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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 March 2013 (07.03.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/033604 A2

(51) **International Patent Classification**:
C12N 1/21 (2006.01) *C12N 15/60* (2006.01)
C12N 15/52 (2006.01) *C12P 7/16* (2006.01)
C12N 15/53 (2006.01)

(21) **International Application Number**:
PCT/US2012/053457

(22) **International Filing Date**:
31 August 2012 (31.08.2012)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
61/529,701 31 August 2011 (31.08.2011) US
61/585,099 10 January 2012 (10.01.2012) US

(71) **Applicant** *(for all designated States except US)*: **THE TRUSTEES OF DARTMOUTH COLLEGE** [US/US]; 11 Rope Ferry Road, Room 6210, Hanover, NH 03755 (US).

(72) **Inventors; and**
(75) **Inventors/Applicants** *(for US only)*: **BHANDIWAD, Ashwini** [IN/US]; 65 Drake Lane, West Lebanon, NH (US). **SHAW, Arthur, Josephus** [US/US]; 95 Olde Farms Road, Grantham, NH 03753 (US). **VAN DER VEEN, Douwe** [NL/NL]; Jearnne d'Arcdreef 58, NL-3561 SE Utrecht (NL). **LYND, Lee, R.** [US/US]; 181 Bean Road, Meriden, NH 03770 (US).

(74) **Agent**: **BARKLEY, Sam, J.**; Lathrop & Gage LLP, 4845 Pearl East Circle, Suite 201, Boulder, CO 80301 (US).

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

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

Published:

— *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))*

WO 2013/033604 A2

(54) **Title**: PRODUCTION OF BUTANOLS IN THERMOPHILIC ORGANISMS

(57) **Abstract**: This disclosure relates to the production of butanols using thermophilic microorganisms.

PRODUCTION OF BUTANOLS IN THERMOPHILIC ORGANISMS

RELATED APPLICATIONS

This application claims the benefit of priority of U.S. Provisional Application Serial No. 61/529,701 filed August 31, 2011 and U.S. Provisional Application Serial No. 61/558,099 filed January 10, 2012 which are both incorporated herein by reference in their entirety.

U.S. GOVERNMENT RIGHTS

This disclosure was made in part with the support of the U.S. Government; the U.S. Government has certain rights in this disclosure as provided for by the terms of Grant #-DE-AC05-00OR-22725 awarded by the Department of Energy.

BACKGROUND

Global energy demand is predicted to increase to 739 quadrillion Btu by 2035 that is approximately a 1.4% increase annually. Out of this total demand, approximately 27% accounts for transportation energy, which, in turn, accounts for 15 billion metric tons of carbon dioxide emissions due to transportation, which is increasing 1.2% annually. Thus the need for sustainable and renewable green energy in the form of biofuels is gaining interest. Most of the existing production of biofuels is for ethanol. However, when compared to ethanol, n-butanol and isobutanol provide several advantages as a fuel source. Ethanol has an energy density of 19 MJ/L while n-butanol has an energy density of 27MJ/L, which is comparable to gasoline with 32 MJ/L. Besides offering higher energy density, n-butanol has a vapor pressure 11 times lower than ethanol. Thus, existing gasoline infrastructure for storage and transportation is conducive for use with n-butanol. Isobutanol offers all the same advantages as n-butanol, but has the additional advantage of having a higher octane number than n-butanol and lower toxicity to organisms used for biosynthesis of the chemical.

There are essentially two ways by which higher alcohols may be produced through biochemical pathways, a catabolic pathway such as the fermentative clostridial pathway and an anabolic pathway known as the Ehrlich pathway. No natural or engineered thermophilic microorganism has been shown to produce significant amounts of butanol.

SUMMARY

The present disclosure provides for genetically engineering an operon that contains all of the genes necessary to produce butanol from acetyl-CoA at thermophilic temperatures. This operon may be transformed into a thermophilic organism for the production of n-butanol at thermophilic temperatures. An n-butanol operon may be transformed into *Thermoanaerobacterium saccharolyticum* and *Clostridium thermocellum* as an expression plasmid or integrated into the native DNA to create thermophilic organisms that are able, either alone or in combination, to produce n-butanol from cellulosic feedstocks.

The clostridial fermentative pathway of *Clostridium acetobutylicum* is outlined in FIG. 1. The five main genes that encode proteins involved in the conversion of acetyl-CoA to butanol are *thl, hbd, crt, bcd* and *adhe* which respectively encode thiolase (Thl), β-hydroxybutyryl CoA dehydrogenase (Hbd), crotonase (Crt), butyryl CoA dehydrogenase (Bcd) that has an electron transfer flavoprotein subunit A+ subunit B (etfAB) complex and aldehyde-alcohol dehydrogenase (Adhe), a gene that expresses a bi-functional enzyme. Thl, Hbd and Crt are able to catalyze their reactions under aerobic conditions, however Bcd and Adhe catalyze their reactions under anaerobic conditions. Thl catalyzes the condensation reaction between two molecules of acetyl CoA for the formation of acetoacetyl CoA. Hbd catalyzes the conversion of acetoacetyl CoA to β-hydroxybutyryl CoA and is a NADH dependant reaction. Crt catalyzes a dehydration reaction that forms crotonyl CoA from hydroxybutyryl CoA. The Bcd+etfAB complex, is a ferredoxin coupled reaction involving two molecules of NADH that catalyzes the reaction of crotonyl CoA to butyryl CoA. Adhe is a NAD(P)H dependent bifunctional enzyme which catalyzes the reduction of butyryl CoA to butyraldehyde and finally to butanol.

T. saccharolyticum is a thermophilic, anaerobic, gram positive bacteria and is capable of growing on almost all sugars from cellulosic biomass. It can degrade complex carbohydrates, xylan, starch, mannan and other pentoses. *T. saccharolyticum* also produces xylanase enzymes. *T. saccharolyticum* ferments xylan as well as almost all soluble biomass sugars. *T. saccharolyticum* does not ferment cellulose and has a branched fermentation pathway with three primary organic compound end products; lactic acid, acetic acid, and ethanol. In order to be suitable for the production of butanols, the central metabolism must be shunted away from making organic acids and ethanol and towards making butanols. This may be achieved by the introduction of heterologous DNA that encodes enzymes that constitute a metabolic pathway for the production of butanols from

acetyl-CoA into chromosomal regions of the host that disrupt the natural production of the three primary organic compound end products of branched fermentation. Thus, instead of expressing enzymes that would shunt the metabolic flux into the branched fermentation pathway native to *T. saccharolyticum*, those enzymes are now replaced with enzymes that channel the flux of metabolites towards the production of butanols.

C. thermocellum is also a thermophilic anaerobic, gram positive bacteria and is able to grow on cellulose or other cellulose containing biomass feedstocks including wood chips, tall grasses, corn stover (residual corn stalks) and sugar cane bagasse. *C. thermocellum* has one of the highest growth rates on cellulose, but does not ferment pentoses such as xylose, and grows poorly on glucose.

In one embodiment, *C. thermocellum* is genetically engineered to efficiently use pentoses and hexoses (e.g., glucose) as well as cellulose for the production of butanols to create a consolidated bioprocessing organism.

In one embodiment, *T. saccharolyticum* engineered for the production of butanols is used with thermophilic cellulases from *C. thermocellum* in the same reaction container since a co-culture of *C. thermocellum* and *T. saccharolyticum* can completely use all carbohydrates found in cellulose biomass and other biomass for the efficient production of butanols through fermentation.

In another embodiment, the disclosure provides for engineering a thermophilic α-KDC into an operon encoding enzymes of the anabolic pathway for the production of butanols. This α-KDC operon may be transformed into *T. saccharolyticum* and *C. thermocellum* as an expression plasmid or integrated into the native DNA to create thermophilic organisms that are able, either alone or in combination, to produce isobutanol from non-cellulosic and/or cellulosic feedstocks.

The Ehrlich pathway is outlined in FIG. 3 and involves the conversion of amino acids into butanols via transamination and decarboxylation reactions. Amino acids that are assimilated by the Ehrlich pathway (valine, leucine, isoleucine, methionine, and phenylalanine) are taken up slowly throughout fermentation. After the initial transamination reaction, the resulting keto acid cannot be redirected into central carbon metabolism. Before keto acids are excreted into the growth medium, microorganisms convert them into alcohols or acids via the Ehrlich pathway. Thus, the production of butanols via the Ehrlich pathway requires the use of a α-ketoacid decarboxylase (α-KDC) and an alcohol dehydrogenase (ADH). Some of the α-KDCs have broad substrate ranges, whereas others are more specific. Therefore, only two steps are needed to produce

biofuels from amino acid biosynthesis pathways to alcohol production, see FIG. 3.

In one embodiment, a recombinant thermophilic microorganism is disclosed which contains a biochemical pathway to produce butanol from a suitable carbon substrate, wherein the biochemical pathway produces acetoacetyl-CoA as an intermediate. In one aspect, the biochemical pathway may contain at least one heterologous polypeptide that is introduced into a parental thermophilic microorganism in order to obtain the disclosed recombinant thermophilic microorganism.

In another embodiment, the disclosed recombinant thermophilic microorganism may further contain elevated expression level of a polypeptide having thiolase activity, wherein the elevated level is in reference to the parental microorganism, and wherein the recombinant microorganism produces an intermediate comprising acetoacetyl-CoA from a substrate comprising acetyl-CoA.

In another embodiment, the polypeptide having thiolase activity may be encoded by a polynucleotide having at least about 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or higher identity to the DNA sequence as set forth in SEQ ID NO:1. In another aspect, the polypeptide having thiolase activity may have at least about 50%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or higher identity to the amino acid sequence as set forth in SEQ ID NO:2.

In another embodiment, the recombinant microorganism may further contain elevated expression of a polypeptide having beta-hydroxybutyryl-CoA dehydrogenase activity, wherein the elevated expression is in reference to the parental microorganism, and wherein the recombinant microorganism produces a metabolite comprising beta-hydroxybutyryl-CoA from a substrate comprising acetoacetyl-CoA.

In another embodiment, the polypeptide having beta-hydroxybutyryl-CoA dehydrogenase activity is encoded by a polynucleotide having at least about 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or higher identity to a sequence as set forth in SEQ ID NO:3. In another aspect, a polypeptide having beta-hydroxybutyryl-CoA dehydrogenase activity is disclosed which has at least about 50%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or higher identity to an amino acid sequence such as SEQ ID NO:4.

In another embodiment, the recombinant thermophilic microorganism may have elevated expression of a polypeptide having crotonase activity, as compared to the parental microorganism, wherein the recombinant microorganism produces a metabolite

comprising crotonyl-CoA from a substrate comprising beta-hydroxybutyryl-CoA. The polypeptide having crotonase activity may be encoded by a polynucleotide having at least about 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or higher identity to a sequence as set forth in SEQ ID NO:5. A polypeptide having crotonase activity is also disclosed which may have at least about 50%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or higher identity to an amino acid sequence such as SEQ ID NO:6.

In another embodiment, the recombinant thermophilic microorganism may have elevated expression of a polypeptide having butyryl-CoA dehydroxygenase activity, wherein the elevated expression is in reference to the parental microorganism, and wherein the recombinant microorganism produces a metabolite comprising butyryl-CoA from a substrate comprising crotonyl-CoA. The polypeptide having butyryl-CoA dehydrogenase activity may be encoded by a polynucleotide having at least about 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or higher identity to a sequence as set forth in SEQ ID NO:7, SEQ ID NO:9, or SEQ ID NO:11. A polypeptide having butyryl-CoA dehydrogenase activity is also disclosed which has at least about 50%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or higher identity to an amino acid sequence such as set forth in SEQ ID NO:8, SEQ ID NO:10, or SEQ ID NO:12.

In another embodiment, the polypeptide having butyryl-CoA dehydrogenase activity is encoded by a bcd gene or homolog thereof, in combination with a etfA gene or homolog thereof and etfB gene or homolog thereof.

In another embodiment, the recombinant thermophilic microorganism has elevated expression of a polypeptide having aldehyde/alcohol dehydrogenase activity, wherein the elevated expression is in reference to the parental microorganism, wherein the recombinant microorganism produces a metabolite comprising butyraldehyde from a substrate comprising butyryl-CoA. In one aspect, the polypeptide having aldehyde/alcohol dehydrogenase activity is encoded by a polynucleotide having at least about 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or higher identity to a sequence as set forth in SEQ ID NO:13. A polypeptide having aldehyde/alcohol dehydrogenase activity is also disclosed which has at least about 50%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or higher identity to an amino acid sequence such as SEQ ID NO:14.

In another embodiment, the recombinant thermophilic microorganism may have

elevated expression of a polypeptide having aldehyde/alcohol dehydrogenase activity, wherein the elevated expression is in reference to the parental microorganism, and wherein the recombinant microorganism produces a metabolite comprising n-butanol from a substrate comprising butyraldehyde, and wherein the polypeptide having aldehyde/alcohol dehydrogenase activity is encoded by a polynucleotide having at least about 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or higher identity to a sequence as set forth in SEQ ID NO:13.

In another embodiment, the recombinant thermophilic microorganism may have elevated expression of a polypeptide having aldehyde/alcohol dehydrogenase activity, wherein the elevated expression is in reference to the parental microorganism, and wherein the recombinant microorganism produces a metabolite comprising n-butanol from a substrate comprising butyryl-CoA, and wherein the polypeptide having aldehyde/alcohol dehydrogenase activity is encoded by a polynucleotide having at least about 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or higher identity to a sequence as set forth in SEQ ID NO:13.

In another embodiment, the recombinant thermophilic microorganism may contain one or more deletions, disruptions or knockouts in a gene encoding an enzyme that catalyzes the conversion of acetyl-coA to n-butanol. In one aspect, the recombinant thermophilic microorganism may contain a deletion, disruption or knockout of adhE, its homologs or variants thereof. In another aspect, the microorganism may contain an increase in the conversion of acetyl-coA to n-butanol in comparison to the parent microorganism.

In another embodiment, a recombinant thermophilic microorganism is disclosed which contains a biochemical pathway to produce n-butanol from fermentation media of a suitable carbon substrate at a temperature of from about 40 °C to about 65 °C, wherein the recombinant biochemical pathway comprises elevated expression of: a) a thiolase as compared to a parental microorganism; b) a hydroxybutyryl-CoA dehydrogenase as compared to a parental microorganism; c) a crotonase as compared to a parental microorganism; d) a butyryl-CoA dehydrogenase as compared to a parental microorganism; and e) a butyraldehyde dehydrogenase as compared to a parental microorganism. Methods for producing n-butanol using such recombinant thermophilic microorganism is also disclosed wherein said method comprises providing to said recombinant thermophilic microorganism a suitable carbon substrate and recovering n-butanol produced by said microorganism.

In another embodiment, contacting said media with said recombinant thermophilic microorganism is performed as a batch fermentation, or as a continuous fermentation. In one aspect, the n-butanol produced by said recombinant thermophilic microorganism is recovered from said fermentation media by distillation, liquid-liquid extraction, adsorption, decantation, pervaporation or combinations thereof. In another aspect, fermentation media may include but are not limited to mixtures of carbon substrates selected from the group consisting of cellulosic feedstocks, hexoses and pentoses.

In another embodiment, a method is disclosed for producing isobutanol at temperatures from about 45 °C to about 65 °C, the method may include providing a fermentation media comprising carbon substrates, and contacting said media with a recombinant thermophilic microorganism expressing an engineered isobutanol biosynthetic pathway wherein said pathway comprises the following substrate to product conversions: i. α-ketoisovalerate to isobutyraldehyde, and ii. isobutyraldehyde to isobutanol, wherein the substrate to product conversion of step (i) is performed by a thermophilic alpha-ketodecarboxylase enzyme, and the substrate to product conversion of step (ii) is performed by a thermophilic alcohol dehydrogenase enzyme whereby isobutanol is produced.

DESCRIPTION OF FIGURES

FIG. 1: Schematic representation of the fermentative clostridial n-butanol pathway.

FIG. 2: Plasmid map of the plasmid harboring the n-butanol pathway for transformation into T. saccharolyticum and C. thermocellum.

FIG. 3: Ehrlich pathway. 1, transamination; 2, decarboxylation; 3, reduction; 4, oxidation.

FIG. 4: Results of the B-fitter program showing the amino acids in α-KDC from L. lactis to be mutated for conferring thermophilicity.

FIG. 5: α-KDC model highlighting the amino acid residues chosen to be saturated for library construction.

FIG. 6: Linear DNA assembled fragment for transformation into T. saccharolyticum strain M0355 as a growth screen for identifying thermophilic α-KDC mutants.

FIG. 7: Chromosomal integration of n-butanol pathway in xynA locus.

FIG. 8: Growth curves showing OD_{600}, substrate consumption and end product formation are depicted in A for wild type and in B for butanol producing strain I2B.

FIG. 9: Carbon balance for WT and strain I2B.

FIG. 10: Plasmid DNA map for transformation into T. saccharolyticum strain M3224 as a growth selection for identifying thermophilic α-KDC variants.

DETAILED DESCRIPTION

The present disclosure provides details related to the expression of one or more heterologous polynucleotides and/or the over-expression of one or more heterologous polynucleotides encoding; (i) a thermophilic polypeptide that catalyzes the production of acetoacetyl-CoA from two molecules of acetyl-CoA; (ii) a thermophilic polypeptide that catalyzes the conversion of acetoacetyl-CoA to 3-hydroxybutyryl-CoA; (iii) a thermophilic polypeptide the catalyzes the conversion of 3-hydroxybutyryl-CoA to crotonyl-CoA; (iv) a thermophilic polypeptide (or thermophilic polypeptide combination) that catalyzes the reduction of crotonyl-CoA to butyryl-CoA; and (v) a thermophilic polypeptide that preferentially catalyzes the conversion of butyryl-CoA to butyraldehyde and butyraldehyde to n-butanol. For example, the disclosure demonstrates that with over-expression of the heterologous thl, hbd, crt, bcd, etfAB, adhE genes in *T. saccharolyticum* and *C. thermocellum* the production of butanols can be obtained.

In another embodiment, the present disclosure provides details related to the production of isobutanol by using a thermophilic organism. Since no thermophilic α-KDC has been identified, the mesophilic *L. lactis* is mutated into a thermophilic α-KDC through site-directed mutagenesis. The site-directed mutants is chosen by using algorithms analyzing three dimensional model of *L. lactis* α-KDC to identify amino acid residues on the mesophilic α-KDC that will impart more stability into the α-KDC and thus make it stable and active at thermophilic temperatures. The thermophilic α-KDC is cloned in both *T. saccharolyticum* and *C. thermocellum* for use in making butanols, especially isobutanol. Native genes in the organisms that use some of the same intermediate metabolites necessary for the production of butanols through the Ehrlich pathway with be deleted or replaced with genes that express engineered proteins dedicated to the production of butanols. These deletions are intended to increase the level of intermediates available for the heterologous butanol pathways.

Definitions

Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the art to which this invention

belongs. The following references provide one of skill with a general definition of many of the terms used in this invention: Singleton et al., Dictionary of Microbiology and Molecular Biology (2nd ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5th Ed., R. Rieger et al. (eds.), Springer Verlag (1991); and Hale & Marham, The Harper Collins Dictionary of Biology (1991). As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

"Open Reading Frame", or "ORF," refers to a series of at least 25 contiguous codons.

"Permissible temperature" refers to a temperature at which a cell may grow and divide, or a protein is capable of retaining its tertiary structure and any innate enzyme activity, or enzymatic activity, the molecule may possess.

"Modulate" refers to the property of being able to quantitatively increase or decrease one or more chemical or physical characteristics of a molecule or process by at least 10% of the initial baseline characteristic in response to an environmental or metabolic change. Modulate may also refer to the ability to qualitatively alter a chemical or physical characteristic of a molecule or process in response to an environmental or metabolic change. Methods for determining modulation of chemical or physical characteristics of a molecule are well known in the art and include, but are not limited to, enzyme assays and spectroscopic analysis.

The term "elevate" refers to the quantitative increase of one or more chemical or physical characteristics of a molecule or process by at least 10% of the initial baseline characteristic in response to an environmental or metabolic change. The term elevate may also refer to an increase in the level of a metabolic product of a biochemical pathway, a product of enzyme catalysis, or any increase in a molecule, gene, polypeptide, nucleic acid, amino acid or other substance versus a beginning state, level or measurement. The term "elevate" can be used in comparison to the level of a substrate, product, metabolite, polypeptide, polynucleotide and/or gene with respect to a wild type microorganism, a parental microorganism, and/or a recombinant microorganism. In some cases the term "elevate" may include a decrease in the level of a particular substrate, product, metabolite, polypeptide, polynucleotide and/or gene with respect to a parental microorganism, but still having an elevated level of the particular substrate, product, metabolite, polypeptide, polynucleotide and/or gene with respect to a wild type microorganism or other non-genetically engineered or non-recombinant microorganism.

The terms "enzyme activity" and "enzymatic activity" are used interchangeably herein. A reference to "displaying (an enzyme activity or enzymatic activity)" refers to a molecular characteristic where a biomolecule such as a protein or nucleic acid catalyzes a chemical reaction. Exemplary enzyme or enzymatic activities are displayed by. The genes that encode for proteins involved in the conversion of acetyl-CoA to butanol which are *thl, hbd, crt, bcd* and *adhe* and which respectively encode for thiolase (Thl), β-hydroxybutyryl CoA dehydrogenase (Hbd), crotonase (Crt), butyryl CoA dehydrogenase (Bcd) that has an electron transfer flavoprotein subunit A+ subunit B (etfAB) complex and aldehyde-alcohol dehydrogenase (Adhe), a gene that expresses a bi-functional enzyme.

The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers, those containing modified residues, and non-naturally occurring amino acid polymer.

The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function similarly to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, γ-carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, e.g., an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, e.g., homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs may have modified R groups (e.g., norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions similarly to a naturally occurring amino acid.

"Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical or associated, e.g., naturally contiguous, sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode most proteins. For instance, the codons GCA, GCC, GCG and GCU all encode the

amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to another of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes silent variations of the nucleic acid. One of skill in the art will recognize that in certain contexts each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, often silent variations of a nucleic acid that encodes a polypeptide is implicit in a described sequence with respect to the expression product, but not with respect to actual probe sequences.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention. Typically conservative substitutions for one another: 1) Alanine (A), Glycine (G); 2) Aspartic acid (D), Glutamic acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W); 7) Serine (S), Threonine (T); and 8) Cysteine (C), Methionine (M).

"Homologous," in relation to two or more peptides, refers to two or more sequences or subsequences that have a specified percentage of amino acid residues that are the same (i.e., about 60% identity, preferably about 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or higher identity over a specified region, when compared and aligned for maximum correspondence over a comparison window or designated region) as measured using a BLAST or BLAST 2.0 sequence comparison algorithms with default parameters described below, or by manual alignment and visual inspection (see, e.g., NCBI web site http://www.ncbi.nlm.nih.gov/BLAST/ or the like). The definition of homologous also includes sequences that have deletions and/or additions, as well as those that have substitutions, as well as naturally occurring, e.g.,

polymorphic or allelic variants, and man-made variants. As described below, the preferred algorithms can account for gaps and the like. Preferably, identity exists over a region that is at least about 25 amino acids in length, or more preferably over a region that is 50-100 amino acids in length.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Preferably, default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, Adv. Appl. Math. 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, J. Mol. Biol. 48:443 (1970), by the search for similarity method of Pearson & Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by manual alignment and visual inspection (see, e.g., Current Protocols in Molecular Biology (Ausubel et al., eds. 1995 supplement)).

Examples of algorithms that are suitable for determining percent sequence identity and sequence similarity include the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al., Nuc. Acids Res. 25:3389-3402 (1977) and Altschul et al., J. Mol. Biol. 215:403-410 (1990). BLAST and BLAST 2.0 are used, with the parameters described herein, to determine percent sequence identity for the nucleic acids and proteins of the invention. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/).

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul, Proc. Nat'l. Acad. Sci. USA 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a peptide is considered similar to a reference sequence if the smallest sum probability in a

comparison of the test peptide to the reference peptide is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001. Log values may be large negative numbers, e.g., 5, 10, 20, 30, 40, 40, 70, 90, 110, 150, 170, etc.

The terms "sequence similarity", "sequence identity", or "percent identity" in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are, when optimally aligned with appropriate nucleotide or amino acid insertions or deletions, the same or have a specified percentage of amino acid residues or nucleotides that are the same (i.e., 50% identity, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or higher identity to an amino acid sequence such as SEQ ID NO:2, or a nucleotide sequence such as SEQ ID NO:1), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. This definition also refers to the compliment of a test sequence. Preferably, the identity exists over a region that is at least about 25 amino acids or nucleotides in length, or even over a region that is at least about 50-100 amino acids or nucleotides in length. These relationships hold, notwithstanding evolutionary origin (Reeck et al., Cell, 50:667 (1987)). When the sequence identity of a pair of polynucleotides or polypeptides is greater or equal to 65%, the sequences are said to be "substantially identical."

Alternatively, substantial identity will exist when a nucleic acid will hybridize under selective hybridization conditions, to a strand or its complement. Typically, selective hybridization will occur when there is at least about 55% homology over a stretch of at least about 14 nucleotides, more typically at least about 65%, 75%, 85% and or even at least about 90%. See, Kanehisa, Nuc. Acids Res., 12:203-213 (1984), which is incorporated herein by reference. The length of homology comparison, as described, may be over longer stretches, and in certain embodiments will be over a stretch of at least about 17 nucleotides, generally at least about 20 nucleotides, ordinarily at least about 24 nucleotides, usually at least about 28 nucleotides, typically at least about 32 nucleotides, more typically at least about 40 nucleotides, 50 nucleotides, and or even at least about 75 to 100 or more nucleotides.

Macromolecular structures such as polypeptide structures can be described in terms of various levels of organization. For a general discussion of this organization, see, e.g., Alberts et al., Molecular Biology of the Cell (3rd ed., 1994) and Cantor and Schimmel, Biophysical Chemistry Part I. The Conformation of Biological

Macromolecules (1980). "Primary structure" refers to the amino acid sequence of a particular peptide. "Secondary structure" refers to locally ordered, three-dimensional structures within a polypeptide. These structures are commonly known as domains. Domains are portions of a polypeptide that form a compact unit of the polypeptide and are typically about 5 to 350 amino acids long. Typical domains are made up of organized sections of peptide such as stretches of β strands (that can interact to form β sheets) and α helices. "Tertiary structure" refers to the complete three-dimensional structure of a polypeptide monomer. "Quaternary structure" refers to the three dimensional structure formed by the non-covalent association of independent tertiary units. A "random coil structure," when referring to the structure of a protein or peptide indicates a lack of higher level (secondary or tertiary) structure, or a relatively disorganized structural sequence between secondary structural motifs, such as β-sheets and α-helices.

The term "recombinant" when used with reference, e.g., to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, e.g., recombinant cells express genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all. By the term "recombinant nucleic acid" herein is meant nucleic acid, originally formed in vitro, in general, by the manipulation of nucleic acid, e.g., using polymerases and endonucleases, in a form not normally found in nature. In this manner, operably linkage of different sequences is achieved. Thus an isolated nucleic acid, in a linear form, or an expression vector formed in vitro by ligating DNA molecules that are not normally joined, are both considered recombinant for the purposes of this invention. It is understood that once a recombinant nucleic acid is made and reintroduced into a host cell or organism, it will replicate non-recombinantly, i.e., using the *in vivo* cellular machinery of the host cell rather than *in vitro* manipulations; however, such nucleic acids, once produced recombinantly, although subsequently replicated non-recombinantly, are still considered recombinant for the purposes of the invention. Similarly, a "recombinant protein" is a protein made using recombinant techniques, i.e., through the expression of a recombinant nucleic acid as depicted above.

The term "operably linked" refers to a linkage of polynucleotide elements in a functional relationship. With regard to the present invention, the term "operably linked" refers to a functional linkage between a nucleic acid expression control sequence (such as

a promoter, or an array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence. Thus, a nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence.

The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, γ-carboxyglutamate, and o-phosphoserine. "Amino acid analog" refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, i.e., a carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, e.g., homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (e.g., norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that function in a manner similar to a naturally occurring amino acid.

Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

The term "amino acid sequence" refers to the positional relationship of amino acid residues as they exist in a given polypeptide or protein.

The term "coding sequence", in relation to nucleic acid sequences, refers to a plurality of contiguous sets of three nucleotides, termed codons, each codon corresponding to an amino acid as translated by biochemical factors according to the universal genetic code, the entire sequence coding for an expressed protein, or an antisense strand that inhibits expression of a protein. A "genetic coding sequence" is a coding sequence where the contiguous codons are intermittently interrupted by non-coding intervening sequences, or "introns." During mRNA processing intron sequences are removed, restoring the contiguous codon sequence encoding the protein or anti-sense strand.

The term "contiguous" in the context of polynucleotide or polypeptide sequences, refers to an uninterrupted sequence of bases or amino acids, each base or amino acid being immediately adjacent to its neighbors in the sequence.

The terms "expression vector" and "expression cassette" include any type of genetic construct containing a nucleic acid capable of being transcribed in a cell. The expression vectors of the invention generally supply sequence elements directing translation of the coding sequence into a protein of the present invention, as provided by the invention itself, although vectors used for the amplification of nucleotide sequences (both coding and non-coding) are also encompassed by the definition. In addition to the coding sequence, expression vectors will generally include restriction enzyme cleavage sites and the other initial, terminal and intermediate DNA sequences that are usually employed in vectors to facilitate their construction and use. The expression vector can be part of a plasmid, virus, or nucleic acid fragment.

The term "fusion gene" refers to the combination of one or more heterologous coding sequences joined in frame to form a single translational/transcriptional unit. Typically the heterologous coding sequences are joined end-to-end. The definition however includes fusion genes where one sequence, or fragment thereof, intervenes in another heterologous sequence.

The term "heterologous" when used with reference to a nucleic acid or protein molecule or fragment thereof indicates that the molecule is non-native to the organism into which it has been introduced.

The terms "primers" or "primer pairs" refer to oligonucleotide probes capable of recognizing and hybridizing to specific nucleotide sequences found in a target gene or sequence to be amplified by polymerase chain reaction (PCR). The degree of complementarity required between the primers and the target sequence determines the specificity, or stringency of conditions required for hybridization of the sequences. A temperature of about 36 °C is typical for low stringency amplification, although annealing temperatures may vary between about 32.degree. C. and about 4 °C depending on primer length. For high stringency PCR amplification, a temperature of about 62 °C is typical, although high stringency annealing temperatures can range from about 50 °C to about 65 °C, depending on the primer length and specificity. Typical cycle conditions for both high and low stringency amplifications include a denaturation phase of about 90 °C to about 95 °C for 30 sec-2 min., an annealing phase lasting 30 sec.-2 min., and an extension phase of about 72 °C for 1-2 min. Protocols and guidelines for low and high stringency

amplification reactions are provided, e.g., in Innis et al., PCR Protocols, A Guide to Methods and Applications, Academic Press, Inc. N.Y. (1990)).

The term "Regulatory sequences" refers to those sequences, both 5' and 3' to a structural gene, that are required for the transcription and translation of the structural gene in the target host organism. Regulatory sequences include a promoter, ribosome binding site, optional inducible elements and sequence elements required for efficient 3' processing, including polyadenylation. When the structural gene has been isolated from genomic DNA, regulatory sequences also include those intronic sequences required to remove of the introns as part of mRNA formation in the target host.

As used herein, the term "metabolically engineered" or "metabolic engineering" involves rational pathway design and assembly of biosynthetic genes, genes associated with operons, and control elements of such polynucleotides, for the production of a desired metabolite, such as an acetoacetyl-CoA or higher alcohol, in a microorganism. A metabolite is a substance produced by metabolism such as a substance necessary for or taking part in a particular metabolic process. A metabolite can be a substrate or product of a biochemical pathway. Metabolically engineered can further include optimization of metabolic flux by regulation and optimization of transcription, translation, protein stability and protein functionality using genetic engineering and appropriate culture condition including the reduction of, disruption, or knocking out of, a competing metabolic pathway that competes with an intermediate leading to a desired pathway. A biosynthetic gene can be heterologous to the host microorganism, either by virtue of being foreign to the host, or being modified by mutagenesis, recombination, and/or association with a heterologous expression control sequence in an endogenous host cell. In one aspect, where the polynucleotide is xenogenetic to the host organism, the polynucleotide can be codon optimized.

The term "biochemical pathway", also referred to as "biosynthetic pathway" or "metabolic pathway", refers to a set of biochemical reactions for converting (transmuting) a chemical species into another chemical species. Gene products belong to the same "metabolic pathway" if they, in parallel or in series, act on the same substrate, produce the same product, or act on or produce a metabolic intermediate (i.e., metabolite) between the same substrate and metabolite end product. The present disclosure provides an isolated thermophilic microorganism that contains a biochemical pathway to produce butanol from a suitable carbon substrate. For purposes of this disclosure, when a living organism (e.g., a microorganism) contains a biochemical pathway, it means that such an organism

contains the one or more functional units required to carry out the set of biochemical reactions for converting at least one chemical species into at least another chemical species. In one embodiment, such functional units are proteins, which may be native or non-native to the thermophilic microorganism. The term "native" is used to refer to a protein (or gene) or fragment thereof that naturally exists in the microorganism. However, for purpose of this disclosure, a native protein (or or its coding gene) may contain one or more mutations in its amino acid sequence (or in the case of a gene, one or more mutations in its nucleotide sequence) as compared to a homologous protein (or gene) in a naturally existing organism. The term "non-native" is used to refer to a protein (or a gene) or fragment thereof that does not naturally exist in the microorganism. It is to be understood that the terms "native" and "non-native" are relative because a protein that is non-native to an organism may be native to another organism.

The term "substrate" or "suitable substrate" refers to any substance or compound that is converted or meant to be converted into another compound by the action of an enzyme. The term includes not only a single compound, but also combinations of compounds, such as solutions, mixtures and other materials which contain at least one substrate, or derivatives thereof. Further, the term "substrate" encompasses not only compounds that provide a carbon source suitable for use as a starting material, such as any biomass derived sugar, but also intermediate and end product metabolites used in a pathway associated with a metabolically engineered microorganism as described herein.

A "biomass derived sugar" includes, but is not limited to, molecules such as glucose, sucrose, mannose, xylose, and arabinose. The term biomass derived sugar encompasses suitable carbon substrates ordinarily used by microorganisms, such as 5 and 6 carbon sugars, including, but not limited to, glucose, lactose, sorbose, fructose, idose, galactose and mannose in either D or L form, or a combination of 6 carbon sugars, such as glucose and fructose, and/or 6 carbon sugar acids including, but not limited to, 2-keto-L-gulonic acid, idonic acid (IA), gluconic acid (GA), 6-phosphogluconate, 2-keto-D-gluconic acid, 5-keto-D-gluconic acid, 2-ketogluconatephosphate, 2,5-diketo-L-gulonic acid, 2,3-L-diketogulonic acid, dehydroascorbic acid, erythorbic acid and D-mannonic acid.

The term "n-butanol" generally refers to a straight chain isomer with the alcohol functional group at the terminal carbon. The straight chain isomer with the alcohol at an internal carbon is sec-butanol or 2-butanol. The branched isomer with the alcohol at a terminal carbon is isobutanol, and the branched isomer with the alcohol at the internal

carbon is tert-butanol. The term "butanol" or "butanols" refers to all of the above isomers either individually or in combination.

Metabolically or genetically "engineered" or "modified" microorganisms are produced via the introduction of genetic material into a host or parental microorganism of choice thereby modifying or altering the cellular physiology and biochemistry of the microorganism. The genetic material introduced into the parental microorganism contains gene(s), or parts of genes, coding for one or more of the enzymes involved in a biosynthetic pathway may also include additional elements for the expression and/or regulation of expression of genes, e.g. promoter sequences.

An engineered, recombinant microorganism, or modified microorganism can also include, in the alternative or in addition to, the introduction of a genetic material into a host or parental microorganism, the disruption, deletion or knocking out of a gene or polynucleotide to alter the cellular physiology and biochemistry of the microorganism. Through the reduction, disruption or knocking out of a gene or polynucleotide, the microorganism acquires new or improved properties (e.g., the ability to produced a new or greater quantities of an intracellular metabolite, improve the flux of a metabolite down a desired pathway, and/or reduce the production of undesirable by-products).

A "metabolite" refers to any substance produced by metabolism, or a substance necessary for or taking part in a particular metabolic process. A metabolite can be an organic compound that is a starting material (e.g., glucose or pyruvate), an intermediate (e.g., acetyl-coA) in, or an end product (e.g., butanols) of metabolism. Metabolites can be used to construct more complex molecules, or they can be broken down into simpler ones. Intermediate metabolites may be synthesized from other metabolites, perhaps used to make more complex substances, or broken down into simpler compounds, often with the release of chemical energy.

Culture conditions suitable for the growth and maintenance of a recombinant microorganism provided herein are described in the examples below. The skilled artisan will recognize that such conditions can be modified to accommodate the requirements of each microorganism. Appropriate culture conditions useful in producing a butanol product comprise conditions of culture medium pH, ionic strength, nutritive content, etc.; temperature; oxygen, carbon dioxide, nitrogen content; humidity; and other culture conditions that permit production of the compound by the host microorganism, i.e., by the metabolic action of the microorganism. Appropriate culture conditions are well known for microorganisms that can serve as host cells.

The term "thermophilic" or "thermostable" refers to a temperature range of from about 45 °C to about 65 °C.

The term "hyperthermophilic" refers to a temperature range of from about 70 °C to about 110 °C.

The term "mesophilic" refers to a temperature range of from about 20 °C to about 40 °C.

It is understood that a range of microorganisms can be modified to include a recombinant thermophilic or hyperthermophilic metabolic pathway suitable for the production of butanols. It is also understood that various microorganisms can act as "sources" for genetic material encoding target enzymes suitable for use in a recombinant microorganism provided herein.

The term "microorganism" includes prokaryotic and eukaryotic microbial species from the domains Archaea, Bacteria and Eucarya, the latter including yeast and filamentous fungi, protozoa, algae, or higher Protista. The terms "microbial cells" and "microbes" are used interchangeably with the term microorganism.

The term "recombinant microorganism" and "recombinant host cell" are used interchangeably herein and refer to microorganisms that have been genetically modified to express or over-express endogenous polynucleotides, or to express non-endogenous sequences, such as those included in a vector. The polynucleotide generally encodes a target enzyme involved in a metabolic pathway for producing a desired metabolite as described above, but may also include protein factors necessary for regulation or activity or transcription. Accordingly, recombinant microorganisms described herein have been genetically engineered to express or over-express target enzymes not previously expressed or over-expressed by a parental microorganism. It is understood that the terms "recombinant microorganism" and "recombinant host cell" refer not only to the particular recombinant microorganism but to the progeny or potential progeny of such a microorganism.

The term "parental microorganism" refers to a wild type or otherwise non-genetically modified microorganism with respect to a particular gene, plasmid, or other recombinant nucleic acids. A parental microorganism can be the microorganism that genetic engineering techniques are performed upon in order to create a recombinant microorganism.

The various components of an expression vector can vary widely, depending on the intended use of the vector and the host cell(s) in which the vector is intended to

replicate or drive expression. Expression vector components suitable for the expression of genes and maintenance of vectors in *T. saccharolyticum* and *C. thermocellum* are described herein.

Recombinant expression vectors contain at least one expression system, which, in turn, is composed of at least a portion of biosynthetic gene coding sequences operably linked to a promoter and optionally to termination sequences that operate to effect expression of the coding sequence in compatible host cells. The host cells are modified by transformation with the recombinant DNA expression vectors of the disclosure to contain the expression system sequences either as extrachromosomal elements or integrated into the chromosome.

Due to the inherent degeneracy of the genetic code, other nucleic acid sequences which encode substantially the same or a functionally equivalent amino acid sequence can also be used to clone and express the polynucleotides encoding such enzymes. As previously noted, the term "host cell" is used interchangeably with the term "recombinant microorganism" and includes any cell type which is suitable for producing e.g., butanols and susceptible to transformation with a nucleic acid construct such as a vector or plasmid.

A nucleic acid of the disclosure can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques and those procedures described in the examples section below. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to nucleotide sequences can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

It is also understood that an isolated nucleic acid molecule encoding a polypeptide homologous to the enzymes described herein can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence encoding the particular polypeptide, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into the polynucleotide by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. In contrast to those positions where it may be desirable to make a non-conservative amino acid substitutions (see above). In some positions it is preferable to make conservative amino acid substitutions.

Recombinant microorganisms provided herein can express a plurality of target

enzymes involved in pathways for the production of butanols from a suitable carbon substrate. Through the introduction of genetic material the parental microorganism acquires new properties, e.g. the ability to produce a new, or greater quantities of, an intracellular metabolite. In an illustrative embodiment, the introduction of genetic material into a parental microorganism results in a new or modified ability to produce butanols. Isomers of butanol include n-butanol, isobutanol, 2-butanol, and tert-butanol. Microorganisms provided herein are preferably modified to produce metabolites in quantities not available in the parental microorganism. In one embodiment the microorganisms produce n-butanol. In another embodiment, microorganisms produce isobutanol.

The term "continuous fermentation" refers to an open system wherein nutrient solution is added to the bioreactor continuously and an equivalent amount of converted nutrient solution with microorganisms, e.g., recombinant microorganisms, is simultaneously removed from the system. Continuous fermentation may refer to a process in which cells or micro-organisms are maintained in culture in the exponential growth phase by the continuous addition of fresh medium that is balanced by the removal of cell suspension from the bioreactor.

The term 'batch fermentation" or "fed-batch fermentation" is a batch process which is based on feeding of a growth limiting nutrient substrate to a culture. Batch production is a technique used in manufacturing, in which the object in question is created stage by stage over a series of workstations. Batch fermentation processes refer to the process that starts with inoculation of a reaction vessel with a microorganism and ends with the retrieval of the product and happens inside a single fermenter with no intermediate steps.

The term "pervaporation" as used herein generally refers to increasing the concentration of a metabolite, product or other molecule in a mixture by removing diluting material. Pervaporation can also refer to a method for the separation of mixtures of liquids by partial vaporization through a non-porous or porous membrane.

The butanols produced by a microorganism provided herein can be detected by any method known to the skilled artisan. Such methods include gas chromatography (GC) mass spectrometry (MS), nuclear magnetic resonance spectroscopy (NMR), capillary electrophoresis, GC/MS, or other spectroscopic methods alone or in combination with chromatographic techniques such as HPLC, for example.

n-Butanol Production in Thermophilic Microorganisms

The present disclosure relates to heterologous production of n-butanol in thermophilic organisms. Using thermophilic organisms and enzymes that grow and have optimum catalytic efficiency at about 45-65 °C instead of mesophilic organisms and enzymes that grow and have optimum catalytic efficiency at about 20-40 °C provides the advantages of reducing contamination issues, reducing cooling costs and providing higher rates of reaction to enable consolidated bio-processing from lignocellulosic and other cellulosic substrates.

In one embodiment of the present disclosure, a method for producing n-butanol is provided. The method includes culturing a recombinant microorganism as provided herein in the presence of a suitable carbon substrate and under conditions suitable for the conversion of the substrate to n-butanol.

In one embodiment, T. saccharolyticum engineered for n-butanol production would be used for the commercial production of n-butanol as a biofuel in conjunction with commercially available cellulases functioning at thermophilic temperatures for cellulosic biomass degradation. In another embodiment *T. saccharolyticum* could be used for the production of n-butanol by feeding it with xylan, starch containing feedstocks and many sugar hydrolysates which it can grow on.

C. thermocellum engineered for n-butanol production could be used as a standalone biocatalyst for conversion of cellulosic biomass to n-butanol biofuel all in one step. Since *C. thermocellum* cannot metabolize pentose sugars and since the engineered *T. saccharolyticum* containing the operon for production of n-butanol cannot metabolize cellulose, in one embodiment, a co-culture containing *C. thermocellum* and *T. saccharolyticum* could be used for increased production of n-butanol from cellulosic feedstocks.

Heterologous Expression of Individual Thermophilic Genes in the n-Butanol Pathway in *T. saccharolyticum*

The genes for the enzymes required to catalyze the five steps for n-butanol formation from acetyl CoA were cloned individually in *T. saccharolyticum* strain M0355. The genes for catalyzing the first four steps (*thl, hbd, crt, bcd+etfAB)* were taken from *Thermoanaerobacterium thermosaccharolyticum* DSM571, and for the last step, the gene (*adhE)* was taken from *Clostridium acetobutylicum* ATCC824. The mutants created using strain M0355 were Athl, Ahbd, Acrt, Abcd-etfAB and Aadh harboring the genes *thl, hbd, crt, bcd+etfAB and adhE.* The specific activities were measured for the first four mutants

and are listed in Table 1 below. The end product determination was determined for the last mutant using precursor substrate and is listed in Table 2.

The heterologous expression and activity of eight thermophilic enzymes involved in the butanol pathway from the pre-cursor acetyl-CoA is demonstrated. The genes were integrated individually into the host chromosome of *T. saccharolyticum* at the pta/ack locus. Each individual gene (or gene cluster when applicable) was driven by the *T. saccharolyticum* native promoter upstream of the *pta/ack* genes. The host strain M0355 used was a homo-ethanologen previously constructed from *T. sachharolyticum* in which the organic acid pathways have been knocked out. The genes encoding proteins that catalyze the first four steps of the pathway (thiolase, 3-hydroxybutyryl-CoA dehydrogenase, crotonase, butyryl-CoA dehydrogenase, etfA and etfB) were taken from the thermophilic anaerobe *T. thermosaccharolyticum* DSM571. However, the *adhE* gene was not very well annotated for C_4 specificity in *T. thermosaccharolyticum*, so the *C. acetobutylicum* ATCC824 *adhE* gene encoding aldehyde/alcohol dehydrogenase AdhE occurring on the sol operon was cloned instead.

T. saccharolyticum was transformed by electroporation as detailed below or without electroporation relying on the natural ability of *T. saccharolyticum* to take up and integrate foreign DNA. The linear DNA that was transformed into M0355 was PCR amplified from constructed plasmids in yeast, containing the upstream and downstream regions of the *T. sachharolyticum* pta/ack locus, gene of interest and kanamycin marker. The plasmid map used as templates for the linear DNA amplification and transformation are shown in FIG. 2.

The testing of functional enzyme activities used protocols as further detailed below. The wild type *T. thermosaccharolyticum* activities of *thl, hbd, crt* and *bcd+etfAB* complex were determined, see Table 1. Moreover, the heterologous activities of these enzymes were demonstrated in *T. saccharolyticum.* The mutants harboring the *thl, hbd, crt* and *bcd+etfAB* genes were Athl, Ahbd, Acrt and Abcd-etfAB, respectively. The enzyme activities were measured at 55°C and are shown in Table 1. The heterologous expression of the genes was driven by a strong tac promoter on a plasmid in *E. coli*. Although the native *T. saccharolyticum* pta/ack promoter was used to drive each of the individual genes of the butanol pathway, replacing the native ribosomal binding site by a consensus sequence (AGGAGG) increased each of the activities 6 – 10 fold.

The thermostability of the AdhE enzyme from *C. acetobutylicum* was verified and the *T. sachharolyticum* mutant harboring the *adhE* gene, Aadh, was tested for end product

formation for the functional activity of the enzyme by addition of 5mM butyryl CoA, the precursor substrate for AdhE, directly into the growth medium, and measuring the n-butanol formation by gas chromatrography (GC) as shown in Table 2.

Table 1: Specific activities of WT *T. thermosaccharolyticum*, M0355 mutants and M0355. Positive control in parenthesis.

Strain	**Enzyme activity (μmol.min^{-1}.mg^{-1} protein)**			
	Thl	***Hbd***	***crt***	***bcd+etfAB***
T. thermosaccharolyticum (positive control)	5.08 (±0.7)	14.43 (±0.65)	11.95 (±1.2)	0.64 (±0.04)
Athl	2.47 (±0.4)	-	-	-
Ahbd	-	6.8 (±0.39)	-	-
Acrt	-	-	2.56 (±0.19)	-
Abcd-etfAB	-	-	-	0.21 (±0.05)
M0355 (negative control)	Nd	Nd	nd	Nd

Table 2: End product determination for Aadh and M0355

Strain	**Pre-cursor substrate added butyryl CoA(mM)**	**End product formation n-butanol (mM)**
Aadh	5	2.3 (±0.2)
M0355 (negative control)	5	0.0

Enzyme Assays

The *T. thermosaccharolyticum* and *T. saccharolyticum* strains were grown in TSC Basel liquid medium or TSD Basel liquid medium at 55°C under anaerobic conditions. 100mL media was inoculated with 1% culture and grown for 14 hours with 200 μg/ml kanamycin where appropriate. The *thl, hbd and crt* assays were carried out aerobically, but anaerobic conditions were maintained for the *bcd+etfAB* assay. For preparation of the cell free extracts, 100mL cultures were centrifuged at 4 °C at 6,000g in a Beckman Coulter Avanti J-25 centrifuge using a JA-10 rotor. The cell pellet was resuspended in 4 mL of appropriate buffer as required for the assay and sonicated using a Misonix

Sonicator 4000 fitted with a microtip, in a 10 mL glass beaker for 8 min with 10s pulse on and 10s pulse off at 50% of the max intensity. Crude cell free extract was obtained by centrifugation at 14,000xg for 25 min and removing the cell debris. Protein concentration was measured using Bradford reagent (Bio-rad). All the subsequent measurements were done using an Agilent 8453 UV-vis spectrophotometer attached to a Peltier temperature controller. All the reactions were measured at 55°C.

The thiolase activity was measured in the reverse direction, with the acetoacetyl CoA cleavage measured at 303nm. The total reaction volume was 1.2mL with – 100mM Tris-HCl (pH 8.0), 1mM DTT, 10 mM $MgCl_2$, 50μM acetoacetyl CoA, 0.2 mM CoA and the cell-free extract. Addition of CoA started the reaction. The molar extinction coefficient used was 14,000 $M^{-1}cm^{-1}$(2)

The β-hydroxybutyryl CoA dehydrogenase activity was measured in the forward direction. Decrease in NADH was measured at 340nm. The total reaction volume was 1.2 mL – 50mM MOPS (pH 7.0), 1mM DTT, 0.2 mM NADH, 75μM acetoacetyl CoA and cell free extract. Addition of acetoacetyl CoA started the reaction. The molar extinction coefficient used was 6,220 $M^{-1}cm^{-1}$(2)

The crotonase activity was measured in the reverse direction. Decrease in crotonyl CoA was measured at 263nm. The total reaction mixture was 1.2mL – 100mM Tris HCl (pH 7.6), 0.15mM Crotonyl CoA and cell free extract. Addition of Crotonyl CoA started the reaction. The molar extinction coefficient used was 6,700 $M^{-1}cm^{-1}$

The butyryl CoA dehydrogenase/etfAB activity was measured in the forward direction. Decrease in NADH was measured at 340nm. The total reaction mixture was 1.2mL - 50 mM Tris-HCl (pH 7.5), 2 mM DTT, 0.1 mM NADH, 0.1 mM crotonyl-CoA, 5μM FAD, 20μM ferredoxin and cell free extract. Addition of crotonyl CoA started the reaction. The molar extinction coefficient used was 6,220 $M^{-1}cm^{-1}$.

The heterologous activity of the AdhE enzyme in the mutant Aadh was determined by end product formation using GC. 5mM of the precursor substrate butyryl CoA was introduced in 1mL TSC basel media and 1% inoculum. Strain M0355 was used as the negative control. The strains were allowed to grow for 14 hours and the supernatant was analyzed using GC for butanol formation.

Heterologous Production of n-Butanol in *T. saccharolyticum* and *C. thermocellum*
Construction of Mutants

In one prophetic example, *T. saccharolyticum* and *C. thermocellum* will be

transformed with a yeast/*E.coli*/ *T. saccharolyticum*/*C. thermocellum* shuttle vector plasmid with all the genes required for the production of n-butanol, see FIG. 2. The promoters chosen have been shown to work previously in both *T. saccharolyticum* and *C. thermocellum*.

The chromosomal integration of the entire n-butanol pathway in *T. saccharolyticum* and *C. thermocellum* will be in the *adhE* locus of both *T. saccharolyticum* and *C. thermocellum*. One advantage in choosing the disruption of native *adhE* is due to the likelihood that the native *adhE* may preferentially act on the C_2 substrate (acetyl CoA) rather than C_4 substrate (butyryl CoA). The genes will be integrated in the reverse order of the pathway, *i.e.*, *adhE→bcd+etfAB→crt→hbd→thl,* such that the precursor substrates can be exogenously added to the media with the mutants for determination of functionality of the pathway.

The construction of the *T. saccharolyticum* and *C. thermocellum* strains harboring a replicative plasmid with all the genes required for the n-butanol pathway as demonstrated in FIG. 2.

The *T. saccharolyticum* and *C. thermocellum* mutants will be tested for product butanol formation using HPLC, GC, and/or GC/MS. Biochemical assays for each of the individual enzymes of the engineered biochemical pathways will also be performed.

Heterologous production of n-butanol in *T. saccharolyticum*

A non-butanol forming wild type *T. saccharolyticum* strain was genetically modified to produce n-butanol. The heterologous n-butanol pathway was engineered on a plasmid harboring all of the genes of interest as is depicted in FIG. 7. However, this cloning strategy failed to generate any transformants in *T. saccharolyticum*, possibly due to the overexpression of the *adhE* gene.

A similar, successful, strategy was used out for chromosomal integration of the n-butanol pathway in *T. saccharolyticum* such that only one copy per gene would be integrated. Because the expressed engineered pathway might be toxic to the host cell, the pathway was integrated into the *xynA* locus of the chromosomal DNA. The *xynA* locus turns on only when xylose is the sole carbon source.

As depicted in FIG. 7, the construction contained an *adhE* gene from *C. acetobutylicum* that was driven by an inducible system such that butanol production wasn't triggered while selecting *T. saccharolyticum* transformants on selective media.

This chromosomal integration strategy construction generated in the order of 10^3 transformants per μg of plasmid DNA. One of the successful transformant clones, I2B, was further studied for production of n-butanol.

To determine the growth characteristics of the wild type (WT) *T. saccharolyticum* strain and recombinant I2B butanol producing, both were grown on TSD minimal media and compared for substrate consumption and product formation as depicted in FIG. 8.

As depicted in FIG. 9, carbon recovery, remaining substrate and end products Xylose, lactic acid (la), acetic acid (aa), ethanol (etoh), and butanol (buoh) were measured by HPLC. The CO_2 generated was calculated based on the predicted stoichiometric relation to HPLC products. The yeast extract in the media and extracellular protein were not considered. The following equation was used to calculate carbon equivalents:

$$C_t = 60/150(X) + 36/90(L) + 36/60(A) + 36/36(E) + 72/74(B) + 12/25.5(CDW)$$

Where C_t is total carbon, X is xylose, L is lactic acid, A is acetic acid, E is ethanol, B is butanol and DCW is dry cell weight. Units are in g/L.

As depicted in FIG. 9, based on pyruvate balance, the carbon recovery for WT fermentation was 101% and 94% for I2B fermentation.

As depicted in Table 3 below, product yields between WT and I2B of butanol, ethanol, acetic acid and lactic acid comparisons are represented as moles of product divided by moles of substrate (xylose).

Table 3:

Strain	Butanol	Ethanol	Acetic acid	Lactic acid
WT	0	1.16	0.4	0.03
I2B	0.16	0.51	0.52	0.13

Isobutanol Production by Thermophilic Microorganisms

In one prophetic embodiment, an engineered thermophilic α-KDC is provided and used in a non-native Erhlich pathway to produce isobutanol. 2-Ketoacid intermediates in amino acid biosynthesis pathways can be converted to aldehydes by α-KDCs and then to alcohols by ADHs. These enzymes catalyze the only two non-native steps are required to produce biofuels by shunting intermediates from amino acid biosynthesis pathways.

One advantage of producing isobutanol instead of n-butanol is that isobutanol is less toxic to the producing microorganism. Isobutanol is produced by the Ehrlich pathway

through shunting excess α-ketoisovalerate into a pathway that uses α-KDC and an ADH. An excess of α-ketoisovalerate can be produced in by engineering organic acid knock-out strains of host ADHs. However, no thermophilic variant of α-KDC is known or has been characterized. The creation of a thermophilic α-KDC is necessary for the production of isobutanol in thermophilic organisms via a α-KDC and ADH pathway.

Amino acid biosynthesis pathways produce various 2-ketoacids. It has been demonstrated that *E. coli* strains with overexpression of α-KDCs and ADHs produced longer chain alcohols including n-propanol, isobutanol, n-butanol, 2-methyl-1- butanol, 3-methyl-1-butanol and 2-phenylethanol depending upon which 2-ketoacid was metabolized to an aldehyde and alcohol. It has also been demonstrated that the addition of specific 2-keto acids to the engineered *E. coli* culture expressing KDC and ADH exhibited specific production of the corresponding alcohols. These prior results indicate that increasing the flux to the 2-keto acids could improve both the productivity and specificity of production of the corresponding alcohols through the Erhlich pathway.

In one prophetic example of the present disclosure, in order to produce isobutanol the valine biosynthesis pathway will be used to generate 2-ketoisovalerate, the precursor to valine, which will then be converted to isobutanol via a decarboxylation step and a reduction step using an α-KDC engineered for thermophilicity.

The approach that will be used for engineering a thermophilic α-KDC from the mesophilic *L. lactis* α-KDC is a combinatorial approach using an iterative saturation mutagenesis strategy.

The mesophilic α-KDC from *L. lactis* is well characterized, including an x-ray crystallographic structure. Mesophilic enzymes can be engineered into thermophilic enzymes by using iterative saturation mutagenesis on amino acid residues identified by the B-FITTER algorithm. The protein structure of α-KDC from *L. lactis* was downloaded from a structural database. B-FITTER was applied to the crystal structure model of α-KDC from *L. lactis* in order to identify residues to mutate in order to introduce thermostability into α-KDC from *L. lactis*.

The results of the B-FITTER analysis of α-KDC from *L. lactis* resulted in identifying ten amino acid residues as targets for saturation mutagenesis using a "B-factor" measurement, see FIG. 4. The B-factor refers to the positional disorder of electrons in an atom with respect to its equilibrium position and hence is the measure of the amino acid residues' flexibility in the protein. The ten amino acids that were targeted occurred in

four clusters. The amino acid residues identified were Leu180, Glu181, Asn188, Thr189, Thr190, Lys342, Gln343, Thr344, Glu345, and Glu346. The targeted residues are mapped on the protein structure in FIG. 5. In order to facilitate the expression of the heterologous protein, sequence was reverse transcribed using a thermoanaerobacterium codon usage table.

Construction of Libraries

A 1.7 Kb gene will be used as the template for generating the α-KDC library and will be ordered as three minigenes and assembled by overlap extension PCR, see FIG. 6. The linear construct of FIG. 6 will be transformed into a homoethanologen strain M0355 (*T. sachharolyticum* Δldh Δpta/ack) that is unable to grow without a functional *adhE* gene. The native *adhE* gene of M0355 will be disrupted and replaced by the linear constructs from the α-KDC library as well a gene (*adhB*) encoding an alcohol dehydrogenase (AdhB) from *T. ethanolicus*. A kanamycin cassette will also be incorporated into the inserted construct as a marker.

In another embodiment, an expression plasmid will be constructed as shown in FIG. 10. The *adhB* gene from *Thermoanaerobacterium ethanolicus* will be cloned into the plasmid. Expression of the genes will be driven by *C. thermocellum* 27405 gapDH promoter.

The expressed, thermophilic α-KDC will react with excess intermediates from the anabolic metabolism of branched chain amino acids and produce aldehydes that can further react with the non-native alcohol dehydrogenase from *T. ethanolicus* to catalyze the reaction from isobutyraldehyde to isobutanol using *C. thermocellum* as a host.

Identifying Functional Variants

In some embodiments, mutations to α-KDC may confer desirable features in functional variants. Functional variants of α-KDC at 55 °C will be identified through growth selection. Since strain M0355 uses solely ethanol as a means of electron sink, if the *adhE* gene is disrupted it can no longer grow. Only cells with functional variants at 55 °C of AdhE that can utilize NADH and make reduced products will be able to grow. The colonies that result from the growth of these mutants will then be screened and sequenced. End product analysis will be done by HPLC, GC and/or GS/MS.

The DNA coding for the α-KDC mutants and AdhE will be isolated from those colonies that exhibit superior growth characteristics. The isolated DNA will then be

transformed in to *C. thermocellum* and inserted into the chromosomal DNA or an expression plasmid through homologous recombination. In one embodiment, the heterologous DNA will be inserted into the native adhE gene of *C. thermocellum*. The transformed *C. thermocellum* will be able to produce isobutanol at 55 °C through the Erhlich pathway using the mutant, thermophilic α-KDC mutants that were generated as well as the AdhB from *T. ethanolicus* through the conversion of the excess amino acids to isobutanol.

The α-KDC gene libraries along with an alcohol dehydrogenase adhB from *T. ethanolicus* and a kanamycin marker will be integrated into the chromosome of *T. saccharolyticum* strain M0355 (a homo-ethanologen) at the adhE gene locus. Since adhE is the enzyme which facilitates electron recycling and production of reduced carbon product, swapping the native adhE gene with the α-KDC gene libraries and adhB as shown in FIG. 6 will force only the mutants expressing functional thermophilic α-KDC to grow. The successful clones will be screened α-KDC activity at thermophilic temperatures. The functionally positive clones will then be sequenced to identify the thermophilic α-KDC candidates.

In another prophetic example, isobutanol is produced by using a *T. saccharolyticum* strain M3224 which produces only acetate and grows very poorly under normal conditions. However, if this strain were to harbor a plasmid as shown in FIG. 7 with the functional variants of α-KDC and an *adhB* gene, the preferential electron cycling will enable the strain to grow better than the native strain. Thus, the strains with the functional variants will be selected through identifying those strains with improved growth characteristics. Similarly, the functionally positive clones will be sequenced to identify the thermophilic α-KDC candidates.

General Procedures

Media, Strains, and Cultivation Conditions

T. saccharolyticum was grown in DSM 122 broth as described elsewhere the concentration of $K_2HPO_4 \cdot 3H_2O$ was reduced from 7.2 g liter^{-1} to 1.9 g liter^{-1}, and glutathione was replaced by cysteine-HCl at a final concentration of 0.5 g liter^{-1}. For growth on plates, 0.58% Difco agar-agar (Becton Dickinson, Sparks, MD) was added to modified DSM 122-cellobiose broth (cellobiose agar). Cellobiose agar additionally contained erythromycin (Em) and lincomycin (Lm), each at 20 μg ml^{-1}, for the selection of transformants of C. thermocellum or 200 μg ml^{-1} of kanamycin (Km) for the selection

of transformants of T. saccharolyticum (all antibiotics were purchased from Sigma, St. Louis, MO). All manipulations associated with the preparation of cultures and subsequent ET were carried out inside an anaerobic chamber equipped with a model 2000 incubator (Coy Laboratory Products, Inc., Grass Lake, MI). *E. coli* Top 10 cells (Invitrogen Corporation, Carlsbad, CA) were grown at 37°C in Columbia broth or on Columbia agar (Columbia broth plus 1.2% Difco agar-agar) containing 200 μg ml^{-1} of ampicillin (Ap) (Sigma, St. Louis, MO) for the selection of transformants, when appropriate.

Plasmid DNA

Shuttle plasmid pIKM1 (6.3 kb; *Apr Emr Lmr Kmr*) was a gift from Juergen Wiegel (University of Georgia, Athens). Plasmid pMU158 was a gift from Mascoma Corporation. pIKM1 DNA was isolated from *E. coli* using a QIAGEN Plasmid Midi kit (QIAGEN Inc., Valencia, CA). The detection of plasmid DNA in presumptive transformants was performed as is well known in the art.

Electrotransformation

Electrotransformation was not generally carried out on *T. saccharolyticum.* Instead, transformation usually relied upon the natural competence of *T. saccharolyticum* cells.

A custom-built pulse generator featuring the cell sample in series with other circuit elements as well as custom-built cuvettes. The cultivation of anaerobic thermophiles prior to electrotransformation (ET), pulse application, post-pulse sample processing, and determination and analysis of cell viability were with the following modification: prior to ET, overnight cultures were diluted 1:7 to 1:10 instead of 1:3 with sterile pre-warmed (58°C) DSM 122 cellobiose broth containing 20 μg ml^{-1} of isoniazid (Sigma, St. Louis, MO). The duration of the electrical pulse applied was as indicated previously. The preparation of *E. coli* cells for ET and the selection of recombinants were performed as is well known in the art. Successfully transformed cells, for example *E. coli* top 10 cells, were grown in LB media containing 100 μg/mL ampicillin and/or 50 μg/mL kanamycin.

Cell suspensions of thermophilic anaerobes or *E. coli* contained approximately 9×10^{10} cells ml^{-1} and were kept on ice until use. For cell-free controls, cell suspensions were replaced either by high-performance liquid chromatography (HPLC)-grade water (Sigma, St. Louis, MO) or by a set of three low-inductance 10 kΩ resistors (TO 220 style thick film power resistors; Ohmite Mfg. Co., Rolling Meadows, IL) (cell sample resistive

equivalent) connected in series. ET was carried out using 90 to 100 μl of cell suspension in 2.0 mL disposable centrifuge tubes with 305 stainless steel electrodes and a single square input pulse of 10 ms. For experiments aimed at investigating the importance of induced oscillations with respect to ET, a custom-built linear filter composed of TO 220 style thick film power resistors (Ohmite Mfg. Co., Rolling Meadows, IL) with a total resistance of 29 kΩ in parallel with a toroidal 107 μH filter inductor (Magnetek, Inc., Los Angeles, CA) was used. The filter was switched in series with the other elements of the above power circuit so that it could be bypassed to eliminate the filtering effect on the pulse current.

The disclosed methods and systems may be modified without departing from the scope hereof. It should be noted that the matter contained in the above description or shown in the accompanying drawings should be interpreted as illustrative and not in a limiting sense. The following claims are intended to cover all generic and specific features described herein, as well as all statements of the scope of the present method and system and reasonable variations thereof, which, as a matter of language, might be said to fall there between.

CLAIMS

1. A recombinant thermophilic microorganism comprising a biochemical pathway to produce butanol from a suitable carbon substrate, said biochemical pathway comprising an acetoacetyl-CoA intermediate.

2. The recombinant thermophilic microorganism of claim 1 comprising a heterologous polypeptide, the heterologous polypeptide being encoded by a heterologous gene, wherein the heterologous gene has been introduced into a parental thermophilic microorganism in order to obtain said recombinant thermophilic microorganism.

3. The recombinant thermophilic microorganism of claim 2, comprising elevated expression of a polypeptide having thiolase activity, as compared to said parental microorganism, wherein the recombinant microorganism produces an intermediate comprising acetoacetyl-CoA from a substrate comprising acetyl-CoA.

4. The recombinant thermophilic microorganism of claim 3, wherein the polypeptide having thiolase activity is encoded by a polynucleotide having at least about 65% identity to a sequence as set forth in SEQ ID NO:1.

5. A polypeptide having thiolase activity having at least about 70% identity to an amino acid sequence as set forth in SEQ ID NO:2.

6. The recombinant thermophilic microorganism of claim 1, comprising elevated expression of a polypeptide having beta-hydroxybutyryl-CoA dehydrogenase activity, as compared to the parental microorganism, wherein the recombinant microorganism produces a metabolite comprising beta-hydroxybutyryl-CoA from a substrate comprising acetoacetyl-CoA.

7. The recombinant thermophilic microorganism of claim 6, wherein the polypeptide having beta-hydroxybutyryl-CoA dehydrogenase activity is encoded by a polynucleotide having at least about 65% identity to a sequence as set forth in SEQ ID NO:3.

8. A polypeptide having beta-hydroxybutyryl-CoA dehydrogenase activity having at least about 70% identity to an amino acid sequence such as SEQ ID NO:4.

9. The recombinant thermophilic microorganism of claim 1, comprising elevated expression of a polypeptide having crotonase activity, as compared to a parental

microorganism, wherein the recombinant microorganism produces a metabolite comprising crotonyl-CoA from a substrate comprising beta-hydroxybutyryl-CoA.

10. The recombinant thermophilic microorganism of claim 9, wherein the polypeptide having crotonase activity is encoded by a polynucleotide having at least about 65% identity to a sequence as set forth in SEQ ID NO:5.

11. A polypeptide having crotonase activity having at least about 70% identity to an amino acid sequence such as SEQ ID NO:6.

12. The recombinant thermophilic microorganism of claim 1, comprising elevated expression of a polypeptide having butyryl-CoA dehydroxygenase activity, as compared to a parental microorganism, wherein the recombinant microorganism produces a metabolite comprising butyryl-CoA from a substrate comprising crotonyl-CoA.

13. The recombinant thermophilic microorganism of claim 12, wherein the polypeptide having butyryl-CoA dehydrogenase activity is encoded by a polynucleotide having at least about 65% identity to a sequence selected from the group consisting of SEQ ID NO:7, SEQ ID NO:9, and SEQ ID NO:11.

14. A polypeptide having butyryl-CoA dehydrogenase activity having at least about 70% identity to an amino acid sequence selected from the group consisting of SEQ ID NO:8, SEQ ID NO:10, and SEQ ID NO:12.

15. The recombinant thermophilic microorganism of claim 12, wherein the polypeptide having butyryl-CoA dehydrogenase activity is encoded by a bcd gene or homolog thereof, in combination with a etfA gene or homolog thereof and etfB gene or homolog thereof.

16. The recombinant thermophilic microorganism of claim 1, comprising elevated expression of a polypeptide having aldehyde/alcohol dehydrogenase activity, as compared to the parental microorganism, wherein the recombinant microorganism produces a metabolite comprising butyraldehyde from a substrate comprising butyryl-CoA.

17. The recombinant thermophilic microorganism of claim 16, wherein the polypeptide having aldehyde/alcohol dehydrogenase activity is encoded by a polynucleotide having at least about 65% identity to a sequence as set forth in SEQ ID NO:13.

18. A polypeptide having aldehyde/alcohol dehydrogenase activity having at least about 70% identity to an amino acid sequence such as SEQ ID NO:14.

19. The recombinant thermophilic microorganism of claim 1, comprising elevated expression of a polypeptide having aldehyde/alcohol dehydrogenase activity, as compared to the parental microorganism, wherein the recombinant microorganism produces a metabolite comprising n-butanol from a substrate comprising butyraldehyde.

20. The recombinant thermophilic microorganism of claim 19, wherein the polypeptide having aldehyde/alcohol dehydrogenase activity is encoded by a polynucleotide having at least about 65% identity to a sequence as set forth in SEQ ID NO:13.

21. The recombinant thermophilic microorganism of claim 1, comprising elevated expression of a polypeptide having aldehyde/alcohol dehydrogenase activity, as compared to a parental microorganism, wherein the recombinant microorganism produces a metabolite comprising n-butanol from a substrate comprising butyryl-CoA.

22. The recombinant thermophilic microorganism of claim 21, wherein the polypeptide having aldehyde/alcohol dehydrogenase activity is encoded by a polynucleotide having at least about 65% identity to a sequence as set forth in SEQ ID NO:13.

23. The recombinant thermophilic microorganism of claim 1, wherein said recombinant thermophilic microorganism comprises one or more deletions, disruptions or knockouts in a gene encoding an enzyme that catalyzes the conversion of acetyl-coA to n-butanol.

24. The recombinant thermophilic microorganism of claim 23, comprising the deletion, disruption or knockout of adhE homologs or variants thereof.

25. The recombinant thermophilic microorganism of claim 23, wherein the microorganism comprises an increase in the conversion of acetyl-coA to n-butanol in comparison to a parent microorganism.

26. The recombinant thermophilic microorganism of claim 23, wherein the thermophilic microorganism is *Thermoanaerobacterium thermosaccharolyticum* or *Clostridium thermocellum*.

27. A recombinant thermophilic microorganism comprising a biochemical pathway to produce n-butanol from fermentation media of a suitable carbon substrate at a temperature of from about 40 °C to about 65 °C, wherein the recombinant biochemical pathway comprises elevated expression of: a) a thiolase as compared to a parental microorganism; b) a hydroxybutyryl-CoA dehydrogenase as compared to a parental microorganism; c) a crotonase as compared to a parental microorganism; d) a butyryl-CoA dehydrogenase as compared to a parental microorganism; and e) a butyraldehyde dehydrogenase as compared to a parental microorganism.

28. A method for producing n-butanol using a recombinant thermophilic microorganism, said recombinant thermophilic microorganism comprising a recombinant biochemical pathway to produce n-butanol from fermentation media of a suitable carbon substrate at a temperature of from about 40 °C to about 65 °C, wherein the recombinant biochemical pathway comprises elevated expression of: a) a thiolase as compared to a parental microorganism; b) a hydroxybutyryl-CoA dehydrogenase as compared to a parental microorganism; c) a crotonase as compared to a parental microorganism; d) a butyryl-CoA dehydrogenase as compared to a parental microorganism; and e) a butyraldehyde dehydrogenase as compared to a parental microorganism, and

wherein said method comprises providing said recombinant thermophilic microorganism a suitable carbon substrate and recovering n-butanol produced by said recombinant thermophilic microorganism.

29. The method of claim 28, wherein contacting said media with said recombinant thermophilic microorganism is performed as a batch fermentation.

30. The method of claim 28, wherein the contacting said media with said recombinant thermophilic microorganism is performed as a continuous fermentation.

31. The method of claim 28, wherein said n-butanol produced by said recombinant thermophilic microorganism is recovered from said fermentation media by distillation, liquid-liquid extraction, adsorption, decantation, pervaporation or combinations thereof.

32. The method of claim 28, wherein said fermentation media comprises mixtures of carbon substrates selected from the group consisting of cellulosic feedstocks, hexoses and pentoses.

33. The recombinant thermophilic microorganism of claim 1, wherein said microorganism is a member of a genus selected from thermophiles, hyperthermophiles and halophiles.

34. A method for producing isobutanol at temperatures from about 45 °C to about 65 °C comprising providing a fermentation media comprising carbon substrates, and contacting said media with a recombinant thermophilic microorganism expressing an engineered isobutanol biosynthetic pathway wherein said pathway comprises the following substrate to product conversions;

i. α-ketoisovalerate to isobutyraldehyde, and

ii. isobutyraldehyde to isobutanol, and

wherein

a) the substrate to product conversion of step (i) is performed by a thermophilic alpha-ketodecarboxylase enzyme; and

b) the substrate to product conversion of step (ii) is performed by a thermophilic alcohol dehydrogenase enzyme whereby isobutanol is produced.

35. The method of claim 34 wherein said thermophilic alpha-ketodecarboxylase enzyme of step a) is made by performing site-directed mutagenesis on amino acid residues of an alpha-ketodecarboxylase enzyme from *L. lactis*, and

wherein said amino acid residues are identified by an in-silico analysis of a crystal structure of an alpha-ketodecarboxylase enzyme from *L. lactis* by an algorithm.

36. The method of claim 34 wherein said algorithm comprises B-fitter, and

wherein B-fitter produces a B value of greater than or equal to about 40 for said amino acid residues.

37. The method of claim 34 wherein said thermophilic alpha-ketodecarboxylase enzyme is encoded in nucleic acid linear insert of heterologous DNA that also encodes for an alcohol dehydrogenase from *T. ethanolicus* and is inserted into the chromosomal DNA locus of *Thermoanaerobacterium thermosaccharolyticum* or *Clostridium thermocellum* that encodes for a native alcohol dehydrogenase.

38. The method of claim 34 wherein said thermophilic alpha-ketodecarboxylase enzyme is encoded in an expression plasmid of heterologous DNA that also encodes for an alcohol dehydrogenase from T. ethanolicus and is transformed into *Thermoanaerobacterium thermosaccharolyticum* or *Clostridium thermocellum*.

39. The method of claim 37 wherein said *Thermoanaerobacterium thermosaccharolyticum* or *Clostridium thermocellum* do not have a functioning native alcohol dehydrogenase.

40. The method of claim 34, wherein contacting said media with a recombinant thermophilic microorganism is performed as a batch fermentation.

41. The method of claim 34, wherein the contacting said media with a recombinant thermophilic microorganism is performed as a continuous fermentation.

42. The method of claim 34, further comprising recovering said isobutanol from said fermentation media by distillation, liquid-liquid extraction, adsorption, decantation, pervaporation or combinations thereof.

43. The method of claim 34, wherein said fermentation media further comprises mixtures of carbon substrates selected from the group consisting of cellulosic feedstocks, hexoses and pentoses.

44. A recombinant thermophilic microorganism of claim 34, comprising elevated expression of a polypeptide having alpha-ketodecarboxylase activity, as compared to a parental microorganism, wherein said recombinant thermophilic microorganism produces a metabolite comprising the corresponding branched chain aldehyde from a substrate comprising the branched chain amino acid and further comprises using an alcohol dehydrogenase that converts said branched chain aldehyde into isobutanol, n-butanol or an isomer of n-butanol.

45. A method of producing isobutanol comprising:

(a) providing the recombinant thermophilic microorganism of claim 34;

(b) growing said recombinant thermophilic microorganism of claim 34 in a culture medium containing a feedstock providing a carbon source;

(c) recovering isobutanol from said culture medium.

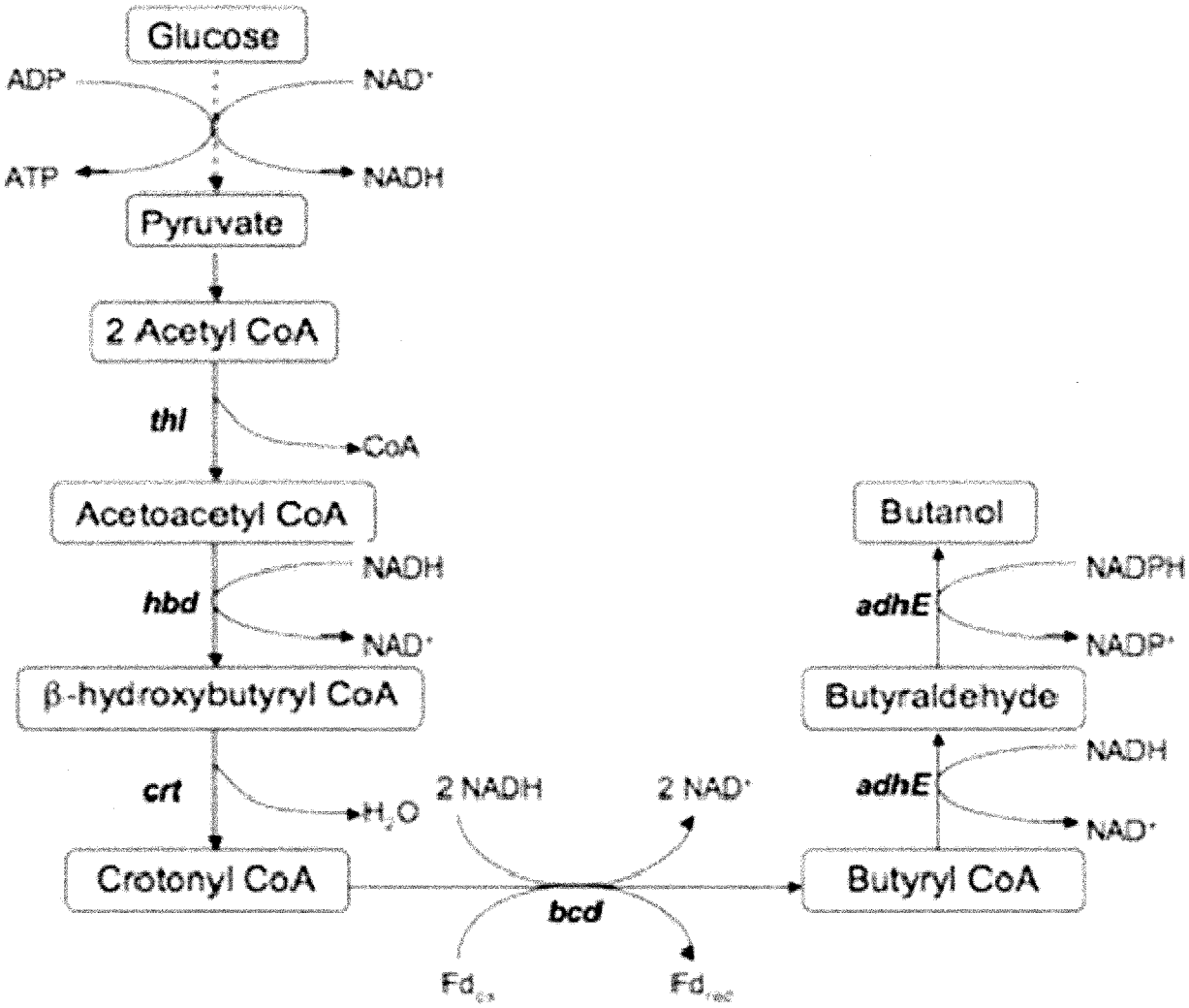
Glucose
ADP
NAD+
ATP
NADH
Pyruvate
2 Acetyl CoA
thl
CoA
Acetoacetyl CoA
NADH
hbd
NAD+
β-hydroxybutyryl CoA
crt
H2O
2 NADH
2 NAD+
Crotonyl CoA
bcd
Fdox
Fdred
Butyryl CoA
Butanol
NADPH
adhE
NADP+
Butyraldehyde
NADH
adhE
NAD+

FIG. 1

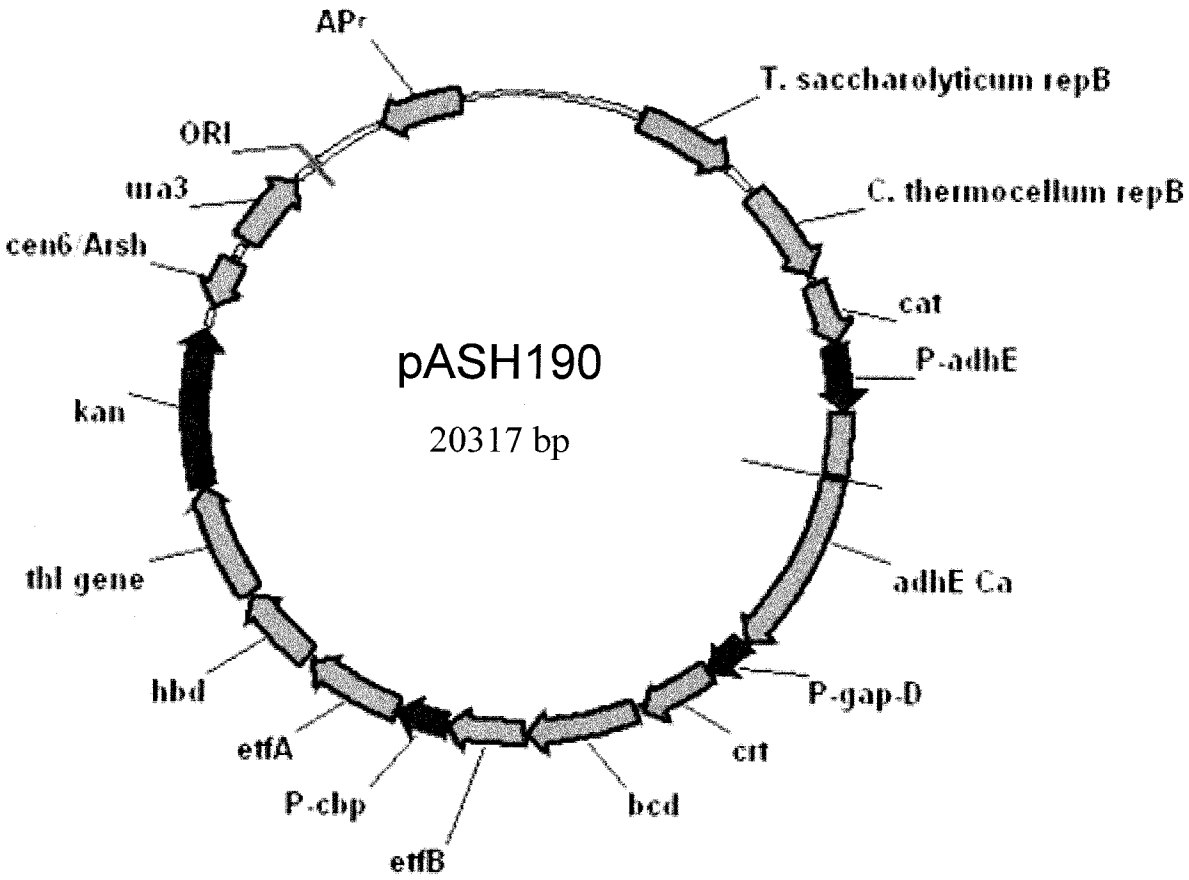
APr
T. saccharolyticum repB
ORI
ura3
C. thermocellum repB
cen6/Arsh
cat
pASH190
20317 bp
P-adhE
kan
adhE Ca
thl gene
P-gap-D
hbd
crt
etfA
P-cbp
bcd
etfB

FIG. 2

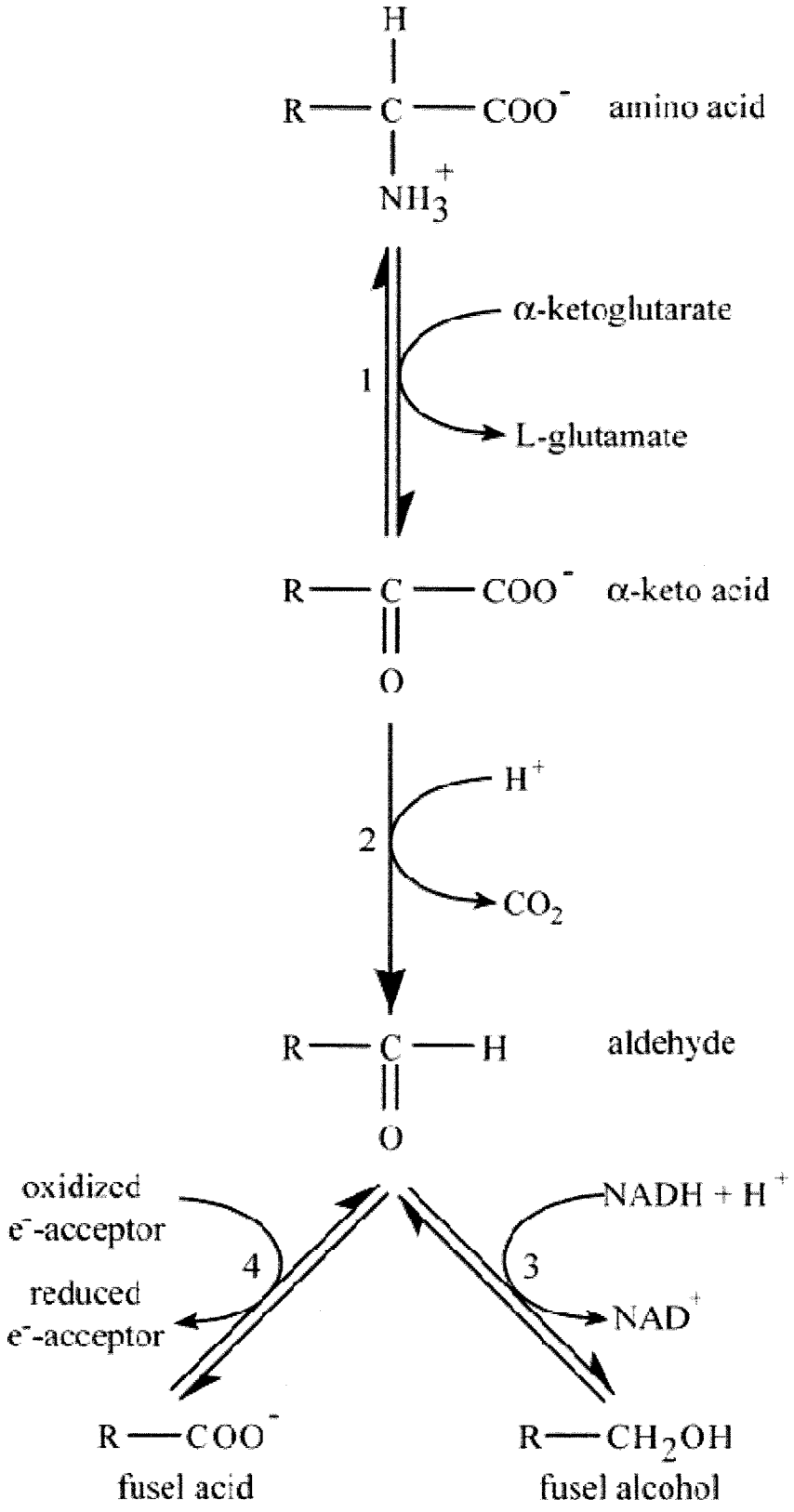
amino acid
α-ketoglutarate
1
L-glutamate
α-keto acid
H+
2
CO2
aldehyde
oxidized
e-acceptor
4
reduced
e-acceptor
NADH + H+
3
NAD+
fusel acid
fusel alcohol

FIG. 3

\\.PSF\.Home\Desktop\2VBF.pdb
Title: THE HOLOSTRUCTURE OF THE BRANCHED-CHAIN KETO ACID
Title: 2 DECARBOXYLASE (KDCA) FROM LACTOCOCCUS LACTIS
(The highest 20 averaged B values are shown only.)

Chain identifier of chain no. 1 : A

Residue Name		Residue seq. no.	B value	Rank
ASN	A	188	49.36	1
GLU	A	181	48.13	2
THR	A	189	47.10	3
GLN	A	343	46.23	4
THR	A	190	43.65	5
LYS	A	342	43.12	6
TYR	A	344	42.73	7
LEU	A	180	42.02	8
GLU	A	346	41.82	9
GLU	A	345	41.80	10
ASP	A	341	40.07	11
SER	A	179	39.62	12
GLN	A	192	39.57	13
GLU	A	191	38.85	14
LEU	A	178	38.66	15
ALA	A	177	38.58	16
GLU	A	329	37.35	17
SER	A	328	37.22	18
LYS	A	175	34.91	19
ASP	A	317	34.89	20

FIG. 4

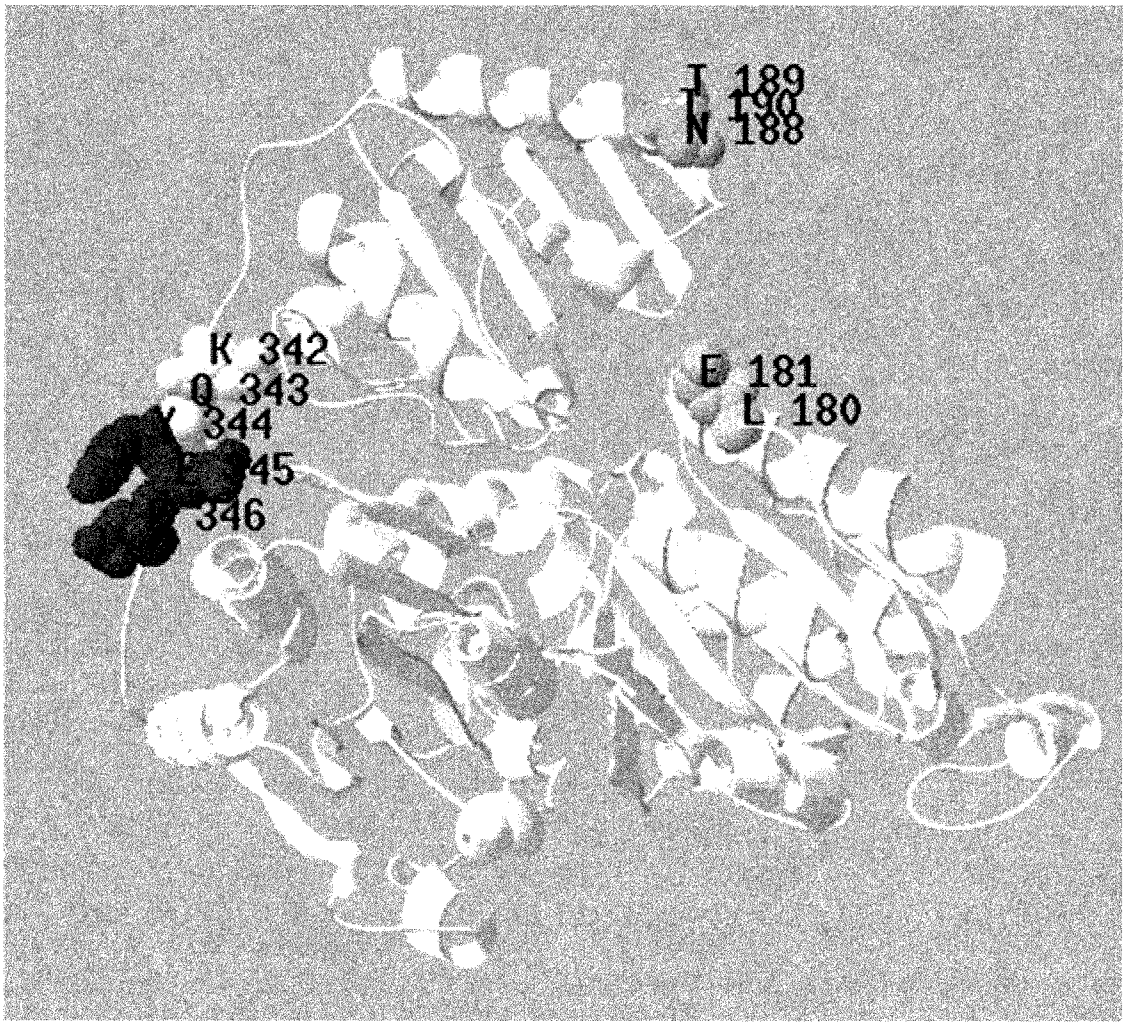

FIG. 5

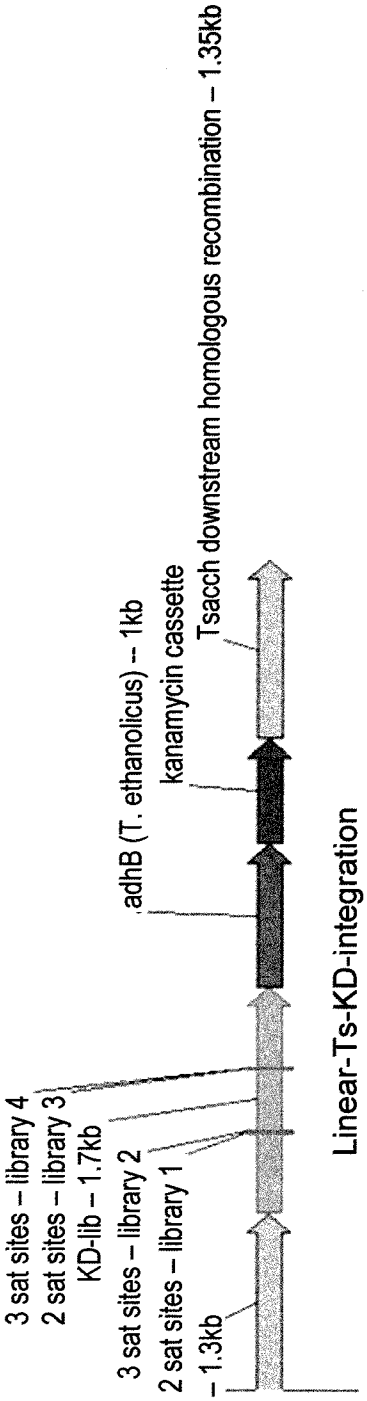
Tsacch upstream homologous recombination – 1.3kb
2 sat sites – library 1
3 sat sites – library 2
KD-lib – 1.7kb
2 sat sites – library 3
3 sat sites – library 4
adhB (T. ethanolicus) – 1kb
kanamycin cassette
Tsacch downstream homologous recombination – 1.35kb
Linear-Ts-KD-integration
6253 bp

FIG. 6

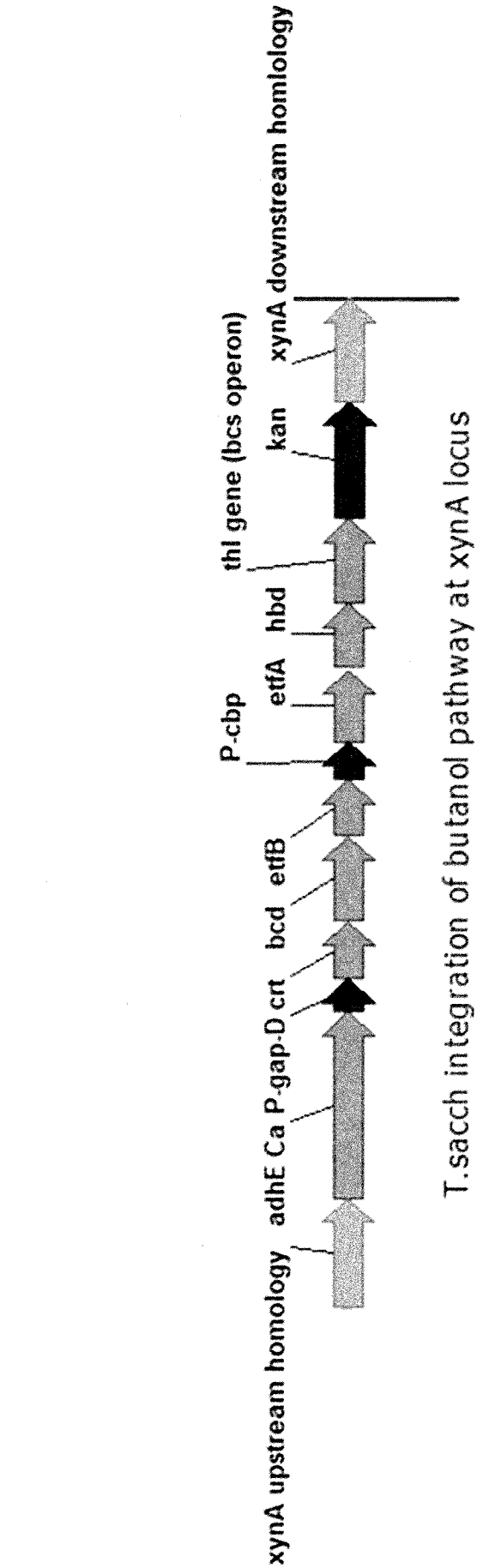
xynA upstream homology
adhE
Ca P-gap-D
crt
bcd
etfB
P-cbp
etfA
hbd
thl gene (bcs operon)
kan
xynA downstream homology
T.sacch integration of butanol pathway at xynA locus
13989 bp

FIG. 7

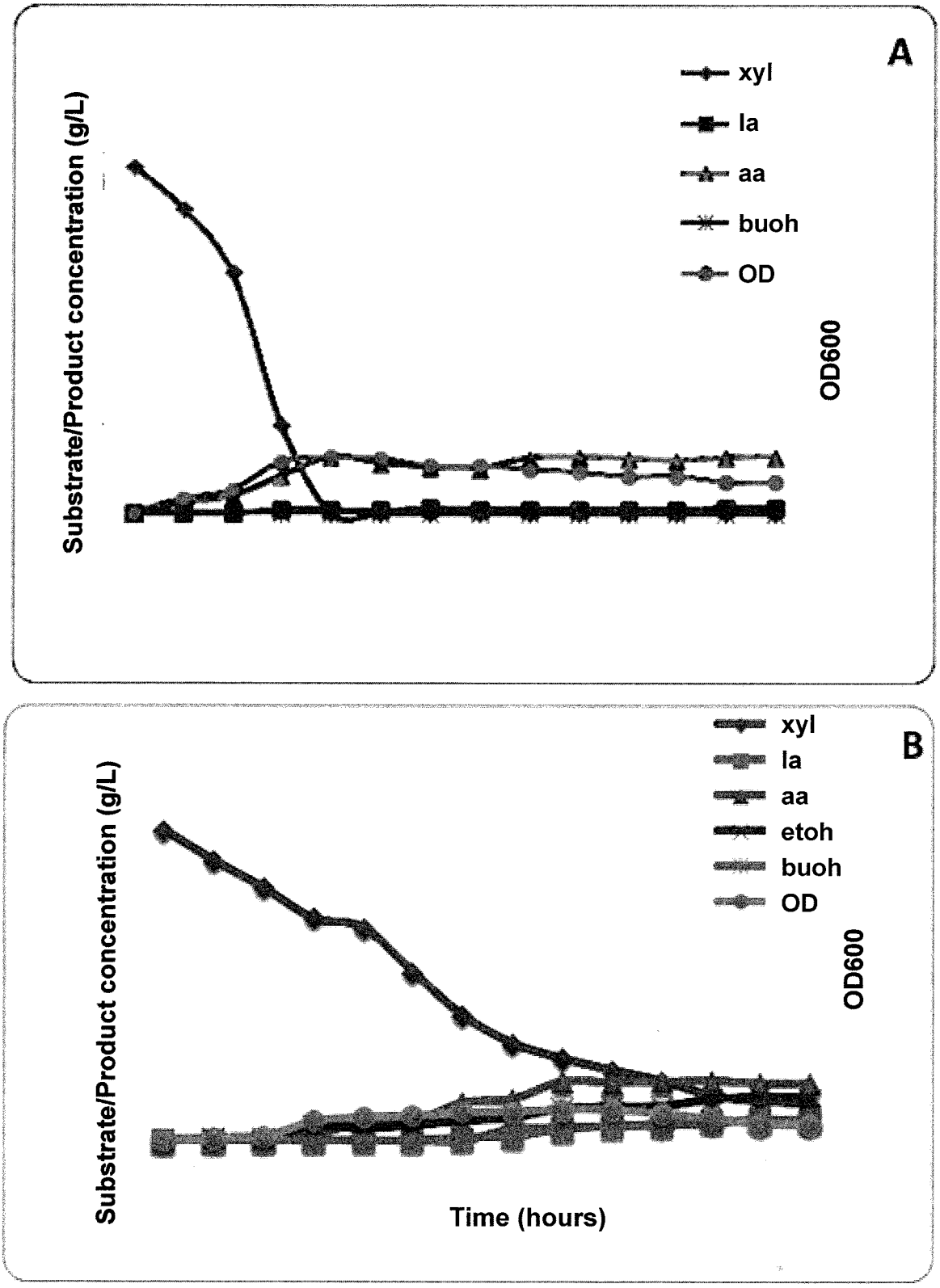
A
xyl
la
aa
buoh
OD
Substrate/Product concentration (g/L)
OD600
B
xyl
la
aa
etoh
buoh
OD
Substrate/Product concentration (g/L)
OD600
Time (hours)

FIG. 8

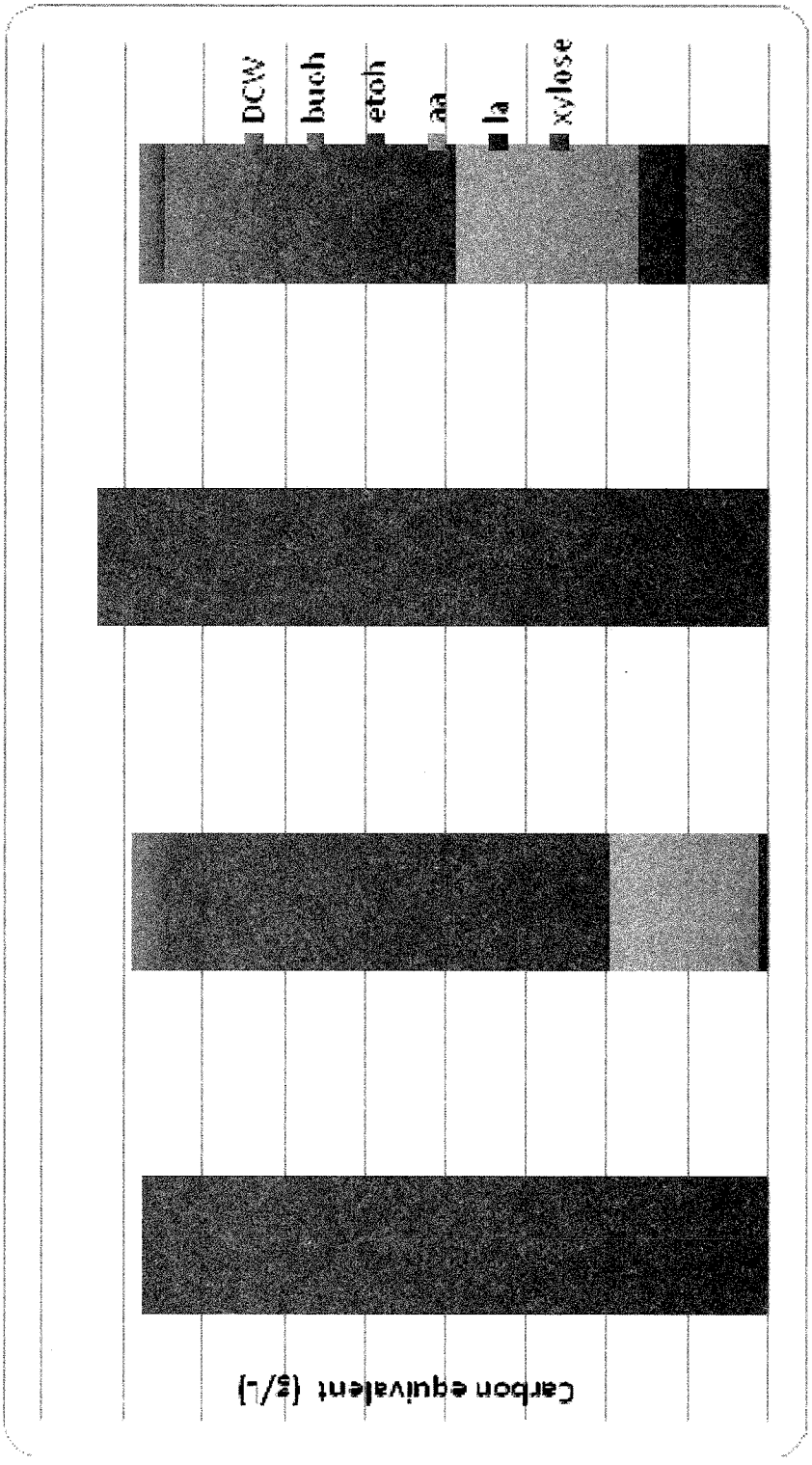

FIG. 9

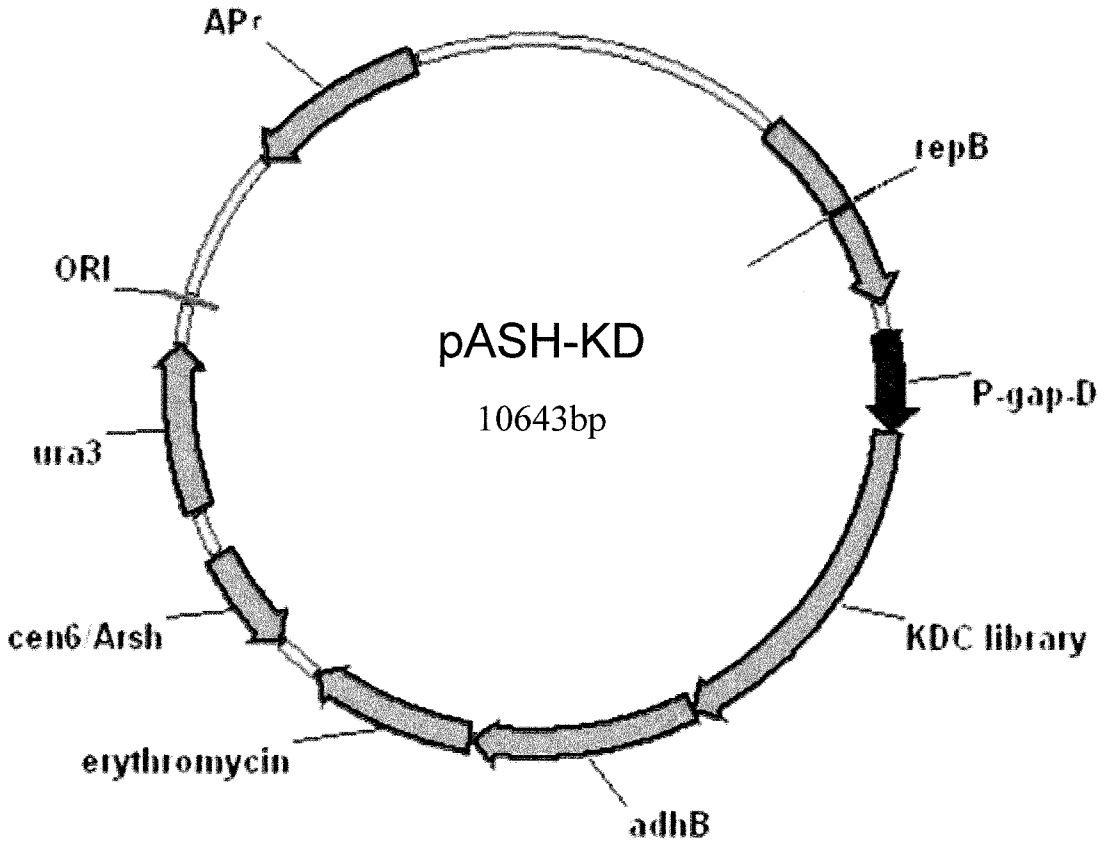
APr
repB
ORI
pASH-KD
10643bp
P-gap-D
ura3
cen6/Arsh
KDC library
erythromycin
adhB

FIG. 10