleotides of the present invention, are polynucleotides which have been altered so that the polypeptides they encode exhibit additional features not found on the native polypeptide. Examples include polynucleotides that encode fusion proteins. Variant polynucleotides may also be referred to herein as "polynucleotide analogs." As used herein a "derivative" of a polynucleotide refers to a subject polynucleotide having one or more nucleotides chemically derivatized by reaction of a functional side group. Also included as "derivatives" are those polynucleotides which contain one or more naturally occurring nucleotide derivatives. For example, 3-methylcytidine may be substituted for cytosine; ribothymidine may be substituted for thymidine; and N4-acetylcytidine may be substituted for cytosine.

A "fragment" is a unique portion of the polynucleotide encoding the ADH enzyme which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides. A fragment used as a probe, primer, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides. Fragments may be preferentially selected from certain regions of a molecule. For example, a polynucleotide fragment may comprise a certain length of contiguous nucleotides selected from the first 100 or 200 nucleotides of a polynucleotide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

In one embodiment of the invention, polynucleotide sequences suitable for expression in recombinant host cells of the invention comprise nucleotide sequences that are at least about 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID
NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, and SEQ ID NO: 20. In another embodiment of the invention, a polynucleotide sequence suitable for expression in recombinant host cells of the invention can be selected from the group consisting of: SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, and SEQ ID NO: 20 or an active variant, fragment or derivative thereof. In one embodiment, polynucleotides have been codon-optimized for expression in a specific host cell.

[0112] In one embodiment of the invention, the polynucleotide sequence suitable for expression in recombinant host cells of the invention has a nucleotide sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to SEQ ID NO: 2. In another embodiment, the polynucleotide comprises the nucleotide sequence of SEQ ID NO: 2 or an active variant, fragment or derivative thereof.

[0113] In one embodiment of the invention, the polynucleotide sequence suitable for expression in recombinant host cells of the invention has a nucleotide sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to SEQ ID NO: 3. In another embodiment, the polynucleotide comprises the nucleotide sequence of SEQ ID NO: 3 or an active variant, fragment or derivative thereof.

[0114] In one embodiment of the invention, the polynucleotide sequence suitable for expression in recombinant host cells of the invention has a nucleotide sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to SEQ ID NO: 11. In another embodiment, the polynucleotide comprises the nucleotide sequence of SEQ ID NO: 11 or an active variant, fragment or derivative thereof.

[0115] In one embodiment of the invention, the polynucleotide sequence suitable for expression in recombinant host cells of the invention has a nucleotide sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to SEQ ID NO: 9. In another
embodiment, the polynucleotide comprises the nucleotide sequence of SEQ ID NO: 9 or an active variant, fragment or derivative thereof.

[0116] As used herein the term "codon degeneracy" refers to the nature in the genetic code permitting variation of the nucleotide sequence without affecting the amino acid sequence of an encoded polypeptide. The skilled artisan is well aware of the "codon-bias" exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a gene for improved expression in a host cell, it is desirable to design the gene such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

[0117] As used herein the term "codon optimized coding region" means a nucleic acid coding region that has been adapted for expression in the cells of a given organism by replacing at least one, or more than one, or a significant number, of codons with one or more codons that are more frequently used in the genes of that organism.

[0118] Deviations in the nucleotide sequence that comprise the codons encoding the amino acids of any polypeptide chain allow for variations in the sequence coding for the gene. Since each codon consists of three nucleotides, and the nucleotides comprising DNA are restricted to four specific bases, there are 64 possible combinations of nucleotides, 61 of which encode amino acids (the remaining three codons encode signals ending translation). The "genetic code" which shows which codons encode which amino acids is reproduced herein as Table 1. As a result, many amino acids are designated by more than one codon. For example, the amino acids alanine and proline are coded for by four triplets, serine and arginine by six, whereas tryptophan and methionine are coded by just one triplet. This degeneracy allows for DNA base composition to vary over a wide range without altering the amino acid sequence of the proteins encoded by the DNA.
Many organisms display a bias for use of particular codons to code for insertion of a particular amino acid in a growing peptide chain. Codon preference or codon bias, differences in codon usage between organisms, is afforded by degeneracy of the genetic code, and is well documented among many organisms. Codon bias often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, inter alia, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization.

Given the large number of gene sequences available for a wide variety of animal, plant and microbial species, it is possible to calculate the relative frequencies of codon usage. Codon usage tables are readily available, for example, at the "Codon Usage Database" available at http://www.kazusa.or.jp/codon/, and these tables can be adapted in a number of ways. See Nakamura, Y., et al. Nucl. Acids Res. 28:292 (2000).
usage tables for yeast, calculated from GenBank Release 128.0 [15 February 2002], are reproduced below as Table 2. This table uses mRNA nomenclature, and so instead of thymine (T) which is found in DNA, the tables use uracil (U) which is found in RNA. The Table has been adapted so that frequencies are calculated for each amino acid, rather than for all 64 codons.

**Table 2: Codon Usage Table for *Saccharomyces cerevisiae* Genes**

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Codon</th>
<th>Number</th>
<th>Frequency per thousand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phe</td>
<td>UUU</td>
<td>170666</td>
<td>26.1</td>
</tr>
<tr>
<td>Phe</td>
<td>UUC</td>
<td>120510</td>
<td>18.4</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>UUA</td>
<td>170884</td>
<td>26.2</td>
</tr>
<tr>
<td>Leu</td>
<td>UUG</td>
<td>177573</td>
<td>27.2</td>
</tr>
<tr>
<td>Leu</td>
<td>CUU</td>
<td>80076</td>
<td>12.3</td>
</tr>
<tr>
<td>Leu</td>
<td>CUC</td>
<td>35545</td>
<td>5.4</td>
</tr>
<tr>
<td>Leu</td>
<td>CUA</td>
<td>87619</td>
<td>13.4</td>
</tr>
<tr>
<td>Leu</td>
<td>CUG</td>
<td>68494</td>
<td>10.5</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>AUU</td>
<td>196893</td>
<td>30.1</td>
</tr>
<tr>
<td>Ile</td>
<td>AUC</td>
<td>112176</td>
<td>17.2</td>
</tr>
<tr>
<td>Ile</td>
<td>AUA</td>
<td>116254</td>
<td>17.8</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Met</td>
<td>AUG</td>
<td>136805</td>
<td>20.9</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>GUU</td>
<td>144243</td>
<td>22.1</td>
</tr>
<tr>
<td>Val</td>
<td>GUC</td>
<td>76947</td>
<td>11.8</td>
</tr>
<tr>
<td>Val</td>
<td>GUA</td>
<td>76927</td>
<td>11.8</td>
</tr>
<tr>
<td>Val</td>
<td>GUG</td>
<td>70337</td>
<td>10.8</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>UCU</td>
<td>153557</td>
<td>23.5</td>
</tr>
<tr>
<td>Ser</td>
<td>UCC</td>
<td>92923</td>
<td>14.2</td>
</tr>
<tr>
<td>Ser</td>
<td>UCA</td>
<td>122028</td>
<td>18.7</td>
</tr>
<tr>
<td>Ser</td>
<td>UCG</td>
<td>55951</td>
<td>8.6</td>
</tr>
<tr>
<td>Ser</td>
<td>AGU</td>
<td>92466</td>
<td>14.2</td>
</tr>
<tr>
<td>Ser</td>
<td>AGC</td>
<td>63726</td>
<td>9.8</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>CCU</td>
<td>88263</td>
<td>13.5</td>
</tr>
<tr>
<td>Amino Acid</td>
<td>Codon</td>
<td>Number</td>
<td>Frequency per thousand</td>
</tr>
<tr>
<td>------------</td>
<td>-------</td>
<td>--------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Pro</td>
<td>CCC</td>
<td>44309</td>
<td>6.8</td>
</tr>
<tr>
<td>Pro</td>
<td>CCA</td>
<td>119641</td>
<td>18.3</td>
</tr>
<tr>
<td>Pro</td>
<td>CCG</td>
<td>34597</td>
<td>5.3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>ACU</td>
<td>132522</td>
<td>20.3</td>
</tr>
<tr>
<td>Thr</td>
<td>ACC</td>
<td>83207</td>
<td>12.7</td>
</tr>
<tr>
<td>Thr</td>
<td>ACA</td>
<td>116084</td>
<td>17.8</td>
</tr>
<tr>
<td>Thr</td>
<td>ACG</td>
<td>52045</td>
<td>8.0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>GCU</td>
<td>138358</td>
<td>21.2</td>
</tr>
<tr>
<td>Ala</td>
<td>GCC</td>
<td>82357</td>
<td>12.6</td>
</tr>
<tr>
<td>Ala</td>
<td>GCA</td>
<td>105910</td>
<td>16.2</td>
</tr>
<tr>
<td>Ala</td>
<td>GCG</td>
<td>40358</td>
<td>6.2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>UAU</td>
<td>122728</td>
<td>18.8</td>
</tr>
<tr>
<td>Tyr</td>
<td>UAC</td>
<td>96596</td>
<td>14.8</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>CAU</td>
<td>89007</td>
<td>13.6</td>
</tr>
<tr>
<td>His</td>
<td>CAC</td>
<td>50785</td>
<td>7.8</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gln</td>
<td>CAA</td>
<td>178251</td>
<td>27.3</td>
</tr>
<tr>
<td>Gln</td>
<td>CAG</td>
<td>79121</td>
<td>12.1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>AAU</td>
<td>233124</td>
<td>35.7</td>
</tr>
<tr>
<td>Asn</td>
<td>AAC</td>
<td>162199</td>
<td>24.8</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>AAA</td>
<td>273618</td>
<td>41.9</td>
</tr>
<tr>
<td>Lys</td>
<td>AAG</td>
<td>201361</td>
<td>30.8</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>GAU</td>
<td>245641</td>
<td>37.6</td>
</tr>
<tr>
<td>Asp</td>
<td>GAC</td>
<td>132048</td>
<td>20.2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>GAA</td>
<td>297944</td>
<td>45.6</td>
</tr>
<tr>
<td>Glu</td>
<td>GAG</td>
<td>125717</td>
<td>19.2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amino Acid</td>
<td>Codon</td>
<td>Number</td>
<td>Frequency per thousand</td>
</tr>
<tr>
<td>------------</td>
<td>-------</td>
<td>--------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Cys</td>
<td>UGU</td>
<td>52903</td>
<td>8.1</td>
</tr>
<tr>
<td>Cys</td>
<td>UGC</td>
<td>31095</td>
<td>4.8</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trp</td>
<td>UGG</td>
<td>67789</td>
<td>10.4</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>CGU</td>
<td>41791</td>
<td>6.4</td>
</tr>
<tr>
<td>Arg</td>
<td>CGC</td>
<td>16993</td>
<td>2.6</td>
</tr>
<tr>
<td>Arg</td>
<td>CGA</td>
<td>19562</td>
<td>3.0</td>
</tr>
<tr>
<td>Arg</td>
<td>CGG</td>
<td>11351</td>
<td>1.7</td>
</tr>
<tr>
<td>Arg</td>
<td>AGA</td>
<td>139081</td>
<td>21.3</td>
</tr>
<tr>
<td>Arg</td>
<td>AGG</td>
<td>60289</td>
<td>9.2</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>GGU</td>
<td>156109</td>
<td>23.9</td>
</tr>
<tr>
<td>Gly</td>
<td>GGC</td>
<td>63903</td>
<td>9.8</td>
</tr>
<tr>
<td>Gly</td>
<td>GGA</td>
<td>71216</td>
<td>10.9</td>
</tr>
<tr>
<td>Gly</td>
<td>GGG</td>
<td>39359</td>
<td>6.0</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stop</td>
<td>UAA</td>
<td>6913</td>
<td>1.1</td>
</tr>
<tr>
<td>Stop</td>
<td>UAG</td>
<td>3312</td>
<td>0.5</td>
</tr>
<tr>
<td>Stop</td>
<td>UGA</td>
<td>4447</td>
<td>0.7</td>
</tr>
</tbody>
</table>

By utilizing this or similar tables, one of ordinary skill in the art can apply the frequencies to any given polypeptide sequence, and produce a nucleic acid fragment of a codon-optimized coding region which encodes the polypeptide, but which uses codons optimal for a given species.

Randomly assigning codons at an optimized frequency to encode a given polypeptide sequence, can be done manually by calculating codon frequencies for each amino acid, and then assigning the codons to the polypeptide sequence randomly. Additionally, various algorithms and computer software programs are readily available to those of ordinary skill in the art. For example, the "EditSeq" function in the Lasergene Package, available from DNASTar, Inc., Madison, WI, the backtranslation function in the VectorNTI Suite, available from InforMax, Inc., Bethesda, MD, and the "backtranslate" function in the GCG-Wisconsin Package, available from Accelrys, Inc., San Diego, CA. In addition, various resources are publicly available to codon-optimize coding region
sequences, e.g., the "backtranslation" function at http://www.entelechonxorn/bioinformatics/backtranslation.php?lang=eng (visited April 15, 2008) and the "backtranseq" function available at http://bioinfo.pbi.nrc.ca:8090/EMBOSS/index.html. Constructing a rudimentary algorithm to assign codons based on a given frequency can also easily be accomplished with basic mathematical functions by one of ordinary skill in the art.

Codon-optimized coding regions can be designed by various methods known to those skilled in the art including software packages such as "synthetic gene designer" (http://phenotype.biosci.urnbc.edu/codon/sgd/index.php).


Alcohol Dehydrogenase (ADH) Enzymes

Alcohol dehydrogenases (ADH) are a broad class of enzymes that catalyze the interconversion of aldehydes to alcohols as part of various pathways in cellular milieu. ADH enzymes are universal and are classified into multiple families based on either the length of the amino-acid sequence or the type of metal cofactors they use.

More than 150 structures are available in the Protein Data Bank (PDB) for a variety of ADH enzymes. The enzymes are highly divergent and different ADHs exist as oligomers with varying subunit compositions. Figures 4 shows the phylogenetic relationship of oxidoreductase enzymes in Saccharomyces cerevisiae, E. coli, Homo sapiens, C. elegans, Drosophila melanogaster, and Arabidopsis thaliana that are related to horse liver ADH and Achromobacter xylosoxidans SadB.

Figure 5 shows the phylogenetic relationship of specific ADH enzyme sequences more closely related to Achromobacter xylosoxidans SadB by sequence.
In one embodiment, ADH enzymes suitable for use in the present invention have a very high $k_{\text{cat}}$ for the conversion of a lower alkyl aldehyde to a corresponding lower alkyl alcohol. In another embodiment, ADH enzymes suitable for use have a very low $k_{\text{cat}}$ for the conversion of a lower alkyl alcohol to a corresponding lower alkyl aldehyde. In another embodiment, ADH enzymes suitable for use have a low $K_M$ for lower alkyl aldehydes. In another embodiment, suitable ADH enzymes have a high $K_M$ for lower alkyl alcohols. In another embodiment, suitable ADH enzymes preferentially use NADH as a cofactor during reduction reactions. In another embodiment, suitable ADH enzymes have one or more of the following characteristics: a very high $k_{\text{cat}}$ for the conversion of a lower alkyl aldehyde to a corresponding lower alkyl alcohol; a very low $k_{\text{cat}}$ for the conversion of a lower alkyl alcohol to a corresponding lower alkyl aldehyde; a low $K_M$ for lower alkyl aldehydes; a high $K_M$ for lower alkyl alcohols; and preferential use of NADH as a cofactor during reduction reactions. In another embodiment, suitable ADH enzymes have a high $\frac{3}{4}$ for lower alkyl alcohols. In another embodiment, suitable ADH enzymes have two or more of the above characteristics.

In one embodiment, ADH enzymes suitable for use in the present invention oxidize cofactors in the presence and absence of a lower alkyl alcohol faster relative to control polypeptides. In one embodiment, the control polypeptide is Achromobacter xylosoxidans SadB having the amino acid sequence of SEQ ID NO: 26.

In another embodiment, suitable ADH enzymes have $K_M$ for a lower alkyl aldehyde that are lower relative to a control polypeptide. In another embodiment, suitable ADH enzymes have a $K_M$ for a lower alkyl aldehyde that is at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 50%, 60%, 70%, 80%, or 90% lower relative to a control polypeptide. In one embodiment, the control polypeptide is Achromobacter xylosoxidans SadB having the amino acid sequence of SEQ ID NO: 26. In one embodiment, the lower alkyl aldehyde is isobutyraldehyde.

In another embodiment, suitable ADH enzymes have a $\frac{3}{4}$ for a lower alkyl alcohol that is higher relative to a control polypeptide. In another embodiment, suitable ADH enzymes have a lower alkyl alcohol $\frac{3}{4}$ that is at least about 10%, 50%, 100%, 200%, 300%, 400%, or 500% higher relative to a control polypeptide. In one embodiment, the control polypeptide is Achromobacter xylosoxidans SadB having the
amino acid sequence of SEQ ID NO: 26. In one embodiment, the lower alkyl alcohol is isobutanol.

[0132] In another embodiment, suitable ADH enzymes have a $k_{cat}/K_M$ for a lower alkyl aldehyde that is higher relative to a control polypeptide. In another embodiment, suitable ADH enzymes have a $k_{cat}/K_M$ that is at least about 10%, 50%, 100%, 200%, 300%, 400%, 500%, 600%, 800%, or 1000% higher relative to a control polypeptide. In one embodiment, the control polypeptide is Achromobacter xylosoxidans SadB having the amino acid sequence of SEQ ID NO: 26. In one embodiment, the lower alkyl aldehyde is isobutyraldehyde.

[0133] In another embodiment, suitable ADH enzymes have two or more of the above characteristics. In another embodiment, suitable ADH enzymes have three or more of the above characteristics. In another embodiment, suitable ADH enzymes have all four of the above characteristics. In one embodiment, suitable ADH enzymes preferentially use NADH as a cofactor.

[0134] In one embodiment, suitable ADH enzymes for use in the present invention catalyze reduction reactions optimally at host cell physiological conditions. In another embodiment, suitable ADH enzymes for use in the present invention catalyze reduction reactions optimally from about pH 4 to about pH 9. In another embodiment, suitable ADH enzymes for use in the present invention catalyze reduction reactions optimally from about pH 5 to about pH 8. In another embodiment, suitable ADH enzymes for use in the present invention catalyze reduction reactions optimally from about pH 6 to about pH 7. In another embodiment, suitable ADH enzymes for use in the present invention catalyze reduction reactions optimally from about pH 6.5 to about pH 7. In another embodiment, suitable ADH enzymes for use in the present invention catalyze reduction reactions optimally at about pH 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, or 9. In another embodiment, suitable ADH enzymes for use in the present invention catalyze reduction reactions optimally at about pH 7.

[0135] In one embodiment, suitable ADH enzymes for use in the present invention catalyze reduction reactions optimally at up to about 70°C. In another embodiment, suitable ADH enzymes catalyze reduction reactions optimally at about 10°C, 15°C, 20°C, 25°C, 30°C, 35°C, 40°C, 45°C, 50°C, 55°C, 60°C, 65°C, or 70°C. In another
embodiment, suitable ADH enzymes catalyze reduction reactions optimally at about 30°C.

[0136] In one embodiment, suitable ADH enzymes for use in the present invention catalyze the conversion of an aldehyde to an alcohol in the presence of a lower alkyl alcohol at a concentration up to about 50 g/L. In another embodiment, suitable ADH enzymes catalyze the conversion of an aldehyde to an alcohol in the presence of a lower alkyl alcohol at a concentration of at least about 10 g/L, 15 g/L, 20 g/L, 25 g/L, 30 g/L, 35 g/L, 40 g/L, 45 g/L, or 50 g/L. In another embodiment, suitable ADH enzymes catalyze the conversion of an aldehyde to an alcohol in the presence of a lower alkyl alcohol at a concentration of at least about 20 g/L. In some embodiments, the lower alkyl alcohol is butanol. In some embodiments, the lower alkyl aldehyde is isobutyraldehyde and the lower alkyl alcohol is isobutanol.

**Recombinant Host Cells for ADH Enzyme Expression**

[0137] One aspect of the present invention is directed to recombinant host cells that express ADH enzymes having the above-outlined activities. Non-limiting examples of host cells for use in the invention include bacteria, cyanobacteria, filamentous fungi and yeasts.

[0138] In one embodiment, the recombinant host cell of the invention is a bacterial or a cyanobacterial cell. In another embodiment, the recombinant host cell is selected from the group consisting of: *Salmonella, Arthrobacter, Bacillus, Brevibacterium, Clostridium, Corynebacterium, Gluconobacter, Nocardia, Pseudomonas, Rhodococcus, Streptomyces, Zymomonas, Escherichia, Lactobacillus, Enterococcus, Alcaligenes, Klebsiella, Serratia, Shigella, Alcaligenes, Erwinia, Paenibacillus, and Xanthomonas*. In some embodiments, the recombinant host cell is *E. coli, S. cerevisiae*, or *L. plantarum*.

[0139] In another embodiment, the recombinant host cell of the invention is a filamentous fungi or yeast cell. In another embodiment, the recombinant host cell is selected from the group consisting of: *Saccharomyces, Pichia, Hansenula, Yarrowia, Aspergillus, Kluyveromyces, Pachysolen, Rhodotorula, Zygosaccharomyces, Galactomyces, Schizosaccharomyces, Torulaspora, Debayomyces, Williopsis, Dekkera, Kloekera, Metschnikowia, Issatchenka*, and *Candida*. 
In one embodiment, the recombinant host cell of the invention produces a lower alkyl alcohol at a yield of greater than about 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, or 90% or higher of theoretical. In one embodiment, the recombinant host cell of the invention produces a lower alkyl alcohol at a yield of greater than about 25% of theoretical. In another embodiment, the recombinant host cell of the invention produces a lower alkyl alcohol at a yield of greater than about 40% of theoretical. In another embodiment, the recombinant host cell of the invention produces a lower alkyl alcohol at a yield of greater than about 50% of theoretical. In another embodiment, the recombinant host cell of the invention produces a lower alkyl alcohol at a yield of greater than about 75% of theoretical. In another embodiment, the recombinant host cell of the invention produces a lower alkyl alcohol at a yield of greater than about 90% of theoretical. In some embodiments, the lower alkyl alcohol is butanol. In some embodiments, the lower alkyl alcohol is isobutanol.

Non-limiting examples of lower alkyl alcohols produced by the recombinant host cells of the invention include butanol, propanol, isopropanol, and ethanol. In one embodiment, the recombinant host cells of the invention produce isobutanol. In another embodiment, the recombinant host cells of the invention do not produce ethanol.

U.S. Publ. No. 2007/0092957 A1 discloses the engineering of recombinant microorganisms for production of isobutanol (2-methylpropan-1-ol). U.S. Publ. No. 2008/0182308 A1 discloses the engineering of recombinant microorganisms for production of 1-butanol. U.S. Publ. Nos. 2007/0259410 A1 and 2007/0292927 A1 disclose the engineering of recombinant microorganisms for production of 2-butanol. Multiple pathways are described for biosynthesis of isobutanol and 2-butanol. The last step in all described pathways for all three products is the reduction of a more oxidized moiety to the alcohol moiety by an enzyme with butanol dehydrogenase activity. The methods disclosed in these publications can be used to engineer the recombinant host cells of the present invention. The information presented in these publications is hereby incorporated by reference in its entirety.

In embodiments, the recombinant microbial host cell produces isobutanol. In embodiments, the recombinant microbial host cell comprises at least two heterologous polynucleotides encoding enzymes which catalyze a substrate to product conversion selected from the group consisting of: pyruvate to acetolactate; acetolactate to 2,3-
dihydroxyisovalerate; 2,3-dihydroxyisovalerate to alpha-ketoisovalerate; alpha-ketoisovalerate to isobutyraldehyde, and isobutyraldehyde to isobutanol. In embodiments, the recombinant microbial host cell comprises at least three heterologous polynucleotides encoding enzymes which catalyze a substrate to product conversion selected from the group consisting of: pyruvate to acetalactate; acetalactate to 2,3-dihydroxyisovalerate; 2,3-dihydroxyisovalerate to alpha-ketoisovalerate; alpha-ketoisovalerate to isobutyraldehyde, and isobutyraldehyde to isobutanol. In embodiments, the recombinant microbial host cell comprises at least four heterologous polynucleotides encoding enzymes which catalyze a substrate to product conversion selected from the group consisting of: pyruvate to acetalactate; acetalactate to 2,3-dihydroxyisovalerate; 2,3-dihydroxyisovalerate to alpha-ketoisovalerate; alpha-ketoisovalerate to isobutyraldehyde, and isobutyraldehyde to isobutanol. In embodiments, the recombinant microbial host cell comprises heterologous polynucleotides encoding enzymes which catalyze the conversion of pyruvate to acetalactate; acetalactate to 2,3-dihydroxyisovalerate; 2,3-dihydroxyisovalerate to alpha-ketoisovalerate; alpha-ketoisovalerate to isobutyraldehyde, and isobutyraldehyde to isobutanol. In embodiments, (a) the polypeptide that catalyzes a substrate to product conversion of pyruvate to acetalactate is acetalactate synthase having the EC number 2.2.1.6; (b) the polypeptide that catalyzes a substrate to product conversion of acetalactate to 2,3-dihydroxyisovalerate is acetoxyhydroxy acid isomeroreductase having the EC number 1.1.1.86; (c) the polypeptide that catalyzes a substrate to product conversion of 2,3-dihydroxyisovalerate to alpha-ketoisovalerate is acetoxyhydroxy acid dehydratase having the EC number 4.2.1.9; and (d) the polypeptide that catalyzes a substrate to product conversion of alpha-ketoisovalerate to isobutyraldehyde is branched-chain alpha-keto acid decarboxylase having the EC number 4.1.1.72.

In embodiments, the recombinant microbial host cell further comprises at least one heterologous polynucleotide encoding an enzyme which catalyzes a substrate to product conversion selected from the group consisting of: pyruvate to alpha-acetolactate; alpha-acetolactate to acetoin; acetoin to 2,3-butanediol; 2,3-butanediol to 2-butanone; and 2-butanone to 2-butanol; and wherein said microbial host cell produces 2-butanol. In embodiments, (a) the polypeptide that catalyzes a substrate to product conversion of pyruvate to acetalactate is acetalactate synthase having the EC number 2.2.1.6; (b) the
polypeptide that catalyzes a substrate to product conversion of acetolactate to acetoin is acetolactate decarboxylase having the EC number 4.1.1.5; (c) the polypeptide that catalyzes a substrate to product conversion of acetoin to 2,3-butanediol is butanediol dehydrogenase having the EC number 1.1.1.76 or EC number 1.1.1.4; (d) the polypeptide that catalyzes a substrate to product conversion of butanediol to 2-butanone is butanediol dehydratase having the EC number 4.2.1.28. In embodiments, (e) the polypeptide that catalyzes a substrate to product conversion of 2-butanone to 2-butanol is 2-butanol dehydrogenase having the EC number 1.1.1.1.

In embodiments, the recombinant microbial host cell further comprises at least one heterologous polynucleotide encoding an enzyme which catalyzes a substrate to product conversion selected from the group consisting of: acetyl-CoA to acetoacetyl-CoA; acetoacetyl-CoA to 3-hydroxybutyryl-CoA; 3-hydroxybutyryl-CoA to crotonyl-CoA; crotonyl-CoA to butyryl-CoA; butyryl-CoA to butyraldehyde; butyraldehyde to 1-butanol; and wherein said microbial host cell produces 1-butanol. In embodiments, (a) the polypeptide that catalyzes a substrate to product conversion of acetyl-CoA to acetoacetyl-CoA is acetyl-CoA acetyltransferase having the EC number 2.3.1.9 or 2.3.1.16; (b) the polypeptide that catalyzes a substrate to product conversion of acetoacetyl-CoA to 3-hydroxybutyryl-CoA is 3-hydroxybutyryl-CoA dehydrogenase having the EC number 1.1.1.35, 1.1.1.30, 1.1.1.157, or 1.1.1.36; (c) the polypeptide that catalyzes a substrate to product conversion of 3-hydroxybutyryl-CoA to crotonyl-CoA is crotonase having the EC number 4.2.1.17 or 4.2.1.55; (d) the polypeptide that catalyzes a substrate to product conversion of crotonyl-CoA to butyryl-CoA is butyryl-CoA dehydrogenase having the EC number 1.3.1.44 or 1.3.1.38; (e) the polypeptide that catalyzes a substrate to product conversion of butyryl-CoA to butyrylaldehyde is butyrylaldehyde dehydrogenase having the EC number 1.2.1.57. In embodiments, (f) the polypeptide that catalyzes a substrate to product conversion of butyrylaldehyde to 1-butanol is 1-butanol dehydrogenase having the EC number 1.1.1.1.

In some embodiments, the recombinant microbial host cell further comprises at least one modification which improves carbon flow to the isobutanol pathway. In some embodiments, the recombinant microbial host cell further comprises at least one modification which improves carbon flow to the 1-butanol pathway. In some
embodiments, the recombinant microbial host cell further comprises a least one modification which improves carbon flow to the 2-butanol pathway.

Methods for Producing Lower Alkyl Alcohols

Another aspect of the present invention is directed to methods for producing lower alkyl alcohols. These methods primarily employ the recombinant host cells of the invention. In one embodiment, the method of the present invention comprises providing a recombinant host cell as discussed above, contacting the recombinant host cell with a fermentable carbon substrate in a fermentation medium under conditions whereby the lower alkyl alcohol is produced and recovering the lower alkyl alcohol.

Carbon substrates may include, but are not limited to, monosaccharides (such as fructose, glucose, mannose, rhamnose, xylose or galactose), oligosaccharides (such as lactose, maltose, or sucrose), polysaccharides such as starch, maltodextrin, or cellulose or mixtures thereof and unpurified mixtures from renewable feedstocks such as cheese whey permeate, cornsteep liquor, sugar beet molasses, and barley malt. Other carbon substrates may include ethanol, lactate, succinate, or glycerol.

Additionally, the carbon substrate may also be a one carbon substrate such as carbon dioxide, or methanol for which metabolic conversion into key biochemical intermediates has been demonstrated. In addition to one and two carbon substrates, methylotrophic organisms are also known to utilize a number of other carbon containing compounds such as methylamine, glucosamine and a variety of amino acids for metabolic activity. For example, methylotrophic yeasts are known to utilize the carbon from methylamine to form trehalose or glycerol (Bellion et al., Microb. Growth CI Compd., [Int. Symp.], 7th (1993), 415 32, Editor(s): Murrell, J. Collin; Kelly, Don P. Publisher: Intercept, Andover, UK). Similarly, various species of Candida will metabolize alanine or oleic acid (Suiter et al, Arch. Microbiol. 755:485-489 (1990)). Hence, it is contemplated that the source of carbon utilized in the present invention may encompass a wide variety of carbon containing substrates and will only be limited by the choice of organism.

Although it is contemplated that all of the above mentioned carbon substrates and mixtures thereof are suitable in the present invention, preferred carbon substrates are glucose, fructose, and sucrose, or mixtures of these with C5 sugars such as xylose and/or
arabinose for yeasts cells modified to use C5 sugars. Sucrose may be derived from renewable sugar sources such as sugar cane, sugar beets, cassava, sweet sorghum, and mixtures thereof. Glucose and dextrose may be derived from renewable grain sources through saccharification of starch based feedstocks including grains such as corn, wheat, rye, barley, oats, and mixtures thereof. In addition, fermentable sugars may be derived from renewable cellulosic or lignocellulosic biomass through processes of pretreatment and saccharification, as described, for example, in U.S. Publ. No. 2007/0031918 Al, which is herein incorporated by reference. Biomass refers to any cellululosic or lignocellulosic material and includes materials comprising cellulose, and optionally further comprising hemicellulose, lignin, starch, oligosaccharides and/or monosaccharides. Biomass may also comprise additional components, such as protein and/or lipid. Biomass may be derived from a single source, or biomass can comprise a mixture derived from more than one source; for example, biomass may comprise a mixture of corn cobs and corn stover, or a mixture of grass and leaves. Biomass includes, but is not limited to, bioenergy crops, agricultural residues, municipal solid waste, industrial solid waste, sludge from paper manufacture, yard waste, wood and forestry waste. Examples of biomass include, but are not limited to, corn grain, corn cobs, crop residues such as corn husks, corn stover, grasses, wheat, wheat straw, barley, barley straw, hay, rice straw, switchgrass, waste paper, sugar cane bagasse, sorghum, soy, components obtained from milling of grains, trees, branches, roots, leaves, wood chips, sawdust, shrubs and bushes, vegetables, fruits, flowers, animal manure, and mixtures thereof.

The carbon substrates may be provided in any media that is suitable for host cell growth and reproduction. Non-limiting examples of media that can be used include M122C, MOPS, SOB, TSY, YMG, YPD, 2XYT, LB, M17, or M9 minimal media. Other examples of media that can be used include solutions containing potassium phosphate and/or sodium phosphate. Suitable media can be supplemented with NADH or NADPH.

The fermentation conditions for producing a lower alkyl alcohol may vary according to the host cell being used. In one embodiment, the method for producing a lower alkyl alcohol is performed under anaerobic conditions. In one embodiment, the method for producing a lower alkyl alcohol is performed under aerobic conditions. In
one embodiment, the method for producing a lower alkyl alcohol is performed under microaerobic conditions.

[0153] In one embodiment, the method for producing a lower alkyl alcohol results in a titer of at least about 20 g/L of a lower alkyl alcohol. In another embodiment, the method for producing a lower alkyl alcohol results in a titer of at least about 30 g/L of a lower alkyl alcohol. In another embodiment, the method for producing a lower alkyl alcohol results in a titer of about 10 g/L, 15 g/L, 20 g/L, 25 g/L, 30 g/L, 35 g/L or 40 g/L of a lower alkyl alcohol.

[0154] Non-limiting examples of lower alkyl alcohols produced by the methods of the invention include butanol, isobutanol, propanol, isopropanol, and ethanol. In one embodiment, isobutanol is produced.

[0155] In embodiments, isobutanol is produced. In embodiments, the method for producing isobutanol comprises:

[0156] (a) providing a recombinant host cell comprising a heterologous polypeptide which catalyzes the substrate to product conversion of isobutyraldehyde to isobutanol and which has one or more of the following characteristics:

[0157] (i) the $K_M$ value of a lower alkyl aldehyde is lower for the polypeptide relative to a control polypeptide having the amino acid sequence of SEQ ID NO: 26;

[0158] (ii) the $\frac{3}{4}$ value for a lower alkyl aldehyde for the polypeptide is higher relative to control polypeptide having the amino acid sequence of SEQ ID NO: 26;

[0159] (iii) the $\frac{k_{\text{cat}}}{K_M}$ value for a lower alkyl aldehyde for the polypeptide is higher relative to a control polypeptide having the amino acid sequence of SEQ ID NO: 26; and

[0160] (b) contacting the host cell of (a) with a carbon substrate under conditions whereby isobutanol is produced.

[0161] In embodiments, 2-butanol is produced. In embodiments, the method for producing 2-butanol comprises:

[0162] (a) providing a recombinant microbial host cell comprising a heterologous polypeptide which catalyzes the substrate to product conversion of 2-butanone to 2-butanol and which has one or more of the following characteristics:
(i) the $K_M$ value for a lower alkyl aldehyde is lower for the polypeptide relative to a control polypeptide having the amino acid sequence of SEQ ID NO: 26;

(ii) the $\frac{k_{cat}}{K_M}$ value for a lower alkyl aldehyde for the polypeptide is higher relative to a control polypeptide having the amino acid sequence of SEQ ID NO: 26; and

(iii) the $k_{cat}/K_M$ value for a lower alkyl aldehyde for the polypeptide is higher relative to a control polypeptide having the amino acid sequence of SEQ ID NO: 26; and

(b) contacting the host cell of (a) with a carbon substrate under conditions whereby 2-butanol is produced.

In embodiments, 1-butanol is produced. In embodiments, the method for producing 1-butanol comprises:

(a) providing a recombinant microbial host cell comprising a heterologous polypeptide which catalyzes the substrate to product conversion of butyraldehyde to 1-butanol and which has one or more of the following characteristics:

(i) the $K_M$ value for a lower alkyl aldehyde is lower for the polypeptide relative to a control polypeptide having the amino acid sequence of SEQ ID NO: 26;

(ii) the $\frac{k_{cat}}{K_M}$ value for a lower alkyl aldehyde for the polypeptide is higher relative to a control polypeptide having the amino acid sequence of SEQ ID NO: 26; and

(iii) the $k_{cat}/K_M$ value for a lower alkyl aldehyde for the polypeptide is higher relative to a control polypeptide having the amino acid sequence of SEQ ID NO: 26; and

(b) contacting the host cell of (a) with a carbon substrate under conditions whereby 1-butanol is produced.

Biosynthetic Pathways

pathway (Maggio-Hall et al, U.S. Patent Publication No. US 20070092957, incorporated herein by reference) have been described in the art. Certain suitable proteins having the ability to catalyze the indicated substrate to product conversions are described therein and other suitable proteins are described in the art. The skilled person will appreciate that polypeptides having the activity of such pathway steps can be isolated from a variety of sources and can be used in a recombinant host cell disclosed herein. For example, US Published Patent Application Nos. US20080261230 and US20090163376, US20100197519, and US Application Serial No. 12/893077 describe acetohydroxy acid isomeroreductases; US20070092957 and US20100081154, describe suitable dihydroxyacid dehydratases.

Equipped with this disclosure, a person of skill in the art will be able to utilize publicly available sequences to construct relevant pathways in the host cells provided herein. Additionally, one of skill in the art, equipped with this disclosure, will appreciate other suitable isobutanol, 1-butanol, or 2-butanol pathways.

Isobutanol Biosynthetic Pathway

Isobutanol can be produced from carbohydrate sources with recombinant microorganisms by through various biosynthetic pathways. Suitable pathways converting pyruvate to isobutanol include the four complete reaction pathways shown in Figure 6. A suitable isobutanol pathway (Figure 6, steps a to e), comprises the following substrate to product conversions:

a) pyruvate to acetylactate, as catalyzed for example by acetylactate synthase,

b) acetylactate to 2,3-dihydroxyisovalerate, as catalyzed for example by acetohydroxy acid isomeroreductase,

c) 2,3-dihydroxyisovalerate to a-ketoisovalerate, as catalyzed for example by acetohydroxy acid dehydratase,

d) a-ketoisovalerate to isobutyraldehyde, as catalyzed for example by a branched-chain keto acid decarboxylase, and

e) isobutyraldehyde to isobutanol, as catalyzed for example by, a branched-chain alcohol dehydrogenase.

Another suitable pathway for converting pyruvate to isobutanol comprises the following substrate to product conversions (Figure 6, steps a,b,c,f,g,e):
a) pyruvate to acetolactate, as catalyzed for example by acetolactate synthase,
b) acetolactate to 2,3-dihydroxyisovalerate, as catalyzed for example by acetohydroxy acid isomeroreductase,
c) 2,3-dihydroxyisovalerate to α-ketoisovalerate, as catalyzed for example by acetohydroxy acid dehydratase,
f) α-ketoisovalerate to isobutyryl-CoA, as catalyzed for example by a branched-chain keto acid dehydrogenase,
g) isobutyryl-CoA to isobutyraldehyde, as catalyzed for example by an acylating aldehyde dehydrogenase, and
e) isobutyraldehyde to isobutanol, as catalyzed for example by, a branched-chain alcohol dehydrogenase.

The first three steps in this pathway (a,b,c) are the same as those described above.

Another suitable pathway for converting pyruvate to isobutanol comprises the following substrate to product conversions (Figure 6, steps a,b,c,h,i,j,e):
a) pyruvate to acetolactate, as catalyzed for example by acetolactate synthase,
b) acetolactate to 2,3-dihydroxyisovalerate, as catalyzed for example by acetohydroxy acid isomeroreductase,
c) 2,3-dihydroxyisovalerate to α-ketoisovalerate, as catalyzed for example by acetohydroxy acid dehydratase,
h) α-ketoisovalerate to valine, as catalyzed for example by valine dehydrogenase or transaminase,
i) valine to isobutylamine, as catalyzed for example by valine decarboxylase,
j) isobutylamine to isobutyraldehyde, as catalyzed for example by omega transaminase, and
e) isobutyraldehyde to isobutanol, as catalyzed for example by, a branched-chain alcohol dehydrogenase.

The first three steps in this pathway (a,b,c) are the same as those described above.

A fourth suitable isobutanol biosynthetic pathway comprises the substrate to product conversions shown as steps k,g,e in Figure 6.

1-Butanol Biosynthetic Pathway
An example of a suitable biosynthetic pathway for production of 1-butanol is disclosed in U.S. Patent Application Publication No. US 2008/0182308 Al. As disclosed this publication, steps in the disclosed 1-butanol biosynthetic pathway include conversion of:

- acetyl-CoA to acetoacetyl-CoA, as catalyzed for example by acetyl-CoA acetyltransferase;
- acetoacetyl-CoA to 3-hydroxybutyryl-CoA, as catalyzed for example by 3-hydroxybutyryl-CoA dehydrogenase;
- 3-hydroxybutyryl-CoA to crotonyl-CoA, as catalyzed for example by crotonase;
- crotonyl-CoA to butyryl-CoA, as catalyzed for example by butyryl-CoA dehydrogenase;
- butyryl-CoA to butyraldehyde, as catalyzed for example by butyraldehyde dehydrogenase; and
- butyraldehyde to 1-butanol, as catalyzed for example by butanol dehydrogenase.

2-Butanol Biosynthetic Pathway

An example of a suitable biosynthetic pathway for production of 2-butanol is described by Donaldson et al. in U.S. Patent Application Publication Nos. US20070259410A1 and US 20070292927A1, and in PCT Publication WO 2007/130521, all of which are incorporated herein by reference. Steps of a suitable 2-butanol biosynthetic pathway comprises the following substrate to product conversions:

- a) pyruvate to alpha-acetolactate, which may be catalyzed, for example, by acetolactate synthase;
- b) alpha-acetolactate to acetoin, which may be catalyzed, for example, by acetolactate decarboxylase;
- c) acetoin to 2,3-butanediol, which may be catalyzed, for example, by butanediol dehydrogenase;
- d) 2,3-butanediol to 2-butanone, which may be catalyzed, for example, by butanediol dehydratase; and
- e) 2-butanone to 2-butanol, which may be catalyzed, for example, by 2-butanol dehydrogenase.
Additional modifications

[0113] Additional modifications that may be useful in cells provided herein include modifications to reduce pyruvate decarboxylase and/or glycerol-3-phosphate dehydrogenase activity as described in US Patent Application Publication No. 20090305363 (incorporated herein by reference), modifications to a host cell that provide for increased carbon flux through an Entner-Doudoroff Pathway or reducing equivalents balance as described in US Patent Application Publication No. 20100120105 (incorporated herein by reference). Yeast strains with increased activity of heterologous proteins that require binding of an Fe-S cluster for their activity are described in US Application Publication No. 20100081179 (incorporated herein by reference). Other modifications include modifications in an endogenous polynucleotide encoding a polypeptide having dual-role hexokinase activity, described in US Provisional Application No. 61/290,639, integration of at least one polynucleotide encoding a polypeptide that catalyzes a step in a pyruvate-utilizing biosynthetic pathway described in US Provisional Application No. 61/380563 (both referenced provisional applications are incorporated herein by reference in their entirety). Additional modifications that may be suitable for embodiments herein are described in US Application Serial No. 12/893089.

[0114] Additionally, host cells comprising at least one deletion, mutation, and/or substitution in an endogenous gene encoding a polypeptide affecting Fe-S cluster biosynthesis are described in US Provisional Patent Application No. 61/305333 (incorporated herein by reference), and host cells comprising a heterologous polynucleotide encoding a polypeptide with phosphoketolase activity and host cells comprising a heterologous polynucleotide encoding a polypeptide with phosphotransacetylase activity are described in US Provisional Patent Application No. 61/356379.

Identification and Isolation of High Activity ADH Enzymes

[0199] The present invention is directed to devising a strategy and identifying several ADH enzymes with superior properties towards the conversion of isobutyraldehyde to isobutanol in a host organism that has been engineered for isobutanol production. The process of ADH candidate selection involves searching among the naturally existing enzymes. Enzymes are identified based on their natural propensity to utilize aldehydes as
preferred substrates and convert them to the respective alcohols with reasonably high $k_{cat}$ and/or low $K_M$ values for the corresponding aldehyde substrates, as documented by literature examples. Once a set of candidates is identified, the strategy involves using this set to isolate closely-related homologues via bioinformatics analysis. Therefore, in one embodiment, the screening method of the invention comprises performing a bioinformatics or literature search for candidate ADH enzymes. In one embodiment, the bioinformatics search uses a phylogenetic analysis.

The protein-encoding DNA sequences of the candidate genes are either amplified directly from the host organisms or procured as codon-optimized synthetic genes for expression in a host cell, such as *E. coli*. Various ADH candidates utilized herein are listed in Table 3.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Polynucleotide SEQ ID NO:</th>
<th>Polypeptide SEQ ID NO:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horse-liver ADH</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td><em>Saccharomyces cerevisiae</em> ADH6</td>
<td>2</td>
<td>22</td>
</tr>
<tr>
<td><em>Saccharomyces cerevisiae</em> ADH7</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td><em>Clostridium acetobutylicum</em> BdhA</td>
<td>4</td>
<td>24</td>
</tr>
<tr>
<td><em>Clostridium acetobutylicum</em> BdhB</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td><em>Achromobacter xylosoxidans</em> SadB</td>
<td>6</td>
<td>26</td>
</tr>
<tr>
<td><em>Bos taurus</em> ARD</td>
<td>7</td>
<td>27</td>
</tr>
<tr>
<td><em>Rana perezi</em> ADH8</td>
<td>8</td>
<td>28</td>
</tr>
<tr>
<td><em>Clostridium beijerinckii</em> ADH</td>
<td>9</td>
<td>29</td>
</tr>
<tr>
<td><em>Entamoeba histolytica</em> ADH1</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td><em>Beijerinckia indica</em> ADH</td>
<td>11</td>
<td>31</td>
</tr>
</tbody>
</table>

Table 3
The present invention is not limited to the ADH enzymes listed in Table 3. Additional candidates can be identified based on sequence homologies to these candidates or candidates can be derived from these sequences via mutagenesis and/or protein evolution. Suitable ADH enzymes include ADH enzymes having at least about 95% identity to the sequences provided herein.

Tables 4 and 5 provide the polynucleotide (codon-optimized for expression E. coli except for SEQ ID NOs. 2, 3, 4, 5, and 6) and polypeptides sequences of the candidate ADH enzymes presented in Table 3, respectively.

<table>
<thead>
<tr>
<th>SEQ ID NO</th>
<th>POLYNUCLEOTIDE SEQUENCE</th>
</tr>
</thead>
</table>
| 1         | atgtcaacagcggtaagttatattaagttaagttgaagcggcagtttgtggaagagaagagaacgcggtagtgcaggaagagtgaagtgcacccaaagaacgcacaggttagaatcaagatgggtgacccacggagaatctgtgatccgacagcaagctgtggtgaggtaggcacacgtctgtggtggtgacccacggagaatctgtgatccgacagcaagctgtggtgaggtaggcacacgtctgtggtggtgacccacggagaatctgtgatccgacagcaagctgtggtgaggtaggcacacgtctgtggtggtgacccacggagaatctgtgatccgacagcaagctgtggtgaggtaggcacacgtctgtggtggtgacccacggagaatctgtgatccgacagcaagctgtggtgaggtaggcacacgtctgtggtggtgacccacggagaatctgtgatccgacagcaagctgtggtgaggtaggcacacgtctgtggtggtgacccacggagaatctgtgatccgacagcaagctgtggtgaggtaggcacacgtctgtggtggtgacccacggagaatctgtgatccgacagcaagctgtggtgaggtaggcacacgtctgtggtggtgacccacggagaatctgtgatccgacagcaagctgtggtgaggtaggcacacgtctgtggtggtgacccacggagaatctgtgatccgacagcaagctgtggtgaggtaggcacacgtctgtggtggtgacccacggagaatctgtgatccgacagcaagctgtggtgaggtaggcacacgtctgtggtggtgacccacggagaatctgtgatccgacagc

Table 4
| 5 | cagcagcgggaacagctgacattatgagtcacacctttgaatcttactttagtggtgttgaaggtgcttatgtgcaggacggtatacgagaagcaatcttaagaacatgtataaagtatggaaaaaatagcaatggagagactgatgattacgaggctagagctaatttgatgtgggcttcaagtttagctataaatggtctattatcac ttggtaaggatagaaatgtgctggggagtaagacagaaacaagatattagaaacagagcgtcctcgaagcctgagaagctgtaagtatacttgaaaaaaacagtattaaattttatgaacttgcaggagtagagccaaatccaagagtaactacagttgaaaaaggagttaaaatatgtagagaaaatggagttgaagtagtactagct atagttggagaagaggtcataaagctgatcggcggggtgatggcttcgagaatgctgcggcttagaagagacctgataggttgatttcgaatattcaataccaactagaatttttttcggtaaagataagataaatgtacttggaagaagctgcaatttttaacacctaattggatggaatatattctaaatgacgatacacttcataaatttg ttcttatggaataaatgtttggggaatagacaaagaaaaaaatcactatgacatagcacatcaagc aatacaaaaaacaagagattactttgtaaatgtactaggtttaccatctagactgagagatgttggaattgaaagaagaaaaattggacataatggcaaaggaatcagtaaagcttacaggaggaaccataggaaacctaagaccagtaaacgcctccgaagtcctacaaatattcaaaaaatctgtgtaa |
| 6 | cagcagcgggaacagctgacattatgagtcacacctttgaatcttactttagtggtgttgaaggtgcttatgtgcaggacggtatacgagaagcaatcttaagaacatgtataaagtatggaaaaaatagcaatggagagactgatgattacgaggctagagctaatttgatgtgggcttcaagtttagctataaatggtctattatcac ttggtaaggatagaaatgtgctggggagtaagacagaaacaagatattagaaacagagcgtcctcgaagcctgagaagctgtaagtatacttgaaaaaaacagtattaaattttatgaacttgcaggagtagagccaaatccaagagtaactacagttgaaaaaggagttaaaatatgtagagaaaatggagttgaagtagtactagct atagttggagaagaggtcataaagctgatcggcggggtgatggcttcgagaatgctgcggcttagaagagacctgataggttgatttcgaatattcaataccaactagaatttttttcggtaaagataagataaatgtacttggaagaagctgcaatttttaacacctaattggatggaatatattctaaatgacgatacacttcataaatttg ttcttatggaataaatgtttggggaatagacaaagaaaaaaatcactatgacatagcacatcaagc aatacaaaaaacaagagattactttgtaaatgtactaggtttaccatctagactgagagatgttggaattgaaagaagaaaaattggacataatggcaaaggaatcagtaaagcttacaggaggaaccataggaaacctaagaccagtaaacgcctccgaagtcctacaaatattcaaaaaatctgtgtaa |
| 11 | atgaaacaaactgtgtttaaccgtgtggcggtgccccagctggcaaatggaaagctgtggaagacagcctgctgtgctgtgcgtgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgctgctgtgct}
Table 5

<table>
<thead>
<tr>
<th>SEQ ID NO</th>
<th>POLYPEPTIDE SEQUENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>MSTAGKVIKCAA VLWEEKKPSIEEVEVAPPKAHEVRIKMAV TGCIRSDDHVSGTLVTPLPVAGHEAAGIVESIGEVTTVPFGD KVIPLFTPQCGKCRVCKHPEGNFCLKNDLSMPRGTMQDGTSRFT CRKPIPHHFLGTSYTVPVEISVAKIADASPLEKVCILGCFS TGYGSAVKVAKVTQGSTCAVFGLGGVGLSVMGCKAAGAARR GVDINDKDFKAKEVGATECVNPQDYKKPIQEVLTEMSNGGVD FSFEVIRGLDTMVTASCCQEAYGVSIVGVVPDSQNSLMNPM LLLSRTWKGAFGGFFSKDSVPKLVADFMAKKFALDPLIHTVLP FKEKNEPFDLLRSGRESITLTF</td>
</tr>
<tr>
<td>22</td>
<td>MSYPEKFEGIAIQSHEDWKNPKKTKYDPFKPYDHIDDIKIEACG VC GDHI CAGHGWGNMKMPLVGHEIVGKVKLGPNSGLKV GVRQVRGVAQVGVSCELDRCRCKNDEPYCTKFVTTYSQPYEDG Y VSQGGYANYVRVHEHVVIDPENPSHLAAPLLCGLGTVYPLV RNGCQPGGKVGQLGLGGSIGMTLIASKMAETYVISRSSKRE DAMKMGADHYATLEEGDWGEKYFTDTLIVCASSLTDIDFNI MPKAMKVGGRIVSISPEQHEMLSLKPYGLKAVSISYSLGSI KE LNQLLKLVSEKDIKSWTLPVEAGVHEAFERMEKGDVYRF TLVGYDKEFSD</td>
</tr>
<tr>
<td>23</td>
<td>MLYPEKFQGIGISNAKDWKPKLVSDFPKPGFDHDVDEIEACG ICSDHFIAVGNWPVQENPQIGHEIVGYKVKGLGSKCQTVG KID RVGVAQALACFECERCKSDKNEQYCTNDHVLTMWTPYKGDY IQ SQGGFASHVRLHEFANIQIPENIPSSPLAAPPCCGTVFSP PLLRNG</td>
</tr>
</tbody>
</table>
[0203] In one embodiment, the method for screening candidate polypeptides having alcohol dehydrogenase activity comprises:

[0204] (a) measuring the rate of cofactor oxidation by a lower alkyl aldehyde for the candidate polypeptides in the presence or absence of a lower alkyl alcohol; and

[0205] (b) selecting only those candidate polypeptides that oxidize a cofactor faster relative to a control polypeptide in the presence or absence of a lower alkyl alcohol. In one embodiment, (b) comprises selecting only those candidate polypeptides that oxidize a cofactor faster relative to a control polypeptide in both the presence and absence of a lower alkyl alcohol. In one embodiment, the cofactor is NADH. In another embodiment, the cofactor is NADPH. In yet another embodiment, the control polypeptide is HLADH having the amino acid sequence of SEQ ID NO: 21. In yet another embodiment, the control polypeptide is Achromobacter xylosoxidans SadB having the amino acid sequence of SEQ ID NO: 26. In another embodiment, step (a) comprises monitoring a change in A₃₄₀ nm.

[0206] In another embodiment, the method for screening candidate polypeptides having alcohol dehydrogenase activity comprises:

[0207] (a) measuring one or more of the following values for the candidate polypeptides:

[0208] (i) the $K_M$ value for a lower alkyl aldehyde;

[0209] (ii) the $\frac{3}{4}$ value for a lower alkyl alcohol; and

[0210] (iii) $\frac{k_{cat}}{K_M}$; and

[0211] (b) selecting only those candidate polypeptides having one or more of the following characteristics:

[0212] (i) the $K_M$ value for a lower alkyl aldehyde is lower relative to a control polypeptide;

[0213] (ii) the $\frac{3}{4}$ value for a lower alkyl alcohol is higher relative to a control polypeptide; and
(iii) the $k_{cat}/K_M$ value for a lower alkyl aldehyde is higher relative to a control polypeptide.

In yet another embodiment, the control polypeptide is *Achromobacter xylosoxidans* SadB having the amino acid sequence of SEQ ID NO: 26. In another embodiment, the selected candidate polypeptides have two or more of the above characteristics. In another embodiment, the selected candidate polypeptides have three or more of the above characteristics. In another embodiment, the selected candidate polypeptides preferentially use NADH as a cofactor.

In one embodiment of the invention, polynucleotide sequences suitable for use in the screening methods of the invention comprise nucleotide sequences that are at least about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, and SEQ ID NO: 20. In another embodiment of the invention, a polynucleotide sequence suitable for use in the screening methods of the invention can be selected from the group consisting of: SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, and SEQ ID NO: 20 or an active variant, fragment or derivative thereof. In one embodiment, polynucleotides have been codon-optimized for expression in a specific host cell.

In one embodiment of the invention, candidate polypeptides suitable for use in the screening methods of the invention have amino acid sequences that are at least about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, and SEQ ID NO: 40. In another embodiment of the invention, a candidate polypeptide suitable for use in the screening methods of the invention has an
amino acid sequence selected from the group consisting of: SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, and SEQ ID NO: 40, or an active variant, fragment or derivative thereof.

In one embodiment, candidate polypeptides suitable for use in the screening methods of the invention have been codon-optimized for expression in a specific host cell.

[0218] In one embodiment of the invention, the polynucleotide sequence suitable for use in the screening methods of the invention has a nucleotide sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to SEQ ID NO: 2. In another embodiment, the polynucleotide comprises the nucleotide sequence of SEQ ID NO: 2 or an active variant, fragment or derivative thereof.

[0219] In one embodiment of the invention, candidate polypeptides for use in the screening methods comprise an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to SEQ ID NO: 22. In another embodiment, the candidate polypeptide comprises the amino acid sequence of SEQ ID NO: 22 or an active variant, fragment or derivative thereof.

[0220] In one embodiment of the invention, the polynucleotide sequence suitable for use in the screening methods has a nucleotide sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to SEQ ID NO: 3. In another embodiment, the polynucleotide comprises the nucleotide sequence of SEQ ID NO: 3 or an active variant, fragment or derivative thereof.

[0221] In one embodiment of the invention, candidate polypeptides for use in the screening methods comprise an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to SEQ ID NO: 23. In another embodiment, the candidate polypeptide comprises the amino acid sequence of SEQ ID NO: 23 or an active variant, fragment or derivative thereof.
In one embodiment of the invention, the polynucleotide sequence for use in the screening methods has a nucleotide sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to SEQ ID NO: 11. In another embodiment, the polynucleotide comprises the nucleotide sequence of SEQ ID NO: 11 or an active variant, fragment or derivative thereof.

In one embodiment of the invention, candidate polypeptides for use in the screening methods comprise an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to SEQ ID NO: 31. In another embodiment, the candidate polypeptide comprises the amino acid sequence of SEQ ID NO: 31 or an active variant, fragment or derivative thereof.

In one embodiment of the invention, the polynucleotide sequence for use in the screening methods has a nucleotide sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to SEQ ID NO: 9. In another embodiment, the polynucleotide comprises the nucleotide sequence of SEQ ID NO: 9 or an active variant, fragment or derivative thereof.

In one embodiment of the invention, candidate polypeptides for use in the screening methods comprise an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to SEQ ID NO: 29. In another embodiment, the candidate polypeptide comprises the amino acid sequence of SEQ ID NO: 29 or an active variant, fragment or derivative thereof.

In another embodiment, the method for screening candidate polypeptides results in selected candidate polypeptides being able to catalyze the conversion of an aldehyde to an alcohol at a temperature up to about 70°C. In another embodiment, the screening method results in selected candidate polypeptides being able to catalyze the conversion of an aldehyde to an alcohol at a temperature of about 10°C, 15°C, 20°C, 25°C, 30°C, 35°C, 40°C, 45°C, 50°C, 55°C, 60°C, 65°C, or 70°C. In another embodiment, the screening method results in selected candidate polypeptides being able to catalyze the conversion of an aldehyde to an alcohol at a temperature of about 30°C.
In another embodiment, the method for screening candidate polypeptides results in selected candidate polypeptides being able to catalyze the conversion of an aldehyde to an alcohol at a pH from about 4 to about 9. In another embodiment, the screening method results in selected candidate polypeptides being able to catalyze the conversion of an aldehyde to an alcohol at pH from about 5 to about 8. In another embodiment, the screening method results in selected candidate polypeptides being able to catalyze the conversion of an aldehyde to an alcohol at a pH from about 6 to about 7. In another embodiment, the screening method results in selected candidate polypeptides being able to catalyze the conversion of an aldehyde to an alcohol at a pH from about 6.5 to about 7. In another embodiment, the screening method results in selected candidate polypeptides being able to catalyze the conversion of an aldehyde to an alcohol at a pH of about 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, or 9. In another embodiment, the screening method results in selected candidate polypeptides being able to catalyze the conversion of an aldehyde to an alcohol at a pH of about 7.

In another embodiment, the method for screening candidate polypeptides results in selected candidate polypeptides that can catalyze the conversion of an aldehyde to an alcohol in the presence of a lower alkyl alcohol at a concentration up to about 50 g/L. In another embodiment, the screening method results in selected candidate polypeptides being able to catalyze the conversion of an aldehyde to an alcohol at a concentration of about 10 g/L, 15 g/L, 20 g/L, 25 g/L, 30 g/L, 35 g/L, 40 g/L, 45 g/L, or 50 g/L. In another embodiment, the screening method results in selected candidate polypeptides being able to catalyze the conversion of an aldehyde to an alcohol at a concentration of at least about 20 g/L.

Non-limiting examples of lower alkyl alcohols that can be used in the screening methods of the invention include butanol, isobutanol, propanol, isopropanol, and ethanol. In one embodiment, the lower alkyl alcohol used in the screening method is isobutanol.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. In case of conflict, the present application including the definitions will control. Also, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. All publications, patents and other
references mentioned herein are incorporated by reference in their entireties for all purposes.

Examples

[0231] The present invention is further defined in the following Examples. It should be understood that these Examples, while indicating embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various uses and conditions.

General Methods


[0233] Materials and methods suitable for the maintenance and growth of bacterial cultures are well known in the art. Techniques suitable for use in the following examples may be found as set out in Manual of Methods for General Bacteriology (Phillipp et al., eds., American Society for Microbiology, Washington, DC, 1994) or by Thomas D. Brock in (Brock, Biotechnology: A Textbook of Industrial Microbiology, Second Edition, Sinauer Associates, Inc., Sunderland, MA (1989). All reagents, restriction enzymes and materials used for the growth and maintenance of bacterial cells were obtained from Sigma-Aldrich Chemicals (St. Louis, MO), BD Diagnostic Systems (Sparks, MD), Invitrogen (Carlsbad, CA), HiMedia (Mumbai, India), SD Fine chemicals (India), or Takara Bio Inc. (Shiga, Japan), unless otherwise specified.

[0234] The meaning of abbreviations is as follows: "sec" means second(s), "min" means minute(s), "h" means hour(s), "nm" means nanometers, "uL" means microliter(s), "mL" means milliliter(s), "mg/mL" means milligram per milliliter, "L" means liter(s), "nm" means nanometers, "mM" means millimolar, "M" means molar, "mmol" means
millimole(s), "µηιο ΐ" means micromole(s), "g" means gram(s), "^g" means microgram(s) and "ng" means nanogram(s), "PCR" means polymerase chain reaction, "OD" means optical density, "OD600" means the optical density measured at a wavelength of 600 nm, "kDa" means kilodaltons, "g" can also mean the gravitation constant, "bp" means base pair(s), "kbp" means kilobase pair(s), "kb" means kilobase, "%" means percent, "% w/v" means weight/volume percent, "% v/v" means volume/volume percent, "HPLC" means high performance liquid chromatography, "g/L" means gram per liter, "^g/L" means microgram per liter, "ng^L" means nanogram per microliter, "pmol^L" means picomol per microliter, "RPM" means rotation per minute, "^mol/min/mg" means micromole per minute per milligram, "w/v" means weight per volume, "v/v" means volume per volume.

Example 1

Selection of Potential Isobutyraldehyde Dehydrogenases for Screening

This example describes the basis for the selection of several ADH candidate enzymes for identifying efficient isobutyraldehyde dehydrogenases. Clostridium acetobutylicum Butanol Dehydrogenase A and B (BdhA and BdhB) were chosen for analysis based on the literature evidence. Achromobacter xylosoxidans was selected by enriching an environmental sludge sample on medium containing 1-butanol. The organism was then cultured and used to purify protein fraction that contained butanol dehydrogenase activity, subsequent to which the gene corresponding to the Secondary Alcohol Dehydrogenase B (SadB) was cloned as described in U.S. Patent Application Publication No. US 2009-0269823 Al. The horse-liver ADH enzyme (HLADH) is commercially available and was reported to have isobutanol oxidation activity by Green et al. in J. Biol. Chem. 268:1192 (1993).

Desirable properties of an ideal isobutyraldehyde dehydrogenase candidate for the isobutanol production pathway have been described above. An extensive literature search identified those candidate ADH enzymes with either a high k_cat and/or low K_M values for isobutyraldehyde or other closely-related aldehydes, or with a lower k_cat and/or higher K_M for isobutanol or other closely-related alcohols. Protein BLAST searches against nonredundant protein sequence database (nr) at
NCBI were performed using horse liver ADH, Achromobacter xylosoxidans SadB, and Saccharomyces cerevisiae ADH6 as queries, respectively. All the BLAST hits were collected and combined, from which sequences with more than 95% sequence identity to each other were removed. Multiple sequence alignment (MSA) was created from the set of remaining 95%-nonredundant sequences and a phylogenetic tree was generated from the MSA using the neighbor-joining method. Similarly, MSA and phylogenetic tree were generated separately for a number of selected ADH enzymes to identify closely-related homologs of each enzyme where the alignment consisted of only the BLAST hits obtained using the target enzyme as the query. These enzymes included Achromobacter xylosoxidans SadB, Saccharomyces cerevisiae ADH6, and Saccharomyces cerevisiae ADH7. Based on these analyses several candidates were selected (Table 3) for evaluation of performance.

Example 2

Cloning, Protein Expression and Purification, and Screening for a Suitable Isobutyraldehyde Dehydrogenase

This example describes preparation of ADH-gene constructs for over-expression/purification and measurement of enzyme activities using a time-course assay. Horse-liver ADH (HLADH; A-6128) was purchased from Sigma. Achromobacter xylosoxidans SadB (SadB), Saccharomyces cerevisiae ADH6 (ScADH6) and ADH7 (ScADH7), Entamoeba histolytica ADH1 (EhADH1), Bos Taurus Aldehyde Reductase (BtARD), Beijerinckia indica subsp. Indica ATCC 9039 (BiADH), Clostridium beijerinckii ADH (CbADH), Rana perezi ADH8 (RpADH8), Rattus norvegicus ADH1 (RnADHI), Thermus sp. ATN1 ADH (TADH), Phenylobacterium zucineum HLK1 ADH (PzADH), Methylocella silvestris BL2 ADH (MsADH), Acinetobacter baumannii AYE ADH (AbADH), Geobacillus sp. WCH70 ADH (GbADH), Vanderwaltozyma polyspora DSM 70294 ADH (VpADH), Mucor circinelloides ADH (McADH), and Rhodococcus erythropolis PR4 ADH (ReADH) were the candidates for which subclones were prepared for protein expression and purification.

Construction of Plasmid Constructs Expressing ADH Candidates
The gene-coding regions of EhADH1, BtARD, CbADH, BiADH, and RpADH8 were synthesized by DNA 2.0 (Menlo Park, CA) and those of RnADH1, TADH, PzADH, MsADH, AbADH, GbADH, VpADH, McADH, and ReADH were synthesized by GENEART AG (Germany) after optimizing the codons for expression in *Escherichia coli*. The amino-acid sequences for these candidates were procured from the Genbank Protein database and provided to DNA 2.0 or Geneart AG for codon optimization.

Each coding region was flanked by Xhol and Kpnl sites at the 5' and 3' ends of the coding sequence, respectively. These constructs were cloned and supplied in either DNA 2.0's vector pJ201 or Geneart's pMA vector.

The plasmids were transformed into chemically competent TOP10 cells (Invitrogen) and amplified by growing the transformants in liquid LB media containing either 25mg/ml Kanamycin or 100 mg/ml Ampicillin. The plasmids, which were purified from overnight cultures (grown at 37°C), were restricted with Xhol (NEB; R0146) and Kpnl (NEB; R0142) and ligated into the corresponding sites in-frame with an N-terminal hexa-histidine tag in the vector pBADHisA (Invitrogen; V43001) using the DNA ligation kit Version 2.1 from Takara Bio Inc. (6022).

The ligation products were transformed into chemically competent TOP10 cells (Invitrogen; C4040-50). The transformed cells were streaked on a plate containing the LB medium plus 100 mg/mL ampicillin. Clones containing the ADH inserts were confirmed by restriction digestion with Xhol/Kpnl. Plasmids with the correct insert contained the expected 1.2 kbp band in each case. The cloned sequence was confirmed via DNA sequencing. The resulting clones were named as pBADHisA::EhADH1, pBADHisA::BtARD, pBADHisA::CbADH, pBADHisA::BiADH, pBADHisA::RpADH8, pBADHisA::RnADH1, pBADHisA::TADH, pBADHisA::PzADH, pBADHisA::MsADH, pBADHisA::AbADH, pBADHisA::GbADH, pBADHisA::VpADH, pBADHisA::McADH, and pBADHisA::ReADH, respectively.

SadB, an enzyme which was previously examined, was PCR-amplified with KOD polymerase enzyme (Novagen), as per the procedure mentioned in the product manual, from pTrc99a::SadB using primers SadBXhoI-f (CCATGGAATCTCGAGATGAAAGCTCTGGTTTACC, SEQ ID NO: 41) and SadBKpnI-r (GATCCCCGGGTACCGAGCTCGAATTC, SEQ ID NO: 42) to introduce
Xhol and Kpnl sites at the 5' and 3' ends, respectively. After confirmation of the PCR product via agarose-gel electrophoresis, the 1.2-kb PCR product was restricted with Xhol and Kpnl and cloned into pBADHisA as described above for the other candidate genes. The genes for ScADH6 and ScADH7 were each amplified from 100 ng of genomic DNA of the yeast wild-type strain BY4741 (ATCC 201388) using primers ADH6_XhoI_f (CAAGAAAAACTCGAGATCATGTCTTATCCTGAG, SEQ ID NO: 43) and ADH6_KpnI_r (GAGCTTGTTACCTAGTCTGAATAATTTTTG, SEQ ID NO: 44) for ScADH6 and ADH7_XhoI_f (CTGAAAAACTCGAGAAAAATGGCTTTACCC, SEQ ID NO: 45) and ADH7_KpnI_r (GAAAAATATTAGGTACCTAGACTATTTATGG, SEQ ID NO: 46) for ScADH7. The strategy and PCR conditions were identical to those used for the amplification of SadB. The genes were then cloned into the Xhol and Kpnl sites of pBADHisA, as per the procedure described above. The plasmids containing SadB, ScADH6 and ScADH7 were labeled as pBADHisA::SadB, pBADHisA::ScADH6 and pBADHisA::ScADH7, respectively.

Expression of Recombinant ADHs in *E. coli*

For the data shown, either BL21-CodonPlus (Invitrogen; 230240) or a proprietary *E. coli* strain were used for the overexpression of ADH enzymes. However, it is believed that commercially available strains, such as BL21-codon plus, are suitable for overexpression of ADH enzymes.

Expression plasmids (pBADHisA plasmids) containing ADH genes were prepared from 3-mL overnight cultures of Top10 transformants using Qiaprep spin miniprep kit (Qiagen, Valencia CA; 27106) following manufacturer's instructions. One ng of each of the plasmid was transformed into either BL21-CodonPlus or proprietary *E. coli* electro-competent cells using a Bio RAD Gene Pulser II (Bio-Rad Laboratories Inc, Hercules, CA) by following the manufacturer's directions. The transformed cells were spread onto agar plates containing the LB medium plus 100 µg/mL of each of ampicillin and spectinomycin. The plates were incubated overnight at 37°C. Colonies from these plates inoculated in 3.0 mL of the LB medium containing 100 µg/mL of each of ampicillin and spectinomycin, at 37°C while shaking at 250 rpm. Cells from these starter cultures (grown overnight) were used to inoculate 1-L media at a dilution of 1:1000. The cells
were induced with 0.02% Arabinose after the culture reached an OD of -0.8. The induction was carried out at 37°C while shaking at 250 rpm overnight. The cells were then harvested by centrifugation at 4000 g for 10 min at 4°C. The cells were lysed by treatment with 40 ml of BugBuster master mix (Novagen; 71456-4), in the presence of Complete, EDTA-free Protease Inhibitor Cocktail tablets (Roche; 11873580001) and 1 mg/ml Lysozyme, by placing on a rocker at 4°C for 30 min. The cell debris was removed by centrifugation at 16,000 g for 20 min at 4°C.

The total protein concentration in samples was measured by the Bradfords Assay using Bradford's dye concentrate (Bio-Rad). The samples and protein standards (Bovine Serum Albumin, BSA) were set up in either individual cuvettes (1-mL reactions) or a 96-well microplate following the manufacturer's protocol. The concentrations of proteins were calculated from absorbance values at 595 nm, measured using either a Cary 100 Bio UV-Visible spectrophotometer (Varian, Inc.) or a SpectraMax plate reader (Molecular Devices Corporation, Sunnyvale, CA).

**ADH Enzyme Purification and Activity Assays**

Cell-free extracts prepared from 1-litre cultures as per the procedure described above, was directly used to purify the various expressed ADH enzymes via IMAC (immobilized metal affinity chromatography) affinity chromatography on 5-mL HisTrap FF columns (GE Healthcare Life Sciences; 175255-01). The entire procedure was carried out using an AKTAexplorer 10 S (GE Healthcare Life Sciences; 18-1145-05) FPLC system. The extracts were mixed with 30mM Imidazole and loaded onto the HisTrap columns. Upon loading, the column was washed with 50mM Sodium phosphate buffer, pH 8.0, containing 30mM Imidazole (approximately -10-20 column volumes) to get rid of unbound and non-specifically bound proteins. The ADH protein was then eluted with a gradient of 30mM to 500mM Imidazole over 20 column volumes. The peak fractions were electrophoresed on 10% Bis-Tris SDS-PAGE gels (Invitrogen; NP0301) using Invitrogen's XCell SureLock Mini-Gel apparatus (EI0001). Upon coomassie staining and destaining, it could be ascertained that the fractions were more than 95% pure and contained only the ADH protein. Activity assays were carried out to ensure that the purified proteins were active.
As a routine practice, the crude extracts and purified proteins were assayed for butanol oxidation activity, in order to ensure that the recombinant proteins were active throughout the purification process. In the reductive direction, isobutyraldehyde reduction assays were carried out with NADH or NADPH as the cofactor and an excess of the isobutyraldehyde substrate (40mM). In each case, enzymatic activity was measured for 1 min at 30°C in 1-ml reactions by following the decrease or increase in the absorbance at 340nm using a Cary Bio 100 UV-Visible spectrophotometer (Varian Inc.), depending on whether the NADH/NADPH is being consumed (absorbance is decreased) or generated (absorbance is increased) in the reaction. Alcohol oxidation activities were carried out in 50mM sodium phosphate buffer at pH 8.8 and aldehyde reduction reactions were assayed in 100mM potassium phosphate buffer at pH 7.0. Depending on the nature of reaction being carried out, the enzyme and cofactor stocks were diluted in the reaction buffers at the respective pHs. Either buffer or cell extract prepared from the proprietary E. coli strain (with no ADH plasmid) was used as the negative control for assays with purified protein and cell-free extracts, respectively.

In initial experiments, there were insufficient levels of protein expression with EhADH1 and RpADH8. Subsequently, the activity assays failed to detect ADH activity in the cell extracts expressing these enzymes. Likewise initially, although the BtARD showed good levels of protein expression and the protein could be purified to homogeneity, it had no detectable activity under the conditions used for the assay. It is believed that one of skill in the art could further optimize expression and assay conditions for these candidates. Sufficient amounts of active protein could be purified with all other enzymes for which data are presented. Cofactor specificities were measured with all these enzymes in isobutyraldehyde reduction reactions (as in proc mentioned above), using either NADH or NADPH as cofactors. In each case, at least a 10-fold difference was observed in the activity numbers, when either NADH or NADPH was used as a cofactor, as against the number corresponding to the other form of the cofactor. Table 6 summarizes the cofactor preferences for some of the ADH enzymes.
Table 6

<table>
<thead>
<tr>
<th>CANDIDATE ADH</th>
<th>COFACTOR PREFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horse-liver ADH</td>
<td>NADH</td>
</tr>
<tr>
<td>Saccharomyces cerevisiae ADH6</td>
<td>NADPH</td>
</tr>
<tr>
<td>Saccharomyces cerevisiae ADH7</td>
<td>NADPH</td>
</tr>
<tr>
<td>Achromobacter xylosidans SadB</td>
<td>NADH</td>
</tr>
<tr>
<td>Beijerickia indica ADH</td>
<td>NADH</td>
</tr>
<tr>
<td>Clostridium beijerinckii ADH</td>
<td>NADPH</td>
</tr>
<tr>
<td>Rattus norvegicus ADH1</td>
<td>NADH</td>
</tr>
<tr>
<td>Thermus sp. ATN1 ADH</td>
<td>NADH</td>
</tr>
<tr>
<td>Phenyllobacterium zucineum HLK1 ADH</td>
<td>NADH</td>
</tr>
<tr>
<td>Methylocella silvestris BL2 ADH</td>
<td>NADH</td>
</tr>
<tr>
<td>Acinetobacter baumannii AYE ADH</td>
<td>NADH</td>
</tr>
<tr>
<td>Geobacillus sp. WCH70 ADH</td>
<td>NADPH</td>
</tr>
<tr>
<td>Mucor circinelloides ADH</td>
<td>NADH</td>
</tr>
</tbody>
</table>

Screening Purified ADH Candidates Using a Semi-Physiological Time-Course Assay

The ideal way to characterize and compare various ADH candidates would be to calculate and compare the full set of kinetic constants, i.e., $k_{cat}$ values for aldehyde reduction and alcohol oxidation, $K_M$ values for isobutyraldehyde, isobutanol, NAD(P) and NAD(P)H, and $\frac{3}{4}$ values for isobutyraldehyde and isobutanol. A detailed characterization for numerous candidates would require considerable expenditure of time, effort and money. Thus, a qualitative assay was developed to allow for quick and efficient comparison of several candidates. A semi-physiological assay was designed to compare the performance of various enzymes. The assays entail the initiation of all reactions with a constant amount of each enzyme. In this case, 1 ug of each enzyme was used to initiate...
reactions that contained isobutyraldehyde and NADH at concentrations 1 mM and 200 µM, respectively. Each reaction's time course was followed for 10 min by measuring the decrease in absorbance at 340nm, as the reaction proceeds towards equilibrium. An enzyme with a high \( k_{\text{cat}} \) would drive the reaction towards equilibrium faster than an enzyme with a lower \( k_{\text{cat}} \). A parallel assay was also carried out under identical conditions, but with the inclusion of 321 mM isobutanol (24 g/L) in the reaction. An enzyme that is relatively uninhibited by this concentration of isobutanol would have a time course that closely mimics the time course in the absence of isobutanol. Figure 1 compares time courses exhibited by the ADH candidate enzymes in these assays.

Based on the results presented in Figure 1, it is inferred that the Beijerickia indica ADH is likely to have the highest \( k_{\text{cat}} \) for the isobutyraldehyde reduction reaction and ADH6 is likely to be the least inhibited by isobutanol in the reaction.

Example 3

Identification of Beijerinckia indica ADH With a High \( k_{\text{cat}} \) and a Low \( K_M \) for Isobutyraldehyde

Kinetic constants of the ADH enzymes were calculated and compared to identify those candidate ADH enzymes with the most desirable properties for the conversion of isobutyraldehyde to isobutanol in the last step of the engineered pathway for isobutanol production. The assays for determining the kinetic constants were carried out using initial rates from the assays described above. Decreases in NADH can be correlated with aldehyde being consumed (Biochemistry by Voet and Voet, John Wiley & Sons, Inc.) However, the amount of a given enzyme used in the reaction was in the range of 0.1 to 5 µg. The concentration of a given enzyme was such that it was conducive for the measurement of initial velocities over a 1-min time course. For each enzyme, Michaelis-Menten plots were generated with a broad range of substrate concentrations. Rough estimates of \( K_M \) were obtained, based on which the assays were redesigned so as to use substrate concentrations in the range 0.5 to 10 times the \( K_M \) value, to be able to obtain the appropriate kinetic constants. Isobutyraldehyde (isobutanal) reduction reactions were carried out at 30°C in 100 mM Potassium phosphate buffer, pH 7.0, containing 200µM NADH. When calculating the ¾ for isobutanol, the same reactions were carried out in the presence of varying concentrations of isobutanol (generally 0-535 mM) in the reaction.
Reactions with isobutanol substrate were performed at 30°C in 50 mM Sodium phosphate buffer, pH 8.8, containing 7.5 mM NAD. The Enzyme kinetics module (Version 1.3) of SigmaPlot 11 (Systat Software, Inc.) was used to fit data to Michaelis-Menten equations and calculate the kinetic constants. Kinetic constants obtained for the indicated ADH enzymes are given in Table 7. The $k_{cat}/K_M$ is derived from the individual numbers of $k_{cat}$ and $K_M$ and not an experimentally determined value. The ratios of the $K_M$, $K_i$, and $k_{cat}/K_M$ for each candidate enzyme as compared to the same parameter for SadB are given in Table 9.

### Table 7

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>$k_{cat}$ (sec$^{-1}$)</th>
<th>$K_M$ (Isobutanal) mM</th>
<th>$K_i$ (Isobutanol) mM</th>
<th>$k_{cat}/K_M$</th>
<th>Other enzymatic properties and cofactor preference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLADH*</td>
<td>8</td>
<td>0.1</td>
<td>2</td>
<td>82</td>
<td>[Isobutanol oxidation: $k_{cat}=5$sec$^{-1}$; $K_M=0.4$mM]</td>
</tr>
<tr>
<td>SadB*</td>
<td>109</td>
<td>1</td>
<td>180</td>
<td>105</td>
<td>$K_M$ (NADH) = 0.02mM [Isobutanol oxidation: $k_{cat}=2$sec$^{-1}$; $K_M=24$mM]</td>
</tr>
<tr>
<td>ScADH6</td>
<td>47</td>
<td>0.6</td>
<td>1170</td>
<td>81</td>
<td>NADPH specific</td>
</tr>
<tr>
<td>ScADH7*</td>
<td>36</td>
<td>0.3</td>
<td>88</td>
<td>120</td>
<td>NADPH specific</td>
</tr>
<tr>
<td>BiADH</td>
<td>283</td>
<td>0.2</td>
<td>36</td>
<td>1252</td>
<td>$K_M$ (NADH) = 0.06mM [Isobutanol oxidation: $k_{cat}=9$sec$^{-1}$; $K_M=4.7$mM]</td>
</tr>
<tr>
<td>CbADH</td>
<td>123</td>
<td>1.5</td>
<td>ND</td>
<td>85</td>
<td>NADPH specific</td>
</tr>
<tr>
<td>TADH</td>
<td>15</td>
<td>1.3</td>
<td>ND</td>
<td>11</td>
<td>NADH specific</td>
</tr>
<tr>
<td>RnADH1</td>
<td>$\sim5$</td>
<td>$\leq0.003$</td>
<td>ND</td>
<td>$\sim1667$</td>
<td>NADH specific</td>
</tr>
</tbody>
</table>

[0252] For those enzymes marked with an asterisk in Table 7, at least 3 assays were performed with separate preparations of the enzyme. All other numbers are values from either one assay or are averages from 2 assays performed with the same enzyme sample.

[0253] The data for *Beijerickia indica* ADH (BiADH) shows the highest number for the $k_{cat}$ and a reasonably high $k_{cat}/K_M$ and is preferred. The enzyme RnADH1 appears to
have a low $K_M$ value for isobutyraldehyde and consequently may have a high catalytic efficiency. However, the low $K_M$ value precludes an accurate determination of its $K_M$ value via spectrophotometric assays. Nevertheless, the enzyme's performance in the isobutanol production host may be limited more by the $k_{cat}$ if the intracellular steady-state levels of isobutyraldehyde are in excess of its $K_M$ value. Comparing BiADH with SadB, the former's catalytic efficiency for isobutyraldehyde reduction is $\sim 12$ times more than that of the latter although it is more sensitive to isobutanol than SadB. With regard to the nucleotide cofactor, SadB has a lower $K_M$ value for NADH when compared with BiADH. ScADH6 has a high $\frac{1}{4}$ value for isobutanol, indicating that this enzyme is likely to function in vivo, unfettered by the presence of isobutanol at concentrations that are expected in an isobutanol production host. Among the candidates analyzed so far, SadB has the least catalytic efficiency for isobutanol oxidation ($\frac{k_{cat}}{K_M} = 0.083$), followed by BiADH (1.91) and HLADH (12.5).

Example 4

Seven additional candidate ADH enzymes were synthesized, expressed, and assayed according to methods such as described in Example 2. Kinetic constants obtained for the indicated ADH enzymes (Phenylobacterium zucineum HLKI ADH (PzADH), Methylocella silvestris BL2 ADH (MsADH), Acinetobacter baumannii AYE ADH (AbADH), Geobacillus sp. WCH70 ADH (GbADH), and Mucor circinelloides ADH (McADH)) are given in Table 8. A comparison of $K_M$, $K_I$, and $\frac{k_{cat}}{K_M}$ for each candidate enzyme as compared to the same parameter for SadB are given in Table 9 as a percentage of the values determined (Table 7) for SadB. Percentages less than 100 indicate a value less than that determined for SadB; percentages higher than 100 indicate a value greater than that determined for SadB. There was no expression for Rhodococcus erythropolis PR4 ADH (ReADH) and no detectable activity for Vanderwaltozyma polyspora DSM 70294 ADH (VpADH) in these assays.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>$k_{cat}$ (sec$^{-1}$)</th>
<th>$K_M$ (Isobutanal) (mM)</th>
<th>$K_I$ (Isobutanol) (mM)</th>
<th>$\frac{k_{cat}}{K_M}$</th>
<th>Other enzymatic properties and cofactor preference</th>
</tr>
</thead>
</table>

- 75 -
<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Indicated parameter as a percentage of the same parameter determined for SadB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k_{\text{cat}}$</td>
</tr>
<tr>
<td>HLADH</td>
<td>7%</td>
</tr>
<tr>
<td>SadB</td>
<td>100%</td>
</tr>
<tr>
<td>ScADH6</td>
<td>43%</td>
</tr>
<tr>
<td>ScADH7</td>
<td>33%</td>
</tr>
<tr>
<td>BiADH</td>
<td>260%</td>
</tr>
<tr>
<td>CbADH</td>
<td>113%</td>
</tr>
<tr>
<td>TADH</td>
<td>14%</td>
</tr>
<tr>
<td>RnADH1</td>
<td>5%</td>
</tr>
<tr>
<td>PzADH</td>
<td>28%</td>
</tr>
<tr>
<td>MsADH</td>
<td>30%</td>
</tr>
<tr>
<td></td>
<td>AbADH 91%</td>
</tr>
<tr>
<td>-------</td>
<td>-----------</td>
</tr>
<tr>
<td>GbADH</td>
<td>29%</td>
</tr>
<tr>
<td>McADH</td>
<td>138%</td>
</tr>
</tbody>
</table>

Example 5

**Construction of *S. cerevisiae* strain PNY2211**

[0255] PNY2211 was constructed in several steps from *S. cerevisiae* strain PNY1507 as described in U.S. Appl. No. 61/380,563, filed September 7, 2010, and in the following paragraphs. First the strain was modified to contain a phosphoketolase gene. Construction of phosphoketolase gene cassettes and integration strains was previously described in U.S. Appl. No. 61/356,379, filed June 18, 2010. Next, an acetolactate synthase gene (*alsS*) was added to the strain, using an integration vector previously described in U.S. Appl. No. 61/308,563. Finally, homologous recombination was used to remove the phosphoketolase gene and integration vector sequences, resulting in a scarless insertion of *alsS* in the intergenic region between pdclA::ilvD (a previously described deletion/insertion of the PDC1 ORF in U.S. Appl. No. 61/308,563) and the native TRX1 gene of chromosome XII. The resulting genotype of PNY2211 is MATa ura3A::loxP his3A pdc6A pdclA::P[PDC1]-DHAD[ilvD_Sm-PDClt-P[FBAI]-ALS][alsS_Bs-CYCl pdc5A::P[PDC5]-ADH]sadB_Ax-PDC5t gpd2A::loxP fra2A adh1A::UAS(PGK1)P[FBAI]-kivD_Ll(y)-ADHl.t.

[0256] A phosphoketolase gene cassette was introduced into PNY1507 by homologous recombination. The integration construct was generated as follows. The plasmid pRS423::CUP1-alsS+FBA-budA (as described in U.S. Publ. No. 2009/0305363 Al) was digested with *NotI* and *Xmal* to remove the 1.8 kb FBA-budA sequence, and the vector was religated after treatment with Klenow fragment. Next, the CUP1 promoter was replaced with a TEF1 promoter variant (M4 variant described by Nevoigt *et al*. Appl. Environ. Microbiol. 72(8): 5266-5273 (2006)) via DNA synthesis and vector construction service from DNA2.0 (Menlo Park, CA). The resulting plasmid, pRS423::TEF(M4)-alsS was cut with *Stul* and *MluI* (removes 1.6 kb portion containing part of the alsS gene and CYC1 termintor), combined with the 4 kb PCR product generated from pRS426::GPD-xpkI+ADH-eutD (SEQ ID NO: 81; the plasmid is described in U.S. Appl. No.
61/356,379) with primers N 1176 and N 1177 (SEQ ID NOs: 47 and 48, respectively) and an 0.8 kb PCR product DNA generated from yeast genomic DNA (ENOl promoter region) with primers N822 and N 1178 (SEQ ID NOs: 49 and 50, respectively) and transformed into S. cerevisiae strain BY4741 (ATCC# 201388; gap repair cloning methodology, see Ma and Botstein). Transformants were obtained by plating cells on synthetic complete medium without histidine. Proper assembly of the expected plasmid (pRS423::TEF(M4)-xpkl+EN01-eutD, SEQ ID No: 51) was confirmed by PCR using primers N821 and N 1115 (SEQ ID NOs: 52 and 53, respectively) and by restriction digest (BglI). Two clones were subsequently sequenced. The 3.1 kb TEF(M4)-xpkl gene was isolated by digestion with Sacl and NotI and cloned into the pUC19-URA3::ilvD-TRXl vector described in U.S. Appl. No. 61/356,379 (Clone A, cut with AflI). Cloning fragments were treated with Klenow fragment to generate blunt ends for ligation. Ligation reactions were transformed into E. coli Stbl3 cells, selecting for ampicillin resistance. Insertion of TEF(M4)-xpkl was confirmed by PCR using primers N 1110 and N 1114 (SEQ ID NOs: 54 and 55, respectively). The vector was linearized with AflI and treated with Klenow fragment. The 1.8 kb KpnI-Hindi genetin resistance cassette described in U.S. Appl. No. 61/356,379 was cloned by ligation after Klenow fragment treatment. Ligation reactions were transformed into E. coli Stbl3 cells, selecting for ampicillin resistance. Insertion of the geneticin cassette was confirmed by PCR using primers N160SeqF5 and BK468 (SEQ ID NOs: 56 and 57, respectively). The plasmid sequence is provided as SEQ ID NO: 58 (pUC19-URA3::pdcl::TEF(M4)-xpkl::kan).

The resulting integration cassette (pdcl::TEF(M4)-xpkl::KanMX::TRXl) was isolated (Ascl and Nael digestion generated a 5.3 kb band that was gel purified) and transformed into PNY1507 using the Zymo Research Frozen-EZ Yeast Transformation Kit (Cat. No. T2001). Transformants were selected by plating on YPE plus 50 µg/ml G418. Integration at the expected locus was confirmed by PCR using primers N886 and N1214 (SEQ ID NOs: 59 and 60, respectively). Next, plasmid pRS423::GALlp-Cre, encoding Cre recombinase, was used to remove the loxP-flanked KanMX cassette (vector and methods described in U.S. Appl. No. 61/308,563). Proper removal of the cassette was confirmed by PCR using primers oBP512 and N160SeqF5 (SEQ ID NOs: 61 and 62, respectively). Finally, the alsS integration plasmid described in U.S. Appl. No. 61/308,563 (pUC19-kan::pdcl::FBA-alsS::TRXl, clone A) was transformed into this
strain using the included geneticin selection marker. Two integrants were tested for acetolactate synthase activity by transformation with plasmids pYZ090AalsS and pBP915 (plasmids described in U.S. Appl. No. 61/308,563, transformed using Protocol #2 in "Methods in Yeast Genetics" 2005. Amberg, Burke and Strathern) and evaluation of growth and isobutanol production in glucose-containing media (methods for growth and isobutanol measurement are described in U.S. Appl. No. 61/308,563 and U.S. Publ. No. 2007/0092957 Al). One of the two clones was positive and was named PNY2218. An isolate of PNY2218 containing the plasmids pYZ090AalsS and pBP915 was designated PNY2209.

[0258] PNY2218 was treated with Cre recombinase and resulting clones were screened for loss of the xpkI gene and pUC19 integration vector sequences by PCR using primers N886 and N160SeqR5 (SEQ ID Nos: 59 and 56, respectively). This leaves only the alsS gene integrated in the pdcl-TRX1 intergenic region after recombination the DNA upstream of xpkI and the homologous DNA introduced during insertion of the integration vector (a "scarless" insertion since vector, marker gene and loxP sequences are lost, see Figure 9). Although this recombination could have occurred at any point, the vector integration appeared to be stable even without geneticin selection and the recombination event was only observed after introduction of the Cre recombinase. One clone was designated PNY2211.

Example 6

Construction of Saccharomyces cerevisiae strain PNY1540

[0259] The purpose of this example is to describe the construction of Saccharomyces cerevisiae strain PNY1540 from strain PNY2211. This strain was derived from CEN.PK 113-7D (CBS 8340; Centraalbureau voor Schimmelcultures (CBS) Fungal Biodiversity Centre, Netherlands) and is described in Example 5 above. PNY1540 contains a deletion of the sadB gene, from Achromobacter xylosidans, which had been integrated at the PDC5 locus in PNY2211. The deletion, which completely removed the entire coding sequence, was created by homologous recombination with a PCR fragment containing regions of homology upstream and downstream of the target gene and a URA3 gene for
selection of transformants. The URA3 gene was removed by homologous recombination to create a scarless deletion.

The scarless deletion procedure was adapted from Akada et al. 2006 Yeast v23 p399. The PCR cassette for the scarless deletion was made by combining four fragments, A-B-U-C, by overlapping PCR. The PCR cassette contained a selectable/counter-selectable marker, URA3 (Fragment U), consisting of the native CEN.PK 113-7D URA3 gene, along with the promoter (250bp upstream of the URA3 gene) and terminator (150bp downstream of the URA3 gene). Fragments A and C, each 500 bp long, corresponded to the 500 bp immediately upstream of the target gene (Fragment A) and the 3’ 500 bp of the target gene (Fragment C). Fragments A and C were used for integration of the cassette into the chromosome by homologous recombination. Fragment B (254 bp long) corresponded to the sequence immediately downstream of the target gene and was used for excision of the URA3 marker and Fragment C from the chromosome by homologous recombination, as a direct repeat of the sequence corresponding to Fragment B was created upon integration of the cassette into the chromosome. Using the PCR product ABUC cassette, the URA3 marker was first integrated into and then excised from the chromosome by homologous recombination. The initial integration deleted the gene, excluding the 3’ 500 bp. Upon excision, the 3’ 500 bp region of the gene was also deleted.

sadB Deletion

The four fragments for the PCR cassette for the scarless sadB deletion were amplified using Phusion High Fidelity PCR Master Mix (New England BioLabs; Ipswich, MA) and CEN.PK 113-7D genomic DNA as template for Fragment U and PNY1503 genomic DNA as template for Fragments A, B, and C. Genomic DNA was prepared with a Gentra Puregene Yeast/Bact kit (Qiagen; Valencia, CA). sadB Fragment A was amplified with primer oBP540 (SEQ ID NO: 63) and primer oBP835 (SEQ ID NO: 64), containing a 5’ tail with homology to the 5’ end of sadB Fragment B. sadB Fragment B was amplified with primer oBP836 (SEQ ID NO: 65), containing a 5’ tail with homology to the 3’ end of sadB Fragment A, and primer oBP837 (SEQ ID NO: 66), containing a 5’ tail with homology to the 5’ end of sadB Fragment U. sadB Fragment U was amplified with primer oBP838 (SEQ ID NO: 67), containing a 5’ tail with homology to the 3’ end
of sadB Fragment B, and primer oBP839 (SEQ ID NO: 68), containing a 5’ tail with homology to the 5’ end of sadB Fragment C. sadB Fragment C was amplified with primer oBP840 (SEQ ID NO: 69), containing a 5’ tail with homology to the 3’ end of sadB Fragment U, and primer oBP841 (SEQ ID NO: 70). PCR products were purified with a PCR Purification kit (Qiagen). sadB Fragment AB was created by overlapping PCR by mixing sadB Fragment A and sadB Fragment B and amplifying with primers oBP540 (SEQ ID NO: 63) and oBP837 (SEQ ID NO: 66). sadB Fragment UC was created by overlapping PCR by mixing sadB Fragment U and sadB Fragment C and amplifying with primers oBP838 (SEQ ID NO: 67) and oBP841 (SEQ ID NO: 70). The resulting PCR products were purified on an agarose gel followed by a Gel Extraction kit (Qiagen). The sadB ABUC cassette was created by overlapping PCR by mixing sadB Fragment AB and sadB Fragment UC and amplifying with primers oBP540 (SEQ ID NO: 63) and oBP841 (SEQ ID NO: 70). The PCR product was purified with a PCR Purification kit (Qiagen).

[0262] Competent cells of PNY2211 were made and transformed with the sadB ABUC PCR cassette using a Frozen-EZ Yeast Transformation II kit (Zymo Research, Orange, CA). Transformation mixtures were plated on synthetic complete media lacking uracil supplemented with 1% ethanol at 30°C. Transformants with a sadB knockout were screened for by PCR with primers Ura3-end (SEQ ID NO: 71) and oBP541 (SEQ ID NO: 72). A correct transformant was grown in YPE (1% ethanol) and plated on synthetic complete medium containing 5-fluoro-orotic acid (0.1%) at 30°C to select for isolates that lost the URA3 marker. The deletion and marker removal were confirmed by PCR with primers oBP540 (SEQ ID NO: 63) and oBP541 (SEQ ID NO: 72) using genomic DNA prepared with a YeaStar Genomic DNA Kit (Zymo Research). The absence of the sadB gene from the isolate was demonstrated by a negative PCR result using primers specific for the deleted coding sequence of sadB, oBP530 (SEQ ID NO: 73) and oBP531 (SEQ ID NO: 74). A correct isolate was selected as strain PNY1540 (BP 1746).

Example 7

Construction of a yeast shuttle vector carrying a gene encoding the B. indica ADH and a negative control vector
The plasmid pLH468 (SEQ ID NO: 75), as described in U.S. Publ. No. 2009/0305363 Al, is an E. coli/yeast shuttle vector that carries 3 chimeric genes encoding enzymes that comprise part of an isobutanol production pathway (dihydroxyacid dehydratase, aKIV decarboxylase and isobutanol dehydrogenase). The existing isobutanol dehydrogenase gene was replaced by the B. indica ADH using gap repair cloning methodology. The B. indica ADH coding region with suitable 5' and 3' flanking sequences was first obtained via DNA synthesis (DNA2.0, Menlo Park, CA) with yeast codon optimization. The sequence is provided (SEQ ID NO: 76). The vector pLH468 was linearized with Bsu36I and transformed along with the B. indica ADH (released from the supplier's cloning vector with EcoRI and BamHI) into yeast strain BY4741. Transformants were plated on synthetic complete medium without histidine (Teknova Cat. No. C3020). Plasmids were prepared from several transformants using a Zymoprep™ Yeast Plasmid Miniprep kit (Zymo Research Cat. No. D2004). PCR (with primers N1092 and N1093, SEQ ID NOs: 77 and 78) and restriction enzyme digestion (with KpnI) were used to confirm incorporation of BiADH in the intended location. This plasmid is referred to as pLH468::BiADH.

A second vector was constructed that eliminated the most of the original isobutanol dehydrogenase gene (hADH) from pLH468. This was done by releasing a 808 bp fragment via digestion with Bsu36I and PacI, filling in the ends of the DNA with Klenow fragment and re-ligating the vector. The ligation reaction was transformed into E. coli Stbl3 cells. Loss of the hADH gene was confirmed by EcoRI digestion of isolated plasmid cones. One successful clone was selected for the experiment described in Example 8, below. The plasmid is referred to as pLH468AhADH.

**Example 8**

Isobutanologen strains carrying BiADH display better glucose-dependent growth, higher glucose consumption and higher isobutanol titer and yield than control strains.

The plasmids pLH468::BiADH and pLH468AhADH were each transformed along with a second isobutanol pathway plasmid (pYZ090AalsS, U.S. Appl. No. 61/380,563) into PNY1540. Transformations were plated on synthetic complete medium lacking
histidine and uracil, containing 1% ethanol as carbon source. Several transformants were patched to fresh plates. After 48 hours, patches (3 of each strain) were used to inoculate synthetic complete medium (minus histidine and uracil) containing 0.3% glucose and 0.3% ethanol as carbon sources. After 24 hours, growth in this medium was similar for all replicates of both strains. Cultures were then sub-cultured into synthetic complete medium (minus histidine and uracil) containing 2% glucose and 0.05% ethanol as carbon sources. Cultures (starting optical density (OD) at 600 nm was 0.2, culture volume was 20 ml in 125 ml tightly-capped flasks) were incubated 48 hours. Samples were collected for HPLC analysis at the time of subculture and again after 48 hours. The final ODs were also determined. The average 48h OD for the BiADH strain was 3.3 (+/-0.1) compared to 2.37 (+/-0.07) for the no ADH control. Thus inclusion of BiADH increased OD by 39% under these conditions. Similarly, glucose consumption (assessed by HPLC compared to samples collected immediately after sub-culturing) was increased by 69% (81 +/-1 mM vs. 47.9 +/-0.6 mM). Isobutanol titers were 4-fold higher and molar yields (i.e. yield of isobutanol per mole of glucose consumed) were doubled as shown in table below. In the no ADH control strain, significant carbon from the isobutanol pathway accumulated as isobutyrate, indicating that aldehyde dehydrogenases were acting upon isobutyraldehyde.

<table>
<thead>
<tr>
<th>TITERS</th>
<th>Isobutanol (mM)</th>
<th>Isobutyrate (mM)</th>
<th>Isobutyraldehyde (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PNY1540/ pLH468::BiADH</td>
<td>32.3 (± 0.6)</td>
<td>10.9 (± 0.3)</td>
<td>ND</td>
</tr>
<tr>
<td>PNY1540/ pLH468ΔADH</td>
<td>6.2 (± 0.2)</td>
<td>18.4 (± 0.4)</td>
<td>2.1 (± 0.4)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MOLAR YIELDS</th>
<th>Isobutanol</th>
<th>Isobutyrate</th>
<th>Isobutyraldehyde</th>
</tr>
</thead>
<tbody>
<tr>
<td>PNY1540/ pLH468::BiADH</td>
<td>0.401 (±0.006)</td>
<td>0.135 (±0.005)</td>
<td>ND</td>
</tr>
<tr>
<td>PNY1540/ pLH468ΔADH</td>
<td>0.129 (±0.004)</td>
<td>0.384 (±0.004)</td>
<td>0.044 (±0.008)</td>
</tr>
</tbody>
</table>
WHAT IS CLAIMED IS:

1. A recombinant microbial host cell comprising:
   a biosynthetic pathway for production of a lower alkyl alcohol, the biosynthetic pathway
   comprising a substrate to product conversion catalyzed by a polypeptide with alcohol
   dehydrogenase activity and one or more of the following characteristics:
   
   (a) the $K_M$ value for isobutyraldehyde is lower for said polypeptide relative to a
   control polypeptide having the amino acid sequence of SEQ ID NO: 26;

   (b) the $K_i$ value for isobutanol for said polypeptide is higher relative to a control
   polypeptide having the amino acid sequence of SEQ ID NO: 26; and

   (c) the $k_{cat}/K_M$ value isobutyraldehyde for said polypeptide is higher relative to a
   control polypeptide having the amino acid sequence of SEQ ID NO: 26.

2. The recombinant microbial host cell of claim 1, wherein the biosynthetic pathway for
   production of a lower alkyl alcohol is a butanol, propanol, isopropanol, or ethanol biosynthetic
   pathway.

3. The recombinant microbial host cell of claim 1, wherein the polypeptide with alcohol
   dehydrogenase activity has at least 95% identity to the amino acid sequence of SEQ ID NO: 21,
   22, 23, 24, 25, 31, 32, 34, 35, 36, 37, or 38.

4. The recombinant microbial host cell of claim 1, wherein the polypeptide with alcohol
   dehydrogenase activity has the amino acid sequence of SEQ ID NO: 31.

5. The recombinant host cell of claim 1 wherein the polypeptide with alcohol
   dehydrogenase activity is encoded by a polynucleotide having at least 85% identity to a
   nucleotide sequence of SEQ ID NO: 1, 2, 3, 4, 5, 6, 11, 12, 14, 15, 16, or 17.

6. The recombinant microbial host cell of claim 1 wherein the polypeptide with alcohol
   dehydrogenase activity preferentially uses NADH as a cofactor.
7. The recombinant microbial host cell of claim 1, wherein said polypeptide having alcohol dehydrogenase activity catalyzes the conversion of isobutyraldehyde to isobutanol in the presence of isobutanol at a concentration of at least about 15 g/L.

8. The recombinant microbial host cell of claim 1, wherein the biosynthetic pathway for production of a lower alkyl alcohol is a butanol biosynthetic pathway.

9. The recombinant microbial host cell of claim 1 wherein the biosynthetic pathway for production of a lower alkyl alcohol is an isobutanol biosynthetic pathway comprising heterologous polynucleotides encoding polypeptides that catalyze substrate to product conversions for each step of the following steps:
   (a) pyruvate to acetolactate;
   (b) acetolactate to 2,3-dihydroxyisovalerate;
   (c) 2,3-dihydroxyisovalerate to α-ketoisovalerate;
   (d) α-ketoisovalerate to isobutyraldehyde; and
   (e) isobutyraldehyde to isobutanol;
   and wherein said microbial host cell produces isobutanol.

10. The recombinant microbial host cell of claim 1 wherein the biosynthetic pathway for production of a lower alkyl alcohol is an isobutanol biosynthetic pathway comprising heterologous polynucleotides encoding polypeptides that catalyze substrate to product conversions for each step of the following steps:
    (a) pyruvate to acetolactate;
    (b) acetolactate to 2,3-dihydroxyisovalerate;
    (c) 2,3-dihydroxyisovalerate to α-ketoisovalerate;
    (d) α-ketoisovalerate to isobutyryl-CoA;
    (e) isobutyryl-CoA to isobutyraldehyde; and
    (f) isobutyraldehyde to isobutanol;
    and wherein said microbial host cell produces isobutanol.

11. The recombinant microbial host cell of claim 1 wherein the biosynthetic pathway for production of a lower alkyl alcohol is an isobutanol biosynthetic pathway comprising
heterologous polynucleotides encoding polypeptides that catalyze substrate to product conversions for each step of the following steps:

(a) pyruvate to acetolactate;
(b) acetolactate to 2,3-dihydroxyisovalerate;
(c) 2,3-dihydroxyisovalerate to a-ketoisovalerate;
(d) a-ketoisovalerate to valine;
(e) valine to isobutylamine;
(f) isobutylamine to isobutyaldehyde; and
(g) isobutyaldehyde to isobutanol;
and wherein said microbial host cell produces isobutanol.

12. A recombinant microbial host cell comprising a biosynthetic pathway for the production of a lower alkyl alcohol and a heterologous polynucleotide encoding a polypeptide with alcohol dehydrogenase activity having at least 85% identity to the amino acid sequence of SEQ ID NO: 21, 22, 23, 24, 25, 31, 32, 34, 35, 36, 37, or 38.

13. The recombinant microbial host cell of claim 12, wherein the biosynthetic pathway for the production of a lower alkyl alcohol is a 2-butanol biosynthetic pathway comprising heterologous polynucleotides encoding polypeptides that catalyze substrate to product conversions for each of the following steps:

(a) pyruvate to alpha-acetolactate;
(b) alpha-acetolactate to acetoin;
(c) acetoin to 2,3-butanediol;
(d) 2,3-butanediol to 2-butanone; and
(e) 2-butanone to 2-butanol;
and wherein said microbial host cell produces 2-butanol.

14. The recombinant microbial host cell of claim 12, wherein the biosynthetic pathway for the production of a lower alkyl alcohol is a 1-butanol biosynthetic pathway comprises heterologous polynucleotides encoding polypeptides that catalyze substrate to product conversions for each of the following steps:

(a) acetyl-CoA to acetoacetyl-CoA;
(b) acetoacetyl-CoA to 3-hydroxybutyryl-CoA;
(c) 3-hydroxybutyryl-CoA to crotonyl-CoA;
(d) crotonyl-CoA to butyryl-CoA;
(e) butyryl-CoA to butyraldehyde; and
(f) butyraldehyde to 1-butanol;
and wherein said microbial host cell produces 1-butanol.

15. The recombinant host cell of claim 12 wherein said polypeptide having alcohol dehydrogenase activity comprises an amino acid sequence with at least 95% identity to the amino acid sequence of SEQ ID NO: 21, 22, 23, 24, 25, 27, 31, 32, 34, 35, 36, 37, or 38.

16. The recombinant host cell of claim 12 wherein said polypeptide having alcohol dehydrogenase activity comprises an amino acid sequence with at least 95% identity to the amino acid sequence of SEQ ID NO: 31.

17. The recombinant host cell of claim 1 or claim 12 wherein the genus of said host cell is selected from the group consisting of: Saccharomyces, Pichia, Hansenula, Yarrowia, Aspergillus, Kluyveromyces, Pachysolen, Rhodotorula, Zygosaccharomyces, Galactomyces, Schizosaccharomyces, Torulaspora, Debaryomyces, Williopsis, Dekkera, Kloekera, Metschnikowia, Issatchenkia, and Candida.

18. A method for producing isobutanol comprising:
   (a) providing a recombinant microbial host cell comprising an isobutanol biosynthetic pathway, the pathway comprising a heterologous polypeptide which catalyzes the substrate to product conversion of isobutyraldehyde to isobutanol wherein the polypeptide has at least 90% identity to the amino acid sequence of SEQ ID NO: 21, 22, 23, 24, 25, 27, 31, 32, 34, 35, 36, 37, or 38; and
   (b) contacting the host cell of (a) with a carbon substrate under conditions whereby isobutanol is produced.
19. The method of claim 18 wherein the heterologous polypeptide which catalyzes the substrate to product conversion of isobutyraldehyde to isobutanol has at least 95% identity to the amino acid sequence of SEQ ID NO: 31.

20. The method of claim 18 wherein the heterologous polypeptide which catalyzes the substrate to product conversion of isobutyraldehyde to isobutanol has the amino acid sequence of SEQ ID NO: 31.

21. A method for producing 2-butanol comprising:
   (a) providing a recombinant microbial host cell comprising a 2-butanol biosynthetic pathway, the pathway comprising a heterologous polypeptide having at least 90% identity to the amino acid sequence of SEQ ID NO: 21, 22, 23, 24, 25, 27, 31, 32, 34, 35, 36, 37, or 38; and
   (b) contacting the host cell of (a) with a carbon substrate under conditions whereby 2-butanol is produced.

22. The method of claim 21 wherein the heterologous polypeptide has at least 95% identity to the amino acid sequence of SEQ ID NO: 31.

23. A method for producing 1-butanol comprising:
   (a) providing a recombinant microbial host cell comprising a 1-butanol biosynthetic pathway, the pathway comprising a heterologous polypeptide having at least 90% identity to the amino acid sequence of SEQ ID NO: 21, 22, 23, 24, 25, 27, 31, 32, 34, 35, 36, 37, or 38; and
   (b) contacting the host cell of (a) with a carbon substrate under conditions whereby 1-butanol is produced.

24. The method of claim 23 wherein the heterologous polypeptide has at least 95% identity to the amino acid sequence of SEQ ID NO: 31.
Figure 1a

Achromobacter xylosoxidans SadB

Horse-liver ADH
C

Saccharomyces cerevisiae ADH6

Absorbance at 340nm

· NADPH Blank
- ScADH6
- ScADH6+iBuOH

Time (min)

D

Saccharomyces cerevisiae ADH7

Absorbance at 340nm

· NADPH Blank
- ScADH7
- ScADH7+iBuOH

Time (min)

Figure 1b
**Figure 1c**

**Beijerinckia indica ADH**

- NADH Blank
- BiADH
- BiADH+iBuOH

**Clostridium beijerinckii ADH**

- Blank
- CbADH
- CbADH+iBuOH
**G**

*Rattus norvegicus* ADH1

![Graph showing absorbance at 340nm over time for Rattus norvegicus ADH1 with different conditions: Blank, RnADH1, RnADH1+iBuOH.](image)

**H**

*Thermus sp. ATN1* ADH

![Graph showing absorbance at 340nm over time for Thermus sp. ATN1 ADH with different conditions: Blank, TADH, TADH+iBuOH.](image)

Figure 1d
Time course assays for measuring Isobutanol inhibition

Reaction conditions:
[Isobutyraldehyde] = 1 mM
[Isobutanol] = 0 or 342 mM
[NADH] = 200 μM
[NAD] = 0 μM
pH = 7.0

Figure 2
Figure 3
Figure 4
Figure 7

**A**

Michaelis-Menten Plots for ADH6

\[ k_{cat} = 43 \text{ sec}^{-1} \]
\[ K_m (\text{IBA}) = 0.37 \text{ mM} \]
\[ K_i (\text{IBOH}) = 1170 \text{ mM} \]

**B**

Michaelis-Menten Plots for BiADH

\[ k_{cat} = 295 \text{ sec}^{-1} \]
\[ K_m (\text{IBA}) = 0.20 \text{ mM} \]
\[ K_i (\text{IBOH}) = 36 \text{ mM} \]
Figure 8
pdc1::ilvD::FBA-alsS::trx1 A locus

6355 bp

Figure 9