

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
9 February 2012 (09.02.2012)

PCT

(10) International Publication Number
WO 2012/019175 A2

(51) International Patent Classification:
C12N 9/88 (2006.01)

(21) International Application Number:
PCT/US2011/046869

(22) International Filing Date:
5 August 2011 (05.08.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/371,582 6 August 2010 (06.08.2010) US

(71) Applicant (for all designated States except US): **MAS-COMA CORPORATION** [US/US]; 67 Etna Road - Suite 300, Lebanon, New Hampshire 03766 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **SILLERS, William Ryan** [US/US]; 229 Mt. Support Road #22, Lebanon, New Hampshire 03766 (US). **TRIPATHI, Shital A.** [IN/US]; 604 Bancroft Way - Suite A, Berkeley, California 94710 (US). **SHAW, IV, Arthur J.** [US/US]; 495 Olde Farms Road, Grantham, New Hampshire 03753 (US). **ARGYROS, Aaron** [US/US]; 74 Runnals Road, White River Junction, Vermont 05001 (US). **HOGSETT, David A.** [US/US]; P.O. Box 1107, Grantham, New Hampshire 03753 (US).

(74) Agents: **JACKMAN, Peter A.** et al.; 1100 New York Avenue N.W., Washington, District of Columbia 20005-3934 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, 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, 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, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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))

(54) Title: PRODUCTION OF MALONYL-COA DRIVED PRODUCTS VIA ANAEROBIC PATHWAYS

(57) Abstract: The present invention provides for novel metabolic pathways to convert biomass and other carbohydrate sources to malonyl-CoA derived products, such as hydrocarbons and other bioproducts, under anaerobic conditions and with the net production of ATP. More specifically, the invention provides for a recombinant microorganism comprising one or more native and/or heterologous enzymes that function in one or more engineered metabolic pathways to achieve conversion of a carbohydrate source to, e.g., long-chain hydrocarbons and hydrocarbon derivatives, wherein the one or more native and/or heterologous enzymes is activated, upregulated, downregulated, or deleted. The invention also provides for processes to convert biomass to malonyl-CoA derived products which comprise contacting a carbohydrate source with a recombinant microorganism of the invention.



WO 2012/019175 A2

PRODUCTION OF MALONYL-COA DERIVED PRODUCTS VIA ANAEROBIC PATHWAYS

BACKGROUND OF THE INVENTION

[0001] Depleting petroleum reserves, recurrent energy crises, increasing demand, and climate change have provided significant impetus in the search for sustainable technologies to replace petroleum as a source of fuels and chemical feedstocks. Long chain fatty acids and other derivatives are commercially attractive as fuel and chemical feedstocks because they can directly replace crude petroleum (as “bio-crude”), which is composed primarily of alkanes, alkenes, and aromatic hydrocarbons. In particular, cellulosic biomass is a preferred source of generating long chain fatty acids and other derivatives for use as fuel and chemical feedstocks, which are compatible with existing petroleum refining and distribution and can substitute for diesel, gasoline, jet fuel, and other derivatives of crude oil.

[0002] Currently, commercial and academic efforts are focused on bio-based petroleum replacement fuels made from microorganisms such as microalgae and that require aerobic microbial production. Algae bio-petroleum can appear as a very attractive option because fuel production occurs directly from sunlight and CO₂. However, algal volumetric productivities are 100-fold lower than fermentative processes, requiring significantly higher biorefinery capital expenditures. See Liliana *et al.*, *Biotechnology and Bioengineering* 102:100-12 (2009). In addition, lower capital algal options, such as open pond culturing, have many technical hurdles to clear before commercial deployment despite decades of research into the issue.

[0003] Other efforts are underway to produce fatty acid compounds from sugars and plant biomass, but all current methods require oxygen to be supplied during fermentation, and are not full consolidated bioprocessing (CBP) processes. Unlike traditional ethanol fermentations, aerobic biofuel synthesis routes feature product formation which is uncoupled from ATP generation and cell growth. Uncoupling of product formation from cell growth simplifies metabolic engineering and has allowed for rapid development of first generation biocatalysts. However, there is a price to be paid for aerobic production when the technology is scaled up to meet industrial needs. First, there are significant costs associated with scaling-up aerobic fermentations, such as, those due to the need for

aeration and heat removal. In practice, these constraints limit the size of aerobic fermentors, with those used in anaerobic fuel ethanol production being an order of magnitude larger. Second, although maximum theoretical product yields from an aerobic process are only slightly lower than an anaerobic process, in practice it is extraordinarily difficult to approach this maximum since there is no biological incentive for microbes to reach high product yields.

[0004] To reach the best aerobic process hydrocarbon yields to date, researchers have resorted to high cell density fermentation, which resulted in product yields between 30-40% of the theoretical maximum. See Tsuruta *et al.*, *PLoS ONE* 4:e4489 (2009); Whited *et al.*, *Industrial Biotechnology* 6:152-163 (2010). While these yields may be quite acceptable for pharmaceutical or specialty chemical production, fuel biorefinery process models have shown that fermentation yields lower than 85% of theoretical result in unattractive process economics. However, an anaerobic, oxygen-free fermentation not only creates higher product yields, but also removes many significant scale-up problems associated with aerobic fermentation. Hydrocarbon fuel production also has process benefits compared to ethanol fuel production, such as a lower product recovery cost and a lower product toxicity to fermenting organisms. The latter could result in smaller fermentation volumes needed to reach equivalent productivities.

[0005] An anaerobic biocatalyst requires a higher degree of metabolic pathway integration to couple product formation with ATP generation, NAD(P)H regeneration, and cell growth. However, once these requirements are met, natural evolutionary forces can be harnessed to increase product yields and productivities, driving them towards theoretical maxima. See Burgard *et al.*, *Biotechnology and Bioengineering* 84:647-57 (2003); Sauer, *Advances in Biochemical Engineering/Biotechnology* 73:129-69 (2001). Higher yields, combined with a lower-cost path for scale-up, make an anaerobic process a preferred option for developing microbes to produce fungible biofuels. The invention describes a method to produce long chain fatty acids and their derivatives in an organism or consortia of organisms in a CBP process that is anaerobic.

[0006] Integral to the process of producing any end product, including those that can be produced using the methods of the invention, is an adequate supply of metabolic substrates. Malonyl-CoA is such a key metabolic precursor for the biological synthesis of various bioproducts, including, but not limited to, fatty acid derived long chain

hydrocarbon compounds such as fatty alcohols, fatty aldehydes, fatty acids, wax esters, and alkanes. However, the biosynthesis of malonyl-CoA is known to occur through only a few mechanisms *in vivo*—namely from acetyl-CoA, carbon dioxide, and ATP by acetyl-CoA carboxylase (*acc*, EC 6.4.1.2) or from malonate, CoA, and ATP by malonyl-CoA synthetase (*matB*) (An and Kim, *Eur. J. Biochem.* 257:395-402 (1998)). Yet, both of these mechanisms require the consumption of ATP to drive the reaction towards malonyl-CoA. In contrast, to produce fatty acid derived hydrocarbons, or any other bioproducts that use malonyl-CoA as a precursor, anaerobically at high yield, the route to malonyl-CoA should result in a net production of ATP. The invention describes recombinant microorganisms, pathways, and methods for producing desired end-products from malonyl-CoA precursors with a net production of ATP.

BRIEF SUMMARY OF THE INVENTION

[0007] The recombinant microorganisms and methods of the invention use metabolic pathways that allow for the production of malonyl-CoA derived products, such as hydrocarbons and hydrocarbon derivatives and other bioproducts, under anaerobic conditions. The metabolic pathways allow for the production of long chain compounds, including, *e.g.*, chain lengths from four carbon atoms up to 40 or more carbon atoms per molecule, and cellular growth in the absence of oxygen or other mechanisms to generate cellular energy (ATP) besides fermentative metabolism.

[0008] An aspect of the invention is the ability to produce long chain compounds at high yield with an anaerobic process rather than with an aerobic process. Anaerobic production results in a higher product yield, easier scalability, and better process thermodynamics. For lignocellulosic biomass conversion, an anaerobic process is even more desirable, as the requirement for oxygen transfer in a medium with suspended solids is highly unattractive from an engineering perspective. Additional advantages include, but are not limited to:

[0009] 1) Production of a direct (fungible) replacement for petroleum;

[0010] 2) Lower separation costs from a dilute aqueous fermentation as a result of the immiscible nature of long chain hydrocarbons compared to fully miscible shorter chain compounds;

[0011] 3) Greater downstream product diversity and flexibility; and

- [0012] 4) Potentially lower product toxicity for fermenting organism which will allow for reduced fermentor volume and lower capital costs in a cellulosic biomass process.
- [0013] One aspect of the invention relates to a recombinant microorganism comprising one or more native and/or heterologous enzymes that function in one or more engineered metabolic pathways to convert a carbohydrate source to a hydrocarbon, wherein the one or more native and/or heterologous enzymes is activated, upregulated, downregulated, or deleted. In certain embodiments, the conversion of a carbohydrate source to a hydrocarbon is under anaerobic conditions. In certain embodiments, the conversion of a carbohydrate source to a hydrocarbon is under microaerophilic conditions.
- [0014] In certain embodiments, the one or more engineered metabolic pathways produce net ATP. In some embodiments, the one or more engineered metabolic pathway produces at least about 0.5 net ATP; at least about 1.0 net ATP; at least about 1.5 net ATP; or at least about 2.0 net ATP. In other embodiments the net ATP production is at least about at least about 0.1 net ATP; at least about 0.2 net ATP; at least about 0.3 net ATP; at least about 0.4 net ATP; at least about 0.5 net ATP; at least about 0.6 net ATP; at least about 0.7 net ATP; at least about 0.8 net ATP; at least about 0.9 net ATP; at least about 1.0 net ATP; 1.1 net ATP; at least about 1.2 net ATP; at least about 1.3 net ATP; at least about 1.4 net ATP; at least about 1.5 net ATP; at least about 1.6 net ATP; at least about 1.7 net ATP; at least about 1.8 net ATP; at least about 1.9 net ATP; or at least about 2.0 net ATP.
- [0015] In particular aspects of the invention, the hydrocarbon produced by the recombinant microorganism is an alkane, an alkene, a hydrocarbon derivative, or a combination of any of these hydrocarbons. In some embodiments, the hydrocarbon produced is selected from the group consisting of an alkane; an alkene; an alkyne; a hydrocarbon derivative; and combinations of these hydrocarbons. In certain aspects, the hydrocarbon derivative is an aldehyde; an alcohol; an ester; a fatty acid; an unsaturated fatty acid; a branched-chain fatty acid; a branched methoxy fatty acid; a multi-methyl branched acid; a divinyl-ether fatty acid; a w-phenylalkanoic acid; or a dicarboxylic acid. In some embodiments, the hydrocarbon derivative is selected from the group consisting of an aldehyde; an alcohol; an ester; a fatty acid; an unsaturated fatty acid; a branched-chain fatty acid; a branched methoxy fatty acid; a multi-methyl branched acid; a divinyl-ether fatty acid; a w-phenylalkanoic acid; a dicarboxylic acid; and combinations of these hydrocarbon derivatives.

- [0016] In certain aspects of the invention, the hydrocarbon or hydrocarbon derivative produced by the recombinant microorganism comprises a carbon backbone of C₄-C₄₀. In some embodiments, the hydrocarbon or hydrocarbon derivative comprises a carbon backbone selected from the group consisting of C₆-C₃₆; C₈-C₃₂; C₁₀-C₂₈; C₁₂-C₂₄; C₁₄-C₂₂; C₁₆-C₂₀; and combinations thereof. In other embodiments, the hydrocarbon or hydrocarbon derivative comprises a carbon backbone selected from the group consisting of C₁₂; C₁₄; C₁₆; C₁₈; C₂₀; C₂₂; C₂₄; and combinations of thereof. In one embodiment, the hydrocarbon or hydrocarbon derivative comprises a carbon backbone of C₁₆.
- [0017] In some aspects of the invention, the carbohydrate source converted to a hydrocarbon is from biomass or from carbohydrates, such as a sugar or a sugar alcohol. In one embodiment, the carbohydrate source converted to a hydrocarbon is a lignocellulosic material. In some embodiments, the carbohydrate is a monosaccharides (e.g., glucose, fructose, galactose, xylose, arabinose, rhamnose, galacturonic acid, xylitol, sorbitol, or ribose), a disaccharide (e.g., sucrose, cellobiose, maltose, or lactose), an oligosaccharide (e.g., xylooligomers, cellodextrins, or maltodextrins), or a polysaccharide (e.g., xylan, cellulose, starch, mannan, or pectin).
- [0018] In a particular aspect of the invention, one of the engineered metabolic pathways in the recombinant microorganism comprises the conversion of oxaloacetate and acetyl-CoA to malonyl-CoA and pyruvate. In one embodiment, the oxaloacetate and acetyl-CoA is converted to malonyl-CoA and pyruvate by a transcarboxylase. In some embodiments, the transcarboxylase is encoded by a heterologous transcarboxylase polynucleotide. In certain embodiments, the transcarboxylase is encoded by a polynucleotide from a *Thermoanaerobacter* species, *P. freudenreichii*, *P. acnes*, or *C. thermocellum*. In one embodiment, the transcarboxylase is genetically modified
- [0019] In another aspect of the invention, one of the engineered metabolic pathways comprises the conversion of phosphoenolpyruvate to oxaloacetate. In one embodiment, the phosphoenolpyruvate is converted to oxaloacetate by a phosphoenolpyruvate carboxykinase. In some embodiments, the phosphoenolpyruvate carboxykinase is encoded by a heterologous phosphoenolpyruvate carboxykinase polynucleotide. In certain embodiments, the phosphoenolpyruvate carboxykinase is encoded by a polynucleotide from a *Thermoanaerobacter* species, *E. coli*, *S. cerevisiae*, or *C. thermocellum*.

[0020] In other aspects of the invention, one of the engineered metabolic pathways further comprises at least one of the following steps: conversion of malonyl-CoA to malonyl-ACP; conversion of malonyl-ACP to an acyl_n-ACP; conversion of an acyl_n-ACP to a β -keto ester_{n+2}-ACP; conversion of a β -keto ester_{n+2}-ACP to a β -D-hydroxyacyl_{n+2}-ACP; conversion of a β -D-hydroxyacyl_{n+2}-ACP to a trans-2-unsaturated acyl_{n+2}-ACP; or conversion of a trans-2-unsaturated acyl_{n+2}-ACP to an acyl_{n+2}-ACP.

[0021] In some aspects of the invention, one of the engineered metabolic pathways further comprises the conversion of pyruvate and CoA-SH into acetyl-CoA and CO₂ and NAD(P)H.

[0022] In some aspects of the invention, one or more of the native enzymes in the engineered metabolic pathways are downregulated or deleted. In certain embodiments, the downregulated or deleted native enzyme is an enzyme involved in central metabolism. In some embodiments, the downregulated or deleted native enzyme is selected from the group consisting of a pyruvate kinase; a hydrogenase; a lactate dehydrogenase; a phosphotransacetylase; an acetate kinase; an acetaldehyde dehydrogenase; an alcohol dehydrogenase; a pyruvate formate lyase; a pyruvate decarboxylase; an enzyme involved in degradation of fatty acids and their derivatives; and combinations of thereof.

[0023] In some aspects of the invention, the microorganism is a thermophilic or a mesophilic bacterium. In certain embodiments, the thermophilic or mesophilic bacterium is a species of the genera *Escherichia*, *Propionibacterium*, *Thermoanaerobacterium*, *Thermoanaerobacter*, *Clostridium*, *Geobacillus*, *Saccharococcus*, *Paenibacillus*, *Bacillus*, *Caldicellulosiruptor*, *Anaerocellum*, *Anoxybacillus*, *Klebsiella*, *Lactobacillus*, *Lactococcus*, or *Corynebacterium*. In other embodiments, the microorganism is a bacterium selected from the group consisting of: *E. coli* strain B, strain C, strain K, strain W, *Shewanella*, *Propionibacterium acnes*, *Propionibacterium freudenreichii*, *Propionibacterium shermanii*, *Propionibacterium pentosaceum*, *Propionibacterium arabinosum*, *Clostridium acetobutylicum*, *Clostridium beijerinckii*, *Thermoanaerobacterium thermosulfurigenes*, *Thermoanaerobacterium aotearoense*, *Thermoanaerobacterium polysaccharolyticum*, *Thermoanaerobacterium zeae*, *Thermoanaerobacterium xylanolyticum*, *Thermoanaerobacterium saccharolyticum*, *Thermoanaerobium brockii*, *Thermoanaerobacterium thermosaccharolyticum*, *Thermoanaerobacter thermohydrosulfuricus*, *Thermoanaerobacter ethanolicus*,

Thermoanaerobacter brocki, *Clostridium thermocellum*, *Clostridium clariflavum*, *Clostridium cellulolyticum*, *Clostridium phytofermentans*, *Clostridium straminosolvens*, *Geobacillus thermoglucosidasius*, *Geobacillus stearothermophilus*, *Saccharococcus caldoxylosilyticus*, *Saccharoccus thermophilus*, *Paenibacillus campinasensis*, *Bacillus flavothermus*, *Anoxybacillus kamchatkensis*, *Anoxybacillus gonensis*, *Caldicellulosiruptor acetigenus*, *Caldicellulosiruptor saccharolyticus*, *Caldicellulosiruptor kristjanssonii*, *Caldicellulosiruptor owensensis*, *Caldicellulosiruptor lactoaceticus*, *Lactobacillus thermophilus*, *Lactobacillus bulgaricus*, *Lactococcus lactis*, and *Anaerocellum thermophilum*. In one embodiment, recombinant microorganism is selected from the group consisting of *Clostridium thermocellum*, and *Thermoanaerobacterium saccharolyticum*.

[0024] Another aspect of the invention relates to a process for converting a carbohydrate source to a hydrocarbon comprising contacting the carbohydrate source with a recombinant microorganism of the invention. In some embodiments, the carbohydrate source comprises lignocellulosic biomass. In certain embodiments, the lignocellulosic biomass is selected from the group consisting of grass, switch grass, cord grass, rye grass, reed canary grass, mixed prairie grass, miscanthus, sugar-processing residues, sugarcane bagasse, sugarcane straw, agricultural wastes, rice straw, rice hulls, barley straw, corn cobs, cereal straw, wheat straw, canola straw, oat straw, oat hulls, corn fiber, stover, soybean stover, corn stover, forestry wastes, recycled wood pulp fiber, paper sludge, sawdust, hardwood, softwood, agave, and combinations thereof. In other embodiments, the carbohydrate source comprises a carbohydrate. In certain embodiments, the carbohydrate is a sugar, a sugar alcohol, or a mixture thereof.

[0025] In some aspects of the invention, the hydrocarbon produced by the recombinant microorganism is secreted.

[0026] Another aspect of the invention relates to an engineered metabolic pathway for producing a hydrocarbon from consolidated bioprocessing media.

[0027] One aspect of the invention relates to a recombinant microorganism comprising a native and/or heterologous enzyme that converts oxaloacetate and acetyl-CoA to malonyl-CoA and pyruvate, wherein said one or more native and/or heterologous enzymes is activated, upregulated, downregulated, or deleted. In some embodiments, the microorganism produces a hydrocarbon. In some embodiments, the enzyme is a

transcarboxylase. In one embodiment, the transcarboxylase is encoded by a polynucleotide from a *Thermoanaerobacter* species, *P. freudenreichii*, *P. acnes*, or *C. thermocellum*. In another embodiment, the transcarboxylase is genetically modified.

- [0028] In some embodiments, the genetic modification produces an altered catalytic activity and/or an altered substrate specificity to improve the conversion of a substrate to a product as compared to the native enzyme. In some embodiments, the genetic modification alters catalytic activity and/or substrate specificity to provide a genetically modified polypeptide that converts a substrate to a product that is not catalyzed by the native enzyme *in vivo*, or is catalyzed at only minimal turnover.

BRIEF DESCRIPTION OF THE DRAWINGS/FIGURES

- [0029] Figure 1A depicts the conversion of phosphoenolpyruvate to oxaloacetate catalyzed by an enzyme from EC 4.1.1.32 or 4.1.1.49.
- [0030] Figure 1B depicts the conversion of oxaloacetate and acetyl-CoA to malonyl-CoA and pyruvate catalyzed by an enzyme from EC 2.1.3.1.
- [0031] Figure 2 depicts three steps in the synthesis of hydrocarbons and hydrocarbon derivatives.
- [0032] Figure 3A depicts the net reaction and a native pathway for the conversion of glucose to butyryl-ACP.
- [0033] Figure 3B depicts the net reaction and a Clostridial pathway for the conversion of glucose to butyryl-ACP.
- [0034] Figure 3C depicts the net reaction and a pathway for the conversion of glucose to butyryl-ACP using a transcarboxylase catalyst.
- [0035] Figure 4 depicts a pathway for the conversion of glucose to acyl_{n+2}-ACP.
- [0036] Figure 5A depicts a pathway for the conversion of a fatty acyl-ACP to a fatty alcohol.
- [0037] Figure 5B depicts a pathway for the conversion of an alcohol and an acid to a wax ester.
- [0038] Figure 6 depicts the Gibbs free energy change for the conversion of glucose into the specified alcohols.
- [0039] Figure 7A is an alignment of the transcarboxylase 5S subunits from *P. freudenreichii*, *P. acnes*, *C. thermocellum*, and *T. saccharolyticum*.

- [0040] Figure 7B is an alignment of the transcarboxylase 1.3S subunits from *P. freudenreichii*, *P. acnes*, *C. thermocellum*, and *T. saccharolyticum*.
- [0041] Figure 7C is an alignment of the transcarboxylase 12S subunit (N-terminus) from *P. freudenreichii*, *P. acnes*, *C. thermocellum*, and *T. saccharolyticum*.
- [0042] Figure 8 depicts the vector pMU433.
- [0043] Figure 9 depicts gene knockout and overexpression strategy to route anaerobic central metabolic flux through oxaloacetate as a key intermediate. Abbreviations: Glc – glucose, Glc-6P – glucose-6-phosphate, PEP – phosphoenolpyruvate, Oxa – oxaloacetate, Fum – fumarate, Mal – malate, Pyr – pyruvate, Ac-CoA – acetyl-CoA, Ac-P – acetyl-phosphate, Aceald – acetaldehyde, Etoh – ethanol.
- [0044] Figure 10 is an agarose gel image showing deletions and overexpressions of target genes in the *E. coli* chromosome to redirect metabolic flux through oxaloacetate.
- [0045] Figure 11 depicts the vector pMU2723 used to construct gene knockouts and chromosomal integrations in *E. coli*.
- [0046] Figure 12 depicts vectors FP45, FP47, FP66, FP67, FP68, and FP75, which are examples of heterologous redox enzymes designed for expression in *E. coli* to modify the native carbohydrate deconstruction pathway.
- [0047] Figure 13 depicts the final step of the anaerobic fatty acid pathway.
- [0048] Figure 14 depicts vectors pMU3061, pMU3062, pMU3063, and pMU3064.
- [0049] Figure 15 depicts the vector pMU2737.
- [0050] Figure 16 depicts the vector pMU2898.
- [0051] Figure 17 depicts the vector pMU2899.
- [0052] Figure 18 depicts the vector pMU2900.
- [0053] Figure 19 depicts the vector pMU2901.
- [0054] Figure 20 demonstrates growth of transformants containing putative transcarboxylases on selective media.
- [0055] Figure 21A depicts polyketide chain synthesis, which proceeds by the addition or condensation of different functional groups to an acyl-ACP chain using a combination of enzymatic activities per two-carbon chain extension.
- [0056] Figure 21B depicts fatty acid chain synthesis, which proceeds by four enzymatic steps per two-carbon chain extension.

- [0057] Figure 22A depicts the total fatty acid content (shown in $\mu\text{g/mL}$) for *E. coli* strain M2933 carrying different acyl-ACP chain termination enzymes.
- [0058] Figure 22B is a graphical representation of the data from Figure 22A.
- [0059] Figure 23 depicts the synthesis of succinate and adipate using omega oxidation.
- [0060] Figure 24 depicts the synthesis of a dicarboxylate using omega oxidation.
- [0061] Figure 25 is a western blot demonstrating the presence of biotinylated enzyme in construct M2557 but not in M2560.
- [0062] Figure 26A depicts a mass spectrum of the transcarboxylase assay products for the negative control sample.
- [0063] Figure 26B depicts a mass spectrum of the transcarboxylase assay products for the transcarboxylase sample.
- [0064] Figure 27 depicts a schematic for the use of the *accC::matBC* *E. coli* strain M2470 to select for more efficient malonyl-CoA production by transcarboxylases.
- [0065] Figure 28 depicts the vector pMU2924.
- [0066] Figure 29 depicts the vector pMU2969.
- [0067] Figure 30 is a phylogenetic tree depicting relatedness between 12S, 5S, 1.3S, and 12S C-term subunits of transcarboxylases from *D. propionicus*, *C. kroppenstedtii*, *P. fuedenreichii*, *G. bemidjiensis*, *C. bescii*, *C. Cellulolyticum*, *C. thermocellum*, and *T. saccharolyticum*.
- [0068] Figure 31 is an alignment of the transcarboxylase subunits from *D. propionicus*, *C. kroppenstedtii*, *P. fuedenreichii*, *G. bemidjiensis*, *C. bescii*, *C. Cellulolyticum*, *C. thermocellum*, and *T. saccharolyticum*.
- [0069] Figure 32 depicts different schematic routes that correspond to co-factor pathway selection presented in Table 10.
- [0070] Figure 33 depicts a pathway for the conversion of glucose to fatty aldehyde or fatty alcohol.
- [0071] Figure 34 depicts the pathways for the conversion of the fermentative metabolism of *S. cerevisiae* from the native pyruvate decarboxylase (*pdh*) based ethanol pathway (A) to an intermediary pyruvate formate lyase and alcohol/aldehyde dehydrogenase (*pfl adhE*) based ethanol pathway (B), and finally to a transcarboxylase based palmitic acid pathway (C).

- [0072] Figure 35 depicts an integration design which deletes FDH1 and replaces the gene with two copies of ADH and two copies of PFL.
- [0073] Figure 36 depicts an integration design which deletes FDH2 and replaces the gene with two copies of ADH and two copies of PFL.
- [0074] Figure 37 depicts an integration design which deletes GPD2 and replaces the gene with two copies of ADH and two copies of PFL.
- [0075] Figure 38 depicts an integration design which deletes GPD1 and replaces the gene with two copies of ADH and two copies of PFL.
- [0076] Figure 39 depicts an integration design which deletes PDC5 and replaces a counter selective gene HSV-TDK and an antibiotic marker (Kan).
- [0077] Figure 40 depicts an integration design which removes the marker shown in Figure 39 resulting in a clean deletion of PDC5.
- [0078] Figure 41 depicts an integration design which deletes PDC6 and replaces a counter selective gene HSV-TDK and an antibiotic marker (Kan).
- [0079] Figure 42 depicts an integration design which removes the marker shown in Figure 41 resulting in a clean deletion of PDC6.
- [0080] Figure 43 depicts an integration design which deletes PDC1 and replaces it with a counter selective gene HSV-TDK and an antibiotic marker (Kan).
- [0081] Figure 44 depicts an integration design which removes the marker shown in Figure 41 resulting in a clean deletion of PDC1.
- [0082] Figure 45 is a phylogenetic tree depicting relatedness between bifunctional malonyl-CoA reductases from *C. aurantiacus*, *C. aurantiacus* J-10-fl, *Chloroflexus* sp. Y-400-fl, *C. aggregans* DSM 9485, *O. trichoides* DG6, *R. castenholzii* DSM 13941, *R.oseiflexus* sp. RS-1, *Erythrobacter* sp. NAP1, and *gamma proteobacterium* NOR51-B.
- [0083] Figure 46 is an alignment of bifunctional malonyl-CoA reductases from *C. aurantiacus*, *C. aurantiacus* J-10-fl, *Chloroflexus* sp. Y-400-fl, *C. aggregans* DSM 9485, *O. trichoides* DG6, *R. castenholzii* DSM 13941, *R.oseiflexus* sp. RS-1, *Erythrobacter* sp. NAP1, and *gamma proteobacterium* NOR51-B.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

- [0084] The indefinite articles "a" and "an" preceding an element or component of the invention are intended to include plurals of the element or component, *e.g.*, one or at least one of the element or component, unless the context is such that only the singular form is intended.
- [0085] The term "heterologous" when used in reference to a polynucleotide, a gene, a polypeptide, or an enzyme refers to a polynucleotide, gene, polypeptide, or an enzyme not normally found in the host organism. "Heterologous" also includes a native coding region, or portion thereof, that is reintroduced into the source organism in a form that is different from the corresponding native gene, *e.g.*, not in its natural location in the organism's genome. The heterologous polynucleotide or gene may be introduced into the host organism by, *e.g.*, gene transfer. A heterologous gene may include a native coding region that is a portion of a chimeric gene including non-native regulatory regions that is reintroduced into the native host. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes.
- [0086] The term "heterologous polynucleotide" is intended to include a polynucleotide that encodes one or more polypeptides or portions or fragments of polypeptides. A heterologous polynucleotide may be derived from any source, *e.g.*, eukaryotes, prokaryotes, viruses, or synthetic polynucleotide fragments.
- [0087] The terms "promoter" or "surrogate promoter" is intended to include a polynucleotide that can transcriptionally control a gene-of-interest that it does not transcriptionally control in nature. In certain embodiments, the transcriptional control of a surrogate promoter results in an increase in expression of the gene-of-interest. In certain embodiments, a surrogate promoter is placed 5' to the gene-of-interest. A surrogate promoter may be used to replace the natural promoter, or may be used in addition to the natural promoter. A surrogate promoter may be endogenous with regard to the host cell in which it is used, or it may be a heterologous polynucleotide sequence introduced into the host cell, *e.g.*, exogenous with regard to the host cell in which it is used.
- [0088] The terms "gene(s)" or "polynucleotide" or "polynucleotide sequence(s)" are intended to include nucleic acid molecules, *e.g.*, polynucleotides which include an open

reading frame encoding a polypeptide, and can further include non-coding regulatory sequences, and introns. In addition, the terms are intended to include one or more genes that map to a functional locus. In addition, the terms are intended to include a specific gene for a selected purpose. The gene may be endogenous to the host cell or may be recombinantly introduced into the host cell, *e.g.*, as a plasmid maintained episomally or a plasmid (or fragment thereof) that is stably integrated into the genome. In addition to the plasmid form, a gene may, for example, be in the form of linear DNA. The term gene is also intended to cover all copies of a particular gene, *e.g.*, all of the DNA sequences in a cell encoding a particular gene product.

[0089] The term "transcriptional control" is intended to include the ability to modulate gene expression at the level of transcription. In certain embodiments, transcription, and thus gene expression, is modulated by replacing or adding a surrogate promoter near the 5' end of the coding region of a gene-of-interest, thereby resulting in altered gene expression. In certain embodiments, the transcriptional control of one or more genes is engineered to result in the optimal expression of such genes, *e.g.*, in a desired ratio. The term also includes inducible transcriptional control as recognized in the art.

[0090] The term "expression" is intended to include the expression of a gene at least at the level of mRNA production.

[0091] The term "expression product" is intended to include the resultant product, *e.g.*, a polypeptide, of an expressed gene.

[0092] The term "polypeptide" is intended to encompass a singular "polypeptide," as well as plural "polypeptides," and refers to a molecule composed of monomers (amino acids) linearly linked by amide bonds (also known as peptide bonds). The term "polypeptide" refers to any chain or chains of two or more amino acids and does not refer to a specific length of the amino acids. Thus, peptides, dipeptides, tripeptides, oligopeptides, "protein," "amino acid chain," "enzyme," or any other term used to refer to a chain or chains of two or more amino acids, are included within the definition of "polypeptide," and the term "polypeptide" may be used instead of, or interchangeably with, any of these terms. A polypeptide may be derived from a natural biological source or produced by recombinant technology. It may be generated in any manner, including by chemical synthesis.

[0093] The term "increased expression" is intended to include an alteration in gene expression at least at the level of increased mRNA production and, preferably, at the level of polypeptide expression. The term "increased production" is intended to include an increase in the amount of a polypeptide expressed, in the level of the enzymatic activity of the polypeptide, or a combination thereof, as compared to the native production of, or the enzymatic activity of, the polypeptide.

[0094] The terms "activity," "activities," "enzymatic activity," and "enzymatic activities" are used interchangeably and are intended to include any functional activity normally attributed to a selected polypeptide when produced under favorable conditions. Typically, the activity of a selected polypeptide encompasses the total enzymatic activity associated with the produced polypeptide. The polypeptide produced by a host cell and having enzymatic activity may be located in the intracellular space of the cell, cell-associated, secreted into the extracellular milieu, or a combination thereof. Techniques for determining total activity as compared to secreted activity are described herein and are known in the art.

[0095] The term "secreted" is intended to include the movement of polypeptides to the periplasmic space or extracellular milieu. The term "increased secretion" is intended to include situations in which a given polypeptide is secreted at an increased level (*i.e.*, in excess of the naturally-occurring amount of secretion). In certain embodiments, the term "increased secretion" refers to an increase in secretion of a given polypeptide that is at least about 10% or at least about 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, 1000%, or more, as compared to the naturally-occurring level of secretion.

[0096] The term "secretory polypeptide" is intended to include any polypeptide(s), alone or in combination with other polypeptides, that facilitate the transport of another polypeptide from the intracellular space of a cell to the extracellular milieu. In certain embodiments, the secretory polypeptide(s) encompass all the necessary secretory polypeptides sufficient to impart secretory activity to a Gram-negative or Gram-positive host cell or to a yeast host cell. Typically, secretory proteins are encoded in a single region or locus that may be isolated from one host cell and transferred to another host cell using genetic engineering. In certain embodiments, the secretory polypeptide(s) are derived from any bacterial cell having secretory activity or any yeast cell having secretory activity. In certain embodiments, the secretory polypeptide(s) are derived from a host cell

having Type II secretory activity. In certain embodiments, the host cell is a thermophilic bacterial cell. In certain embodiments, the host cell is a yeast cell.

[0097] The term "derived from" is intended to include the isolation (in whole or in part) of a polynucleotide segment from an indicated source or the purification of a polypeptide from an indicated source. The term is intended to include, for example, direct cloning, PCR amplification, or artificial synthesis from or based on a sequence associated with the indicated polynucleotide source.

[0098] By "thermophilic" is meant an organism that thrives at a temperature of about 45°C or higher.

[0099] By "mesophilic" is meant an organism that thrives at a temperature of about 20-45°C.

[0100] Certain embodiments of the present invention provide for the "insertion," (*e.g.*, the addition, integration, incorporation, or introduction) of certain genes or particular polynucleotide sequences within thermophilic or mesophilic microorganisms, which insertion of genes or particular polynucleotide sequences may be understood to encompass "genetic modification(s)" or "transformation(s)" such that the resulting strains of said thermophilic or mesophilic microorganisms may be understood to be "genetically modified" or "transformed." In certain embodiments, strains may be of bacterial, fungal, or yeast origin.

[0101] In certain embodiments, the polynucleotide sequences of the invention are genetically modified such that the encoded enzyme is engineered to alter catalytic activity and/or alter substrate specificity to improve the conversion of a substrate to a product as compared to the native enzyme. In certain aspects, the genetic modification alters catalytic activity and/or substrate specificity to provide an encoded enzyme that converts a substrate to a product that is not catalyzed by the native enzyme *in vivo*, or is catalyzed at only minimal turnover. Techniques to genetically modify polynucleotides are known in the art and include, but are not limited to, alteration, insertion, and/or deletion of one or more nucleic acids in the polynucleotide. Such techniques to alter, insert, and/or delete nucleic acids include, but are not limited to, random, site-directed, or saturating mutagenesis.

[0102] Certain embodiments of the present invention provide for the "inactivation" or "deletion" of certain genes or particular polynucleotide sequences within thermophilic or

mesophilic microorganisms, which "inactivation" or "deletion" of genes or particular polynucleotide sequences may be understood to encompass "genetic modification(s)" or "transformation(s)" such that the resulting strains of said thermophilic or mesophilic microorganisms may be understood to be "genetically modified" or "transformed." In certain embodiments, strains may be of bacterial, fungal, or yeast origin.

[0103] The term "consolidated bioprocessing" or "CBP" is intended to include a processing strategy for cellulosic biomass that involves consolidating into a single process step, four biologically-mediated events: enzyme production, hydrolysis, hexose fermentation, and pentose fermentation. Implementing this strategy requires development of microorganisms that both utilize cellulose, hemicellulosics, and other biomass components while also producing a product of interest at sufficiently high yield and concentrations. The feasibility of CBP is supported by kinetic and bioenergetic analysis. See van Walsum and Lynd (1998) *Biotech. Bioeng.* 58:316.

[0104] The term "CBP organism" is intended to include microorganisms of the invention, e.g., microorganisms that have properties suitable for CBP.

[0105] In one aspect of the invention, the genes or particular polynucleotide sequences are inserted to activate the activity for which they encode, such as the expression of an enzyme. In certain embodiments, genes encoding enzymes in the metabolic production of fatty acids may be added to a mesophilic or a thermophilic organism.

[0106] In one aspect of the invention, the genes or particular polynucleotide sequences are partially, substantially, or completely deleted, silenced, inactivated, or down-regulated in order to inactivate the activity for which they encode, such as the expression of an enzyme. Deletions provide maximum stability because there is no opportunity for a reverse mutation to restore function. Alternatively, genes can be partially, substantially, or completely deleted, silenced, inactivated, or down-regulated by insertion of nucleic acid sequences that disrupt the function and/or expression of the gene (e.g., P1 transduction or other methods known in the art). The terms "eliminate," "elimination," and "knockout" are used interchangeably with the terms "deletion," "partial deletion," "substantial deletion," or "complete deletion." In certain embodiments, strains of thermophilic or mesophilic microorganisms of interest may be engineered by site directed homologous recombination to knockout the production of organic acids. In still other

embodiments, RNAi or antisense DNA (asDNA) may be used to partially, substantially, or completely silence, inactivate, or down-regulate a particular gene of interest.

[0107] In certain embodiments, the genes targeted for deletion or inactivation as described herein may be endogenous to the native strain of the microorganism, and may thus be understood to be referred to as "native gene(s)" or "endogenous gene(s)." An organism is in "a native state" if it has not been genetically engineered or otherwise manipulated by the hand of man in a manner that intentionally alters the genetic and/or phenotypic constitution of the organism. For example, wild-type organisms may be considered to be in a native state. In other embodiments, the gene(s) targeted for deletion or inactivation may be non-native to the organism.

[0108] Similarly, the enzymes of the invention as described herein can be endogenous to the native strain of the microorganism, and can thus be understood to be referred to as "native" or "endogenous."

[0109] The term "upregulated" means increased in activity, *e.g.*, increase in enzymatic activity of the enzyme as compared to activity in a native host organism.

[0110] The term "downregulated" means decreased in activity, *e.g.*, decrease in enzymatic activity of the enzyme as compared to activity in a native host organism.

[0111] The term "activated" means expressed or metabolically functional.

[0112] As used herein, the term "hydrocarbon" is intended to include compounds containing only carbon and hydrogen, such as aliphatic hydrocarbons and aromatic hydrocarbons. Examples of hydrocarbons include, but are not limited to, alkanes, alkenes, or alkynes.

[0113] As used herein, the term "hydrocarbon derivative" is intended to include compounds formed by the addition of at least one functional group to a hydrocarbon. Examples of hydrocarbon derivatives include, but are not limited to, aldehydes, alcohols, esters, fatty acids, unsaturated fatty acids, branched-chain fatty acids, branched methoxy fatty acids, multi-methyl branched acids, divinyl-ether fatty acids, w-phenylalkanoic acids, dicarboxylic acids.

[0114] The term "carbohydrate source" is intended to include any source of carbohydrate including, but not limited to, biomass or carbohydrates, such as a sugar or a sugar alcohol. "Carbohydrates" include, but are not limited to, monosaccharides (*e.g.*, glucose, fructose, galactose, xylose, arabinose, or ribose), sugar derivatives (*e.g.*, sorbitol, glycerol,

galacturonic acid, rhamnose, xylitol), disaccharides (*e.g.*, sucrose, cellobiose, maltose, or lactose), oligosaccharides (*e.g.*, xylooligomers, cellodextrins, or maltodextrins), and polysaccharides (*e.g.*, xylan, cellulose, starch, mannan, alginate, or pectin).

[0115] As used herein, the term "microaerophilic" is intended to include conditions in which oxygen is present at lower concentrations than atmospheric oxygen content. A microaerophilic organism is one that requires a lower concentration of oxygen for growth than is present in the atmosphere. Microaerophilic conditions include those in which oxygen is present at less than about 5%, less than about 10%, less than about 15%, less than about 20%, less than about 25%, less than about 30%, less than about 35%, less than about 40%, less than about 45%, less than about 50%, less than about 55%, less than about 60%, less than about 65%, less than about 70%, less than about 75%, less than about 80%, less than about 85%, less than about 90%, less than about 95%, or less than about 99% of atmospheric oxygen concentration.

[0116] As used herein, the term "malonyl-CoA derived product" or "malonyl-CoA derived bioproduct" is intended to include those products that are synthesized from, derived from, or are used as an intermediate in their synthesis from, malonyl-CoA. The term includes products such as hydrocarbons, hydrocarbon derivatives, polyketides, organic acids, including but not limited to adipic acid and 3-hydroxypropionate, and any other products from which malonyl-CoA can serve as a precursor.

Metabolic Pathway Engineering

[0117] Many bacteria have the ability to ferment simple hexose sugars into a mixture of acidic and pH-neutral products via the process of glycolysis. The glycolytic pathway is abundant and comprises a series of enzymatic steps whereby a six carbon glucose molecule is broken down, via multiple intermediates, into two molecules of the three carbon compound pyruvate. This process results in the net generation of ATP (biological energy supply) and the reduced cofactor NADH.

[0118] Pyruvate is an important intermediary compound of metabolism. For example, under aerobic conditions pyruvate may be oxidized to acetyl coenzyme A (acetyl CoA), which then enters the tricarboxylic acid cycle (TCA), which in turn generates synthetic precursors, CO₂ and reduced cofactors. The cofactors are then oxidized by donating hydrogen equivalents, via a series of enzymatic steps, to oxygen resulting in the

formation of water and ATP. This process of energy formation is known as oxidative phosphorylation.

- [0119] Under anaerobic conditions (no available oxygen), fermentation occurs in which the degradation products of organic compounds serve as hydrogen donors and acceptors. Excess NADH from glycolysis is oxidized in reactions involving the reduction of organic substrates to products, such as lactate and ethanol. In addition, ATP is regenerated from the production of organic acids, such as acetate, in a process known as substrate level phosphorylation. Therefore, the fermentation products of glycolysis and pyruvate metabolism include a variety of organic acids, alcohols and CO₂.

Biomass

- [0120] Biomass can include any type of biomass known in the art or described herein. The terms "lignocellulosic material," "lignocellulosic substrate," and "cellulosic biomass" mean any type of biomass comprising cellulose, hemicellulose, lignin, or combinations thereof, such as but not limited to woody biomass, forage grasses, herbaceous energy crops, non-woody-plant biomass, agricultural wastes and/or agricultural residues, forestry residues and/or forestry wastes, paper-production sludge and/or waste paper sludge, waste-water-treatment sludge, municipal solid waste, corn fiber from wet and dry mill corn ethanol plants, and sugar-processing residues. The terms "hemicellulosics," "hemicellulosic portions," and "hemicellulosic fractions" mean the non-lignin, non-cellulose elements of lignocellulosic material, such as but not limited to hemicellulose (*i.e.*, comprising xyloglucan, xylan, glucuronoxylan, arabinoxylan, mannan, glucomannan, and galactoglucomannan), pectins (*e.g.*, homogalacturonans, rhamnogalacturonan I and II, and xylogalacturonan), and proteoglycans (*e.g.*, arabinogalactan-protein, extensin, and proline-rich proteins).

- [0121] In a non-limiting example, the lignocellulosic material can include, but is not limited to, woody biomass, such as recycled wood pulp fiber, sawdust, hardwood, softwood, and combinations thereof; grasses, such as switch grass, cord grass, rye grass, reed canary grass, miscanthus, or a combination thereof; sugar-processing residues, such as but not limited to sugar cane bagasse; agricultural wastes, such as but not limited to rice straw, rice hulls, barley straw, corn cobs, cereal straw, wheat straw, canola straw, oat straw, oat hulls, and corn fiber; stover, such as but not limited to soybean stover, corn stover; succulents, such as but not limited to, Agave; and forestry wastes, such as but not

limited to, recycled wood pulp fiber, sawdust, hardwood (*e.g.*, poplar, oak, maple, birch, willow), softwood, or any combination thereof. Lignocellulosic material may comprise one species of fiber; alternatively, lignocellulosic material may comprise a mixture of fibers that originate from different lignocellulosic materials. Other lignocellulosic materials are agricultural wastes, such as cereal straws, including wheat straw, barley straw, canola straw and oat straw; corn fiber; stovers, such as corn stover and soybean stover; grasses, such as switch grass, reed canary grass, cord grass, and miscanthus; or combinations thereof.

[0122] Paper sludge is also a viable feedstock for lactate or acetate production. Paper sludge is solid residue arising from pulping and paper-making, and is typically removed from process wastewater in a primary clarifier. At a disposal cost of \$30/wet ton, the cost of sludge disposal equates to \$5/ton of paper that is produced for sale. The cost of disposing of wet sludge is a significant incentive to convert the material for other uses, such as conversion to ethanol. Processes provided by the present invention are widely applicable. Moreover, the saccharification and/or fermentation products may be used to produce ethanol or higher value added chemicals, such as organic acids, aromatics, esters, acetone and polymer intermediates.

Xylose Metabolism

[0123] Xylose is a five-carbon monosaccharide that can be metabolized into useful products by a variety of organisms. There are two main pathways of xylose metabolism, each unique in the characteristic enzymes they utilize. One pathway is called the "Xylose Reductase-Xylitol Dehydrogenase" or XR-XDH pathway. Xylose reductase (XR) and xylitol dehydrogenase (XDH) are the two main enzymes used in this method of xylose degradation. XR, encoded by the *XYL1* gene, is responsible for the reduction of xylose to xylitol and is aided by cofactors NADH or NADPH. Xylitol is then oxidized to xylulose by XDH, which is expressed through the *XYL2* gene, and accomplished exclusively with the cofactor NAD⁺. Because of the varying cofactors needed in this pathway and the degree to which they are available for usage, an imbalance can result in an overproduction of xylitol byproduct and an inefficient production of desirable ethanol. Varying expression of the XR and XDH enzyme levels have been tested in the laboratory in the attempt to optimize the efficiency of the xylose metabolism pathway.

[0124] The other pathway for xylose metabolism is called the "Xylose Isomerase" (XI) pathway. Enzyme XI is responsible for direct conversion of xylose into xylulose, and does not proceed via a xylitol intermediate. Both pathways create xylulose, although the enzymes utilized are different. After production of xylulose both the XR-XDH and XI pathways proceed through enzyme xylulokinase (XK), encoded on gene XKS1, to further modify xylulose into xylulose-5-P where it then enters the pentose phosphate pathway for further catabolism.

[0125] Studies on flux through the pentose phosphate pathway during xylose metabolism have revealed that limiting the speed of this step may be beneficial to the efficiency of fermentation to ethanol. Modifications to this flux that may improve ethanol production include a) lowering phosphoglucose isomerase activity, b) deleting the GND1 gene, and c) deleting the ZWF1 gene. See Jeppsson *et al.*, *Appl. Environ. Microbiol.* 68:1604-09 (2002). Since the pentose phosphate pathway produces additional NADPH during metabolism, limiting this step will help to correct the already evident imbalance between NAD(P)H and NAD⁺ cofactors and reduce xylitol byproduct. Another experiment comparing the two xylose metabolizing pathways revealed that the XI pathway was best able to metabolize xylose to produce the greatest ethanol yield, while the XR-XDH pathway reached a much faster rate of ethanol production. See Karhumaa *et al.*, *Microb Cell Fact.* 6:5 (February 5, 2007); see also International Publication No. WO2006/009434, incorporated herein by reference in its entirety.

Arabinose Metabolism

[0126] Arabinose is a five-carbon monosaccharide that can be metabolized into useful products by a variety of organisms. L-Arabinose residues are found widely distributed among many heteropolysaccharides of different plant tissues, such as arabinans, arabinogalactans, xylans and arabinoxylans. *Bacillus* species in the soil participate in the early stages of plant material decomposition, and *B. subtilis* secretes three enzymes, an endo-arabanase and two arabinosidases, capable of releasing arabinosyl oligomers and L-arabinose from plant cell.

[0127] Three pathways for L-arabinose metabolism in microorganisms have been described. Many bacteria, including *Escherichia coli*, use arabinose isomerase (AraA; E.C. 5.3.1.4), ribulokinase (AraB; E.C. 2.7.1.16), and ribulose phosphate epimerase (AraD; E.C. 5.1.3.4) to sequentially convert L-arabinose to D-xylulose-5-phosphate

through L-ribulose and L-ribulose 5-phosphate. *See, e.g.,* Sa-Nogueira I., *et al., Microbiology* 143:957–69 (1997). The D-xylulose-5-phosphate then enters the pentose phosphate pathway for further catabolism. In the second pathway, L-arabinose is converted to L-2-keto-3-deoxyarabonate (L-KDA) by the consecutive action of enzymes arabinose dehydrogenase (ADH), arabinolactone (AL), and arabinonate dehydratase (AraC). *See, e.g.,* Watanabe, S., *et al., J. Biol. Chem.* 281: 2612-2623 (2006). L-KDA can be further metabolized in two alternative pathways: 1) L-KDA conversion to 2-ketoglutarate via 2-ketoglutaric semialdehyde (KGSA) by L-KDA dehydratase and KGSA dehydrogenase or 2) L-KDA conversion to pyruvate and glycolaldehyde by L-KDA aldolase. In the third, fungal pathway, L-arabinose is converted to D-xylulose-5-phosphate through L-arabinitol, L-xylulose, and xylitol, by enzymes such as NAD(P)H-dependent aldose reductase (AR), L-arabinitol 4-dehydrogenase (ALDH), L-xylulose reductase (LXR), xylitol dehydrogenase (XylD), and xylulokinase (XylB). These, and additional proteins involved in arabinose metabolism and regulation may be found at http://www.nmpdr.org/FIG/wiki/rest.cgi/NmpdrPlugin/SeedViewer?page=Subsystems;subsystem=L-Arabinose_utilization, visited July 29, 2011, which is incorporated by reference herein in its entirety.

[0128] AraC protein regulates expression of its own synthesis and the other genes of the Ara system. *See* Schleif, R., *Trends Genet.* 16(12):559-65 (2000). In *E. coli*, the AraC protein positively and negatively regulates expression of the proteins required for the uptake and catabolism of the sugar L-arabinose. Homologs of AraC, such as regulatory proteins RhaR and RhaS of the rhamnose operon, have been identified that contain regions homologous to the DNA-binding domain of AraC (Leal, T.F. and de Sa-Nogueira, I., *FEMS Microbiol Lett.* 241(1):41-48 (2004)). Such arabinose regulatory proteins are referred to as the AraC/XylS family. *See also*, Mota, L.J., *et al., Mol. Microbiol.* 33(3):476-89 (1999); Mota, L.J., *et al., J. Bacteriol.* 183(14):4190-201 (2001).

[0129] In *E. coli*, the transport of L-arabinose across the *E. coli* cytoplasmic membrane requires the expression of either the high-affinity transport operon, *araFGH*, a binding protein-dependent system on the low-affinity transport operon, *araE*, or a proton symporter. Additional arabinose transporters include those identified from *K. marxianus* and *P. guilliermondii*, disclosed in U.S. Patent No. 7,846,712, which is incorporated by reference herein.

- [0130] In some embodiments, the recombinant microorganisms of the invention have the ability to metabolize arabinose using one or more of the above enzymes.

Vectors and Host Cells

- [0131] The present invention also relates to vectors which include genes encoding for enzymes of the present invention, as described above, as well as host cells which are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques.
- [0132] Host cells are genetically engineered (transduced or transformed or transfected) with the vectors of this invention which can be, for example, a cloning vector or an expression vector. The vector can be, for example, in the form of a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the genes of the present invention. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.
- [0133] The DNA sequence in the expression vector is operatively associated with an appropriate expression control sequence(s) (promoter) to direct mRNA synthesis. Any suitable promoter to drive gene expression in the host cells of the invention can be used. Additionally, promoters known to control expression of genes in prokaryotic or lower eukaryotic cells can be used. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector can also include appropriate sequences for amplifying expression, or can include additional regulatory regions.
- [0134] The vector containing the appropriate selectable marker sequence as used herein, as well as an appropriate promoter or control sequence, can be employed to transform an appropriate thermophilic host to permit the host to express the protein.
- [0135] Host cells useful in the present invention include any prokaryotic or eukaryotic cells; for example, microorganisms selected from bacterial, algal, and yeast cells. Among host cells thus suitable for the present invention are microorganisms, for example, of the genera *Aeromonas*, *Aspergillus*, *Bacillus*, *Escherichia*, *Kluyveromyces*, *Pichia*, *Rhodococcus*, *Saccharomyces* and *Streptomyces*.

[0136] In some embodiments, the host cells are microorganisms. In one embodiment the microorganism is a yeast. According to the present invention the yeast host cell can be, for example, from the genera *Saccharomyces*, *Kluyveromyces*, *Candida*, *Pichia*, *Schizosaccharomyces*, *Hansenula*, *Kloeckera*, *Schwanniomyces*, and *Yarrowia*. Yeast species as host cells may include, for example, *S. cerevisiae*, *S. bulderi*, *S. barnetti*, *S. exiguus*, *S. uvarum*, *S. diastaticus*, *K. lactis*, *K. marxianus*, or *K. fragilis*. In some embodiments, the yeast is selected from the group consisting of *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Candida albicans*, *Pichia pastoris*, *Pichia stipitis*, *Yarrowia lipolytica*, *Hansenula polymorpha*, *Phaffia rhodozyma*, *Candida utilis*, *Arxula adeninivorans*, *Debaryomyces hansenii*, *Debaryomyces polymorphus*, *Schizosaccharomyces pombe* and *Schwanniomyces occidentalis*. In one particular embodiment, the yeast is *Saccharomyces cerevisiae*. In another embodiment, the yeast is a thermotolerant *Saccharomyces cerevisiae*. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

[0137] In some embodiments, the host cell is an oleaginous cell. The oleaginous host cell can be an oleaginous yeast cell. For example, the oleaginous yeast host cell can be from the genera *Blakeslea*, *Candida*, *Cryptococcus*, *Cunninghamella*, *Lipomyces*, *Mortierella*, *Mucor*, *Phycomyces*, *Pythium*, *Rhodosporidium*, *Rhodotorula*, *Trichosporon* or *Yarrowia*. According to the present invention, the oleaginous host cell can be an oleaginous microalgae host cell. For example, the oleaginous microalgae host cell can be from the genera *Thraustochytrium* or *Schizochytrium*. Biodiesel could then be produced from the triglyceride produced by the oleaginous organisms using conventional lipid transesterification processes. In some particular embodiments, the oleaginous host cells can be induced to secrete synthesized lipids. Embodiments using oleaginous host cells are advantageous because they can produce biodiesel from lignocellulosic feedstocks which, relative to oilseed substrates, are cheaper, can be grown more densely, show lower life cycle carbon dioxide emissions, and can be cultivated on marginal lands.

[0138] In some embodiments, the host cell is a thermotolerant host cell. Thermotolerant host cells can be particularly useful in simultaneous saccharification and fermentation processes by allowing externally produced cellulases and ethanol-producing host cells to perform optimally in similar temperature ranges.

- [0139] Thermotolerant host cells can include, for example, *Issatchenkia orientalis*, *Pichia mississippiensis*, *Pichia mexicana*, *Pichia farinosa*, *Clavispora opuntiae*, *Clavispora lusitaniae*, *Candida mexicana*, *Hansenula polymorpha* and *Kluyveromyces* host cells. In some embodiments, the thermotolerant cell is an *S. cerevisiae* strain, or other yeast strain, that has been adapted to grow in high temperatures, for example, by selection for growth at high temperatures in a cytostat.
- [0140] In some particular embodiments, the host cell is a *Kluyveromyces* host cell. For example, the *Kluyveromyces* host cell can be a *K. lactis*, *K. marxianus*, *K. blattae*, *K. phaffii*, *K. yarrowii*, *K. aestuarii*, *K. dobzhanskii*, *K. wickerhamii*, *K. thermotolerans*, or *K. waltii* host cell. In one embodiment, the host cell is a *K. lactis*, or *K. marxianus* host cell. In another embodiment, the host cell is a *K. marxianus* host cell.
- [0141] In some embodiments, the thermotolerant host cell can grow at temperatures above about 30° C, about 31° C, about 32° C, about 33° C, about 34° C, about 35° C, about 36° C, about 37° C, about 38° C, about 39° C, about 40° C, about 41° C or about 42° C. In some embodiments of the present invention the thermotolerant host cell can produce ethanol from cellulose at temperatures above about 30° C, about 31° C, about 32° C, about 33° C, about 34° C, about 35° C, about 36° C, about 37° C, about 38° C, about 39° C, about 40° C, about 41° C, about 42° C, or about 43 °C, or about 44 °C, or about 45 °C, or about 50° C.
- [0142] In some embodiments of the present invention, the thermotolerant host cell can grow at temperatures from about 30° C to 60° C, about 30° C to 55° C, about 30° C to 50° C, about 40° C to 60° C, about 40° C to 55° C or about 40° C to 50° C. In some embodiments of the present invention, the thermotolerant host cell can produce ethanol from cellulose at temperatures from about 30° C to 60° C, about 30° C to 55° C, about 30° C to 50° C, about 40° C to 60° C, about 40° C to 55° C or about 40° C to 50° C.
- [0143] In some embodiments, the host cell has the ability to metabolize xylose. Detailed information regarding the development of the xylose-utilizing technology can be found in the following publications: Kuyper M., *et al.*, *FEMS Yeast Res.* 4: 655-64 (2004); Kuyper M., *et al.*, *FEMS Yeast Res.* 5:399-409 (2005); and Kuyper M., *et al.*, *FEMS Yeast Res.* 5:925-34 (2005), which are herein incorporated by reference in their entirety. For example, xylose-utilization can be accomplished in *S. cerevisiae* by heterologously expressing the xylose isomerase gene, *XylA*, *e.g.*, from the anaerobic fungus *Piromyces*

sp. E2, overexpressing five *S. cerevisiae* enzymes involved in the conversion of xylulose to glycolytic intermediates (xylulokinase, ribulose 5-phosphate isomerase, ribulose 5-phosphate epimerase, transketolase and transaldolase) and deleting the GRE3 gene encoding aldose reductase to minimize xylitol production.

[0144] The host cells can contain antibiotic markers or can contain no antibiotic markers.

[0145] Aspects of the present invention relate to the use of thermophilic and thermotolerant microorganisms as hosts. Their potential in process applications in biotechnology stems from their ability to grow at relatively high temperatures with attendant high metabolic rates, production of physically and chemically stable enzymes, elevated yields of end products, and lower susceptibility to microbial contamination. Major groups of thermophilic bacteria include eubacteria and archaeobacteria. Thermophilic eubacteria include: phototropic bacteria, such as cyanobacteria, purple bacteria, and green bacteria; Gram-positive bacteria, such as *Bacillus*, *Clostridium*, Lactic acid bacteria, and Actinomyces; and other eubacteria, such as *Thiobacillus*, Spirochete, *Desulfotomaculum*, Gram-negative aerobes, Gram-negative anaerobes, and *Thermotoga*. Within archaeobacteria are considered Methanogens, extreme thermophiles (an art-recognized term), and *Thermoplasma*. In certain embodiments, the present invention relates to Gram-negative organotrophic thermophiles of the genera *Thermus*, Gram-positive eubacteria, such as genera *Clostridium*, and also which comprise both rods and cocci, genera in group of eubacteria, such as *Thermosipho* and *Thermotoga*, genera of Archaeobacteria, such as *Thermococcus*, *Thermoproteus* (rod-shaped), *Thermofilum* (rod-shaped), *Pyrodictium*, *Acidianus*, *Sulfolobus*, *Pyrobaculum*, *Pyrococcus*, *Thermodiscus*, *Staphylothermus*, *Desulfurococcus*, *Archaeoglobus*, and *Methanopyrus*.

[0146] Some examples of thermophilic or mesophilic (including bacteria, procaryotic microorganism, and fungi), which may be suitable for the present invention include, but are not limited to: *Clostridium thermosulfurogenes*, *Clostridium cellulolyticum*, *Clostridium thermocellum*, *Clostridium thermohydrosulfuricum*, *Clostridium thermoaceticum*, *Clostridium thermosaccharolyticum*, *Clostridium tartarivorum*, *Clostridium thermocellulaseum*, *Clostridium phytofermentans*, *Clostridium straminosolvens*, *Thermoanaerobacterium thermosaccharolyticum*, *Thermoanaerobacterium saccharolyticum*, *Thermobacteroides acetoethylicus*, *Thermoanaerobium brockii*, *Methanobacterium thermoautotrophicum*, *Anaerocellum*

thermophilum, *Pyrodictium occultum*, *Thermoproteus neutrophilus*, *Thermofilum librum*, *Thermothrix thioparus*, *Desulfovibrio thermophilus*, *Thermoplasma acidophilum*, *Hydrogenomonas thermophilus*, *Thermomicrobium roseum*, *Thermus flavas*, *Thermus ruber*, *Pyrococcus furiosus*, *Thermus aquaticus*, *Thermus thermophilus*, *Chloroflexus aurantiacus*, *Thermococcus litoralis*, *Pyrodictium abyssi*, *Bacillus stearothermophilus*, *Cyanidium caldarium*, *Mastigocladus laminosus*, *Chlamydothrix calidissima*, *Chlamydothrix penicillata*, *Thiothrix carnea*, *Phormidium tenuissimum*, *Phormidium geysericola*, *Phormidium subterraneum*, *Phormidium bijahensi*, *Oscillatoria filiformis*, *Synechococcus lividus*, *Chloroflexus aurantiacus*, *Pyrodictium brockii*, *Thiobacillus thiooxidans*, *Sulfolobus acidocaldarius*, *Thiobacillus thermophilica*, *Bacillus stearothermophilus*, *Cercosulcifer hamathensis*, *Vahlkampfia reichi*, *Cyclidium citrullus*, *Dactylaria gallopava*, *Synechococcus lividus*, *Synechococcus elongatus*, *Synechococcus minervae*, *Synechocystis aquatilis*, *Aphanocapsa thermalis*, *Oscillatoria terebriformis*, *Oscillatoria amphibia*, *Oscillatoria germinata*, *Oscillatoria okenii*, *Phormidium laminosum*, *Phormidium parparasiens*, *Symploca thermalis*, *Bacillus acidocaldarias*, *Bacillus coagulans*, *Bacillus thermocatenalatus*, *Bacillus licheniformis*, *Bacillus pamilas*, *Bacillus macerans*, *Bacillus circulans*, *Bacillus laterosporus*, *Bacillus brevis*, *Bacillus subtilis*, *Bacillus sphaericus*, *Desulfotomaculum nigrificans*, *Streptococcus thermophilus*, *Lactobacillus thermophilus*, *Lactobacillus bulgaricus*, *Bifidobacterium thermophilum*, *Streptomyces fragmentosporus*, *Streptomyces thermonitrificans*, *Streptomyces thermovulgaris*, *Pseudonocardia thermophila*, *Thermoactinomyces vulgaris*, *Thermoactinomyces sacchari*, *Thermoactinomyces candidas*, *Thermomonospora curvata*, *Thermomonospora viridis*, *Thermomonospora citrina*, *Microbispora thermodiastatica*, *Microbispora aerata*, *Microbispora bispore*, *Actinobifida dichotomica*, *Actinobifida chromogena*, *Micropolyspora caesia*, *Micropolyspora faeni*, *Micropolyspora cectivugida*, *Micropolyspora cabrobrunea*, *Micropolyspora thermovirida*, *Micropolyspora viridinigra*, *Methanobacterium thermoautothropicum*, *Caldicellulosiruptor acetigenus*, *Caldicellulosiruptor saccharolyticus*, *Caldicellulosiruptor kristjanssonii*, *Caldicellulosiruptor owensensis*, *Caldicellulosiruptor lactoaceticus*, *Clostridium clariflavum*, *E. coli strain B*, *strain C*, *strain K*, *strain W*, *Shewanella*, *Propionibacterium acnes*, *Propionibacterium freudenreichii*, *Propionibacterium shermanii*, *Propionibacterium pentosaceum*, *Propionibacterium arabinosum*, *Clostridium*

acetobutylicum, *Clostridium beijerinckii*, *Lactobacillus thermophilus*, *Lactobacillus bulgaricus*, *Lactococcus lactis*, variants thereof, and/or progeny thereof.

[0147] In particular embodiments, the present invention relates to thermophilic bacteria selected from the group consisting of *Clostridium cellulolyticum*, *Clostridium thermocellum*, and *Thermoanaerobacterium saccharolyticum*.

[0148] In certain embodiments, the present invention relates to thermophilic bacteria selected from the group consisting of *Fervidobacterium gondwanense*, *Clostridium thermolacticum*, *Moorella sp.*, and *Rhodothermus marinus*.

[0149] In certain embodiments, the present invention relates to thermophilic bacteria of the genera *Thermoanaerobacterium* or *Thermoanaerobacter*, including, but not limited to, species selected from the group consisting of: *Thermoanaerobacterium thermosulfurigenes*, *Thermoanaerobacterium aotearoense*, *Thermoanaerobacterium polysaccharolyticum*, *Thermoanaerobacterium zeae*, *Thermoanaerobacterium xylanolyticum*, *Thermoanaerobacterium saccharolyticum*, *Thermoanaerobium brockii*, *Thermoanaerobacterium thermosaccharolyticum*, *Thermoanaerobacter thermohydrosulfuricus*, *Thermoanaerobacter ethanolicus*, *Thermoanaerobacter brockii*, variants thereof, and progeny thereof.

[0150] In certain embodiments, the present invention relates to microorganisms of the genera *Geobacillus*, *Saccharococcus*, *Paenibacillus*, *Bacillus*, and *Anoxybacillus*, including, but not limited to, species selected from the group consisting of: *Geobacillus thermoglucosidasius*, *Geobacillus stearothermophilus*, *Saccharococcus caldxylosilyticus*, *Saccharococcus thermophilus*, *Paenibacillus campinasensis*, *Bacillus flavothermus*, *Anoxybacillus kamchatkensis*, *Anoxybacillus gonensis*, variants thereof, and progeny thereof.

[0151] In certain embodiments, the present invention relates to mesophilic bacteria selected from the group consisting of *Saccharophagus degradans*; *Flavobacterium johnsoniae*; *Fibrobacter succinogenes*; *Clostridium hungatei*; *Clostridium phytofermentans*; *Clostridium cellulolyticum*; *Clostridium aldrichii*; *Clostridium termitididis*; *Acetivibrio cellulolyticus*; *Acetivibrio ethanolgignens*; *Acetivibrio multivorans*; *Bacteroides cellulosolvens*; *Alkalibacter saccharofomentans*, variants thereof, and progeny thereof. In certain embodiments, the present invention relates to mesophilic bacteria selected from the group consisting of *Escherichia coli*, *E. coli* strain

B, strain *C*, strain *K*, strain *W*, *Shewanella*, *Propionibacterium acnes*, *Propionibacterium freudenreichii*, *Propionibacterium shermanii*, *Propionibacterium pentosaceum*, *Propionibacterium arabinosum*, *Clostridium acetobutylicum*, *Clostridium beijerinckii*, variants thereof, and progeny thereof.

Codon-Optimized Polynucleotides

[0152] The polynucleotides encoding heterologous polypeptides can be codon-optimized. As used herein the term "codon-optimized coding region" means a nucleic acid coding region that has been adapted for expression in the cells of a given organism by replacing at least one, or more than one, or a significant number, of codons with one or more codons that are more frequently used in the genes of that organism.

[0153] In general, highly expressed genes in an organism are biased towards codons that are recognized by the most abundant tRNA species in that organism. One measure of this bias is the "codon adaptation index" or "CAI," which measures the extent to which the codons used to encode each amino acid in a particular gene are those which occur most frequently in a reference set of highly expressed genes from an organism.

[0154] The CAI of codon optimized sequences of the present invention corresponds to between about 0.8 and 1.0, between about 0.8 and 0.9, or about 1.0. A codon optimized sequence may be further modified for expression in a particular organism, depending on that organism's biological constraints. For example, large runs of "As" or "Ts" (e.g., runs greater than 3, 4, 5, 6, 7, 8, 9, or 10 consecutive bases) can be removed from the sequences if these are known to effect transcription negatively. Furthermore, specific restriction enzyme sites may be removed for molecular cloning purposes. Examples of such restriction enzyme sites include PacI, AscI, BamHI, BglII, EcoRI and XhoI. Additionally, the DNA sequence can be checked for direct repeats, inverted repeats and mirror repeats with lengths of ten bases or longer, which can be modified manually by replacing codons with "second best" codons, *i.e.*, codons that occur at the second highest frequency within the particular organism for which the sequence is being optimized.

[0155] Deviations in the nucleotide sequence that comprise the codons encoding the amino acids of any polypeptide chain allow for variations in the sequence coding for the gene. Since each codon consists of three nucleotides, and the nucleotides comprising DNA are restricted to four specific bases, there are 64 possible combinations of nucleotides, 61 of which encode amino acids (the remaining three codons encode signals

ending translation). The "genetic code" which shows which codons encode which amino acids is reproduced herein as Table 1. As a result, many amino acids are designated by more than one codon. For example, the amino acids alanine and proline are coded for by four triplets, serine and arginine by six, whereas tryptophan and methionine are coded by just one triplet. This degeneracy allows for DNA base composition to vary over a wide range without altering the amino acid sequence of the proteins encoded by the DNA.

TABLE 1: The Standard Genetic Code

	T	C	A	G
T	TTT Phe (F) TTC " TTA Leu (L) TTG "	TCG Ser (S) TCC " TCA " TCG "	TAT Tyr (Y) TAC " TAA Ter TAG Ter	TGT Cys (C) TGC TGA Ter TGG Trp (W)
C	CTT Leu (L) CTC " CTA " CTG "	CCT Pro (P) CCC " CCA " CCG "	CAT His (H) CAC " CAA Gln (Q) CAG "	CGT Arg (R) CGC " CGA " CGG "
A	ATT Ile (I) ATC " ATA " ATG Met (M)	ACT Thr (T) ACC " ACA " ACG "	AAT Asn (N) AAC " AAA Lys (K) AAG "	AGT Ser (S) AGC " AGA Arg (R) AGG "
G	GTT Val (V) GTC " GTA " GTG "	GCT Ala (A) GCC " GCA " GCG "	GAT Asp (D) GAC " GAA Glu (E) GAG "	GGT Gly (G) GGC " GGA " GGG "

[0156] Many organisms display a bias for use of particular codons to code for insertion of a particular amino acid in a growing peptide chain. Codon preference or codon bias, differences in codon usage between organisms, is afforded by degeneracy of the genetic code, and is well documented among many organisms. Codon bias often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, *inter alia*, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis.

Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization.

[0157] Given the large number of gene sequences available for a wide variety of animal, plant and microbial species, it is possible to calculate the relative frequencies of codon usage. Codon usage tables are readily available, for example, at <http://www.kazusa.or.jp/codon/> (visited July 30, 2010), and these tables can be adapted in a number of ways. See Nakamura, Y., *et al.* "Codon usage tabulated from the international DNA sequence databases: status for the year 2000," *Nucl. Acids Res.* 28:292 (2000). Codon usage tables for yeast, calculated from GenBank Release 128.0 [15 February 2002], are reproduced below as Table 2. This table uses mRNA nomenclature, and so instead of thymine (T) which is found in DNA, the tables use uracil (U) which is found in RNA. The table has been adapted so that frequencies are calculated for each amino acid, rather than for all 64 codons.

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

Amino Acid	Codon	Number	Frequency per hundred
Phe	UUU	170666	26.1
Phe	UUC	120510	18.4
Leu	UUA	170884	26.2
Leu	UUG	177573	27.2
Leu	CUU	80076	12.3
Leu	CUC	35545	5.4
Leu	CUA	87619	13.4
Leu	CUG	68494	10.5
Ile	AUU	196893	30.1
Ile	AUC	112176	17.2
Ile	AUA	116254	17.8
Met	AUG	136805	20.9
Val	GUU	144243	22.1
Val	GUC	76947	11.8
Val	GUA	76927	11.8
Val	GUG	70337	10.8
Ser	UCU	153557	23.5
Ser	UCC	92923	14.2

Amino Acid	Codon	Number	Frequency per hundred
Ser	UCA	122028	18.7
Ser	UCG	55951	8.6
Ser	AGU	92466	14.2
Ser	AGC	63726	9.8
Pro	CCU	88263	13.5
Pro	CCC	44309	6.8
Pro	CCA	119641	18.3
Pro	CCG	34597	5.3
Thr	ACU	132522	20.3
Thr	ACC	83207	12.7
Thr	ACA	116084	17.8
Thr	ACG	52045	8.0
Ala	GCU	138358	21.2
Ala	GCC	82357	12.6
Ala	GCA	105910	16.2
Ala	GCG	40358	6.2
Tyr	UAU	122728	18.8
Tyr	UAC	96596	14.8
His	CAU	89007	13.6
His	CAC	50785	7.8
Gln	CAA	178251	27.3
Gln	CAG	79121	12.1
Asn	AAU	233124	35.7
Asn	AAC	162199	24.8
Lys	AAA	273618	41.9
Lys	AAG	201361	30.8
Asp	GAU	245641	37.6
Asp	GAC	132048	20.2
Glu	GAA	297944	45.6
Glu	GAG	125717	19.2
Cys	UGU	52903	8.1
Cys	UGC	31095	4.8

Amino Acid	Codon	Number	Frequency per hundred
Trp	UGG	67789	10.4
Arg	CGU	41791	6.4
Arg	CGC	16993	2.6
Arg	CGA	19562	3.0
Arg	CGG	11351	1.7
Arg	AGA	139081	21.3
Arg	AGG	60289	9.2
Gly	GGU	156109	23.9
Gly	GGC	63903	9.8
Gly	GGA	71216	10.9
Gly	GGG	39359	6.0
Stop	UAA	6913	1.1
Stop	UAG	3312	0.5
Stop	UGA	4447	0.7

[0158] By utilizing this or similar tables, one of ordinary skill in the art can apply the frequencies to any given polypeptide sequence, and produce a nucleic acid fragment of a codon-optimized coding region which encodes the polypeptide, but which uses codons optimal for a given species. Codon-optimized coding regions can be designed by various different methods.

[0159] In one method, a codon usage table is used to find the single most frequent codon used for any given amino acid, and that codon is used each time that particular amino acid appears in the polypeptide sequence. For example, referring to Table 2 above, for leucine, the most frequent codon is UUG, which is used 27.2% of the time. Thus all the leucine residues in a given amino acid sequence would be assigned the codon UUG.

[0160] In another method, the actual frequencies of the codons are distributed randomly throughout the coding sequence. Thus, using this method for optimization, if a hypothetical polypeptide sequence had 100 leucine residues, referring to Table 2 for frequency of usage in the *S. cerevisiae*, about 5, or 5% of the leucine codons would be CUC, about 11, or 11% of the leucine codons would be CUG, about 12, or 12% of the leucine codons would be CUU, about 13, or 13% of the leucine codons would be CUA,

- 34 -

about 26, or 26% of the leucine codons would be UUA, and about 27, or 27% of the leucine codons would be UUG.

[0161] These frequencies would be distributed randomly throughout the leucine codons in the coding region encoding the hypothetical polypeptide. As will be understood by those of ordinary skill in the art, the distribution of codons in the sequence can vary significantly using this method; however, the sequence always encodes the same polypeptide.

[0162] When using the methods above, the term "about" is used precisely to account for fractional percentages of codon frequencies for a given amino acid. As used herein, "about" is defined as one amino acid more or one amino acid less than the value given. The whole number value of amino acids is rounded up if the fractional frequency of usage is 0.50 or greater, and is rounded down if the fractional frequency of use is 0.49 or less. Using again the example of the frequency of usage of leucine in human genes for a hypothetical polypeptide having 62 leucine residues, the fractional frequency of codon usage would be calculated by multiplying 62 by the frequencies for the various codons. Thus, 7.28 percent of 62 equals 4.51 UUA codons, or "about 5," *i.e.*, 4, 5, or 6 UUA codons, 12.66 percent of 62 equals 7.85 UUG codons or "about 8," *i.e.*, 7, 8, or 9 UUG codons, 12.87 percent of 62 equals 7.98 CUU codons, or "about 8," *i.e.*, 7, 8, or 9 CUU codons, 19.56 percent of 62 equals 12.13 CUC codons or "about 12," *i.e.*, 11, 12, or 13 CUC codons, 7.00 percent of 62 equals 4.34 CUA codons or "about 4," *i.e.*, 3, 4, or 5 CUA codons, and 40.62 percent of 62 equals 25.19 CUG codons, or "about 25," *i.e.*, 24, 25, or 26 CUG codons.

[0163] Randomly assigning codons at an optimized frequency to encode a given polypeptide sequence, can be done manually by calculating codon frequencies for each amino acid, and then assigning the codons to the polypeptide sequence randomly. Additionally, various algorithms and computer software programs are readily available to those of ordinary skill in the art. For example, the "EditSeq" function in the Lasergene Package, available from DNASTAR, Inc., Madison, WI, the backtranslation function in the VectorNTI Suite, available from InforMax, Inc., Bethesda, MD, and the "backtranslate" function in the GCG--Wisconsin Package, available from Accelrys, Inc., San Diego, CA. In addition, various resources are publicly available to codon-optimize coding region sequences, *e.g.*, the "backtranslation" function at

<http://www.entelechon.com/bioinformatics/backtranslation.php?lang=eng> (visited July 30, 2010) and the "backtranseq" function available at <http://emboss.bioinformatics.nl/cgi-bin/emboss/backtranseq> (visited July 30, 2010). Constructing a rudimentary algorithm to assign codons based on a given frequency can also easily be accomplished with basic mathematical functions by one of ordinary skill in the art.

[0164] A number of options are available for synthesizing codon optimized coding regions designed by any of the methods described above, using standard and routine molecular biological manipulations well known to those of ordinary skill in the art. In one approach, a series of complementary oligonucleotide pairs of 80-90 nucleotides each in length and spanning the length of the desired sequence is synthesized by standard methods. These oligonucleotide pairs are synthesized such that upon annealing, they form double stranded fragments of 80-90 base pairs, containing cohesive ends, *e.g.*, each oligonucleotide in the pair is synthesized to extend 3, 4, 5, 6, 7, 8, 9, 10, or more bases beyond the region that is complementary to the other oligonucleotide in the pair. The single-stranded ends of each pair of oligonucleotides is designed to anneal with the single-stranded end of another pair of oligonucleotides. The oligonucleotide pairs are allowed to anneal, and approximately five to six of these double-stranded fragments are then allowed to anneal together via the cohesive single stranded ends, and then they are ligated together and cloned into a standard bacterial cloning vector, for example, a TOPO[®] vector available from Invitrogen Corporation, Carlsbad, CA. The construct is then sequenced by standard methods. Several of these constructs consisting of 5 to 6 fragments of 80 to 90 base pair fragments ligated together, *i.e.*, fragments of about 500 base pairs, are prepared, such that the entire desired sequence is represented in a series of plasmid constructs. The inserts of these plasmids are then cut with appropriate restriction enzymes and ligated together to form the final construct. The final construct is then cloned into a standard bacterial cloning vector, and sequenced. Additional methods would be immediately apparent to the skilled artisan. In addition, gene synthesis is readily available commercially.

[0165] In additional embodiments, a full-length polypeptide sequence is codon-optimized for a given species resulting in a codon-optimized coding region encoding the entire polypeptide, and then nucleic acid fragments of the codon-optimized coding region, which encode fragments, variants, and derivatives of the polypeptide are made from the

original codon-optimized coding region. As would be well understood by those of ordinary skill in the art, if codons have been randomly assigned to the full-length coding region based on their frequency of use in a given species, nucleic acid fragments encoding fragments, variants, and derivatives would not necessarily be fully codon optimized for the given species. However, such sequences are still much closer to the codon usage of the desired species than the native codon usage. The disadvantage of this approach is that synthesizing codon-optimized nucleic acid fragments encoding each fragment, variant, and derivative of a given polypeptide, although routine, would be time consuming and would result in significant expense.

Transposons

[0166] To select for foreign DNA that has entered a host it is preferable that the DNA be stably maintained in the organism of interest. With regard to plasmids, there are two processes by which this can occur. One is through the use of replicative plasmids. These plasmids have origins of replication that are recognized by the host and allow the plasmids to replicate as stable, autonomous, extrachromosomal elements that are partitioned during cell division into daughter cells. The second process occurs through the integration of a plasmid onto the chromosome. This predominately happens by homologous recombination and results in the insertion of the entire plasmid, or parts of the plasmid, into the host chromosome. Thus, the plasmid and selectable marker(s) are replicated as an integral piece of the chromosome and segregated into daughter cells. Therefore, to ascertain if plasmid DNA is entering a cell during a transformation event through the use of selectable markers requires the use of a replicative plasmid or the ability to recombine the plasmid onto the chromosome. These qualifiers cannot always be met, especially when handling organisms that do not have a suite of genetic tools.

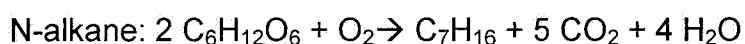
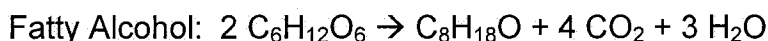
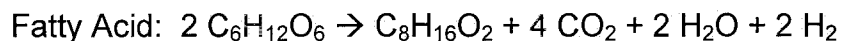
[0167] One way to avoid issues regarding plasmid-associated markers is through the use of transposons. A transposon is a mobile DNA element, defined by mosaic DNA sequences that are recognized by enzymatic machinery referred to as a transposase. The function of the transposase is to randomly insert the transposon DNA into host or target DNA. A selectable marker can be cloned onto a transposon by standard genetic engineering. The resulting DNA fragment can be coupled to the transposase machinery in an *in vitro* reaction and the complex can be introduced into target cells by electroporation. Stable insertion of the marker onto the chromosome requires only the

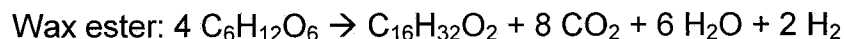
function of the transposase machinery and alleviates the need for homologous recombination or replicative plasmids.

- [0168] The random nature associated with the integration of transposons has the added advantage of acting as a form of mutagenesis. Libraries can be created that comprise amalgamations of transposon mutants. These libraries can be used in screens or selections to produce mutants with desired phenotypes. For instance, a transposon library of a CBP organism could be screened for the ability to produce more ethanol, or less lactic acid and/or more acetate.

Hydrocarbon Synthesis

- [0169] Hydrocarbons consist of carbon and hydrogen and include aliphatic hydrocarbons and aromatic hydrocarbons. Non-limiting examples of hydrocarbons include, alkanes, alkenes, alkynes, and hydrocarbon derivatives. The latter of which includes those compounds formed by the addition of at least one functional group to a hydrocarbon. Examples of hydrocarbon derivatives include, but are not limited to, aldehydes, alcohols, esters, fatty acids, unsaturated fatty acids, branched-chain fatty acids, branched methoxy fatty acids, multi-methyl branched acids, divinyl-ether fatty acids, w-phenylalkanoic acids, dicarboxylic acids.
- [0170] Hydrocarbons produced by the recombinant microorganisms and methods of the invention include carbon backbones of at least 4 carbons and up to 40 or more carbons. Such chain lengths are referred to as long-chain hydrocarbons. In certain aspects, the chain lengths include C₆-C₃₆; C₈-C₃₂; C₁₀-C₂₈; C₁₂-C₂₄; C₁₄-C₂₂; or C₁₆-C₂₀. In some embodiments, the chain length comprises a carbon backbone of C₁₂, C₁₄, C₁₆, C₁₈, C₂₀, and/or C₂₂. In further embodiments, the chain length comprises a carbon backbone of C₁₆.
- [0171] To produce hydrocarbons and hydrocarbon derivatives according to the invention, the following stoichiometric equations provide examples of an electron-balanced process.



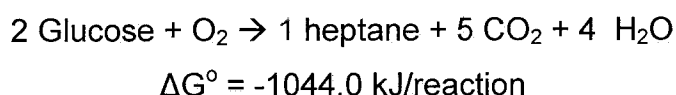


[0172] The synthesis of hydrocarbons becomes more thermodynamically favorable as the chain length increases (*see* Figure 6 (values derived from Stull *et al.*, The Chemical Thermodynamics of Organic Compounds, Wiley, New York, NY (1969))). For example, the stoichiometry for the production of a fatty alcohol proceeds according to the following equation, where *n* is the number of glucose molecules and *x* is the number of carbon atoms in the saturated fatty alcohol.

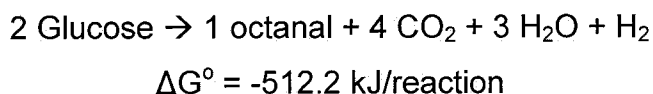


[0173] As can be seen, the number of H₂O molecules generated increases as chain length increases. This helps contribute to a more overall thermodynamically favorable reaction. Gibbs free energy changes per 2 glucose molecules (*n*=2) for specific alcohols are shown in Figure 6.

[0174] The Gibbs free energy change for the production of heptane, accounting for the requirement of elemental oxygen for the conversion of a fatty aldehyde to alkane by aldehyde decarbonylase (Li *et al.*, *JACS*, 133:6158-6161 (2011)) is:



[0175] The Gibbs free energy change for the production of octanal is:



[0176] Other sugars, including, but not limited to, xylose or arabinose, have a similar Gibbs free energy change as glucose. While some steps in the production of hydrocarbons or hydrocarbon derivatives can be slightly unfavorable, *e.g.*, aldolase or triosephosphate isomerase in glycolysis, the overall reaction will be thermodynamically

favorable when the final steps include chain termination steps, *e.g.*, acid, aldehyde, alcohol, and/or ester formation. The very low aqueous concentrations of the final hydrocarbons or hydrocarbon derivatives will further drive the thermodynamic equilibrium towards product formation.

Polyketide Synthesis

[0177] Polyketides are a structurally and functionally diverse family of natural products that possess a wide range of biological and pharmacological properties. Such properties include, but are not limited to, antibiotic, antitumor, antifungal, and immunosuppressive activities. Jenke-Kodama, H., *et al.*, *Mol. Biol. Evol.* 22(10):2027-39 (2005). Polyketides are synthesized as secondary metabolites in bacteria, fungi, plants, and animals by different classes of polyketide synthases (PKS), which resemble the classes of fatty acid synthases. *Id.* Polyketide synthesis proceeds by the addition or condensation of different functional groups to an acyl-ACP chain. *See* Figure 21. And while fatty acid elongation includes four enzymatic steps per two carbon chain extension (KS (ketosynthase), KR (ketoreductase), DH (dehydratase), ER (enoyl reductase)) (Figure 21B), PKS elongation can include a combination of enzymatic activities, *e.g.*, (KS), (KS, KR), (KS, KR, DH), or (KS, KR, DH, ER), at each step (Figure 21A). Malonyl-CoA produced by the recombinant microorganisms and pathways of the invention can be used as a metabolic precursor for polyketides.

Organic Acid Synthesis

[0178] Malonyl-CoA produced by the recombinant microorganisms and pathways of the invention can be used as a metabolic precursor for number of bioproducts. For example, the organic acid 3-hydroxypropionic acid ("3-HP"), also known as 3-hydroxypropanoate, is used in the production of various industrial chemicals such as renewable polyesters, acrylic acid, malonic acid, and co-polymers with lactic acid. Although 3-HP can be produced by organic chemical synthesis, it is desirable to use bio-alternative methods that allow for more cost effective, efficient, and renewable production. While some microorganisms are known to produce 3-HP (*see, e.g.*, WO 01/16346; WO 02/42418; US 2011/0144377; US 2011/0125118, each of which is incorporated by reference herein), few biological systems have been developed that would result in its efficient production. Production of malonyl-CoA at high yield via transcarboxylase in an anaerobic process

would allow for efficient high yield 3-hydroxypropionic acid production using a suitable enzymatic pathway from malonyl-CoA to 3-hydroxypropionic acid and a suitable redox system to generate NADPH during carbohydrate deconstruction. See, e.g., redox systems are “F” and “G” in Table 10 .

[0179] Enzymes employed for the production of 3-HP by the recombinant microorganisms and methods of the invention include 1) malonyl-CoA reductase (EC 1.2.7.5), 2) 3-hydroxypropionate dehydrogenase (EC 1.1.1.59 and EC 1.1.1.298), and 3) a bifunctional enzyme which harbors aldehyde dehydrogenase and alcohol dehydrogenase domains (Hügler *et al.*, *J. Bacteriol.* 184:2402-2410 (2002)).

[0180] The following example pathways demonstrate the production of 3-HP from a malonyl-CoA metabolic precursor using the above-referenced enzymes:

[0181] 1) Malonyl CoA Reductase (EC 1.2.1.75)

[0182] Malonate semialdehyde + coenzyme A + NADP(+) $\rightleftharpoons$ malonyl-CoA + NADPH

[0183] 2a) 3-Hydroxypropionate Dehydrogenase (EC 1.1.1.59)

3-hydroxypropanoate + NAD(+) $\rightleftharpoons$ Malonate semialdehyde + NADH

2b) 3-Hydroxypropionate Dehydrogenase (EC 1.1.1.298)

3-hydroxypropanoate + NADP(+) $\rightleftharpoons$ Malonate semialdehyde + NADPH

[0184] 3) bifunctional dehydrogenase (aldehyde-alcohol)

malonyl-CoA + NADPH + H⁺ $\rightarrow$ malonate semialdehyde + NADP⁺ + CoA

malonate semialdehyde (3-oxopropanoate) + NADPH + H⁺ $\rightarrow$ 3-hydroxypropionate + NADP⁺

[0185] The sequence of a malonyl-CoA reductase from *Chloroflexus aurantiacus* is provided below:

C. aurantiacus Malonyl-CoA Reductase (amino acid sequence;
>gi|42561982|gb|AAS20429.1) (SEQ ID NO:1)

MSGTGRLAGKIALITGGAGNIGSELTRRFLAEGATVIISGRNRAKLTALAERMQA
EAGVPAKRIDLEVMDGSDPVAVRAGIEAIVARHGQIDILVNNAGSAGAQRRLAEI
PLTEAELGPGAEETLHASIANLLGMGWHLMRIAPHMPVGSVINVESTIFSRAEY

YGRIPYVTPKAALNALSQLAARELGARGIRVNTIFPGPIESDRIRTVFQRM DQLKG
RPEGDTAHHFLNTMRLCRANDQGALERRFPSVGDVADAAVFLASAESAALSGE
TIEVTHGMELPACSETSLARTDLRTIDASGRITTLICAGDQIEEVMALTGMLRTCG
SEVIIGFRSAAALAQFEQAVNESRRLAGADFTPPIALPLDPRDPATIDAVFDWGAG
ENTGGIHA AAVILPAT SHEPAPCVIEVDDERVLNFLADEITGTIVIASRLARYWQSQ
RLTPGARARGPRVIFLSNGADQNGNVYGRISAAIGQLIRVWRHEAELDYQRAS
AAGDHVLPVWVANQIVRFANRSLEGLEFACAWTAQLLHSQRHINEITLNIPANIS
ATTGARSASVGWAESLIGLHLGKVALITGGSAGIGGGQIGRLLALS GARVMLAAR
DRHKLEQM QAMIQSELA EVG YTDVEDRVHIAPGCDVSSEAQLADLVERTLSAFG
TVDYLINNAGIAGVEEMVIDMPVEGWRHTLFANLISNYS LMRKLAPLMKKQGS
YILNVSSYFGGEKDAAIPYPNRADYAVSKAGQRAMAEVFARFLGPEIQINAIAPG
PVEGDRLRGTGERPGLFARRARLILENKRLNELHAALIAAARTDERSMHVELL
LPNDVA ALEQNPAAPTALRELARRFRSEGDPAASSSSALLNRSIAAKLLARLHNG
GYVLPADIFANLPNPPDPFFTRAQIDREARKVRD GIMGMLYLQRMPTFEFDVAMA
TVYYLADRNVSGETFHPSGGLRYERTPTGGELFGLPSPERLAELVGSTVYLIGEH
LTEHLNLLARAYLERYGARQVVMIVETETGAETMRLLHDHVEAGRLMTIVAG
DQIEAAIDQAITRYGRPGPVVCTPFRPLPTVPLVGRKDS DWSTVLSEAEFAELCEH
QLTHHFRVARKIALSDGASLALVTPETTATSTTEQFALANFIKTTLHAFTATIGVE
SERTAQRILINQVDLTRRARAEPRDPHERQQELERFIEAVLLVTAPLPPEADTRY
AGRIHRGRAITV

- [0186] Additional malonyl-CoA reductase enzyme examples include, but are not limited to, those from *Chloroflexus* sp., *Oscillochloris* sp., *Roseiflexus* sp., and marine gamma proteobacterium. See, e.g., Hügler et al., *J. Bacteriol.* 184:2402-2410 (2002); Rathnasingh, C., et al., *Biotech. Bioeng.* 104(4) (2009); Rathnasingh, C., et al., "Production of 3-hydroxypropionic acid via malonyl-CoA pathway using recombinant *Escherichia coli* strains," *J. Biotech.* (Epub June 23, 2011). A phylogenetic tree and an alignment of several malonyl-CoA reductase enzymes is shown in Figures 45 and 46. These alignments were made using malonyl-CoA reductase enzymes from *Chloroflexus aurantiacus* (GenBank Accession No. AAS20429; SEQ ID NO:1); *Chloroflexus aurantiacus* J-10-fl (GenBank Accession No. YP_001636209; SEQ ID NO:291); *Chloroflexus* sp. Y-400-fl (GenBank Accession No. YP_002570540; SEQ ID NO:292); *Chloroflexus aggregans* DSM 9485 (GenBank Accession No. YP_002462600; SEQ ID NO:293); *Oscillochloris trichoides* DG6 (GenBank Accession No. ZP_07684596; SEQ ID NO:294); *Roseiflexus castenholzii* DSM 13941 (GenBank Accession No. YP_001433009; SEQ ID NO:295); *Roseiflexus* sp. RS-1 (GenBank Accession No. YP_001277512; SEQ ID NO:296); *Erythrobacter* sp. NAP1 (GenBank Accession No. ZP_01039179; SEQ ID NO:297); gamma proteobacterium NOR51-B (GenBank Accession No. ZP_04957196 SEQ ID NO:298).

- [0187] Another product that can be produced from a malonyl-CoA metabolic precursor, and/or as an end-product of the fatty acid syntheses described herein, is adipic acid. Adipic acid is a six-carbon dicarboxylic acid, which is used as a chemical intermediate in the synthesis of polymers, such as polyamides (nylons), polyurethanes, and plasticizers, as well as a food acidulant. Chemical synthesis of adipic acid uses various noxious chemicals for oxidation and/or hydration of ketoalcohols or cyclohexanes, which present environmental safety and energy input concerns. Engineering a biological system to produce adipic acid from a carbohydrate source can avoid these concerns and provide a renewable means for producing adipic acid-derived products.
- [0188] Attempts at the bioproduction of adipic acid have used alternative synthetic pathways, catalysts, substrates, intermediates, and/or recombinant microorganisms. *See, e.g.*, WO2011/003034, WO1995/007996, WO2009/151728, and WO2010/144862, each of which is incorporated by reference herein. In particular, WO2011/003034 discloses the synthesis of adipic acid from, *inter alia*, fatty acids, fatty alcohols, alkanes, and oils, but does not, however, disclose the synthesis of adipic acid from a malonyl-CoA metabolic precursor. The pathways of the invention for producing malonyl-CoA can be used to produce a C₁₂ fatty acid or fatty alcohol, which can be further engineered to produce adipic acid via omega oxidation using. *See, e.g.*, Figures 23 and 24 (“At” is *Arabidopsis thaliana*; “Cc” is *Candida cloacae*); WO2011/003034; Vanhanen S., *et al.*, *J. Biol. Chem.* 275(6):4445-52 (2000); Picataggio, S., *et al.*, *Bio/Technology* 10(8):894-98 (1992). To accommodate the oxidation of the fatty acid or fatty alcohol, either a facultative anaerobe (*e.g.*, *E. coli* or *S. cerevisiae*) can be engineered to include an adipic acid pathway that can be switched to aerobic conditions after a pool of malonyl-CoA or fatty acids/alcohols is synthesized, or a facultative anaerobe or aerobe comprising an adipic acid pathway can be engineered to use in tandem or in series with a recombinant microorganism of the invention that produces fatty acids or fatty alcohols.
- [0189] To generate adipic acid from a fatty acid or fatty alcohol using omega oxidation pathway, enzymes such as, *e.g.*, a mixed function oxidase to hydroxylate the omega carbon and alcohol and aldehyde dehydrogenases to oxidate the introduced hydroxyl group, can be used.

Phosphoenolpyruvate carboxykinase

[0190] Phosphoenolpyruvate carboxykinase (PEPCK) includes those enzymes that catalyze the conversion of phosphoenolpyruvate (PEP) to oxaloacetate (*see* Figure 1A) and that correspond to Enzyme Commission Number 4.1.1.49 or 4.1.1.32. *See, e.g.,* Matte, A., *et al., J. Biol. Chem.* 272:8105-08 (1997). The reaction is reversible and is used in succinic acid producing bacteria to convert PEP to oxaloacetate. *E. coli* can mutate to use PEPCK when flux is directed primarily to succinic acid. PEPCK requires Mg^{2+} as a co-factor. A number of ATP and GTP using enzymes have been described, including, *e.g.,* a GTP-utilizing PEPCK in *C. thermocellum* and ATP-utilizing PEPCK in *T. saccharolyticum*, *T. tengcongensis*, *E. coli* and *S. cerevisiae*.

[0191] PEPCKs have been classified according to nucleotide specificity, *i.e.,* those that are ATP-dependent and those that are GTP- or ITP-dependent. Within each group, the species show significant amino acid sequence identity, in the range of 40-80%, and share similar nucleotide and oxaloacetate binding "consensus motifs" between the groups, including key conserved residues at or near the active sites. *See* Matte, A., *et al., J. Biol. Chem.* 272:8105-08 (1997). Additional structural characterizations have been described in, *e.g.,* Matte, A., *et al., J. Biol. Chem.* 272:8105-08 (1997). Examples of PEPCK sequences include:

[0192] *C. thermocellum* PEPCK (GTP)

[0193] >Cthe_2874 (SEQ ID NO:2)

```
atgacatcaacaacatgacaaaaacaaaaactgctggattgggttaaggaaatggctgaaatgtgtcagcctgatgaaattt
attggtgcgatggttcggaggaagaaaaatgagcgccttgataaagttgatggtggattcagggttggtacgcctttgaatctgaa
aagcgacctggatgttatctcttcgcagcgatccgtccgacgttgcccgtgttgaggacagaactttattgcacccaaaccaa
agaagatgcaggacctacaacaactggatagatccggttgagctcaaggcaactatgaaagagttgtacaagggtgtatgaa
gggaagaacaatgtatgttattccttttccatgggacctatcggttcacccatttcaaaaatcggcgttgaaattgaccgacagccct
tatgttgtgttaacatgcgcattatgactcgcataaggcaaggctgtgttgatcagctcggagaagacggagattttgtacctgtc
tccactcagtcggtgctccgctcaaagagggagaaaaggataaaggttgccatgcgcaccaatcgaagaaatacataagc
cacttcccgaagaaaggactatatggtcatatggttcggatacgggtgaaatgcgcttttaggaaagaaatgctttgcattcgt
attgcattctgttatggcacgtgacgaaggttggttgcgtgaacacatgcttatccttcgcataacagaccctgaaggaaacaagac
atatgttacaggtgctttccaagcgcatgcggaagacgaacctggctatgcttattcctacaattcccggatggaaggtgaaa
caatcggtgacgatattgatgagatttgaaagacggccgtttgtatgtatcaaccctgaagcaggattctttgtgttg
ctccgggtacatccatggattcaaatccgaacgcaatgcatacaattaagaaaaatactatatttacaacggttcattgactgatg
acggcgatgtttggtgggaaggcatcggaactgaaccgccggctcatctcatagactggcagggtaagactggactcctgatt
ccggaactttggcagcacatcccaacggacgtttacagcacctgcaagtcagtgcctgtaatgtcctgaattgggaggatcc
ggaaggtgtgccgatttcagcaatccttatcggtggacgccgtccgaacaccattccgcttgcattgaaagctttgactggaacc
```

- 44 -

atgggtgattcatgggttcaatcatgggttctgaaattacggctgccgcaatttcaaacaaaatcggacaggtacgccgtgacccg
 ttgctatgctgcccttcataggctacaacgtaaatgactatttcagcactgggtgaacatgggtaccaagactgaccaagcaag
 ctcccaagatattctatgtaaaactgggtccgcaaggacagcaacggtaaatgggttggtgacctggatacggtgaaaacagccgtgt
 tctcaagtggattgttgaagagtcacggaaaaggtaagcagtaagacacctataggatatatgcttacagttgacgctatc
 gacacaaccggccttgatgtaagcaagaggatatggaagaactcttgagcgttaacaaagaacagtggtccaggaagttga
 gtcaataaaagaacattataagtcatacgggagaaaaactgccgaaagaattgtgggcacaattggaggctcttgaacaacgttg
 aaagagtataacgggtaa

[0194] *T. saccharolyticum* PEPCK

[0195] >or2173 (SEQ ID NO:3)

ATGATTATGAAAAAATCAAAGAAATGTTTCAATCTGAATATTGACGACAAAG
 AAACCTTGAATACTTTTGGAAAGTTCGAGAGGAGAATTGTTTATGATAGATTTA
 GATGATGTATTTAAAAATTCTGGCAGTATTCTTTACAATTTACCTGTTTCAGA
 TTTGATAGAGGAAGCCATAAGAAATAATGAAGGGAAATTGTTAGAAAATGGT
 GCATTAGATGTTTTTACAGGTAAATATACGGGAAGAATACCAAAAGATAAAT
 ACATTGTAAATGAAGAATCTATTTCATAATGATATTTGGTGGGAAAATAATAA
 TTCAATGGAAAAAGAAAATTTTATTAGAGTTTTAAACAGAGTAATTGATTATT
 TAAAAAAGAGCAGAAAATTGTATGTTTTTAAAGGTTTTGTTGGCGCAGACCC
 GCGATATAGATATCAAGTAACCGTTATTAATGAATATGCCTATCAAAACGCTT
 TTGTACATCAATTATTTATTAATCCTAAAAATGAAGAAGAACTTAAAAAGGA
 ATCCGATTTTACAGTTATTTGTGTGCCGAATTTTTTAGCTGATCCAATTTATGA
 TGGAACATAATTCTGAGGCATTTATTATTATAAGTTTTGAAGAAAAATTAATTT
 TAATTGGTGGAAACAAGATATTCAGGAGAAATAAAAAAATCTGTCTTCACAAT
 GATGAATTATTTGATGTTAAAAAGGAATGTACTGCCTATGCATTGTGCAGCTA
 ATATAGGTTCCAATAATGATACAGCGCTTTTTTTGGGTTGTCCGGAACCGGC
 AAGACAACCTTTATCAACGGATCCAGAAAGATTTTTAATTGGCGACGATGAAC
 ATGGATGGTCTTCACATGGAATTTTTAATTTTGAGGGTGGATGCTATGCAAAG
 TGTATAAATTTATCCCATATAATGAACCTGAAATATGGAATGCAATTAGATT
 TGGAACAATTTTAGAAAATGTTATTTATGATGTAAATAATATGCCAGTCTATA
 CAAGTAGTAAAATAACTGAAAATACAAGAGCTTCATATCCACTTGAGTACAT
 CCCTAGGAAAGCGTCAAATGGCATTGGCGGTAATCCTAAAATTATTTTTTCT
 TGGCAGCCGATGCTTTTGGAGTATTGCCTCCAATTTCTAAGCTGACAAATGAA
 CAGGCTGTTGACTATTTCTTATTAGGATATACGAGCAAAATACCAGGAACAG
 AAAAGGGAATTTGCCGAACCACAAGCAACGTTTTTCATCATGTTTTTGGAGCACC
 ATTTTTGCCATCATATCCAATGAGGTATGCTGAATTGTTAAAGAAAAAATCG
 CAGAAAATGATTCAGTTGTTTATTTAATAAATACTGGATGGATAGGTGGACA
 TTATGGAATTGGCAAAGGATAGATTTAAATAACACAAGAGAAATCATAAAA
 AATGTTTTAAATGGTGAATTGGAAAAAGCAAAATTTAAAAAAGATACAGTAT
 TTGATTTGATGATACAGAAAAGTGCAATAACATTCCAGATGAATTATTAGA
 TCCTATAAAAAACATGGGAAGACAAAAATGATTACTTCCAACTGCTAATAAT
 TTATTATCTGCATTTAAAGCGAGATTAGATTATATAAAAAATGGGATTCATCA
 ATAA

[0196] *E. coli* K12 PEPCK (ATP) (SEQ ID NO:4)

- 45 -

ATGCGCGTTAACAATGGTTTGACCCCGCAAGAACTCGAGGCTTATGGTATCA
 GTGACGTACATGATATCGTTTACAACCCAAGCTACGACCTGCTGTATCAGGA
 AGAGCTCGATCCGAGCCTGACAGGTTATGAGCGCGGGGTGTAACTAATCTG
 GGTGCCGTTGCCGTCGATAACCGGGATCTTCACCGGTCGTTACCAAAAAGATA
 AGTATATCGTCCGTGACGATAACCACTCGCGATACTTTCTGGTGGGCAGACAA
 AGGCAAAGGTAAGAACGACAACAAACCTCTCTCTCCGAAACCTGGCAGCAT
 CTGAAAGGCCTGGTGACCAGGCAGCTTTCCGGCAAACGTCTGTTTCGTTGTCG
 ACGCTTTCTGTGGTGGCAACCCGGGATACTCGTCTTTCCGTCCGTTTCATCACC
 GAAGTGGCCTGGCAGGCGCATTTTGTCAAAAACATGTTTATTCGCCCCGAGCG
 ATGAAGAACTGGCAGGTTTCAAACCAGACTTTATCGTTATGAACGGCGCGAA
 GTGCACTAACCCGCAGTGGAAGAACAGGGTCTCAACTCCGAAAACCTTCGTG
 GCGTTTAAACCTGACCGAGCGCATGCAGCTGATTGGCGGCACCTGGTACGGCG
 GCGAAATGAAGAAAGGGATGTTCTCGATGATGAACTACCTGCTGCCGCTGAA
 AGGTATCGCTTCTATGCACTGCTCCGCCAACGTTGGTGAGAAAGGCGATGTT
 GCGGTGTTCTTCGGCCTTTCCGGCACCGGTAAAACCACCCTTTCCACCGACCC
 GAAACGTCGCCTGATTGGCGATGACGAACACGGCTGGGACGATGACGGCGTG
 TTAACTTCGAAGGCGGCTGCTACGCAAAAACCTATCAAGCTGTGAAAGAAG
 CGGAACCTGAAATCTACAACGCTATCCGTCGTGATGCGTTGCTGGAAAACGT
 CACCGTGCGTGAAGATGGCACTATCGACTTTGATGATGGTTCAAAAACCGAG
 AACACCCGCGTTTCTTATCCGATCTATCACATCGATAACATTGTTAAGCCGGT
 TTCCAAAGCGGGCCACGCGACTAAGGTTATCTTCCTGACTGCTGATGCTTTTCG
 GCGTGTTGCCGCCGGTTTCTCGCCTGACTGCCGATCAAACCCAGTATCACTTC
 CTCTCTGGCTTCACCGCCAAACTGGCCGGTACTGAGCGTGGCATCACCGAAC
 CGACGCCAACCTTCTCCGCTTGCTTCGGCGCGGCATTCTGTGCTGCACCCG
 ACTCAGTACGCAGAAGTGCTGGTGAAACGTATGCAGGCGGCGGGCGCGCAG
 GCTTATCTGGTTAACACTGGCTGGAACGGCACTGGCAAACGTATCTCGATTA
 AAGATACCCGCGCCATTATCGACGCCATCCTCAACGGTTCGCTGGATAATGC
 AGAAACCTTCACTCTGCCGATGTTTAACTGGCGATCCCAACCGAACTGCCG
 GGCGTAGACACGAAGATTCTCGATCCGCGTAACACCTACGCTTCTCCGGAAC
 AGTGGCAGGAAAAAGCCGAAACCCTGGCGAAACTGTTTATCGACAACCTTCGA
 TAAATACACCGACACCCCTGCGGGTGCCGCGCTGGTAGCGGCTGGTCCGAAA
 CTGTAA

[0197] *S. cerevisiae* PEPCK (ATP) (SEQ ID NO:5)

ATGTCCCCTTCTAAAATGAATGCTACAGTAGGATCTACTTCCGAAGTTGAACA
 AAAAATCAGACAAGAATTGGCTCTTAGTGACGAAGTCACCACCATCAGACGC
 AATGCTCCAGCTGCCGTTTTGTATGAAGATGGTCTAAAAGAAAATAAACTG
 TCATTTTCATCAAGCGGTGCATTGATCGCTTATTCCGGTGTTAAAACCGGAAGA
 TCTCCAAAGGACAAACGTATTGTTGAAGAACCTACCTCGAAAGACGAAATTT
 GGTGGGGTCCGGTCAATAAACCATGTTCTGAAAGAACATGGTCTATCAACCG
 TGAAAGAGCTGCAGATTACTTGAGAACAAAGAGACCACATTTATATTGTCGAT
 GCATTTGCAGGATGGGATCCAAAATACAGAATCAAAGTCCGCGTTGTTTGTG
 CCAGGGCTTACCACGCTTTATTCATGACAAATATGCTTATTAGACCTACAGAA
 GAAGAATTAGCCCATTTTGGAGAACCTGATTTTACTGTCTGGAACGCTGGTCA
 GTTCCCAGCCAATTTACACACCCAGGATATGTCTTCAAAGAGTACTATAGAA
 ATTAACCTCAAAGCAATGGAAATGATCATTTTAGGTACCGAATACGCCGGTG
 AAATGAAAAAAGGTATTTTCACAGTTATGTTTTACTTGATGCCTGTGCACCAT
 AACGTTTTAACTTTGCACTCTTCCGCCAACCGGGTATTCAAACCGGTGACGT

- 46 -

TACTTTATTCTTTGGCCTAAGTGGTACCGGGAAAACCACTTTATCCGCAGACC
CACATAGATTGTTGATCGGCGATGATGAACATTGTTGGTCCGACCATGGTGTGTC
TTCAATATCGAAGGTGGTTGTTACGCCAAGTGTATTAATTTATCTGCCGAAAA
GGAGCCTGAAATTTTCGACGCTATCAAGTTTGGTTCTGTATTAGAAAACGTTA
TCTATGACGAGAAGTCGCATGTAGTCGACTATGACGACTCTTCTATTACTGAA
AATACTAGATGTGCCTACCCAATTGACTACATTCCAAGTGCCAAGATTCCATG
TTTGGCGGACTCTCATCCAAAGAACATTATCCTGCTAACTTGTGATGCTTCGG
GTGTTTTACCACCAGTATCTAAATTGACTCCTGAACAAGTCATGTACCATTTTC
ATCTCTGGTTACACTTCTAAAATGGCTGGTACTGAGCAAGGTGTCACTGAACC
TGAACCAACATTTTCATCTTGTTCGGACAACCCTTCCTAGCCTTGCACCCTAT
TAGATACGCAACCATGTAGCTACAAAGATGTCTCAACATAAAGCTAATGCG
TACTTAATCAACACCGGCTGGACTGGTTCTTCCTACGTATCTGGTGGTAAACG
TTGCCCATTTGAAGTACACAAGGGCCATTCTGGATTCTATTTCATGATGGTTCGT
TAGCCAATGAAACGTACGAACTTTACCGATTTTCAATCTTCAAGTACCTACC
AAGGTTAACGGTGTTCAGCTGAGCTTTTGAATCCTGCTAAAACTGGTCTCA
AGGTGAATCCAAATACAGAGGTGCAGTTACCAACTTGGCCAACTTGTTTGTTC
AAAATTTCAAGATTTATCAAGACAGAGCCACACCAGATGTATTAGCCGCTGG
TCCTCAATTCGAGTAA

Transcarboxylase

- [0198] The conversion of oxaloacetate and acetyl-CoA to pyruvate and malonyl-CoA allows for the anaerobic high yield production of fatty acid derived hydrocarbons. This reaction has not been reported to occur *in vivo*. However, an *in vitro* substrate specificity study for fraction-purified (S)-methylmalonyl-CoA:pyruvate carboxytransferase (a transcarboxylase, “Me-TC,” E.C. 2.1.3.1) showed the ability of this enzyme to utilize oxaloacetate and acetyl-CoA as substrates. See Wood and Stjernholm, *PNAS* 47:289-303 (1961). The *in vitro* reaction occurred at one half the velocity of the enzyme’s natural substrates, oxaloacetate and propionyl-CoA, however, and the ability of the enzyme to produce malonyl-CoA in its native organism (*Propionibacterium shermanii*) was not determined. Me-TC enzymes are known to be present in other *Propionibacteria* (e.g., *Propionibacterium freudenreichii* and *Propionibacterium acnes*), which ferment carbohydrates and lactate to propionate and acetate, and in obligately syntrophic bacteria such as *Pelotomaculum thermopropionicum*, *Candidatus Cloacamonas acidaminovorans*, and *Geobacter bemidjiensis*, which convert propionate and other medium chain organic acids and alcohols to acetate and hydrogen or reduced metals. Falentin *et al.*, *PLOS one* 5(7): e11748 (2010); Kosaka *et al.*, *Genome Res.* 18:442-448 (2008); Pelletier *et al.*, *J. Bact.* 190:2572-2579 (2008); Aklujkar *et al.*, *BMC Genomics* 11:490 (2010).

- [0199] As used herein, transcarboxylase (TC) includes enzymes that catalyze the conversion of oxaloacetate and acetyl-CoA to malonyl-CoA and pyruvate (see Figure 1B)

- 47 -

and that correspond to Enzyme Commission Number 2.1.3.1 (methylmalonyl-CoA carboxyltransferase). *In vivo*, TC also catalyzes the conversion of methylmalonyl-CoA and pyruvate to oxaloacetate and propionyl-CoA. The reaction is reversible and requires co-factors such as Biotin, Co, or Zn^{2+} . TC consists of 3-4 subunits encoding domains for: a 5S subunit, a 12S subunit, and a 1.3S subunit; a 12S C-terminal subunit may also be present. See Carey *et al.*, *IUBMB Life* 56:575-83 (2004). TC enzymatic activity has been observed in *Propionibacterium* species such as *Propionibacterium freudenreichii* and *Propionibacterium acnes*, *Bacteroides fragilis*, *Veillonella parvula*, *Veillonella gazogenes*, *Pelotomaculum thermopropionicum*, *Candidatus Cloacamonas acidaminovorans*, and *Geobacter bemidjiensis*. See Falentin *et al.*, *PLoS One* 5:e11748 (2010); Kosaka *et al.*, *Genome Res.* 18:442-448 2008 (2008); Pelletier *et al.*, *J. Bact.* 190:2572-2579 (2008); and Akujkar *et al.*, *BMC Genomics* 11:490 (2010). Based on similarity to TC enzymes, high similarity TC genes have been identified in *Thermoanaerobacter* strains (*T. saccharolyticum* or0945, or0947, and or1888), *C. thermocellum* (Cthe_0699, Cthe_0700, and Cthe_0701), *Caldicellulosiruptor bescii*, *Clostridium cellulolyticum*, and *Corynebacterium kroppenstedtii*. Protein engineering, either across all subunits or on a specific subunit, using techniques known to those in the art, can be employed to increase enzymatic activity towards malonyl-CoA generation.

[0200] An alignment of *C. thermocellum* and *T. saccharolyticum* homologs of TC from *Propionibacterium freudenreichii* CIRM-BIA1 and *Propionibacterium acnes* is shown in Figures 7A-7C. Additional sequences of TC include:

[0201] *Propionibacterium freudenreichii* subsp. *shermanii* CIRM-BIA1
Transcarboxylase

>PFREUD_18840 (1.3S subunit nucleotide sequence)

Antisense strand: (SEQ ID NO:6)

```
tcagccgac ttgatgagac cctgaccgcc ctgcacggcg tcacgtctct tgacaaggac cttctcgacc ttgccgtcgg
tgggagcgtt gatctcggtc tccatcttca tggcctcgag aacgagcacg gtctgaccag ccttgaccgt gtcaccctec
ttcacgagga tcttgagac ggtgccggcc agcggagcgg gaatctcgcc ctctccggcc ttaccggcgc
ctgcgccacc tgetgcgcgc ggtgccggcg cgccgccggt gccgccgccg aacaggatgg tgcccatcgg
gtttcgtgt gactgtcga cgtaacgtc aacgtcatc gcagtgcctg tgactgttac ctcagttc at
```

Sense strand: (SEQ ID NO:7)

- 48 -

atgaaactgaaggtaacagtcaacggcactgcgtatgacgttgacgttgacgtcgacaagtcacacgaaaacccgatgggcac
 catcctgttcggcgccggcaccggcgccggcgccggcaccgcgcgcagcaggtggcgccagggcgccgtaaggccggaga
 gggcgagattcccgtccgctggccggcaccgtctccaagatcctcgtgaaggagggtgacacgggtcaaggctggtcagacc
 gtgctcgttctcaggccatgaagatggagaccgagatcaacgctcccaccgacggcaaggctcgagaaggctcctgtcaagga
 gcgtgacgccgtgcaggcggtcagggtctcatcaagatcggc

>PFREUD_18840 (1.3S subunit amino acid sequence) (SEQ ID NO:8)

MKLKVTVNGTAYDVDVDVDKSHENPMGTILFGGGTGGAPAPRAAGGAGAGKA
 GEGEIPAPLAGTVSKILVKEGDTVKAGQTVLVLEAMKMETEINAPT DGKVEKVL
 VKERDAVQGGQGLIKIG

>PFREUD_18870 (5S subunit nucleotide sequence)

Antisense strand: (SEQ ID NO:9)

tcacgcctgc tgaacgggtga cttcgcggac ggttcgccc acgttcacgt ttaggtgac gggaccggcc
 acggcgagcg acttctcgtc gccctcggcc tcggccttca gctgggcac ggtgagagcc acgtgtgacg
 ggccctcggc gcatgctcg aagaagaccg gagcgacctg cgggaacagt gcataggtga gcacgtcctc
 gtcgggtccg ttgaagccct tgaggggcgc ggccctcttg gactgtcct cccactcggg gggcagcaga
 tcggccgggc gctgggtgat cggttcttg ccgactgct cctcggccaa ctgaccacc ttcggatcgc gatcgccgg
 gctggcgccg tagtagccga gcatgatgc ggcgaaactc cgggtcatcc tctgtactc gccatcacc acgttgaaca
 cggcctgcgt gccacgac tggtcggac gggtgaccag gggcggggaag ccggcgccct tgcggacgcg
 cggcacctct gccatgacct cgtccatct gtctcggcg cctcggcgcg gcagctgcga ctccatgttg gagagcatgc
 cgccggggat ctgcgacttg aagatcgagg tgcgacaag cgtcttcgac tcgaactct tgtactcgg gcggatggcc
 ttgaagtgat cgcgcatct gtgcaggcga tcgtagtcaa ggttggtggt gtaccgggtg cctcagagca tctcggcaac
 cgactcggtg ggggttggtg ccgggcccag cgacatggac gagatggcgg tgcgacgac gtcgacgccg
 gccctgatgg ccttcgatg ggagacctc gtgacaccg tgggtggagt gcagtcagg ttgatctgc tctctggcc
 gtaggtgtcc ttgatggcct tgatgatgc gtaggcccgc tgcggcttga gcaggcgccg catgtcttc agggcgatgg
 aatcagcacc catgtcgagc agctgaccag caagcttgac atagccctca acggtgtgga ccgggctgat
 cgtgtagcaa atgggtccct gcgctgctt gccggccttc ttacggcag ccatggcgtg cgccatgttg cggggatcat
 tcatggcgtc gaagacacgg aacacgtcca tgccgttctc agcggacttg tcgacgaagc gatcgacgac ctgctcgttg
 tagtgccgtt aaccagcag gttctggcca cgcagcagca tctggagacg gctgttgggc atcagcttgc
 ggaacgtgcg cagacgtcc caagatcct cgttgaggaa gcggatacac gactcatagc tggcaccacc
 ccaacactcc actgaccagt acccggcagc atcaatgtc gcacaggcgc cgaccatgc ttccattgcc attcgtgtgg
 ccatcaggct ctgatgggca tcgcgcagca cgagctcggg gataccaacc tcgcgggct cggaaacctc aattctcgc
 ggactcat

Sense strand: (SEQ ID NO:10)

atgagtcgcgagaaattgaggttccgagccgcgcgaggttggtatcaccgagctcgtgctgcgcgatgccatcagagcct
 gatggccacacgaatggcaatggaagacatggcggcgctgtgcagacattgatgctgccgggtactggcagtgagtggt
 ggggtggtgccagtatgactcgtgtatccgttctcaacgaggatccttgggagcgtctgcgcagcttccgaagctgatgcc
 caacagccgtctccagatgctgctgcgtggccagaacctgctgggttaccgccactacaacgacgaggtcgtcgatcgttctg
 cgacaagtcgctgagaacggcatggacgtgttcgtgtcttcgacgccatgaatgatccccgaacatggcgacgccatgg
 ctgccgtaagaaggccggcaagcacgcgcaggccaccattgtacacgatcagcccggtccacaccgttgagggtatgt
 caagcttgctggtcagctgctcgacatgggtgtgattccatcgccctgaaggacatggccgacctgtcaagccgcagccg
 cctacgacatcatcaaggccatcaaggacacctacggccagaagacgcagatcaacctgcactgccactccaccacgggtgtc
 accgaggtctccctcatgaaggccatcaggccggcgctcgacgtcgtcgacaccgccatctcgtccatgtcgtcggccggg

- 49 -

ccacaacccccaccgagtcggttgccgagatgctcgagggcaccgggtacaccaccaaccttgactacgatcgctgcacaag
 atccgcgatcacttaaggccatccgcccgaagtacaagaagttcgagtcgaagacgcttgacacctcgatctcaagtcgc
 agatccccggcgcatgctctccaacatggagtcgagctgcgcgcccagggcgccgaggacaagatggacgaggtcatgg
 cagaggtgccgcgctccgaaggccgcccgttcccggccctggtcaccccgctccagccagatcgtcggcagcgagggcg
 tgttaacgtgatgatggcgagtagacaagaggtatgaccggcgagttcgccgacatcatgctcggctactacggcgccagccccg
 gccgatcgcgatccgaaggtggtcaagttggccgaggagcagtcggcgaagaagccgatcaccagcgccccggcgatctg
 ctgccccccgagtgaggagcagtcgaaggaggccgcccgtcaagggttaacggcaccgacgaggacgtgctcac
 ctatgcactgttcccgcaggtcgtctcgtctcttcgagcatcgccgagggcccgacagcgtggctctcaccgatgccc
 gctgaaggccgagggcgagggcgacgagaagtcgtcgcggtggccggtcccgctacctaacaagtgacgtggcgga
 ccgctccgcaagtcaccgttcagcagggcgtga

>PFREUD_18870 (5S subunit amino acid sequence) (SEQ ID NO:11)

MSPREIEVSEPREVGITELVLRDAHQSLMATRMAMEDMVGACADIDAAGYWSV
 ECWGGATYDSCIRFLNEDPWERLRTFRKLMPNSRLQMLLRGQNLLGYRHYNDE
 VVDRFVDKSAENGMDVFRVFDAMNDPRNMAHAMA AVKKAGKHAQGTICYTIS
 PVHTVEGYVKLAGQLDMGADSI ALKDMAALLKPQPAYDIIKAIKDTYGQKTQI
 NLHCHSTTGVTESLMKAIEAGVDVVDTAISSMSLGP GHNPTESVAEMLEGTGY
 TTNLDYDRLHKIRDHFKAIRPKYKKFESKTLVDT SIFKSQIPGGMLS NMESQLRA
 QGAEDKMDEVMAEVPRVRKAAGFPPLVTPSSQIVGTQAVFNVMMGEYKRMTG
 EFADIMLGYYGASPADRDPKVVKLAEEQSGKKPITQRPADLLPPEWEEQSKEAA
 ALKGFNGTDEDVLTALFPQVAPVFFEHRAEGPHSVALTDAQLKAEAEGDEKSL
 AVAGPVTYNVNVGGTVREVTVQQA

>PFREUD_18860 (12S subunit nucleotide sequence)

Antisense strand: (SEQ ID NO:12)

tcagcagggg aagtttccat gcttcttcgc cggcggggtc tgacgcttg tggcgtacat ctccagggcg gaagcaatct
 ttcgacgggt atcagccggg tcaatcacgt cgtcgacctg acccgggcg gggccacgt acggcgtgtt
 gaacgcgttc tggactcct cgatcttctc ggcgcgcgatg gctcgggat cgtcggcagc ctgatctcc ttgcggaaga
 tcacattgc cgcacctcg gcgccatca ccgcaatctc ggcgctgggc caggcgta ca cggcgctggc
 accaaggtea cggttgcaca tggccagga ggagccggc taggccttc ggagcaccac ggtgatcttc
 ggcacggtgg cctcggagta ggcgtacagc atcttcgcg catggcgaat gatgcgcgc tactctgct
 gcacgccggg caggaagccc ggcacgtcga ccagctgcac cagcgggatg tgaacgaat cgcagaaatt
 cacgaattcg gcggcctgt cagaggcgt gatgtcgagg caaccgcaca tcaccgacgg ctgattggcc
 acgatgccc ccgaacgacc attgaccgg gcgaaggcg tcacgaggt ggtggcatag ccggcctga
 cctcagggta gtcacccag tcgacgatct tggcaatgac atcgcgcacg tcataccct tcttgccgta aatcggaacg
 atgtcgcgca gtcgggtatt ggggctgacg tcattgttcg ggttgacgaa ggatgcttc tcaagtgtt tctgcggaag
 gaagtcagc agcttcttg caatgagctc cgcggcgctc tctctcgg ccacgaagt gatattgcc
 gagatggcca tatgggctc agcgcaccg agttcgtcag cgtgacatc ctgcgggtg accgactga
 tgacctgggg gccggtgat aacatatgg ccttcttgg catgatgat aagtcagtca gtccggcgca atacgagggc
 ccaccggcac aggggcggc aatgatggcg atctcgcca cgaccccga cagcttcacg ttggcgaaga
 acatcttgcc gtaaccgctc agcgagtcga tgccctctg gatccggcg ccgccgaat cgtagaagaa
 caggaagggc gtccgggtga gcagcgctg ttccatctc tcgacgacct tcgtggactg cgtctcgcca
 gccgaaccac ccatgacct gaagtcttg gacggcgct gcacgggacg accaaggtat gtgccacggc
 cgggtgaccac gccatctgc gggacgagc cttgtccat gccgaacaac gtggtgcggt gcttgcgga
 agcggcgacc tcgtgaacg aatggggatc gagcaggtg ttcaggcgt cagagcggt ctgcttacc tgggaatgt
 gcttctgac gcgacgttc ccgccaccg ctctgatcac ctggcgctc tctcgagct gtcacgcg accttcacg
 gtgctggcga gcttcaatt gttgtttca gccat

- 50 -

Sense Strand: (SEQ ID NO:13)

atggctgaaaacaacaattgaagctcgccagcaccatggaaggctcgctggagcagctcgagagcagcgccaggtgatcg
aagccggtggcggaacgtcgctcgagaagcaacattcccagggttaagcagaccgctcgtagcgctgaacaacctgc
tcgatccccattcgttcgacgaggtcggtgctttccgcaagcaccgcaccacgttgctggcatggacaaggccgctgctccgg
cagatggcggtggtcaccggccgtggcaccatccttggtcgtcccgtgcacgccgctcccaggacttcacggctcatgggtggt
cggctggcgagacgcagtcacgaaggtcgctgagacgatggaacaggcgctgctcaccggcagcccttctgttcttac
gattcggggcggtcgccggatccaggagggcatcgactcgctgagcggttacggcaagatgttcttcgccaacgtgaagctgc
ggcgctcggtccgcagatcgccatcattgcccggccccctgtgccggtggcgccctgtattcgccggcactgactgactcactc
atgaccaagaaggcccatatgttcacacggggccccaggtcatcaagtcggtcaccggcgaggatgtcaccgtgacgaact
cgggtggcgctgaggccccatagccatctcgggcaatatccacttcgtggcgaggacgacgacggcgaggatcattgcc
aagaagctgctgagcttcttcgcagaacaactgaggaagcatccttcgcaaccgcaacaatgacgtcagcccaatacc
gagctgcgcgacatcgtccgattgacggcaagaagggtatgacgtgcgcgatgtcattgccaagatcgtcgactgggtga
ctacctgaggtcaaggccggtatgccaccaacctcgtgaccgccttcgcccgggtcaatggtcgttcgggtgggcatcgtggc
caatcagccgctcggtgatgtcggttgccctgcacatcaacgcctctgacaaggccgcccgaattcgtgaattctcgcgattcgtca
acatcccgtggtgcagctggtgcaggtgccgggcttctgcccggcggtgcagcaggagtacggcggtcatcattgccatggc
gcgaagatgctgtacgcctactccgaggccaccgtgccgaagatcaccgtggtgctccgcaaggcctacggcggtcctacct
ggccatgtgcaaccgtgaccttggtgccgacggcggtgacgctggcccagcgccgagattcggtgatggcgccgagggt
gcggaatgtgatcttcgcaaggagatcaagggtgccgacgatcccgacgcatgcgcgccgagaagatcaggagtac
cagaacgcgttcaacacgccgtacgtggccgcccgcggtcaggctgacgacgtgattgaccgggtgataccggtcgaa
agattgcttcgccccggagatgtacgccaccaagcgtcagaccgcccggcgagaagcatggaacttccccctgc

>PFREUD_18860 (12S subunit amino acid sequence) (SEQ ID NO:14)

MAENNNLKLASTMEGRVEQLAEQRQVIEAGGGERRVEKQHSQGKQTARERLNN
LLDPHSFDEVGAFRKHRTTLFGMDKAVVPADGVVTGRGTILGRPVHAASQDFTV
MGGSAGETQSTKVVETMEQALLTGTPFLFFYDSGGARIQEGIDSLSGYGKMFFA
NVKLSGVVPQIAIIAGPCAGGASYPALTDFIIMTKKAHMFITGPQVIKSVTGEDV
TADELGGAEAHMAISGNIHFVAEDDDAAELIAKKLLSFLPQNNTTEEASFVNPND
VSPNTELRDIVPIDGKKGYDVRDVIKIVDWGDYLEVKAGYATNLVTAFARVNG
RSVGIVANQPSVMSGCLDINASDKAAEFVNFCDNFNIPLVQLVDVPGFLPGVQQE
YGGIIRHGAKMLYAYSEATVPKITVVLKAYGGSYLAMCNRDLGADAVYAWPS
AEIAVMGAEGAANVIFRKEIKAADDPDAMRAEKIEEYQNAFNTPYVAAARGQV
DDVIDPADTRRKIASALEMYATKRQTRPAKKHGNFPC

>*P.freudenreichii* (12S_C-term nucleotide sequence) (SEQ ID NO:15)

atggctgatgaggaagagaaggacctgatgatcgccacgtcaacaagcgctcgctcattggagtct
gagttgggttcactccagagcgatacccagggtgtcaccgaggacgtactgacggccatttcggccgcc
gttgccgctatctcggcaacgatggatcggtgaggtcgtccatttcgccccgagccgaactgggtcc
gcgagggtcgtcggtctcgcagaaccattccattcgt

>*P.freudenreichii* (12S_C-term amino acid sequence) (SEQ ID NO:16)

MADEEEKDLMIATLNKRVASLESELGSLQSDTQGVTEVDVLTASAAVAAYLGND
GSAEVVHFAPSPNWVREGRRALQNHSIR

[0202]

Propionibacterium acnes SK137 Transcarboxylase

P. acnes (12S subunit) (SEQ ID NO:17)

atggctgagaagaacaaatcaagctggccgataccatggccggccggatcgagcagctcgccgacgagcgccacgctgtg
gagcttggcggggcgaggatcgccctgcaaaagcagcgtgacaggggcaagcagacagcccgtgaacggatcgacaacct
cgttgatgcttattccttcgatgaggtgggtgcgttcgtaagcaccgcaccacccttttcggcatggacaaggccgaagtcccg
ccgacggcgtagtcaccggctgctgcgaccatccatggtcgcccgggtccacatcgctctcaggacttcaccgtcatgggtgggt
cggctggcgagaccagtcgacgaaggctcgtcgagacgatggaacagtcctgctgaccggcactccgtttctgttctatga
ctcggggcgcgcccgaattcaagaaggcatcactcgtgtccgggtacggcaagatgttctacggaacgtcaagctgtcgg
gctcgtgcccagatcgccatcattgctggccctgcgcggcgccctctattccccggcctgaccgacttcacatcat
gacgaagaaggcccatgtttcattacggggccccggagtcacaaagtcggttaccgggtgaggaggtgactgtgacgacctgg
gtgtgctggtatgcgcacatgtccacctcgggcaatatccacttcgtggccgaagatgacgacgccgagtgctcatcgcgag
aagtgctgagcttctcgccgaaaacaactgaggacgccagatctccaaccccaatgacgatgtctccccgcagcctgag
ctcgcgacatcggtccgctggatggtaagaagggtacgacgtccgcgacgtcatctcaagatcgctgactggggcgacta
cctagaggtaaggccggttggcgaccaacatcgtcaccgcctttgccgggtcaatggctgtaaccgtcggcatcgtggcca
accagccgaagggtgatgtcgggtgccttgacatcaatgcttcgggacaaggctgccgagttcattaccttctgcgactcgttcaat
attcgttgggtcagttggtgacgttctggttctgctggtgtccagcaggagtacggcgccatcatccgccacggcgcgga
agatgctgtatgcctactccgaggccaccgtcccgaagatcaccgtggtgctgcgtaaggcttacggcggtccttaccttgccat
gtgcaaccgtgacctgggtgctgacgccgtctatgcctggcgagcgcgagattgcggtgatgggtgccgatggcgctgcc
aacgtcattttcgtcgccagatcaaggactctgaggatcccgcagccaccgtgccgcgaagatcgaggagtaccgcaacgc
cttaacacgccttacgtggctgcccggcgtggacaggttgacgacgtgatcccgcggaacccgtcgcaagatcaccg
ccgctctggagacctacgccactaagcgtcagtcctccgtccggccaagaagcacggcgctatgccttgcgtga

P. acnes (5S subunit) (SEQ ID NO:18)

atgagtcacgaaagattggcggttaccgagctcgtgctccgcgacgcgcacatcagagcctgcttgcactcgcattggccatgga
ggacatgggttgatgcctgtgccgacattgatgcggcaggcttctggtccggttaatgctggggcgaggctaccttcgattcttga
tccgattcctcaacgaagaccatgggagcgtctgcgtactttccgcaagctgctgccgaactcccgggttcagatgctgtcgcg
tgcccaaaaccttctgggtaccgccactacaacgacgaggtcgtcgacaagtgttcgagaagtcggccgagaacggcatgg
acgtgttccgggtgttcgacgtctgaacgatcctcgaaccttgagcacgcgatggcagccgtcaagaagaccggcaagcac
gccagggcaccatctgtacaccacttccccgattcacaccccagagagcttcgtcaagcaggccgatcgtctcatcgacatg
gggtccgactcgtatgccttcaaggacatggctgcttgcgaagccgcagcctgcttacgacatcaagggcattaaggag
aacatccggacgtgcagatcaacctgcaactgccactccaccacggcgctacccttggtcacctgcagaaggccatcgagg
ctggtgtcgacgtcgtcgacaccgtatctcgtcgatgtcgtcggccccggggcacaacccaaccgagcttcttggtcgagatgct
cgagggcaccgagtagaccaccggcctcgacatggatgcctgctcaagatccgcgaccacttcaagaaggtgcgtccgaag
tacaagaagttcgagtcgaagacgttggtcaacaccaactcttccagtccagatcccggcggaatgctctcaacatggag
tcccagctcgaggccagggtgctggagaccgatggatgaggtcatgaaggaggtgccgcggttcgtaaggatgccggct
accgccgctggtcacccgctcctccagatcgtgggaacccaggcggtgttcaacgtcctgatgggcaatggttcgtacaaga
acctactgccgagtttgccgacctcatgcttggctactacggcaagcccattggcgagctcaatcccagatcgttgagatggc
caagaagcagaccggcaaggagccgatcactgccgtcccgcgacctgctcgagcctgagtgggaccagctggtcgagca
ggccaagagtcttgagggttcgacggctccgacgaggacgttcttaccacgcctgttcccgggagttgccccgaagttcct
caaggaacgcgcagggccccgaagagcgtcgcgatgaccgaggcacagctgaaggccgagaagggaaggcaccggcgcc
tgccggcatcgccggaccggtcaactacaacgtgacggctcgggtggcaacaccaccaggtgaccgtcgagcctgcgtga

P. acnes (1.3S subunit) (SEQ ID NO:19)

atgaagctcaaggtagaccgtcaatgacgtcgatacgacgttgacgttgataagaccgccaatgcgccgatggcgcc
gatcctcttgggtggcgggcgccggcgcccgatgaaggcatccgggtggcgcgccggtaaggccgggtgagggcgaggttcc
cgcaccgctagctgggactgttgccaagatcctggtggccgaaggagatgccgtcaaggccgggtcaggtgctcctgacctcg
aggccatgaagatggagaccgagatcaatccccggcgacggaaccgtcaaggggatcctggtgctgtcgggtgacggcg
tccagggtggtcagggcctggtgctctgggctga

[0203] *C. thermocellum* Transcarboxylase

>Cthe_0699 (12S subunit nucleotide sequence) (SEQ ID NO:20)

atggacaaagtagacaagatcggccttctccgtgaaaaactggcccaggtgaacagggcggaggagctgaaaaatcgcaa
aacagcatgatgccggaaaaatgacagcaagagaaagaatccaggctttattgatgaaaacagctttgttgagatcgacacatt
gttgagacaagaagcattgacttcgatatgcaaaaaaagaagtcggggagacgggtgtgtaacagggtatggtccatagac
ggacgtctggtctttgttgcggcgcaggactttactgtaatcggtgggtctttgggtgaaatgcatccgcaaaaatcacaaagt
aatggacatggcaatgaaaatgggcgcaccgtttataagcattaatgattccggcggtgcaagaattgaagaagggaattgacgc
actcaagggatttgagatatctctacagaaatactttggcttcagggtgtaattccccagatttcagttatcatgggacatgcgca
ggcggagcgggtatattctcctgcaataaccgactttatatattgttgacaaaaccagtcagatgtttataacgggaccccaggta
attaagtccgtaaccggagaaagacgtgacttttgaaaaacttggcggtgcggaaaccacaactccataagcgggtgttgcact
tcagaagtcaagtgaaaaagaatgtatagagcaaatcaaaaagcttattagttatcttctgataacaatcttccgatgttccgatt
gttccaactcaggatgacataaacagaattactgacaacctgtcgatattccgcaggactccaacaagccttatgacatgat
ggaaataatcacttccgtagtgtgacaacgggtgactttttgaaattcaaaaagactttgcaaaaacattataataggttcggcaga
atgaacggcggaaccgtcggtatagtggaatcagccaaaagttgccgcaggggttttgatgtgaactcctctgacaaagcc
gcaaggtttgtcgtttctgtgatgcgttcaacattccaattataaccttaccgatgtaccggggtatctgcccgagtagggcagg
agcacagcggagtaataagacacgggtgcaaagcttcttatgcttctctgaagccaccgttccaaaatcaatgttattgtcagaa
aagcttacggcggtgcatatattgccatgaacagcaagcaccttggagcggacatgtattgtcgtggccttcggcggaattgc
agttatgggaccgggaagggtgcggcaaacatcatttcaagaaagatatagctgctgccgatgaccaatggaaacaagaaaga
ggctcattgaagaatatcgtaaaaattctcaatccgtatgttgacagcttcaaggggttatgttgatgatgaattgatccggcaac
aacaaggataagactgattagtgcccttgaaatgcttgcaagtaagagagaaaacagacctgccaaaagcatggaaatattcc
attataa

>*C. thermocellum*_ (12S subunit amino acid sequence) (SEQ ID NO:21)

MDKVDKIGLLREKLAQVEQGGGAEKIAKQHDAGKMTARERIQALFDENSFVEID
TFVETRSIDFDMQKKKVPDGVVTGYGSIDGRLVFVAAQDFTVIGGSLGEMHAA
KITKVMMDAMKMGAPFISINDSGGARIEEGIDALKGFGDIFYRNTLASGVIPQISV
IMGPCAGGAVYSPAITDFIFMVDKTSQMFIITGPQVIKSVTGEDVTFEKLGAETH
NSISGVAHFRSSSEKECIEQIKKLISYLPDNNLSDVPIVPTQDDINRITDNLVDIIPQD
SNKPYDMMEIITSVVDNGDFFEIQKDFAKNIIIGFGRMNGGTGVGIVANQPKVAAG
VLDVNSSDKAARFVRFCDAFNIPITFTDVPGYLPVGVGQEHSGVIRHGAULLYAFS
EATVPKINVIVRKAYGGAYIAMNSKHLGADMVFAWPSAEIAVMGPEGAANIIFK
KDIAAADDPMETRKLIEEYREKFSNPYVAASRGYVDDVIDPATRIRLISALEM
LASKRENRPAAKHGNIPL

>Cthe_0700 (1.3S subunit nucleotide sequence) (SEQ ID NO:22)

atgaaaaagttttgataaaggtaaacggaaatcaatatgaggttgaggtgaagaaatcagagacgggtgcttcagcaccacagg
ttacttcagcacaccttcggctgcacctgcgccttcaccggcaccggctcaggaaacgaaaacagctgcaccaaagaaagac
agcacagtaccggcagggtgtacggcaattaaagctccgatccgggtaccatactgacattcgtgtaaatcaaggggatacg
gtaaagaaaggccaagttctttaattctgaagcaatgaagatggaaaatgaaatagttgctccaaatgacgggtacagttgcatca
attaatgtttcaaaggggtgcatctgtaaactcgagagaggttctgtctcattaaaatag

>*C. thermocellum*_ (1.3S subunit amino acid sequence) (SEQ ID NO:23)

MKKFLIKVNGNQYEVEVEEIRDGASAPQVTLSTPSAAPAPSPAPAQETKTAAPKK
DSTVPAGATAIKAPMPGTILDIRVNQGD TVKKGQVLLILEAMKMENEIVAPNDG
TVASINVSKGASVNVGEVLVSLK

>Cthe_0701 (5S subunit nucleotide sequence) (SEQ ID NO:24)

atggctaaggtaaaaattaccgaaacggcgctgagggatgccatcaatctctcattgcaacaagaatgagaatagaagagatg
cttctatcatagataaaactggacgagatcggttatcattctttggaggtatggggcggtgcaacctttgatgcctgcctgagatttt
gaatgaagacccgtgggaaaggcttagaattataaaaagccactgcaagaaaactcccccttcaaatgcttttaaggccagaat
ctttgggttacaagcattatgccgatgacgttggtgagttgtgtacaaaagagcgttgcaaacgggtataaacataaagaatttt
cgacgccttgaatgacaccagaaatagaaactgcaatcaaaagcctgcaaaaaagaaggcggtcatgctcagggaaacggat
gttatacaataagtcctgtcacatcttgaactttttgtcaaaagatgcaaaagactcttggtgaaatgggagctgactccatatgcgt
aaaggatatggcaggacttctgcttccatatgttgcatatgacctatcaaaagcattaaaagaaaacgtaaaagtgcgattcaact
cataccactatacagcggtgttgcttcaatgacatatctgaaggcaattgagcgaggtgctgatgttggtgactgcgctatctc
accaatgtcaatgggaacatcccagcctccgacagaacctcttggtgcaaccttaaaaggcacgcctgacgataccggacttga
cctggataaattaagtgaatcgcagactacttcagacctctcaagaaaaagtataattcagaaggacttcttgatgtaaaggttatg
ggagttgacgtaaacactctcaataaccagggtacccgggtggaatgctttcaaacctggtgtctcagttaaagcagtcctaatgcggt
tgataaattcgaagaggtctgaaagaagtgcgaagagtaagagaagacttcggatactcctcgttggttacacctacaagccag
attgtaggtactcaggcagttttaaattgtgtaacgggtgaaagatacaaaatggttccaaaagaatccaaggcactgatcaagg
gtgaatacggcagaacaccggctccgggtcaaccctgaagtcagaagaagattttaaagatgaagagccgattacaggttagac
ctgctgatttgatagagcccagcttgacaagatcagaatgaaatgaaagaatacctggaacaagacgaggacgtttgtccta
tgcactgttcccgaggtggcagagaagttctccaatacaggaaagctcaaaaataagatagaaccggacatggtcgattac
gaaaacaggggttcacccggttaa

>*C.thermocellum*_ (5S subunit amino acid sequence) (SEQ ID NO:25)

MAVKKITETALRDAHQSILATRMRIEEMLPIDKLDEIGYHSLEVWGGATFDACL
RFLNEDPWRLRIKSHCKKTPLQMLLRGQNLLGYKHYADDVVEYFVQKSVAN
GINIIRIFDALNDTRNIETAIKACKKEGHAQGTVCYTISPVHNLELFVKDAKTLV
EMGADSICVKDMAGLLLPYVAYDLIKALKENVKVPIQLHHTYSGVASMTYLK
AIEAGCDVVDCAISPMMSGTSQPPEPLVATLKGTPYDTGLDLKLSEIADYFRP
LKEYYISEGLLDVKVMGVDVNTLKYQVPGMLSNLVSQLKQSNVAVDKFEEVLK
EVPRVREDFGYPLVTPTSQIVGTQAVLNVVTGERYKMPKESKALIKGEYGRTP
APVNPEVQKKILKDEEPTVRPADLIEPELDKIRNEMKEYLEQDEDVLSYALFPQV
AEKFFQYRKAQKYKIEPDMVDYENRVHPV

>*C.thermocellum*_ (12S_C-term nucleotide sequence) (SEQ ID NO:26)

atgaaagagcaaataaatgaagaaattattctggcaatcagcgccattgctgctttggaacaagaccggatacaagcttgt
agtaagatcatttaaaagaataccccaactctcctgtatggtccgctacaggaaaaatcgagagaatcagaagaagtatg

>*C.thermocellum*_ 12S_C-term amino acid sequence (SEQ ID NO:27)

MKEQINEEIIAISAAIAALETRPGYKLVVRSFKRIPQTSPVWSATGKIERIRRS

[0204] T. saccharolyticum Transcarboxylase

>or0945 (12S subunit nucleotide sequence) (SEQ ID NO:28)

atgtcaatagatgataggattgaagaccttcttagaagaagagagatggttttagaaggcggtggtttagataaagtagagaaaca
acacaaaagggaagcttaccgcaagagagaggatatacaagcttttagatgaagatagctttgtggaaatagatgcgtatgtt
gagcacaggtgtattgactttggcatgaaaagcaaaggatacctggcgaggcgtagtacagggtatgggacgatagatgg
aaggcttgtctacgtttatgcacaggattttacggttttagaggagatcattaggcgagtatcatgcaagaaaatcacaaaatcat
ggatatggctttaagatgggagcaccgctcattggattaaatgattccggaggtgccagaatacaggaaggcgctgatgctttat
cgggatatggcaacatatatttcagaaacacgctggcatcaggcgtaataccgcaaatatcgggtgataatggggcccagcgctg

- 54 -

gaggtgcagttattcgcctgctcttactgactttatattcatggtagacaagacaagtcagatgtttataactggaccgcaggtcata
aaagccgtcacaggtgaagatgttcggcagaggagcttggtgatcgattactcacagcacgaaaagcgggtgtggcgcat
agggctgaaaacgacgaagagtggtgaagatggtgaggaagctattaagtaccttccatcaaacatttgaagatccgccac
agttggcgacagatgacgacataaacagattttccgataggcttattgagataatcccagatagtcctaataagccatacgaatga
aagaagtaatttcggaatagtgatgaaggcgtgtatttgaatcacaggcaatgtatgcgcaaacataataacggcatttgca
aggcttaatggaaggacggtagggataatagcaaatcagcctaagtttggctggatgtctcgacatcaatgcgtctgataagg
catcgagggtttataaggttttcgatgcatttaacatcccgtcttcaatatagtatagttccaggattttgcctggaacgaatcaa
gagtacgggtggaataatacggcatggggcaaatgtttgtacgttactctgaggctacagtgccaaaagtgaacttcattgtga
ggaaagcttatggcggtgcttaccttgccatgtgcagcaaaagacttaggagctgattttgtttggcatggcctactgctgaaatag
cggctatgggacctgatggggcagcaaacatcgtgtttaaaaatgaataaaaatcgtctgatgatcctgtggctgcaagaaatga
aaagataaatgagtacagggaatttcgcaaatccatacagggcagcagcgagaggatagtagatgatgtagttctgccgca
agagacgagacctgcctcatctcggcgttcgatatgcttatgagcaaaaggagtgcaaggccagcaaaaagcatggcaattt
tcctgtttaa

>*T. saccharolyticum*_ (12S subunit amino acid sequence) (SEQ ID NO:29)

MSIDDRIEDLLRRREMVLEGGGLDKVEKQHKGKLTARERIYKLLDEDSFVEIDA
YVEHRCIDFGMEKQRIPGEGVVTGYGTIDGRLVYVYAQDFTVLGGS LG EYHAKK
ITKIMDMALKMGAPLIGLNDSSGARIQEGVDALSGYGNIFRNTLASGVIPQISVI
MGPSAGGAVYSPAL TDFIFMVDKTSQM FITGPQVIKAVTGEDVSAEELGGSITHS
TKSGVAHFRAENDEECLKMVRKLLSYLPSNNLEDPPQLATDDDINRFS DR LIEIIP
DSPNKP YDMKEVISEIVDEGVYFESQAMYA QNIITAFARLNGRTVGIHANQPKVL
AGCLDINASDKASRFIRFCDAFNIPLLNIVDVPGLPGTNQEYGGIIRHGAKMLYA
YSEATVPKVTLIVRKAYGGAYLAMCSKDLGADFVLAWPTAEIAVMGPDGAANI
VFKNEIKSSDDPVAARNEKINEYRENFANPYRAAARGYVDDVVL PQETR PRLISA
FDMLMSKRESRPSKKHGNFPV

>or0947 (1.3S subunit nucleotide sequence) (SEQ ID NO:30)

atgaaaaattatagtaactgtcaatggaaaaatac gatgtggaagtagaagaagtaaaagtcgacgtggcaagt gagaaa
aaagcaaaagaagatactgctgtaaaaatgcgtcagatgcaagtgtaaaaagcaaacagggtgaagtaaaaacgaagtcaa
agacgggtttc caatcaatgcaccgatgccgggaactatattggatgtcaaaaataagccaaggccagactgtcagacgaggcga
tgtgtctttaatactggaagccatgaagatggaaaatgaaatcacgtcaccttacgatggcacaataataatccataaatgtttcaaaa
gggtgcctctgtaaatacaggcgatgtgctttgtacttaaatga

>*T. saccharolyticum*_ (1.3S subunit amino acid sequence) (SEQ ID NO:31)

MKKFIVTVNGKKYDVEVEEVKVDVASEKKAKEDTAAKNASDASVSKSQVEVK
NEVKDGF SINAPMPGTILDVKISQGQTVRRGDVLLILEAMKMENEITSPYDGTIISI
NVSKGASVNTGDVLLYLK

>or1888 (5S subunit nucleotide sequence) (SEQ ID NO:32)

atgtctaagataaaaaataacggagactgttttaagagatgcacatcaatcgttgctggcaaccagaatgacaaccgatgaaatgct
tcctatagcagaaaaattagatgaagttgggtttttctcgctggaagcatggggcggtgtacatttgatgcatgtatgagattttga
atgaagaccatgggaaagattaagacttttaagaaggcgattaagaagacaccttcaaatgttttaagagggtcaaaatttac
tcggatataaacactatcccgatgatgtcgtaaatgaatttataataaaaatcgttgaaaatggatatagataataagaattttgatg
cgttaaatgatgtgagaaatttagaagtgccaataaaaatctgcaaaaagtgcaggtgctcatgtacaggcagctattgtatatacag
ttagtctgtacataatacagatcattatttgaaagtggcaaggtctcttcaagatatgggtgcggattccatgcatgaaggaatg

- 55 -

tctggaatattatcacccctatgttgcatagcatttgattaaatctctgaaaagagcactttacacgccaattcaactgcatagccattat
 acagcaggactggcttcaatgacttatttaaaagccatagaagctggtgtagacggggttgatacagctatttctcgttgcccttag
 gaacatcacaccagctacagaatcaatcgtggctgcattgaaagatacagaatatacagggctagatttaaaattgcttgct
 gagatagctcagcattttaatgtagtcaaacagaatcacaaaaatgacagcgatatgtctttgcttatgtctgttgatgttaaagcatt
 agaaagtc aaataaccagggggaatgttatcaaatttggtttcacagctaaagcagcagaatgcattaaacaaatatcaagacgtct
 tgaaagaagttccaagggtacgcgaagatttggtatatacctctcttctgttactccaatgagccagatggttggaaccaggtgttt
 taaatgttattacaggggagagatataaaatcgttctaaagaataaagattatgtcaaagggttatatgggatgccaccagctcc
 aatttcagattctatacgaagaaaataatcgccgatgaagaagtaatttcaaagaggccagcagatttactaagtctcaattgga
 tgaatttaaaatagagataaaggaatttatagagcaagatgaagatgtttatcatatgcattatttctcaagtagcaagaagatttt
 cgagtataggcaagccaaaaatacagaattgattcaacattataaatatcgaagaagggttcacccgatataa

>*T. saccharolyticum*_(5S subunit amino acid sequence) (SEQ ID NO:33)

MSKIKITETVLRDAHQSLLATRMTTDEMLPIAEKLDEVGFFSLEAWGGATFDAC
 MRFLNEDPWERLRLKKAIKKTPLQMLLRGQNLLGYKHYPDDVVNEFIKSVEN
 GIDIIRJFDALNDVRNLEVPIKSAKSAGAHVQAAIVYTVSPVHNTDHYLKVAKSL
 QDMGADSICIKDMSGILSPYVAYDLIKSLKRALYTPQLHSHYTAGLASMTYLKA
 IEAGVDGVDTAISSALGTSQPATESIVAALKDTEYDTGLDLKLLAEIAQHFNVV
 KQNHKNDSDMSLLMSVDVKALESQIPGGMLSNLVSQKQNALNKYQDVLKE
 VPRVREDLGYPLVTPMSQMVGTQAVLNVTGERYKIVPKEIKDYVKGLYGMPP
 APISDSIRKKIIGDEEVISKRPADLLSPQLDEFKNEIKEFIEQDEDLVSYALFPQVAR
 RFFEYRQAKKYRIDSTLLNIEERVHPI

>*T. saccharolyticum*_(12S_C-term nucleotide sequence) (SEQ ID NO:34)

atggaagagataaatgaagaaatagttgctgtcattgaagctgcgatttacgcggcatttggtcagtagcaaaagaatttccgcatt
 caaggtataaaagagagtggtgactcaaatatgccggatggagaaaagctggcctttacaatcagatgagatag

>*T. saccharolyticum*_(12S_C-term amino acid sequence) (SEQ ID NO:35)

MEEINEEIVAVIEAAIYAAFGQYEKNFRIKVIKRVDSNMPEWRKAGLYNQMR

[0205] Caldicellulosiruptor bescii DSM 6725 Transcarboxylase

>*C. bescii*_(12S subunit nucleotide sequence) (SEQ ID NO:36)

atgacaaacaagctcagagagctcaagcaaaagagagaaagaataactaaagcttggtggagaagataaaataaaaaaacagc
 atgatagcaaaaaacttactgttagagagagaatagaatatttacttgaccctggaagcttcaatgaaatagatatgtttgttgaca
 cagatgtcaagaatttgatagaaagatacattgtccctgtgatggtgttgtaacgggttatggaacaatcaatggcagaaaaagt
 tttgtttatgctcaagattttacttcgataggcgggttctcttggtgagatgcattgcaaaaaagatttgtaaagtttggacttagcattaa
 aatatggtgtccagtgtataggataaatgattctggtggtgcaagaattcaagaaggtgttgatgcattagcaggatatggtgaaa
 tcttctatagaaataccatggcatcaggtgtaatccacaaattgcagctataatgggacctgtgcaggtggagctgtatactctcc
 tgctattatggattttttttatgggtggacaaaaccagccaaatgtttgttacaggacctcaggtataaaagctgtgactggagagg
 agatatcctttgaagagcttggtggcgcttactcacagctcaaagagtgagggtgtgctattttattgcagaggatgagtatcacct
 acttgatagataaagtatttattgtcgtttataccttcaataacatggaagaccacctttataatgtcatctgattcagaaaaaga
 tttgttcccgagctcgaaaaataataattccgcaagagccaaacaaagcttatgatgtaaaagaaataattataaagtagtagacaac
 caagaatttttagaagtacaaccttattttgtcaaaaatgctgttgtaggatttggtagaatagggggcttagcgttaggaattgtagc
 aaatcagcccaaagtgaacgtggagtgttgattatgattcgtctgacaagatagcacgatttgaagattttgtgatgcttttaata
 tccccataataacatttacagacgtgcctggattttgccaggtgttaaccaagagcacaatggaataatcgtcatggggcctaagg
 tttgtatgcatactcagaggcaacagttccaagataaatgtaattttgagaaaagcatatggtggggcttaccattgcaatgagca
 gcaaacacattggtgcagactttgtttgtcatggccaactgccgagatagctgttatgggaccagatggcgagcaaatattata
 ttagaaaaagagatacaaaagcgtcaaaatcccgaagaggaaagaaaaagaaggatagaagagtatactcaaaagtttgcaaat

- 56 -

ccatacattgcagctgccgtgggtatgttgacgatgtgattgagccacagcttaccgtaacaaaatcattgagcgctcaaaat
ttccattacaaaaagagagcaaaggcccccaaaaagcatggcaatattccatta

>*C. bescii*_(12S subunit amino acid sequence) (SEQ ID NO:37)

MTNKLRELKQKRERILKLGEDKIKKQHDSKKLTCRERIEYLLDPGSFNEIDMFV
EHRCQEFDMDKDTFVPCDGVVTGYGTINGRKVFVYAQDFTSIGGSLGEMHAKKIC
KVLDLALKYGCPVIGINDSGGARIQEGVDALAGYGEIFYRNTMASGVIPQIAAIM
GPCAGGAVYSPAIMDFIFMVDKTSQMFVTGPQVIKAVTGEEISFEELGGAYTHSS
KSGVAHFIAEDEYHLLDMIKYLLSFIPSNMEDPPFIMSSDSEKRFVPELENIIPQE
PNKAYDVKEIYKVVNDQEFLEVQPYFAQNAVVGFGGRIGGFSVGIVANQPKVNA
GVLDYDSSDKIARFVRFCDAFNIPITFTDVPGLPGVNQEHNGIIRHGAKVLYAY
SEATVPKINVILRKAYGGAYIAMSSKHIGADVFVAWPTAEIAVMGPDGAANIIFR
KEIQSAQNPEEERKRRIEEYTQKFANPYIAAARGYVDDVIEPQLTRNKIIEALKISI
TKREQRPPKKHGNIP

>*C. bescii*_(1.3S subunit nucleotide sequence) (SEQ ID NO:38)

atgagaaagtcaaggtgaagatcaatagccaagaattgtgtagaagtgaagaataggagtgaaaatgctacttctgtcgt
gccaaggcctaagattggccattttgagccaaaacaggaaaaacatgaggataaaacaaaacaaagccctgtactttcttctgat
aaaaattcggttggtgccagctccgggtactattgtaaggctgctaaaaagtgaagggtgatgtgttgatgcaaatgaacctgttt
aattctgaagccatgaaaatgaaaaatgaaataactgcacctgtcaaaggaaaaatgaaagaatacatgtaaaggaagggca
gaaggtgacaaaaggagatttgctatttgaaatagag

>*C. bescii*_(1.3S subunit amino acid sequence) (SEQ ID NO:39)

MRKFKVKINSQEFVVEVEEIGVENATSVVPRPKIGHFEPKQEKHEDKTKQSPVLS
SDKNSVVAQLPGTIVRLKSEGDVVDANEPVLILEAMKMENEITAPVKGKIKRIH
VKEGQKVAKGDLLFEIE

>*C. bescii*_(5S subunit nucleotide sequence) (SEQ ID NO:40)

atgggggtaaaaataacagaaacaatactcagagatgctcatcagtcactcattgcaacccgcatgacaactgaacagatgcttg
agattgctcctgtgcttgaccaagttggttatttctcggtgagtgctggggcggtgctacatttgatgcgtgctgagggttttcaatg
aagacctatgggaaagattaaaaagactgagaactgctttaaaaagacaaagctccagatgctcttcgaggacaaaatcttggt
gggtatagacattattctgatgatgtgttgagagttgtaaaaagggccatatactatggcattgatattataagaatatttgatgca
cttaatgacatccggaatattgaaatggctctaaaaataacaaaaaaagaaaaaggacatgccaggttgccatatcatactgt
ctcaccttatcatactattgaaaactatgtaaattggcacaacaaatagaagaacttggggcagactcaattgtataaaagacatg
gctgggcttctctccatttgatgcttataaactgtaaaagcgtaaaagagcaggtaaaacttctattcatcttcatacacactac
accacaggatttgatcaatgacatatttgaaagctgtcgaagcaggtgtggatggattgacacggcttatctccgcttgactg
ggcacatcccagctccaaccgaaacaattgtatatgcacttgaaaatacagaatatgctccaaaacttgatttagaaaagatcaa
cgaggcaagcgaatattttaagttactcagagaagaatatataagaaaagggtcttgacccgaaagtattaagtgtgatataa
acgctcttcattatcaatacctggtggaatgctatcaaatcttatttctcagctaaaagaacaagggcaggaagacaagttgatg
aggttttaaaagggtacctgaggttcgaaaagattttggatatccgccacttgtaactcctacgagtcaaatgtgggaacacaag
ctgttttgatgttatagcaggtgagagatacaaaactgtcacaaaagaacaaaaagcatattttaagggtgagtatgggaaacctc
cagctcctgtgaatgaagaggtaaaaagaaaaatcttgaaagacgaaaaagagataacctgcagacctgcagatttgatttggc
agagcttgaaaaatgcaaaaagaaagattaaggagtattgaaaatgatactgatgtgtaacttactgtttatccctcaacttgca
gaaaatttttcaaattaagggttcgaaaaaaatacaagggtgacgctgatcttggtcagggttaacaaagtgtatcctgtg

>*C. bescii*_(5S subunit amino acid sequence) (SEQ ID NO:41)

- 57 -

MGVKITETILRDAHQS LIATRMTTEQM LEIAPVLDQVGYYSVECWGGATFDACL
 RFFNEDPWERL KRLRTAFKKT KLQMLLRGQNLVGYRHYSDDVVEEFVKKAIYY
 GIDIIRIFDALNDIRNIEMALKITKKEKGHAQVAISYTVSPYHTIENYVNLAQIEE
 LGADSICIKDMAGLLSPFDAYKL VKALKEQVKLPIHLHTHYTTGFGSMTY LKAV
 EAGVDGIDTALSPLALGTSQPPTETIVYALENTEYAPKLDLEKINEASEYFKVLR E
 EYIRKGLLDPKVLSVDINALHYQIPGGMLS NLISQLKEQGQEDKLDEV LKEVPEV
 RKDFGYPP LVTPTSQIVGTQAVLNVIAGERYKLVTKETKAYFKGEY GKPPAPVN
 EEVKRKILKDEKEITCRPADLILPELENAKEKIKEYIENDTDVV TYCLFPQLAENFF
 KLRFAKKYKVDADLVQGNKVYPV

>*C. bescii*_ (12S_C-term nucleotide sequence) (SEQ ID NO:42)

atgtatgtcaggtcagtactatttcaaccattacaaaagaagaacttgcttgattgtgcatgtctgcacattgtgatgggtgaagg
 tcaatataaaattaccaataactaaacagcaaaacaagtggggtcaaaggtgcaagagaaatgatgtcaatcagtcacagatg
 ttttatagatggagg

>*C. bescii*_ (12S_C-term amino acid sequence) (SEQ ID NO:43)

MYAQVSTISTITKEELACICAC LHIVMGEGQYKITNITKQ QNKWVKGAREMMLN
 QSQM FYRWR

[0206] *Clostridium cellulolyticum* H10 ATCC 35319 Transcarboxylase

>*C. cellulolyticum*_ (12S subunit nucleotide sequence) (SEQ ID NO:44)

atgtcacaaattgaaaagatacaaaatttaaaaaacatgaaaaaactatagctaaaggcggcgagagaagagaaaatagcaaaa
 agacacgcagatggaaagctttctgccagagaaaagaatccattgtgttgatgaaaacagttttgttgaggtgatgcattcatag
 aatccagatgctttgactttggtatgcagaagaagaactccaggtgacgggggtgttaccggttacggaacagttaatggcaga
 aaggctctttgttcacacaggactttactgttataggcgggtcattgggagagatgcacgcaaagaaaattacaaaggttatggata
 tggctctgaaaaatgggagcaccgttcatagccattaatgattccggcgagctcgtattgaggaaggtctggatgctcttcaggtt
 acggagatatttttacaggaatactcttgcacagggcgttatccgcagatatcagtaataatggggccatgtgcaggtggtgcgg
 tatattccccggccataactgattttatattcatggtggaaaaaacaagtcagatgtttattacaggcccacaggtataaaagctgtt
 acgggtgaagatgtatcagttgaaaatctgggaggtgcagatgttcatactgctacaagcgggtgtagcacatttcaaatctcaag
 cgaagaagagtgtatagaagataaaagaggcttttaagttttattcccagataataatgtatcagatactatgtactacggagtgtctg
 atgctgccgacagattagccgaaagcctcaacagcattattccagaagagtc aaacaagccatgatgacatgtttgacgtaatagc
 agaaglagtagatgatggagatttcttgaagttcagagttattctctcagaatataataatcggttgcaagaatgaatggcaga
 agtgttggtattgttgc aaaccagcctaagataatggcagggctactagatatgaacgggctgataaggcggcacgtttcgttcg
 tttctgtgatgcatttaattctgtcgtttcattaaccgatgtacctgcattcctgcccgggtagcccaggagcataacggcataa
 tacgtcacgggtgaaaactcctatatgctttctgaagcaacagtaccaaagataaatgtattcttagaaaggcatatggaggag
 cataatattgctatgaacagtaaaacaataggtgccgatatggtttggcatggccatcagctgaaattgcagttatgggacctgacg
 gagcagcaaatattatatttaaaaaggatattgctgcgtcggaagatccagcagaaaccagaaaggaaaagattgcggaatata
 gagataaattctcaaatccttatgtagcagcatcaagagggtatattgatgatgttatcgagccttctgaaaccagagtaaaaattat
 aactgctctggaatgtctggatacaaaagagggaacaggccttcaaaaaacatggaaacattccgcta

>*C. cellulolyticum*_ (12S subunit amino acid sequence) (SEQ ID NO:45)

MSQIEKI QNLKNM KKTIAKGGGEEKIAKRHADGKLSARERIHLLFDENSFVEVDA
 FIESRCFDFGMQKKKLPGDGVVTGYGT VNGRKVFVSSQDFTVIGGSLGEMHAKK
 ITKVMDMALKMGA PFIAINDSGGARIEEGLDALSGYGDIFYRNTLASGVIPQISVI
 MGPCAGGAVYSPAITDFIFMVEKTSQMFITGPQVIKSVTGEDVSVENLGGADVHT

- 58 -

ATSGVAHFKSSSEEEEDIKRLLSFIPDNNVSDTMYYGVS DAADRLAESLNSIPE
 ENSKPYDMFDVIAEVVDDGDFFEVQSYFSQNIIGFARMNGRSVGIVANQPKIMA
 GSLDMNAADKAARFVRFCDAFNIPVVS LTDVPAFLPGVAQEHNGIIRHGAKLLY
 AFSEATVPKINVILRKAYGGAYIAMNSKTIGADMVLA WPSAEIAVMGPDGAANII
 FKKDIAASEDPAETRKEKIAEYRDKFSNPYVAASRGYIDDVIEPSETRVKIITALE
 MLDTKRENRP SKKHGNIPL

>*C. cellulolyticum*_(1.3S subunit nucleotide sequence) (SEQ ID NO:46)

atgagtaatatataataaaggtaaacggaactccttatgaagtagaggtgaagaagtggcgggggaaggccatttcagct
 gctccaaagctaagagctaccaagccgggacatacctctgctgcaaaagcagcacagccgaggcaggtaaagcaggtgat
 gttgctgctcaatgccgggaactgtttaaaggtaaggttgctatcggtgatgaagtaagaaggggcaggtacttttaatactt
 gaagctatgaaaatggagaatgaaatagttgctccggctgacggtaaaagttacggcgtaaacgtcgaggccggaaagtctgtt
 actgctggagaactaatggtgtctatagcc

>*C. cellulolyticum*_(1.3S subunit amino acid sequence) (SEQ ID NO:47)

MSKYIHKVNGTPYEVEVEEVGGGRPISAAPKLRA TKPGHTSAAKAAQPQAGKAG
 DVAAPMPGTVLKV KVAIGDEVKKGQVLLILEAMKMENEIVAPADGKVTALNVE
 AGKSVTAGELMVSIA

>*C. cellulolyticum*_(5S subunit nucleotide sequence) (SEQ ID NO:48)

atgccaggcgtaagaattacggaaacagttttaagagatgctcaccagtccttatagcaaccagaatgaagaccgaagaaatg
 ctccaattgttgagaagcttgacaatattggttaccattcactggaagcttggggcgagctacttttgactcatgtatgagattttg
 aatgaagatccatggatgagacttagaaaaataaaagatgttgcaagaaaacacctctgcaaatgcttcttagggccagaacc
 ttttaggatacaaacactatgccgatgatagattgagtactttgttcagaaggctgttgcaaacggcatggacattatgagaatattc
 gatgcactaaatgatgccaggaatatcgagacggcaattaaggcatgtaaaaaggaaggcgccatgctcagggctgtatttgc
 tatactataagtctgttcacaatcttgagctttttgtaaaagatgcaaagcagttggagagcatgggagcagattctatctgtataaa
 agacatggccggactctgtgtgccgtatcaggcttatgaactggttaaaggctttgaaagaaagtgtaaagataccgatacaattgc
 acactcactatactagcgggttagcatctatgacgtatttgaaggctatagaagcaggtatagatattgttgactgtgcaatttcacct
 atgtcaatgggaacgtcacagccgcctacagagcctttggtggcaactttaagggaactgatttcgatactggactggatttggga
 aaaactcagtgaaattgcagactatttcagacccttaagaaaaatatattgagagcggactattagacgttaaggtaattgggtgt
 tgacgttaacactctttttatcaggctacgttggaatgctttcaaatctgtttcacaattgaagcagtc aaatgctttggataaatat
 gaagagggttctcaaggaagttccagagtaagagccgatttcggctatcctccgcttgaacaccatcaagtcagatagttggtac
 ccaagcggtagcttaattgtattgactggtgagagatacaagatggtacaaagggaatcaaaaggcgttgtaaggggggaatacgg
 taaaaccctgcacctattagtgtatgaaataaaagctaagattctggcgatgaaaagcctataacatgcagacctgctgacctta
 ttgaacctgagcttgaaaagattagagaagctgttaaggattatataagcaggtatgaagatgtactttcatacgaatgcttcctc
 aggttgccgagaagttctttaacagcgtattgaggatagaataaggctactgcacccgcacagacgaaataaaacccgaag
 ttgtagcggcaatatcagccgtagtaaacgaaatgggcgaaagagacggcacacagtacagaatcggaatatctctaagttga
 accagaatcagaacagatggagctctgtatggtatgcttgatagattcagaacaaaaatt

>*C. cellulolyticum*_(5S subunit amino acid sequence) (SEQ ID NO:49)

MPGVRIETVLRDAHQS LIATRMKTEEMPLIVEKLDNIGYHSLEAWGGATFDSC
 MRFLNEDPWMRLRKIKDVAKKTPLQMLLRGQNLLGYKHYADDIVEYFVQKAV
 ANGMDIMRIFDALNDARNIETAIKACKKEGGHAQGCICYTISP VHNLELFVKDAK
 QLESMGADSICIKDMAGLLVPYQAYELVKALKESVKIPIQLH THYTSGVASMTYL
 KAIEAGIDIVDCAISPM SMGTSQPPTPEPLVATLKGTD FDTGLDLEKLSEIADYFRPL
 KEKYIESGLLDVKVMGVDVNTLIYQVPGGMLSNLVSQ LKQSNALDKYEEVLKE
 VPRVRADFGYPPLVTPSSQIVGTQAVLNVLTGERYKMVPKESKGVVKGEY GKTP

APISDEIKAKILGDEKPITCRPADLIEPELEKIREAVKDYIEQDEDVLSYAMLPQVA
EKFFKQRIEDRNKATAPASDEIKPEVVAAISAVVNEMGERDGTQYRIGNISKLNQ
NQNRWSLYGMLDRFRTKI

[0207] Corynebacterium kroppenstedtii DSM 44385 Transcarboxylase

>*C. kroppenstedtii*_ (12S subunit nucleotide sequence) (SEQ ID NO:50)

atgagtgagcaacctcacgateccagcatgcctgagcgctcgacagctggaagaagaaagaaaccgcatccgactcggc
ggcggggcaggcacgcctggacaagcagcacgaccgcggaagatgaccggcgagcgcacccaagcttgctgacg
aagacacgttcaggaaaccggaatgttcgccaagcaccggacaacgcacttcggcatggacaaggccgacgccccg
acggcgctgacacggatccggcggtctacggacggcagtgacatcgctccaggacttcagcgctatggcggtatc
tgctggcgaaatgcagtccaacaaagtggctgcatgatgaaggcgctcgacacacggcacccttcgtctttatcaacga
ctccggcgagctcgtgtccaagaggcgcactccctctccggatacggcgctgtttacaacaacgtgctgctctccgg
actcgtaccgcaggtctccatcatcgccggcccggtgctggtggcgctactcgccggcactgacggacttcacatcca
gaccgcgaaggccaacatgttcacacggcccaaggtcatcgagtcgacggcgaaaaagtcacggccgacgaactc
ggtggtgccgatgccacatgagcacagctggcaacattcacttcgctcgccgacgatgacgagcaagccatcctgatcgca
gaagtcctgagcttctgcccgaacaaacaccgaagagccgccatcgctgatccggacgaggtgtcgagcccgacgatt
ccctccgacatcgctcccgctgatggcgcaagggtacgacgtccgcgatatcatccgcaagatcgctgactacggcgac
ttcctcgaggtcaggccggatagcccaaacctcgtggctggattgcccgcgtggtggccgacagtcggtatcgctgct
aaccagtcgaagtgtatcgccggtcttgacatcaactcgctggacaaggcgcaagcttcgttcgcttcgactcctca
atattccgctcctaccctcgctgacgtccccggttcacgaggtgctgcacaagagcatggcggaatcattcgccacggcg
cgaagatgtgttcgctactcggcgccacgtgcccgaagctgacgtggtcctccgcaaatcctatggcggtatcgtacctgg
ccatgtgctccaaggaccttgccgacgacgctggtggcgccaccgctgaaattgcggtcatgggtgacgacggagc
cgtgaacgtcgttccgtaaggaaatcaagaaagccaggaagaggggtggcgacgaagccgtgcagcaagaagagcga
actcgtccagctctacaagacaccttcgacgccatacatggcggtcccgaggcctcgtgatgacatcatcgacccgc
ggacacacgtcgcaaatgtctggtccctggagttgctgaccaacaagcgtgagaaccggcgtccaagaagcacggcctg
gcacccaac

>*C. kroppenstedtii*_ (12S subunit amino acid sequence) (SEQ ID NO:51)

MSEQPHDPSMPERLGQLEEERNRIRLGGGQARLDKQHDRGKMTARERITKLVDE
DTFQETGMFAKHRTTHFGMDKADAPADGVVTGSGAVYGRPVHIASQDFSVMG
GSAGEMQSNKVVAMMKASATTGTPFVFINDSGGARVQEGIDSLSGYGRVFYNN
VLLSGLVPQVSIIAGPCAGGAAYSPALTDFFIIQTRKANMFITGPKVIESVTGEKVT
ADELGGAHAMSTAGNIHFVADDDEQAIIAQKLLSFLPQNNTTEPPIVDPDEVV
EPDDSLRDIVPVDGRKGYDVRDIIRKIVDYGDFLEVQAGYAQNLLVVGFAVVGR
TVGIVANQSQVMSGVLDINSSDKGASFVRFCDSFNIPLLTLVDVPGFMPGVAQEH
GGIIRHGAKMLFAYSAATVPKLTVVLRKSYGGSYLAMCSKDLGADRVWAWPT
AEIAVMGADGAVNVVFRKEIKKAQEEGGDEAAAKKSELVQLYKDTFSTPYMA
ASRGLVDDIIDPADTRREIALALELLTNKRENRPSSKKHGLAPN

>*C. kroppenstedtii*_ (1.3S subunit nucleotide sequence) (SEQ ID NO:52)

atgaaactgacagttaccgtcaacggcgctcccctattccgtggacgtagaagttgaacacgaagaacgccccacactcggcac
catcatcactggtggcaacagcaacgggccaacacccaccgcgcgaccacatcatgtccagggtgtcagcgccaattcgg
tcacggcaccctggctggttccgctcagcaaggtgctgtggaggaaggccaagccatcacggccggcggaagtatcgtgtc
cttgaagccatgaagatgaaaccgaaattacggccccaacgacggcaccgtcaccgcgttcacgtgcaaccggcgacg
ccgttcagggtggacagctctgctggagatcggggac

>*C. kroppenstedtii* (1.3S subunit amino acid sequence) (SEQ ID NO:53)

MKLTVTVNGVPYSVDVEVEHEERPTLTGTHITGGNSNGPTPTAPTSSVQGVANS
VTAPLAGSVSKVLVEEGQAITAGEVIVVLEAMKMETEITAPNDGTVTALHVQPG
DAVQGGQSLEIGD

>*C. kroppenstedtii* (5S subunit nucleotide sequence) (SEQ ID NO:54)

atgaccacgcgaaaaattggagtgaccgaactggctctgcgtgatgctcaccagagcctcatggcaacacgcatggccctcga
agacatggctcgatgcctgtgaggatagacaaagccgggtactggagcgtggaatgtggggcggggcaaccttcgacgcc
tgcatcgccttctgaacgaagacccgtgggagagactgcgcacattccgcaagctcatgccaaactcacgccttcagatgctg
cttcgtggccagaatcttctgggataccgtcactacgaggacggcgtcgtcgataagttgtgaaaaatccgtgaaaacggcat
ggacgtcttcaggggttttcgacgcgtcaacgacccccgcaacctcgagcacgccatgcaagctgtgaaaaagtgggcaagc
acgcgcagggcaccatctgctacaccgtgtccccgtacacgacgtgcagggtacattgatctagcagggcgtttgctggaca
tgggcgcggattcgatcgcgctcaaagacatggcagcgtgctcaaaccacagccccctacgacattatccgtggcattaag
gacacctacggcgaggatagcagatcaacgtccactgccactccaccaccggcgctcacgatggtcaccctcatgaaggccat
tgaagccggtgcggatgctcgcacaccgccatttctccatgtccctcggccagggcataaccccaccgagtcctcgtcga
aatgctcgaaggaaccgactacgagaccgggcttgacatggatcggtcattaacatccgcgaccacttcaagacagtgcgcc
cgaagtacgcggagtttgagtcgaaaacactgggtcaacaccaatattttccaatcgagattccggggcggaatgctctccaacat
ggatcgcagctcaaagcccaggcgcgggcgaccgtatcgacgaggtcatgaaagaagtccccgtcgttcggaaagctgc
cggatacccgccgttggtgacgccatgctccagatgctcgaccaccaggccgtgtcaacgtgctgatgggccgtacaaag
tactcacggctgaattcgccgacctccctcctcggtactacggcggaagcaccaggtgagagggtataaagacctcatcgagcaa
gccaagaagcagaccggcaaagagcccatcaccgagcgtcctgctgacctcctgagcccgaatgggacaacctggttgagg
aagctgacgaactcgacggcaccgacgggtccgacgaagacgtcctcacaacgccctgttcccgaggtcgcgcggcggtt
cttcaagactcgccccgacggccccgaagaacgtcggcaagactaaggaacagctcgagcgcgaagagggcgaaggcctccg
gcgacgccactgccatccgcgaaccgattatgtacaaagtcaccacaggcgccgcagccacactgtctccgtggaaccgc
a

>*C. kroppenstedtii* (5S subunit amino acid sequence) (SEQ ID NO:55)

MTTRKIGVTELALRDAHQSLMATRMALEDMDACEDIDKAGYWSVECWGGAT
FDACIRFLNEDPWERLRTFRKLMPNSRLQMLLRGQNLLGYRHYEDGVVDKFVE
KSAENGMDVFRVFDALNDPRNLEHAMQAVKKVGKHAQGTICYTVSPLHDVQG
YIDLAGRLLDMGADSIALKDMAALLKPQPAYDIIRGIKDTYGEDTQINVHCHSTT
GVTMVTLMKAIEAGADVVDTAISSMSLGPGHNPTESLVEMLEGTDYETGLDMD
RLINIRDHFKTVRPKYAEFESKTLVNTNIFQSQIPGGMLSNMESQLKAQGAGDRI
DEVMKEVPVVRKAAGYPPLVTPSSQIVGTQAVFNVLMGRYKVLTAEFADLLLG
YYGEAPGERDKDLIEQAKKQTGKEPITERPADLLEPEWDNLVEEADELDGTDGS
DEDVLTNALFPQVAPGFFKTRPDGPKNVGKTKEQLEREEAKASGDATAIREPIM
YKVTGGRSHTVSVEPA

>*C. kroppenstedtii* (12S C-term subunit nucleotide sequence) (SEQ ID NO:56)

atgaatacagacaatgcactctgctgaactcagtcagttgttgcccgctgtccaaccaggtagaaaaactctcccgcaacg
tcaccaagctcgaaaaatgaagttgcggcactgaagcagcgtctgacgaggaaattcctgaagatgtcttgatcgatcagtcg
ggcgtatccgcctacatgggtaaccgcggaaccgtgcgcgcagttcattcttgcccatcgacgtggtcacagcaaggtc
ggcaggcagttcagcacaaggcgaaatggcaa

>*C. kroppenstedtii* (12S C-term subunit amino acid sequence) (SEQ ID NO:57)

MNTDNASSAELSQLLARLSNQVEKLSRNVTKLENEVAALKQRSDEEIPEDVLIAI
SAAVSAAYMGNRGTVRAVHFLRHSWSQQGRQAVQHKAKWQ

[0208] *Geobacter bemidjiensis* BEM(^T) Transcarboxylase

>*G. bemidjiensis*_ (12S subunit nucleotide sequence) (SEQ ID NO:58)

atgtccatagaagagaagataaaagcgctgaacgacaagaagagcaagctgaagctgggcgggcggtcgaagatcgac
cagcagcacgcccaggggaagcctgaccgcccgggagcggatagaggcgctgggtggacaaggacagctccaggaaatcg
gcatcttcgccaggcaccgctgcaccaatttcggcatggccgggaaggaactgccggccgaagggtggtcaccggcgag
ggagcgtgggcgggaggatggtgcacctggcgagccaggattaccgctgccgggggagtcggcgggcgaggtgcacagc
gacaagatcgtgcaggcgatgctggggtcgtgaagaccggaaccccccttctctcatgaacgattccggcgggcgccaggat
ccaggaagggatcgactcgttagccggctacggcaaggctcttaccacaacgtgatgctcagcgggggtggtgccgcagatct
cgctcatctgcggccccctgtgccggggcgcgccctacagccccggcgctcaccgattcatcatccagaccgccaaggcgcg
catgttcatcaccggcccccttcgtgatcaaggaggcgaccggcggaagagatcagcgccgaggagctgggagggccactgtc
gcagatgaacctatagcggcgtagcccatctgtggcgggagaacgacctggtggcgcttcgcatctgaagaagctcttcccta
cctccccccaacaacatcgaggaccgcccagttggaaagcgacgacgtcatcgtccggacaagacgtgaacagcatc
gtgccgtcggagcagaagaaggcctacgacgtgaggaaacgtgacacgcccgtgacgagggcgcgacttctggaggtg
cagcctctgttcgtgccaacatcgtggtcgggttcggcaggatactcggcgagcgtcggcatcgtcgcaatcagccgtc
ggtcttggcgggggcgctggacatcaacgcttcggacaaggggagccaggttcgtccggttctgaacgccttcaacatcccgt
ggtgacctggtggagcttcgggtttctccccgggtacagcaggagaagggggggatcatccggcacggcgccaagatg
ctcttcgcctaccccgggccaccgtcccgaagataaccgtcatcatcgcaaggcgtagggcgcgcccttctcgccatgtg
cggcaaggagttggagaccgatcgggttttcgctggcccagcgccgagatcgcggtcatggaccgcaggagcggtcaa
cgctcatctccggaacgagatcgcccaggcggaagatcccaagaaaagcgcgacgagctgatcgttcttaccagggaacct
tcgccactccctatcgggcgcgggcacggcgatgtggacgacatcatcgagcccgccgatacagggcgccacctcgccat
gacgctggacatcctgagcaccaagcggaattcaggcccatgaagaagcatggcctcattccgctg

>*G. bemidjiensis* (12S subunit amino acid sequence) (SEQ ID NO:59)

MSIEEKIKALNDKSKLKLGGGRSKIDQQHAQGSLTARERIEALVDKDSFQEIGIF
ARHRCNTNFGMAGKELPAEGVVTGAGSVGGRMVHLASQDFTVAGGSAGEVHSD
KIVQAMLGSLKTGTPFVFMNDSGGARIQEGIDSLAGYGKVFYHNVMLSGVVPQI
SLICGPCAGGAA YSPAL TDFIIQTAKARMFITGPSVIKEATGEEISAEELGGPLSQM
NHSGVAHFVAENDLVALRICKKLLSYLPSNNIEDPPQLESDDVIVPDKTLNSIVPS
EQKKA YDVRNVITRLIDGGDFLEVQPLFAANIVVGFRILGRSVGIVANQPSVLA
GALDINASDKGARFVRFCNAFNIPLVTLVDVPGFLPGVQQEKGGIIRHGAKMLFA
YAAATVPKITVIMRKAYGGAFLAMCGKELETDRVFAWPSAEIAVMGPQGAVNV
IFRNEIAQAEDPKKKRDELIASYQGT FATPYAAAARRDVDDIIEPADTRRHLAMT
LDILSTKREFRPMKKHGLIPL

>*G. bemidjiensis*_ (1.3S subunit nucleotide sequence) (SEQ ID NO:60)

gtgcaactgaccatgaccattgacggaaagaaataccgggtggacgtagaagtcgaggaaggggaagaggtgcgtacggaa
ggggccttccctcccaccgcactatgcaggcgatcccggtgtattcggcgcatccaaccgcgaccccgccgctggccgcgc
cgaccccgccctccagttcgaaaagatctgccgcagtcctcgcgggggtggttttaagatcgtggcgaggtgggtcaa
cacctggagatgaacgacctgctggtcgtcctcagggcgatgaagatggagaccaacatcaccgcgcacatgtccgggaagg
tggaagagattctggttccgtgggcgaagcggtgcagcctggacaggcaattgccgaattgcc

>*G. bemidjiensis*_ (1.3S subunit amino acid sequence) (SEQ ID NO:61)

VQLTMTIDGKKYRVDVEVEEGEEVRTEGAFPTATMQAYPVYSAHPTATPPLAA
PTPASSSEKICRSPIAGVVFKIVAQVGQHLEMNDDL VVLEAMKMETNITAHMSG
KVEKILVSVGEAVQPGQAIAEFA

>*G. bemidjiensis*_ (5S subunit nucleotide sequence) (SEQ ID NO:62)

atggaccgcattatcgacataaccgaactggctctgcgcgacgcgcaccagagccttatcgctacgaggctcgggatagacga
catggttccggtgtgcgaggacctggaccaggcgggctactggtccatcgagtctggggcggggccacatgacgcctgc
atccgctttctcaacgaagatccgtgggtgaggcttaggaccttaaggagctgatgccgaaaaccccgctgcagatgctttgc
gggggcagaacctttgggataccggcattaccaggacgaggtggtggaccggttcgtccagaagagcgccgagaacggcat
cgacgtgttccggtatcttcgatgcgtgaacgatctgaggaacctggagcggtcggtccaggcgggtgaagcagtcgggaaag
cacgcgcaggtcgccatctctataccatcagccccattcacaccacggcgaaattcgtggagcaggcgaagcgcttggtcga
catgggggtgcgactccatctgcatcaaggacatggcggcgctgatcaagccgcacgcgacatacgacctggtgagagggatc
aaagaggcctgcggcgaccggtatccggatacagctgcatgcgcacgccaccagcggcgtgacctggtgagttacatgaag
gcggtggaggcggggcgtggacggcggtgagtcacatgagcctcgggccccggacacaacccgacggaga
gctttgcggagatgctggaaaatacgggctacaccacgcgcacacaccccgccgggtgaacaaggtgaaggagcatttcgc
caaggtgctccccaggtactcagaattcctctccaccatcaccggcgccggagacggagatcttcaggagccagattccaggcg
ggatgctttcaacatggagagccagttgaagcagcagggggctggggaccggatgcgcgacgtgctggaagagataccgc
tggtgagaaaggacacgggatacgtcccgctggttaacccccgaccagccagatcgctgggacccaggcgggtgctgaacgtatt
gatggggcgctacaaggtgctgaccggcgagttcggcgacctgatgctcggtactacggcctcagccgggagaaacggaa
cccgagggtggtggagcaggcgcgcggccacgcgaataaggagccgatagagtccgccccgcagatctattggagccgg
aatggggcaagctgcggggcgggcgctcccttggaggggtgcgacggcagcagaggacgtgctcacctacgccctctt
tccgcaggtggcgccgaagtcttcgccacgaggagtgaaggacccccgaaacctggggcgcgatcccgctaccggagcttcg
gaaaccagcattcccgaagggcaccccggaagatcaccggccccgtcacctacacggtcaccttgagcgggcagccgcac
aaggtgacgggtgcaccctacggccaggaat

>*G. bemidjiensis*_ (5S subunit amino acid sequence) (SEQ ID NO:63)

MDRIIDITELALRDAHQSLIATRLGIDDMVPVCELDQAGYWSIECWGGATYDA
CIRFLNEDPWVRLRTFKELMPKTPLQMLLRGQNLLGYRHYQDEVVDRFVQKSA
ENGIDVFRIFDALNDLRNLSVQAVKQCGKHAQVAISYTIPIHTTAKFVEQAK
RLVDMGCDSICIKDMAALIKPHATYDLVRGIKEACGDRIRIQLHAHATSGVTMVS
YMKAVEAGVDGVDTA VSSMSLGP GHNPTESFAEMLENTGYTTTRIDLGRV NKVK
EHFAKVLPRYSEFLSTITGAETEIFRSQIPGGMLSNMESQLKQQGAGDRMRDVLE
EIPLVRKDTGYVPLVTPTSQIVGTQAVLNVL MGRYKVL TGEFADLMLGY YGLTP
GERNPEVVEQARRHANKEPIECRPADLLEPEWGLRAAALPLEGCDGSDVDLT
YALFPQVAPKFFATRSEGPRNLGRDPVTGASETSIPEGHPGKITGPV TYTVTL SGQ
PHKVTVAPY GQE

>*G. bemidjiensis*_ (12S_C-term subunit nucleotide sequence) (SEQ ID NO:64)

gtggacgaagagatggagcaggaacacgatccggaaatcacgccgaactgctgatggtgatgtccgccgcatagccgcgt
atctgggcaagaccgtgaggataaggcggggccagggttcgtcgacccgaatctgatcaacgcctggggacagtcgagccgcgt
gggtgctgcaggcgtcgcacaaacttgaggaga

>*G. bemidjiensis*_ (12S_C-term subunit amino acid sequence) (SEQ ID NO:65)

VDEEMEQEHDPEITPELLMVMSAAIAAYLGKTVRIKRRARFVDPNLINAWGQSSR
VVLQASHNLRR

[0209] *Desulfobulbus propionicus* DSM 2032 Transcarboxylase

>*D. propionicus* DSM 2032_ (12S subunit nucleotide sequence) (SEQ ID NO:66)

atgagcacaaggaaaaattagagcagctaaagcaaaaaagggccaaagccttgctgggcgcggtcaggataaaatcgac
aagateccactcccaggggcaaatataccgcccgtgagcgtattcaactcctcctcgaccaggcaccttcgaggaatacgaatgctt
tcaagctccatcgctgctacaacttcggcatggaaaaaatcaagttttcggcgacgggtatcgtcaccggatggcaagctggc
cggccggcggtttatatftacgcgcaggacttttcggctcctcgcgggttctcttccggaaccttggtgaaaaaatatgcaaaat
catggatctgggcatgaaaaacggcattccggtcacggttgacgactccggtggcgcccgtatccaggaaggtatcgagg
ccctggcaggatataccgaaatcttaccgcgtaatgttctcgttcgggtgtgttccccagatttcgggtgttttcggacctgcgc
cggtggcgcccgtttactctcctgcctgaccgactcatcatccagggtcaagatccagtcctacatgttctgacaggtcccaagg
tcgttaagactgtgttaaacgaggacgtcaccaccgagcagttgggtggtgcggccatgcataccaccaagtccggcgctcacc
gactatgctgccgagaacgaggacgacgccattcagtcacatcaaggatctgatgagctatttggcgagaacaatctggagaat
cctccggatgccccctgcgacgatccgatcaccgcccgtccgaactgtcaacgacatcattccggacaacccgaatgccgc
ctacgacatgaaaaaggtcatcaccgagacggcagacaacggtatcttcttgaatcaagaagaatttcgctccgaacatcgctc
atcggttttggcgttatggtggcaaggctattggcatcgttgccaaccagccgctctactacgccggtgttctcgacatcgattcct
cgatcaaaaggtgcccgttcatccgcttctgcgactgctcaacattccgatccttacctctgcgacgtccctggcttctgcccg
gcactgcacaggaattcggcggttatccgcaacggcgccaagatgctgtatgcctacgccgaatcgacagtgccaaaggtga
acgattattaccgtaaatcctatggcggcgcctactgcgctatgtcgtccaagcacctgcgaaccgatataactactcctggcc
gaccggtgaaatcgccgttatgggtccaaaggcgcggtcgaagtcctgcacgccaagggcgctaaagcagcagaagatcc
cagagcgttctggccgaaaaagaaaacgagtacaacgagcagttctccaatccatattgtgcccggagcgtggctatatcga
cgatgtcattgaaccggccgaaaccaggtaccgtatcatcaacgcgtttgagtcgatctctggaagcgtgacacgatcccgatg
aagaaacacggcaatatcccgtg

>*D. propionicus* DSM 2032_(12S subunit amino acid sequence) (SEQ ID NO:67)

MSTKEKLEQLKQKRAKALLGGGQDKIDKIHSQGYTARERIQLLLDPGTFEEYD
AFKLHRCYNFGMEKIKFFGDGIVTGYGKLAGRPVYIYAQDFSVLAGSLSGTLAE
KICKIMDLGMKNGIPVIGLNDSSGGARIQEGIEALAGYTEIFTRNVLASGVVPQISG
VFGPCAGGAVYSPALTDFFIIVKIQSYMFLTGPVKVKTVLNEDVTTEQLGGAAM
HTTKSGVTDYAAENEDDAIQYIKDLMSYLPQNNLENPPDAPCDDPITRRSELLND
IIPDNPNAAYDMKKVITETADNGIFFEIKKNFAPNIVIGFARYGGKAIGIVANQPSY
YAGVLDIDSSIKGARFIRFCDCFNIPILTFVDVPGFLPGTAQEFGGVIRNGAKMLY
AYAESTVPKVTIITRKSYYGGAYCAMSSKHLRTDINYSWPTGEIAVMGSKGAVEV
LHAKGAKAAEDPRAFLAEKENEYNEQFSNPYCAAERGYIDDVIEPAETRYRIINA
FESISGKRDTIPMKKHGNIPL

>*D. propionicus* DSM 2032_(12S_c-terminal nucleotide sequence) (SEQ ID NO:68)

atggcaaaaatgaacaaaaaatggctgcggcccttcgagccgttaatgcctacctgatgcaggaagaggaggcgccatacca
ggcccagttgctggctgccaatctgttgaccagccgggccaagcttatgggcaattgccggccgtcaggatatcatgaatttc
cgcaggctgattcaaatgaaagccttc

>*D. propionicus* DSM 2032_(12S_c-terminal amino acid sequence) (SEQ ID NO:69)

MAKMNKMAAALAAVNAYLMQEEEAAYQAQLLAAKSVAPAGPSLWAIAGRQ
DMMNFRRLIQMKAF

>*D. propionicus* DSM 2032_(5S_1.3S_fusion nucleotide sequence) (SEQ ID NO:70)

atgagcgaccaagtgaaaatgaccgcatgaattatgcaactgaccggcctgctgcagaaaatccgggtcaaaagttaggacttg
agccttcgtgacggccaccagctctgttcgccaccgcggcgccaccgaggacatgattccgatcgcgaaatgatggacga
gatcggcttctgggcagttgagacctggggtggcgccacctttgacaccatgcaccgcttctcaacgaggaccctggggagc
gtctccgcacctgaaacgttacatcaagaagaccccccttctccatgttgctgcgcgcgcagaacctggttgataccgtaactat
gccgatgacttggccaccgctttgttgagcgcgctgccgagaacgggtatggatatcttccggacctttgacgccctcaacgatta

- 64 -

ccgtaacttcgagaccgttgtaaacagatcaagaagagcggcaagcacttcagggttgatttgcattcgtgaccgaaccg
 cgtctggcggggatgttatgacctgaagtactatgtcgaccgcgcaaagcgttgacgacatggcgctgactccatctgc
 atcaaggacatggccggtctgatcgcccatagcgcctacgccatcgtaaggctatcaaggaaagtcaccaagaccccgat
 ccactgcacagccacttcacctctggtatggcgctcatgagtcactctgaaggccattgaggcgtggcgtagatatcgttgacacct
 gcatgaccccgtagctttccgtaccgcccacccgcccacgcagccgttggtcatggccctgctcggcaccacccgcgacacc
 ggtttcgacatcaagaaactggccgcatcaacgaggtgctagagaaagaggttatgccgaaatacaagcacctcatggatga
 ctccaagtgtcaatcatcgatatcaacgttcttccatcagacccggcgccatgctctccaacctgggtcaaccagttgcgtg
 agatggatgctctggacaagatcgatcaggtctacaaagagctgccgaaagttcggaaagacctcggccagattccgctggtta
 ccccgaccagccagatcgttggcatccagaccgtgaacaacgtgctgtttgacactcctgatgagcgtacaagatgatcaccg
 ccaggtcaaaagacctgtgctacggtctctatggtaaaaccgctgtgcccgatcaacctgaactgcagaagaaggctctgaaag
 gctatccgcgcggtgaagagccgatcacctgccgtccggcagaggtgcttgagcccagttggaaaaggccaagaagagat
 tggcgatctcgcaaggatatcgatgacttggactctacgccatctaccggtcaccgggaagaagttccttgagtggaaagtat
 ggcatatccccggcaccgcccgaagtaagccgctcacccttgaggatgtcaagaagcgtgatgaactgggtggccaaggcca
 aggttggaagctcatcgagcccaagcccgtgctccggagaagaccgctaacttcggaccttcaactgtcttcgtcgacggt
 gagtattcaacgttgaggtcgacccgaccggtgacttccagccgatggctgccgctgctccgcggcctgccgcacctgccgct
 gcaccgaaagctgctgcacctgccgctgctgcacctgctgccgcgccaaggctgctgcacctgccgcgcccgtccggctc
 cagccgctgttgagggaggaaccccgctgttgcccccatgcccgcatgatcgtaagaatctggtcaatgttggtgatcgcg
 tcaaagctggcgaccccatcctcgttcttgaggccatgaagatggagaacaatctcggttctccgtgcatggtactgtgaaggc
 gcttaattttggcagcggtgactcggttgccaaggataccgctcctggcaatcatcgga

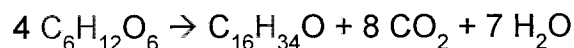
>*D. propionicus* DSM 2032_(5S_1.3S_fusion amino acid sequence) (SEQ ID NO:71)

MSDQVKMTAMNYATDRPAAENPVKVMDLSLRDGHQSLFATRGRTEDMIPIAE
 MMDEIGFWAVETWGGATFDTMHRFLNEDPWRLRTLKRYIKKTPFSMLLRAQN
 LVGYRNYADDLATAFVERAAENGMDIFRTFDALNDYRNFEVTVVKQIKKSGKHF
 QGCICYSLTEPRLGGDVYDLKYYVDRAKALDDMGADSICIKDMAGLIAPYDAY
 AIVKAIKEVTKTPIHLHSHFTSGMASMSHLKAIEAGVDIVDTCMTPYAFRTAHPAI
 EPLVMALLGTNRDTGFDIKKLAINEVLEKEVMPKYKHLMDDSKCSIIDINVLLH
 QTPGGMLSNLVNQLREMDALDKIDQVYKELPKVRKDLGQIPLVTPTSQIVGIQTV
 NNVLFDTPDERYKMITAQVKDLCYGLYGKTAVPINPELQKKALKGYPRGEEPIT
 CRPAEVLEPELEKAKKEIGDLAKDIDDLVLYAIYPVTGKKFLEWKYGITPAPPEV
 KPLTLEDVKKRDELVAKAKAGKLIKPAAPEKTANVRTFNVFVDGEYFNVEVD
 PTGDFQPMVAAAPRPAAPAAAPKAAAPAAAAPAAAPKAAAPAAAAPAPA AVE
 GGTPLLAPMPGMIVKNLVNVGDAVKAGDPILVLEAMKMENNLGSPCDGTVKAL
 NFGSGDSVAKDTVLAIIIG

Engineered Pathways to Produce Hydrocarbons and Other Malonyl-CoA Derived Products

[0210] Production of a bio-product at high yield requires a balanced chemical equation describing the conversion of substrate to product and a thermodynamically feasible reaction with a negative change in Gibbs free energy. Long chain hydrocarbons, *e.g.*, those that have carbon backbones of at least four carbons and up, derived from fatty acids satisfy both of these requirements. For example, production of a C₁₆ fatty alcohol can be described by the following equation:

- 65 -



[0211] Production of a C₁₆ fatty alcohol results in a Gibbs free energy change of -285 kJ/mol glucose. For comparison, production of ethanol results in a Gibbs free energy change of -208 kJ/mol glucose.

[0212] The present invention describes the engineering of a recombinant microorganism to convert a native fatty acid biosynthetic pathway into a fermentative pathway, *i.e.*, one that generates net positive ATP and is redox neutral. As shown below, a native fatty acid pathway generates zero net ATP, which stems from the mechanism of producing malonyl-CoA, the acyl-ACP chain precursor used to increase chain length. Malonyl-CoA is formed from the conversion of one glucose into two acetyl-CoA, which produces two ATP and four NAD(P)H. However, ATP is required to produce malonyl-CoA from acetyl-CoA, which results in a net zero ATP balance. In the synthetic route shown below, malonyl-CoA formation is accomplished without the concomitant use of ATP.

Native Pathway: $\text{Glucose} + \text{CoA} \rightarrow 2 \text{ Malonyl-CoA} + 2 \text{ NADH} + 2 \text{ NAD(P)H}$

Synthetic Pathway: $\text{Glucose} + \text{CoA} \rightarrow 2 \text{ Malonyl-CoA} + 2 \text{ ATP} + 2 \text{ NADH} + 2 \text{ NAD(P)H}$

[0213] In either case, the NAD(P)H produced during malonyl-CoA synthesis is balanced via reduction of the growing acyl-ACP chain.

[0214] The synthetic pathways described herein proceed according to three steps: chain initiation, chain extension, and chain termination (*see* Figure 2) and can be carried out in aerobic or anaerobic conditions. In some embodiments, the synthetic pathways produce a hydrocarbon and/or a hydrocarbon derivative under anaerobic conditions. In some embodiments, the synthetic pathways produce a polyketide or an organic acid under aerobic or anaerobic conditions. Chain initiation can proceed by one of several options that are ATP positive and in which NAD(P)H is balanced by chain termination and H₂ generation (*see* Figure 3A-3C).

[0215] In native cells, *e.g.*, *E. coli*, chain extension proceeds from pyruvate to acetyl-CoA to malonyl-CoA. *See* Steen *et al.*, *Nature* 463:559-562 (2010). To conserve ATP during the generation of malonyl-CoA, two enzymes are introduced into the central metabolic network for chain extension: a phosphoenolpyruvate carboxykinase (PEPCK) to convert

phosphoenolpyruvate to oxaloacetate and a transcarboxylase (TC) to convert oxaloacetate and acetyl-CoA to malonyl-CoA and pyruvate (*see* Figure 4 or Figure 33). The introduction of these enzymes results in the production of 2 NADH, 2 NAD(P)H, and 2 ATP, resulting in a net production of ATP per carbohydrate, such as but not limited to, glucose. For example, for glucose, the net production of ATP per 6 carbons is about 2. For xylose, the net production of ATP per 5 carbons is about 1.67 ATP. When considering the net production of ATP per hydrocarbon produced rather than sugar consumed, for every 4 carbons of hydrocarbon, the net is about 2 ATP. Thus, for a C₁₆ fatty acid, the net ATP is about 8. The conversion of phosphoenolpyruvate to oxaloacetate using PEPCK results in the net production of ATP. *See* Figure 4. For example, the net production of ATP in the recombinant microorganisms of the invention includes at least about 0.5 net ATP; at least about 1.0 net ATP; at least about 1.5 net ATP; or at least about 2.0 net ATP during anaerobic growth. The conversion of oxaloacetate and acetyl-CoA to malonyl-CoA and pyruvate by TC then requires the regeneration of acetyl-CoA from the TC-generated pyruvate. The recycling of pyruvate by conversion of pyruvate and CoA-SH into acetyl-CoA and CO₂ and NAD(P)H not only facilitates flux in the direction of producing malonyl-CoA, but also generates the reduced NAD(P)H needed to balance redox. Enzymes that can be used to catalyze this pyruvate recycling pathway include, but are not limited to, a pyruvate dehydrogenase, a pyruvate:ferredoxin oxidoreductase and ferredoxin:NAD(P)H oxidoreductase, or a pyruvate formate lyase and formate dehydrogenase.

[0216] In addition, competing metabolic pathways can be removed or attenuated. These include, but are not limited to, pyruvate kinase, hydrogenase, lactate dehydrogenase, phosphotransacetylase, acetate kinase, acetaldehyde dehydrogenase, alcohol (ethanol) dehydrogenase, pyruvate formate lyase, pyruvate decarboxylase, and native enzymes involved in the degradation of fatty acids and their derivatives.

[0217] PEPCK and TC can be derived from *C. thermocellum* and *T. saccharolyticum* or other organisms. Engineering of these enzymes into the recombinant microorganism of the invention may require alteration of substrate specificity to minimize undesirable side reactions. In addition, cofactor specificity in the overall metabolic pathway can be modified, which has been done with other, similar proteins. To increase flux to malonyl-

CoA production, native pathways for organic acid and ethanol production can be modified. Each of these engineering steps is within the abilities of those skilled in the art.

[0218] The acyl-ACP chain can be extended through the fatty acid biosynthesis (Fab) enzymes present in all organisms that produce fatty acids. These include FabB, FabF, FabG, FabZ, and FabI. Overexpression of these enzymes can benefit hydrocarbon formation; however, the native biosynthetic pathway is largely regulated by the availability of the malonyl-CoA precursor and the accumulation of long-chain fatty acyl-ACP compounds. See Li *et al.*, *Journal of Bacteriology* 175:332-340 (1993); Davis *et al.*, *Journal of Biological Chemistry* 275:28593-28598 (2000); Davis and Cronan, *Journal of Bacteriology* 183; Heath and Rock, *Journal of Biological Chemistry* 271:1833-1836 (1996)). Supply of sufficient precursor and removal of fatty acyl-ACP via chain termination steps allows for sufficient flux through this chain extension pathway.

[0219] Once an acyl-ACP chain has reached its desired length, the reaction is terminated and the hydrocarbon product is excreted from the cell. Many chain termination options are available in the art to produce hydrocarbon products or hydrocarbon derivative products, including, but not limited to, fatty acids, alcohols, aldehydes, wax esters, or alkanes (see Figure 5A and 5B). See Steen *et al.*, *Nature* 463:559-562 (2010); Sukovich, *et al.*, *Applied and Environmental Microbiology* 76:3850-62 (2010); Kalscheuer and Steinbüchel, *Journal Biological Chemistry* 278:8075-82 (2003); Reiser and Somerville, *Journal of Bacteriology* 179:2969-2975 (1997); Kalscheuer *et al.*, *Microbiology* 152:2529-36 (2006); Beller *et al.*, *Applied and Environmental Microbiology* 76:1212-23 (2010). The termination steps, in concert with chain extension, impart properties on the final compound to mimic petroleum based diesel, gasoline, or jet fuel. For example, production of C₁₄₋₁₈ fatty alcohols and esters as first generation products can be directly blended to create cellulosic diesel, or serve as a bio-crude that could be converted into other fuels with conventional catalysis technology. Production of fatty alcohols requires expression of a fatty acyl reductase and a fatty aldehyde reductase. See Reiser and Somerville, *Journal of Bacteriology* 179:2969-75 (1997); Steen *et al.*, *Nature* 463:559-562 (2010). Some organisms, such as *E. coli*, have native fatty aldehyde reductase activity, while enzymes such as the jojoba acyl reductase is a bifunctional acyl-ACP/aldehyde reductase. See Reiser and Somerville, *Journal of Bacteriology* 179:2969-75 (1997). Wax esters can be produced via an acyltransferase in the presence of ethanol

or a long-chain alcohol. See Kalscheuer and Steinbüchel, *Journal Biological Chemistry* 278:8075-82 (2003); Reiser and Somerville, *Journal of Bacteriology* 179:2969-2975 (1997); Kalscheuer *et al.*, *Microbiology* 152:2529-36 (2006).

[0220] The chain length of the hydrocarbon product or hydrocarbon derivative product is controlled based on, *e.g.*, the specificity of the native organism. See Wang *et al.*, *Extremophiles* 10:347-56 (2006); van Beilen *et al.*, *Microbiology* 147:1621-30 (2001). Based on techniques known in the art, termination enzymes can be screened and engineered to develop hydrocarbon products or hydrocarbon derivative products with the desired chain length. See Steen *et al.*, *Nature* 463:559-562 (2010); Sukovich, *et al.*, *Applied and Environmental Microbiology* 76:3850-62 (2010); Kalscheuer and Steinbüchel, *Journal Biological Chemistry* 278:8075-82 (2003); Reiser and Somerville, *Journal of Bacteriology* 179:2969-2975 (1997); Kalscheuer *et al.*, *Microbiology* 152:2529-36 (2006); Beller *et al.*, *Applied and Environmental Microbiology* 76:1212-23 (2010).

[0221] Hydrocarbon products or hydrocarbon derivative products can exit the cell through a membrane "flip" mechanism. In such a mechanism, the polar hydrophilic-hydrophobic compound enters the lipid bi-layer on the intracellular side with the hydrophilic head pointing towards the inside of the cell, flips over so that the hydrophilic head points outside of the cell, and then exits the bi-layer into the extracellular environment. See Black and DiRusso, *Microbiology and Molecular Biology Reviews* 67:454-472 (2003). Alternatively, to ensure efflux from the recombinant microorganism, high efficiency hydrophobic compound efflux transporters can be engineered, although at a cost of one ATP per molecule extruded. See Kieboom *et al.*, *Journal of Biological Chemistry* 273:85-91 (1998). Such mechanisms allow for collection of the hydrocarbon products or hydrocarbon derivative products in the fermentation medium, in addition to other products naturally secreted or expelled by the host cell.

[0222] As hydrocarbon products or hydrocarbon derivative products accumulate in the fermentation media, the products can form a 2-phase organic layer after saturating the aqueous fermentation volume. See Neumann *et al.*, *Applied and Environmental Microbiology* 71:6606-612 (2005). At saturating concentrations, toxicity correlates to the "minimum membrane concentration" of a compound, which is a function of the

octanol/water partition coefficient and the aqueous solubility. Generally, as chain length increases, compounds become less toxic.

[0223] Product recovery and product toxicity are independent of substrate concentration. This provides the advantages that either a minimal pretreatment can be run at low fermentor solids or, when using refined material, the refined material can be run at very high solids without product toxicity to the fermenting organisms. In addition, because the hydrocarbon products are insoluble, product recovery can be at low cost. This means that the hydrocarbon products can be readily purified for use in fuels and chemical feedstocks.

EXEMPLIFICATION

[0224] The invention now being generally described, it will be more readily understood by reference to the following examples, which are included merely for purposes of illustration of certain aspects and embodiments of the present invention, and are not intended to limit the invention.

Example 1

[0225] The present prophetic example describes the engineering of a recombinant microorganism to convert a native fatty acid biosynthetic pathway into a fermentative pathway, *i.e.*, one that generates net positive ATP and is redox neutral during anaerobic growth.

[0226] 1.1 Production of Hexadecanol in *T. saccharolyticum*

[0227] Gene overexpression and gene deletion followed by evolutionary engineering will be performed to create a strain producing 1-hexadecanol.

[0228] The strain *T. saccharolyticum* JW/SL-YS485 has an established transformation system based on a natural competence protocol. See Shaw *et al.*, *Applied and Environmental Microbiology* 76:4713-4719 (2010). Recombinant DNA, either linear or plasmid based, can be introduced with the following protocol.

[0229] 1.1.1 *T. saccharolyticum* Transformation Protocol

[0230] Prior to use, petri dishes, 50 mL and 15 mL conical falcon tubes, and pipet tips are all placed in the anaerobic chamber at least overnight. Transformations are performed in an anaerobic chamber by inoculation of 10 mL liquid medium M122 (pH 6.1 or 6.7 — there is less precipitation at pH 6.1 and it facilitates OD measurement, but kanamycin selection is better at pH 6.7) with 1-3 μ L of a frozen working stock culture of *T. saccharolyticum*, which has been frozen-down when in exponential growth. After mixing, 1 mL aliquots of the 10 mL medium are transferred to tubes containing between 0.25 μ g – 1 μ g DNA. The tubes are then incubated at 55°C for 16-18 hours (overnight) to an OD of 0.6-1. Maintaining cells past 18 hours in stationary phase can dramatically reduce transformation efficiencies.

[0231] Next, 100 μ L and 500 μ L aliquots of the transformant culture are mixed with 25 mL liquid medium M122 pH 6.7 at 55°C containing 1.2% agar and kanamycin at 200 μ g/mL. The mixture is poured into petri dishes and allowed to solidify at room temperature for 30 minutes, or until completely solid, and the petri dishes are incubated at 55°C in a moisture retaining container until colony formation (24-48 hours).

[0232] 1.1.2 Gene Deletion

[0233] Gene deletions will be performed with a marker removal system, which allows for clean genomic deletions and marker recycling. The plasmid pMU433 (*see* Figure 8) contains the marker removal cassette, namely the *pta* and *ack* genes linked to a kanamycin resistance gene. Genes targeted for deletion include L-lactate dehydrogenase, bifunctional alcohol dehydrogenase *adhE*, pyruvate kinase, pyruvate formate lyase, and phosphotransacetylase and acetate kinase. To construct deletion vectors, homology regions are generated from the target gene sequences from *T. saccharolyticum* and cloned into pMU433.

[0234] L-lactate dehydrogenase (SEQ ID NO:72)

```
atgagcaaggtagcaataataggatctggttttaggtgcaacatcggcatttacgctggcattaagtgggactgtgacagatat
cgtgctggttgattaaacaaggacaaggctataggcgatgcactggacataagccatggcataaccgctaatacagcctgtaaat
gtgtatgcaggtgactacaaagatgtgaaaggcgcagatgtaatagttgtgacagcaggtgctgctcaaaagccgggagagac
acggcttgacctgtaaagaaaaatacagccatatttaagtcacatgatacctgagcttttaaagtacaatgacaaggccatatttg
attgtgacaaatcccgtagatatactgacgtacgttacatacaagatttctggacttccatggggcagagttttggttctggcaccg
```

ttcttgacagctcaagggttagataccttttaagcaagcactgcaatatagatccgagaaatgtccacggaaggataatcggcgag
catggtgacacagagtttgagcatggagcataacaaacatatcggttatatcatttaagtactgcagcatatcgggacgcgt
ctgcaacacaaatttcagaaaggaagtagaagaagaagtcgtaaagtctgttacaagataatagacaaaaagggtctacata
ctatgctgtggcagttgcagtaagaaggattgtggagtgcatcttaagagatgaaaattccatctcacagtatcatctccattaat
ggacagtacggcgtaagagatgtttcattaagcgtgccatctatcgtaggcaggaatggcggtgccaggattttgacttgcccttat
ctgacgaagaagtgagaagtttaggcattcagcaagtgtcatggcagatgtcataaaacaattagatata

[0235] Bifunctional alcohol dehydrogenase *adhE* (SEQ ID NO:73)

atggcaacgacaaaaacggaattagacgttcagaagcagatagatctacttgtgtcaagagcacaagaggctcagaaaaattc
atgtcttacacgcaagagcaaatcgacgcaatgtaaggcaatggctttagcaggcgttgacaaacacgtagagctggcaaag
atggcgtagcaagagacaaaaatgggtgtatcgaagataagataacaaaaatctcttcgcaacagagtacgtgtaccacgac
ataaaaaatgaaagactgtaggaatcataaacgagaacatagaagaaaactacatggaagtggcagaaccgtaggcgtaat
tgccggtgtcacacctgtcacaaaccaacatctaccacgatgtttaatgcttaatatccataaagcgcgaaatcctataatattc
agcttccatccaaaggcaataaagtgcagcatcgagcagccaaagtgtatgtatgaagctgactaaaggcaggcgacactga
aggatgcataggtgtagataaaacgccatcaattgaggccacacagcttctcatgacacatccaggcgtagctgctgatccttgca
acggcggtgcaggaatggtaaaagcggcatacagctcaggaaaaccggcattaggcgtaggtcctggcaatgtgccatgct
acatcgaaaaatcagcaaacataaagagggtgtatcggatctcatactaaagcaagacattgacaatggagtaatatgcgcatc
agagcaggccgtaataatagacgaggaaatagcagatgaagtcaaaaagcttatgaaagaatacggctgctacttcttaacaa
agatgaaataaagaagcttgagaaattgcaattgatgagcaaaagctgcgccatgagccctgcagtggtaggtcagccagcgg
cgaagattgctgaaatggcaggttcaaagtccccgaaggcacaagatattagtggcagagtacgaaggagtaggtccaaaa
tatcctctatcaaggagaaactaagccgattctgtgtgtacacgtcaaagactacaatgaaggaaatcaaaaagtgcgagg
aaatgactgaattcggaggttaggccactctgtgttaatacactctgaaaatcaaaacgtcataaatgaattgcaaggcgagtc
gcacaggaagacttatcgtaaatccatcatcacaggggagcaataggagatatatacaatacaaacacgccatcacttacatta
ggctgtgtgtctatgggaagaaactcaacgacagacaatgtaagcgtcaagaacctttgaaatattaagcgtgtcgtgataaggaa
tgatagaatgaaatggttcaagattccaccgaagatttactttgaaagcgggtcactccagtagctgtgcaaaagtcaaaaagaaaa
aagcgtttatcgacagatccattcatgttaagcttggttcgtagacaaaagtacatatcaattagacaaagcaaacatcgat
acgaaatattctcagaagtagagccagatccatctgttgacacagtcataaacggcgtaaaaataatgaattcgtacaatcctgac
ttaataatcgctgtaggcggtggtctgtcaatagacgcagcaaaagggaatgtggcttttctacgaatatcctgatacagagtttgaa
acattgaggcttaaatgtcagacatcagaaaaaggcatttaagtcccagaacttggcaaaaagcgctattcatcgcaatacc
gacaacaagcggcacaggctcagaagtacagcatttccgtaataaccgacaaaaagagaaacatcaagtatccactggca
gactacgaacttacactgacatagccataatagatcctgacattacaagactgtaccgccatctgtaacagcagacacaggca
tggatgtgtgacacacgccatagaagcatacgtatcagtaattggcatcagactacacagatgcactggcggaagggctataa
agatcgtattgaaatcctgccaagggctataaaaacggcaatgatgaagaagccgcgaaaagatgcacaatgcttctgcat
ggctgtgtatggcattcacaatgcattcttaggaataaaccacagcatggcacacatactggcggaaggttccacataccacac
ggaagagcaaatgcaacttctgccgtatgtaataaggtacaatgcagaaaaacctacaagttgtggcattccacaatacga
atatcaaaaagcagcagaaagatatgcggaaatcgccaaattcttaggactgcctgcttcaactgttgaaagaaggcgtagaag
cttaatagaagctataaagaacctcatgaaagagcttaacattccgcttacactaaagacgcggcatcaaaaagaacagttg
aaaaagaaatagaggaaatgtcagacatcgcttcaacgatcagtcacagggacaaaaccgagaatgcctctcacaaaagaa
attgcagagatctacagaaaagcatacgggtgca

[0236] Pyruvate kinase (SEQ ID NO:74)

atgcgtagaactaagataatatgcacgattggtcctgccagtgaataatagatatattgaaagagcttatagaaagcggcttaat
atttgagggtgaattttcacatggggatcatgaagagcatggaagcagaatagacaatattataaagattagagaagaacttaag
ctgcctattgcaattatgcttgatacaaaaggcctgaataaggactggcagatttaaaaggcgggtgtgcagagcttaaaag
gccagacatttacgataacatcaagggaattgaaggagataacactatttctgtttcatacaaggggtcctcctcaagatgtgg
agagagggtctgcataattgattgatgacggattagtagcattgaaagtcaatgacgtaaaagggtgaagatatagtagtgcactgtgg

agaattctggtacaataggtgatcacaaaggtgtaaatgtacctggtacaaagcttaattgcctgccataacgcaaaaagacgtg
gatgatataagatttgaataaaaaaggaatcgacatgattgcagcgctctttgtcagaaaagcagcagatgtaattgccataag
gagattgtagaagacaatgacgctggccatatacttatcatatcaaaaattgaaaatcggaaggcgtagaaaatattgacgaaa
taatcaaaagtctctgatggcataatgtagcccgcggcgatttgggtgtcgaaattcctatagaggaaatacctatcgttcagaaaa
ggataattgaaaaatgcaacaaagcaggtaaaccagtagttactgctacacagatgcttgactctatgataagaaatccaaggcc
aacaagggcagaagtaacagatgtagccaatgctatatgttgatggcactgatgcgataatgttctgttggtgaaacagcgcaagg
caaatatcctgtagaggctttaagacgatgtcaaagatagctgaaaagattgagacgtatataaattacaaagaaaatttagataa
aatgtggtattacaatatttctatgacaaatgccataagccatgctacgtgcactaccgcgagagatataggcgcaactgccattat
tacatctacaatatcaggttatactgcgagaatgggtgtctaagtatagaccgtcagcacctataatagcagtgacgccaacaaag
atgttgcaagaaggcttagcatcgtgtgggtgtacatccattgatatacaggaagtcaattctacagatgaaatgatagaagtat
cagtaaataccggctttaaatgaaggattaattcgaaatggcgatattgtagtaatatcggcaggaatacctgtcgcgactacaggc
acaacaaatatgtgaaggttcatattgtgggagatgtaatagtaaaaggcacaggcataggcactaaatccataagtggtgtgtt
tccatcataagagatccatacaaggacaaagataagttcagagaaggagatatcatcgttgcctcaaaaaactgaaagggtattat
gcctataattgagaaggcttcagctatcataacagaagagggtgactaacgtcccattgctgcaatagttgattgaactatggatt
acctgtcattgtaggctgtgaaggagtaacttcaaagcttaaagatggatgacggtaactctcgatactgccagaggattggct
acaaagggtatagtgaatataaaatag

[0237] Pyruvate formate lyase (SEQ ID NO:75)

atgatcaatgaatggcgcggtttcaggagggcaaatggcaaaagactattgacgttcaagattttatccagaaaaattacacatt
atacgaaggcgatgatagtttttagaaggccctacagaaaagactattaagctttggaacaaagttcttgagctaataaggaag
aactgaaaaaagggtgtgttagatattgatacaaaaactgtatcgtctataacatcccatgatgcggggtatatagacaaagattg
aggaaatagttgattgcagacagacaaaacctcttaaaagagctataatgccttacgggtggcataagaatggtaaaaaagcttg
cgaagcttatggatataaagtggacccaaaagtagaagagatattfacgaagtacagaaaagaccacaatgatgggtgtatttga
gcataactccagaaataagagcagcaagacatgccggcataataacagggtcttccagatgcataaggcagaggaagaatcata
gggtgattacagaagagttgctctttatggaattgatagactcatcgaagaaaaggaaaaagaaaaacttgagcttgattacgatga
atttgatgaagcaactattcgttgagagaagaattgacagaacagataaaagcattaaacgaaatgaagagatggccttaagt
acggttatgacatatcaaaagcctgcaaaaaatgcaaaagaagctgtgcagtggaacttacttgccttctgctgtataaaggaa
aaaatgggtgccgctatgtcgtgggcagagtatctactttttagatatatacattgaaagagatcttaagaaggaaacattgacag
agaaacaagcacaagagttaatggatcactttgtcatgaagcttagaatgggtgaggttctaaggactcctgattacaatgaactatt
tagtggcgatcctgttgggtgactgaatcaattggcggttaggcgtagacggaagaccttctgactaaaaattcattcaggat
attaaatactttatataacttaggtcctgcacctgagccaaacttgacggttttatgggtccaaaaaccttctgaaaactttaaaagatt
ctgtgccaaaggtatcaatagatacaagttctattcaatatgaaaatgacgacttaattgaggccaatatacaatgacgactatagcat
cgctgtctgtgtcagctatgaagacgggagaaacagatgcaatttttggagcaagggaatctcgcgaaggcgctactgtat
gctataaacggcggtatcgtgaaaggtataaaacgcaagtgccacaaaatttaactctataacgtctgagtatttagactacgat
gaggtaatggcagcatatgacaatatgttagagtggttgcaaaagtgtatgttaaagctatgaatataactacatgcacgat
aaatacgttatgaaagatcccttatggcttgcagtagagacatcgttaaggacgatggcttttggaatcgaggtcttctgttg
cggcagattcgtaagcgccataaagtatgctaaagtaaaagccataagagatgaaaatggcatagcaatagattatgaagtga
aggagatttccctaagtttggcaatgatgatgacaggggtgactcaatagcagttgacattgtagaaagattcatgaataagcttaa
aaagcacaagacttacagaaactctataccaactgtctgttttgacaataacgtcaaatgtggtgtacggcaaaaagacgggt
gctacacctgacggaagaaaagcgggagaaacttttgcgccaggcgcaaatccgatgcacggcagagatacaaaagggtgcc
atagcatcaatgaattcagtatcaaaaataccttatgacagttcattggatggtatatcatcacatttacgattgtacaaatgcgctt
ggcaaggatgacgaagataaaattaataatctgttaggactattagatggatgcatttaatgcggggcaccacataaacatcaa
tgttttaaacagagatatgttgcctgatgctatggagcatcctgaaaaatatccgcagcttactataagggtttcagggtatgctgtca
atttcaataaattaacgagagagcaacagttggagggtatatcccgcacttttcacgaatctatg

[0238] Phosphotransacetylase and acetate kinase (SEQ ID NO:76)

gtgtatacaatatatttcttcttagtaagaggaatgtataaaaataaatattttaaggaaggacgatcttatgagcattattcaaaac
 atcattgaaaaagctaaaaagcgataaaaagaaaattgttctgccagaaggtgcagaacccaggacattaaaagctgctgaaata
 gttttaaagaagggttgcagatttagtcttcttgaaatgaagatgagataagaaatgctgcaaaagacttggacatatccaaa
 gctgaaatcattgaccctgtaaagtctgaaatgtttgataggtatgctaattatttctatgagttaaggaagaacaaggaatcacgt
 tggaaaaagccagagaaacaatcaaggataatactatttggatgtatgatggttaaagaagggttatgctgatggattggtatctg
 gcgctattcatgctactgcagatttattaagacctgcatttcagataaattaaaacggctccaggagcaaagatagtatcaagcttttt
 ataatggaagtgcctaattgtgaatatggtgaaaatggtgtattcttgttgcgtgattgtgcggtcaacccatcgccctaatgcagaag
 aacttgccttattgccgtacaatctgctaatactgcaaagaatttgttgggcttgaaccaaagttgccatgctatcattttctacaa
 aaggtatgcatcacatgaattagtagataaagtaagaaaagcgacagagatagcaaaagaattgatgccagatgttgcatacga
 cggatgaattgcaattggatgctgctcttgttaaagaagttgcagagctaaaagcgcgggaagcaaagttgcgggatgtgcaaat
 gtgcttatattcccgatttacaagctggaatataggaataagcctgtacagaggttagctaaggcaaatgcaattggacctataa
 cacaaggaatgggtgcaccgggttaattgatttatcaagaggatgcagctatagagatattgttgacgtaataagcaacaacagctgtg
 caggctcaataaatgtaaagtatggaggatgaaaattatgaaaatactggttattaattgcgggaagtcttcgctaaaatatcaact
 gattgaatcaactgatggaaatgtgttgcaaaaggccttgcgtgaaagaatcggcataaatgattccatgttgacacataatgctaa
 cggagaaaaaatcaagataaaaaaagacatgaagatcacaaaagacgcaataaaattggttttagatgctttggttaaacagtgc
 tacggcggtataaaagatatgtctgagatagatgctgttaggacatagagtgttcacggaggagaatctttacatcatcagttctca
 taaatgatgaagtgttaaagcgataacagattgcatagaattagctccactgcacaatcctgctaataatagaaggaattaaagctt
 gccagcaaatcatgccaacgttccaatggtggcggtatttgatacagcctttcatcagacaatgcctgattatgcatactttatcc
 aataccttatgaatactacacaaagtacaggattagaagatatggatttcatggcacatcgcataaatatgtttcaaatagggtgcga
 gagattttgaataaacctatigaagattgaaaatcataacttgcattcttggaaatggctccagcattgctgctgtcaaatatggtaa
 atcaattgacacaagcatgggatttacaccattagaaggtttggctatgggtacacgatctggaagcatagacccatccatcatttc
 gtatcttatggaaaaagaaaatataagcgctgaagaagtagtaaatatattaaataaaaaaatctggtgtttacgggtatttcaggaata
 agcagcgatttttagagacttagaagatgccgcctttaaaaatggagatgaaagagctcagttggctttaaattgtgttgcatatcga
 gtaaagaagacgattggcgcttatgcagcagctatgggaggcgctgatgtcattgtatttacagcaggtgttggtgaaatgggtcc
 tgagatacgagaatttatacttgatggattagagtttttaggggttcagcttgataaagaaaaataaagtcagaggaaaagaaac
 tattatatctacgccgaattcaaaagttagcgtgatggtgtgcctactaatgaagaatacatgattgctaagataactgaaaagatt
 gtaaagagtataaaa

[0239] For knockout vector construction, the 0.8-1.2 kb flanking regions (with primers) on both sides of target are first identified. Once identified, the new flanking regions are used to replace the L-ldh flanking regions in pMU433 using *in silico* analysis. Yeast-mediated ligation primers (4 total) for the two new flanking regions are made by adding to the targeting primers 5' regions homologous to DNA segments labeled "X01648," "X01649," "X01654," and "X01655" on pMU433 shown in Table 3. Total primer length should be about 55-65 bp.

[0240] Next, the flanking regions from *T. saccharolyticum* YS485 genomic DNA are PCR amplified. PCR cleanup is not necessary if correct product was highly amplified.

X01648	(SEQ ID NO:77)	<u>GTCTTTCGACTGAGCCTTTCGTTTTATTG</u> <u>ATGCCTGG</u>	pMU433 construction
X01649	(SEQ ID NO:78)	<u>AATTGTAGAATACAATCCACTTCACAATG</u> <u>GGCACG</u>	pMU433 construction
X01654	(SEQ ID NO:79)	<u>AGGGGTCCCGAGCGCCTACGAGGAATT</u> <u>GTATCG</u>	pMU433 construction
X01655	(SEQ ID NO:80)	<u>CCGTCAGTAGCTGAACAGGAGGGACAGC</u> <u>TGATAGA</u>	pMU433 construction

Table 3. Primers for Knockout Vector Construction

[0241] About 100-200 ng pMU433 per yeast transformation is then digested with BamH1/BspE1. Allowing digestion to proceed to completion helps reduce background during yeast transformation.

[0242] The digested DNA is transformed into *ura3-* *S. cerevisiae* (Invitrogen INVSc1 cat#C81000 or equivalent) following the "Lazy bones" yeast transformation protocol. *See* Shanks *et al.*, *Applied and Environmental Microbiology* 72:5027-5036 (2006). Briefly, about 100 ng digested plasmid and 10-50 μ L of each PCR amplified flanking region are mixed. Prior purification is not necessary for either plasmid or PCR unless there are BamH1/BspE1 sites in the flanking regions. Other yeast transformation protocols can suitably be used. To control for background, a plasmid only control can be used.

[0243] The transformed yeast are plated on SD-URA plates (SD Medium-URA MP Biomedicals #4812-075 or equivalent) and incubated at 30 °C. The plates are incubated for 3-5 days and then yeast total DNA is harvested from plates containing colonies. If cell mass is low, the colonies can be streaked on a new plate to increase the number of colonies. Yeast DNA is isolated using the "Smash and Grab" protocol (*see* Shanks *et al.*, *Applied and Environmental Microbiology* 72:5027-5036 (2006)), or an equivalent protocol.

[0244] Next, competent *E. coli* are transformed with 1-5 μ L of yeast total DNA and selected on 50 or 100 Kan LB plates. Colonies are screened to verify the constructs. 2-5 μ g total plasmid DNA is then used for *T. saccharolyticum* transformations.

[0245] A second vector for gene deletion/marker removal is constructed using *in silico* analysis to place the two flanking regions adjacent to each other. Overlapping regions are added to the two adjacent primers on the flanking regions to obtain about 40 kb of homology between the regions when amplified.

[0246] Using two rounds of PCR amplification, the flanking regions can be connected. The first PCR amplification is a traditional amplification, and the second amplification is a dilution of the first round products to approximately 1 ng/μL. This dilution is used as a template; and the upstream flanking region 5' primer and downstream flanking region 3' primer are used for amplification. If necessary, optimization of annealing temperature or MgCl₂ can be performed. Alternatively, TOPO cloning (Invitrogen) or other known techniques can be used to make the second construct.

[0247] Following a PCR clean-up, 2-3 μg of the vector product is then used to transform *T. saccharolyticum*.

[0248] 1.1.3 Gene Insertion

[0249] To create a metabolic route to 1-hexadecanol, native and/or recombinant genes are overexpressed. The native PEPCK and TC genes are overexpressed via insertion of high level promoters in front of the coding sequence for these genes. This is accomplished through the pMU433-based marker cycling system, except that the recombinant promoter region will remain behind after the marker is removed. High expression level promoter regions can be chosen, without limitation, from any of the following promoters:

[0250] adhE promoter (SEQ ID NO:81)

tcataataagtgttaagggtgattgttaaatgaataacaaaaattatttacatcacacagtccaaaattcaattcattcaagcgaatttctgttgaatgcttgaaaaactgatacaatcacctgaaatgtagagattattgttaataaattaacacggagggtgtttatt

[0251] cbp promoter (SEQ ID NO:82)

gagtcgtgactaagaacgtcaaagtaattaacaatacagctatttttctcatgcttttaccctttcataaaaatttaattttatcggtatcataaaaattatagacgttatattgcttgccgggatatagtgctgggcattcggttgcaaaatgttcggagtaaggtggatattgattgcatgttgatctattgcattgaaatgattagttatccgtaaatattaattaatcataatcataaattaattatataataattgtttgacgaatgaaggttttgataaattatcaagtaaaaggaacgctaaaaattttggcgtaaaatatcaaaatgaccactgaattaatatggttaaagt agatataatattttgtaaacatgccttcagcaaggtagattagctgtttccgtataaattaaccgtatggttaaaccggcagtcagaaaaataagtataagattccgttatgaaaataatacttcggtagtaataataagagatatgaggttaagagatacaagataagagataaaggtagcaatgtataagatggtgcttttaggcacactaaataaaaaacaaataaacgaaaattttaaggaggacgaaag

[0252] pta promoter (SEQ ID NO:83)

gtattctacaattaaacctaatacgtcataatatgcgcctttctaaaaaattattaattgtacttattttataaaaaaatatgttaaatgtataatgtgtatacaatatatttctttagtaagaggaatgtataaaaaataatattttaagggaaggacgatctt

[0253] hyd promoter (SEQ ID NO:84)

ataagcgaaagggtaaattgctttgatttagatgattgaatatgtagtcgactggatgtgcaagtaaagaaaacatatcaaattag
tcgggattatcagaaaaataaaaaattttattttaactgttaaaaaataattaacatatggtataataattatgtcctattttgcaattt
aaagattaattttttaaaaggagggtattag

[0254] hfs promoter (SEQ ID NO:85)

gctgtaattgtccttgatgacgataggaagataaacattccaacaaaatatctcccagcaatattgctgaagaagatgccatagat
atttcattggatgtcaatgaaagaggacgaaaattaaaaaagtgtgattgaagaatcaagggaggaagactaatttttaatttttaa
cgtaattgttaataaataactattgtttacacactttctttatgtaataaaataattgtatacagtatacgg

[0255] ech promoter (SEQ ID NO:86)

tactgaatggagaaactgcacaaaaagctgttgacggcagcagaggagattatcctctgctattttgtgggaaaaactgcaaa
attcattgaaatattgtiaataataaacaaaattaattaataataacaattgacttatcatttaattagattataatcaaatgggtat
ttaaaatgtatacaatatataatattcattaaatgaaataaagaaggagtgaiaaaa

[0256] Next, recombinant genes encoding a fatty acyl-ACP reductase and hexadecanal dehydrogenase from organisms such as *Acinetobacter calcoaceticus* and *Geobacillus thermodenitrificans* (see Reiser and Somerville, *Journal of Bacteriology* 179:2969-2975 (1997); Liu *et al.*, *Microbiology* 155:2078-2085 (2009)) are identified (see below). These recombinant genes can be integrated into the genome, driven by a high level expression promoter, or expressed via a replicating plasmid such as pMU131 (see WO 2009/035595).

[0257] Nucleotide sequence of *Acinetobacter* acrl fatty acyl-ACP (-CoA) reductase (GenBank# U77680) (SEQ ID NO:87)

cagaagatat ggttcggtta tcggttgga ttgaacatat tgatgattg attgcagatctggaacaagc attggccaca
gtttgagcgt aaattttata aaaaacctct gcaatttcag aggtttttt atatttgctt tattatcgta tgatgttcatt aattgatcta
gcaaataata aaaattagag caattactct aaaaacattt gtaatttcag atacttaaca ctgattttt taaccaaatc
acttttagatt aactttagtt ctggaaattt tatttcctt taaccgtctt caatccaaat acaataatga cagcctttac
agtttgatat caatcaggga aaaacgcgtg aacaaaaaac ttgaagctct ctccgagag aatgtaaaag gtaaaagtgc
tttgatcact ggtgcatcta gtggaatcgg ttgacgatt gcaaaaagaa ttgctgcggc aggtgctcat gtattattgg
ttgccgaac ccaagaaaca ctggaagaag tgaaagctgc aattgaacag caagggggac aggcctctat tttccttgt
gacctgactg acatgaatgc gattgaccag ttatcacaac aaattatggc cagtgtcgat catgtcgatt tectgatcaa
taatgcaggg cggtcgatc gccgtgccgt acacgagtcg ttgatcgct tccatgatt tgaacgcacc atgcagctga
attactttgg tgcggtacgt ttagtgtaa atttactgcc acatatgattaagcgtaaaa atggccagat catcaatc

agctctattg gtgtattggc caatgacacccgttttctg cttatgtcgc gtctaaagct gcgctggatg cttcagtcg
 ctgtctttca gccgaggtac tcaagcataa aatctcaatt acctcgattt atatgccatt ggtgcgtacc ccaatgatcg
 caccaccaa aatttataaa tacgtgccca cgctttcccc agaagaagcc gcagatctca ttgtctacgc cattgtgaaa
 cgccaacac gtattgcgac gcacttgggt cgctggcgt caattaccta tgccatcgca ccagacatca ataatttct
 gatgtcgatt ggatttaacc tattccaag ctcaacggct gcaactgggtg aacaggaaaa attgaatcig ctacaacgtg
 cctatgcccg ctgtttccca ggcgaacact ggtaaaattt ataaaagaag cctctcatalc cgagaggctt tttatgggt
 acgacatca gccagattta gaggaaattg acttttctg tttttacatc ataaatcgca ccaacaatat caatttctt
 gcgatccagc atatctttaa gtacagaact atgtgaata atgtattgaa tattatagtg aacattcata gcagtcacct
 gatcaataaa tgcttggctt aattcacgcg gttgcataat atcaaatata ctgccaaccg aatgcatgag tggcccaagc
 acgtattgga tgtgtggcat ttctgaata tcggaaatct gcttatgttg caatcttaac tggcatgcgc tggtagccgc
 accacagtcg gtatgtccca aaaccagaat cactttggaa cctttggctt gacaggcaaa

[0258] Nucleotide sequence of *Geobacillus thermodenitrificans* NG80-2 adh1 (Genbank
 Nucleotide CP000557, Protein ABO67118) (SEQ ID NO:88)

ttacgccttg tgcggctcta cgategtccc ggcaaacgcg gcttcgtaaa tcgcacggat gtcggttcc aacagcggca
 acggactgcg ggcaagcaaa cgtttctgtt ggacagcatc ttctgtaag ctttctagcg cgctttcggg gatgccaaat
 cccccaatg ttttcggaat gccgacatcg gcgacgaacc gttctagttc ctcgacgcac cgataagacg ctccacttc
 ggacaaaaaa cttgagttgc cgccaagcgc gttgaaaata tcggccattc tcttcgtaca gctttgacgg atgtagccca
 tcacatacgg caacagcaca gcattcgatt caccatgagc gatatgaac tgaccaccga gcggataagc
 gagcgcacgc acaccggcta ccccggcgtt gaaaaatgcc aagccggcca aataactgcc gttcgccata
 tcaatgcgcg cctgtttgtc cgaaccgttg gccaccgctt tgcgcagtga gcgtgaaatc agccgaatag cggcaacggc
 caatccatcc gatgttgggc tcgcattgac cgacacatac gctcaactg catgggtgag tgcataat cccgttgcgg
 ccgttaccgg cgggtggaacg gaaacggta gctgcccgtc aacgatcgcg acgtcggcca ataagtaatc
 gtgcgtcacg acatctttcg tcgtttccaa agacaagaca gagatgtttg tcaattccga cccggtgccc gatgtcgttg
 gaatcaaaat tttcggcaac cctttttct caagtgttcg cgttcctgtc aaatttaa atgtcagcgc cgagccatca
 tgcaccgcca aaacagccgc cagtttcgcc aaatccagcg cgctgccacc accaacaccg atgacaaggt
 caaactttcc gtcggggca aacgccactg cctttcccc tgtctcaagc ggcggctctg gcacaacatc cgtatacaca
 tgcacgtat acccttcttg acggagcggg gacgtcactt gatgcactag gccgatctc acaagcatcg ggtcggtaat
 caccaaaata tgttttgtc ccaaccgctt cacttcagga actaactggt caagcgtcc ccagccgaca tggctgagcg
 gcggaaagac aatgcgggct acatcat

[0259] Nucleotide sequence of *Geobacillus thermodenitrificans* NG80-2 adh1 (Genbank
 Nucleotide CP000557, Protein ABO68223) (SEQ ID NO:89)

ttataaagac gcacgcaaaa tggcgagcac atcatcacga ttaacgttt tgaacggcc aaactacca aacgccatcg
 ctttatccgc catcagctcg agattttct cgccgatgcc ataacagcc aatcgagacg gcgccccgag gctcgaccaa
 aacgcgcgca accgctcgat gccctcaagc gccacgtcgc gctccgtttt gcccgctgga tcgacgtcaa
 agacgcgcac cgccagttgg gcgaacggc tgacatttc atcaagcaca tgttcatcc aattcgggaa caaaatggcc
 aatccccgg cgtagcggat atcgtataca gcagagaccg catgctcgat atcatgcgtc gcccaatcac
 cgcgcacgcc catttgcaaa aagcgtttta aggcgatcgt gcccgagtag atgacgtct cgcgagctc gtatgtctct
 aagtcgtcaa ccaattttgg cgccgctca atgaccgttt ttaacactgc ctcgcacatc cggctttgca gcggcgtgtt

cggcgatga tggaaatatt gctcaaacac atgggacatc atatcgacga tgccgtaaac ggtatggtct ttcggcacccg
 tcatcgtgta cgtcggatcc aaaatcgaaa attgcgggaa tgcaccggg ctgccccagc cgtattttc ttcgtctcc
 caattggtga tcaccgatcc ggcgttcatt tccgagccgg tcgctgccag cgtcaggacc gtcccaaacg gcaacgcctc
 agtgacagtc gctttttcg taatgaactc ccacggatcg ccatcaaact tcgcgccggc tgcaatcgct ttcgtacagt
 cgatcacact gccgccgcca acggcaagca aaaattcaat tcctcccgct ctgcaaagt ctacccctt tttgacggc
 gaaaggcgcg ggttcgggtc gacgcctggc agttcaacga ctccggcgcc aatgtccgtc aataggctca tgacttcac
 atatagtcg ttcgtttaa tgcgcgccg cccatagaca agcagcactt tttgccata ttcggcact tctctttga
 gtgctcaat ttgtctctc ccaaaaatga gtttggtcgg attgcgaaac gtaaaattt gcat

[0260] 1.1.4 Selection and Optimization of Engineered Strains

[0261] The engineered strain is cultured continuously via any of several methods, including chemostat, pH-auxostat, or serial batch transfer, to select for naturally occurring mutations that impart a benefit upon cellular growth and 1-hexadecanol formation. Because ATP generation and NAD(P)H regeneration are both coupled to 1-hexadecanol formation in the engineered strain, evolutionary forces will select for cells that are better able to carry out this conversion.

[0262] 1.1.5 Detection of 1-Hexadecanol

[0263] 1-hexadecanol formation in cultured engineered strains is detected via gas chromatography-mass spectrometry (GC/MS) with or without an extraction step prior to analysis. See Steen *et al.*, *Nature* 463:559-562 (2010); Aldai *et al.*, *Journal of Chromotography* 1110:133-139 (2006).

Example 2

[0264] 2.1 Diverting central metabolic flux through oxaloacetate in *E. coli*

[0265] This example describes engineering the central metabolic flux in *Escherichia coli* so that the majority of glycolytic flux passes from phosphoenolpyruvate to oxaloacetate rather than from phosphoenolpyruvate to pyruvate. See Figure 9. This is accomplished via a series of gene deletions that inactivate competing pathways and gene overexpressions that activate desired pathways. Target genes are shown in Table 4. A set of minimal target gene deletions is shown in Table 5.

Table 4. Targets for Gene Inactivation

<i>E. coli</i> gene name	description	locus tag ^a
<i>edd</i>	6-phosphogluconate dehydratase	b1851
<i>ldhA</i>	lactate dehydrogenase	b1380
<i>pta</i>	phosphate acetyltransferase	b2297
<i>adhE</i>	acetaldehyde dehydrogenase / alcohol dehydrogenase	b1241
		b4151-
<i>frdABCD</i>	fumarate reductase (anaerobic)	b4154
<i>fdhF</i>	formate dehydrogenase-H	b4079
<i>ppc</i>	phosphoenolpyruvate carboxylase	b3956
<i>pykA</i>	pyruvate kinase	b1854
<i>pykF</i>	pyruvate kinase	b1676
<i>mdh</i>	malate dehydrogenase	b3236
<i>maeA</i>	malic enzyme NADH	b1479
<i>maeB</i>	malic enzyme NADPH	b2463
<i>fadE</i>	acyl coenzyme A dehydrogenase	b0221
<i>ptsI</i>	PEP-protein phosphotransferase of PTS system	b2416
<i>pflB</i>	pyruvate formate lyase	b0903
<i>aceEF</i>	pyruvate dehydrogenase	b0114
<i>poxB</i>	pyruvate oxidase	b0871
<i>mgsA</i>	methylglyoxal synthase	b0963
<i>ppsA</i>	phosphoenolpyruvate synthase	b1702

^a locus tag numbers are given for the genome sequence of *E. coli* MG1655, which can be accessed via Genbank (Accession No. U00096) or the Kyoto Encyclopedia of Genes and Genomes (KEGG).

Table 5. Minimal Targets for Gene Inactivation

gene name	description	locus tag
<i>ldhA</i>	lactate dehydrogenase	b1380
<i>pta</i>	phosphate acetyltransferase	b2297
<i>adhE</i>	acetaldehyde dehydrogenase / alcohol dehydrogenase	b1241
<i>pykA</i>	pyruvate kinase	b1854
<i>pykF</i>	pyruvate kinase	b1676
<i>mdh</i>	malate dehydrogenase	b3236
<i>ptsI</i>	PEP-protein phosphotransferase of PTS system	b2416

[0266] 2.1.1 Deletion and Overexpression of Target Genes

[0267] In order to perform gene modifications (either deletion or overexpression) in *E. coli* to redirect metabolic flux through oxaloacetate, 500 bp to 2000 bp flanking regions upstream and downstream of a target gene were amplified via PCR using primers (Table 6) and ligated into pMU2723 (Figure 11) using standard molecular biology methods. See Shanks *et al.*, *AEM* 72: 5027-5036 (2006). The *pta::fdh* and Pspc *pckA* promoter exchange modifications were performed by placement of the heterologous DNA (*fdh* or Pspc) between the two homologous flanking regions of the target gene, with appropriate design to allow either expression of *fdh* from the native *pta* promoter, or heterologous expression of the native *pckA* gene. The protocol used is described briefly below.

[0268] The starting strain, M2162 or subsequent progeny, was grown overnight in 8 mL of LB medium at 37°C. Two 500 mL baffled flasks, each containing 150 mL of LB, were pre-incubated at 37°C and then inoculated with 2 mL of the overnight culture. These cultures were incubated at 37°C with shaking until the OD reached 0.5 to 0.8 (checked OD every 20 min. after 2 hrs). The flasks were then placed in an ice bath for about 15 minutes after which the cultures were transferred to six 50mL conical tubes. The tubes were spun at 4000 rpm for 8 minutes in a clinical swinging bucket centrifuge at 4°C. Following centrifugation, the supernatant was removed, about 10mL of ice cold water was added to each tube, and the pellets were resuspended and transferred to two 50mL tubes which were balanced to 50mL with ice cold water. The tubes were centrifuged for 8 minutes in the conditions described above. The supernatants were removed and the pellets were resuspended with about 200µL of cold water, after which 80 µL of the resuspended cells were transferred to a cold 1 mm gap cuvette which contained 2-4 µL of pre-added plasmid DNA targeting the gene of interest. The cuvette was electropulsed using an exponential decay pulse, 1.8 kV voltage, 25 µF capacitance, 200 Ω resistance, and a 1 mm gap cuvette method. 1 mL of SOC medium was added to the cuvette and the entire volume was then transferred to a 14 mL falcon tube and incubated at 37°C for 1 hour. 250µL of cells were removed, plated on LB plates containing 50 µg/mL kanamycin, and incubated at 37°C for 24-48 hours. Colony PCR was performed on kanamycin resistant colonies using one internal and one external primer to the site of integration with primers listed in Table 8. Two positive colonies were re-streaked on 50

µg/mL kanamycin plates and incubated overnight at 37°C. Two colonies were selected and grown in 5 mL of LB medium, either for 8 hours or overnight at 37°C. Serial dilutions of 1:10, 1:100, and 1:1000 of each LB culture were prepared, and 20 µL of each dilution was plated on 10% w/v sucrose + 500 µg/mL streptomycin plates. The plates were incubated overnight at either 37°C or 42°C. Colony PCR was performed on 7 colonies from each initial LB culture with two primers, as listed in Table 8, external to the site of integration. Two positive colonies were re-streaked on 500 µg/mL streptomycin plates and incubated at 37°C overnight. One colony from each plate was selected and re-patched on a kanamycin 50 µg/mL plate and a streptomycin 500 µg/mL plate. The patches that grew on the streptomycin but not the kanamycin plates were then used to make culture collection stocks.

Table 6. Primers Used to Create Gene Deletion and Gene Overexpression Plasmids for Routing Metabolic Flux Through Oxaloacetate

Primer #	SEQ ID NO:	Primer sequence	Description
X12312	90	TTTTGCTGCTTAAATTTTGGTATCATTAAGGATCTATGTAAACCCAGGAAGCGG CAA	pta 1
X12313	91	ACGAGATTACTGCTGCTGTGCAGACTTTTGCGTTCCATTGCACGGATCA	pta 2
X12314	92	TGATCCGTGCAATGGAAACGCAAGTCTGCAACAGCAGCAGTAATCTCGT	pta 3
X12315	93	GATAACAAATTTTACACACAGGAAACAGCTATGACCATACGGCCTCTTCTCCCATACC AAAT	pta 4
X12316	94	TTTTGCTGCTTAAATTTTGGTATCATTAAGGAACGCAGTTGCTGGATATCAGA GGT	ldh 1
X12317	95	TACTGGTCAGAGCTTCTGCTGTCAACTCGTTCAACCITGTTGCAGGTACT	ldh 2
X12318	96	AGTACCTGCAACAGGTGAACGAGTTGACAGCAGAAGCTCTGACCCAGTA	ldh 3
X12319	97	GATAACAAATTTTACACACAGGAAACAGCTATGACCAATTTGGGATGTGTGCAATTACCC AACG	ldh 4
X12320	98	TTTTGCTGCTTAAATTTTGGTATCATTAAGGATACTGGTAAACGTCTGCCGACC AA	edd 1
X12321	99	ACAGCTTAGCGCCTTCTACAGCTTCGCGCGAACGTTCAATGATTCGAT	edd 2
X12322	100	ATCGAATCATTGAACGTTCCGCGGAAGCTGTAGAAAGCGCTAAGCTGT	edd 3
X12323	101	GATAACAAATTTTACACACAGGAAACAGCTATGACCAATGCTGACATTTGGCTATCCCTG CAT	edd 4
X12324	102	TTTTGCTGCTTAAATTTTGGTATCATTAAGGAGCGGGTCAATTTCCAGATAAAG CA	frd 1
X12325	103	TCAGGAACAGGAATAACGCAACCAAGATCGGCTTGAAAGGTTTGCACGA	frd 2
X12326	104	TCGTGCAAAACCTTTCAAGCCGATCTTGGTCGCGTATTCCTGTTCCTGA	frd 3
X12327	105	GATAACAAATTTTACACACAGGAAACAGCTATGACCAATGCGAAACATGCACCTGCCTTA CCT	frd 4

- 83 -

X12328	106	TTTTGTCTGCTTAAATTTTGGTATCATTTATAGGATGGACCGAATGGACGATGGAG TTT	pfl 1
X12329	107	AGAAATGCCCTTTCACGCGTTCCATGTCTGTTGCTTTATAGACACCCGCCCT	pfl 2
X12330	108	AGCGGGTGTCTATAAAGCAACGACATGGAACGCGTGAAAGGCATTCT	pfl 3
X12331	109	GATAACAATTTTCACACAGGAAACAGCTATGACCATTTCCGTTAACGATACGCTTC GGGT	pfl 4
X12332	110	TTTTGTCTGCTTAAATTTTGGTATCATTTATAGCAATTCAAACGTTATGCCCGACGC TG	ppc 1
X12333	111	AGCGGTCGGTGTAATAATTCCGTTCTTGATGGTTTCTCCAGCACT	ppc 2
X12334	112	AGTGCTGGGAGAAACCATCAAGGAACGGAATATTTACACCGACCCGCT	ppc 3
X12335	113	GATAACAATTTTCACACAGGAAACAGCTATGACCATTTGAAATTAGCCAGTGGCGG CAAG	ppc 4
X12336	114	TTTTGTCTGCTTAAATTTTGGTATCATTTATAGGACAGCCGCTACATTAAAGGCAC CAA	ptsl 1
X12337	115	CCAGCAGCGGCAGATCAAAATTCAAATGGCGGTTGCACTTTAGCCTGTAT	ptsl 2
X12338	116	ATACAGGCTAAAGTCGAACCGCCATTGAATTTGATCTGCCGCTGCTGG	ptsl 3
X12339	117	GATAACAATTTTCACACAGGAAACAGCTATGACCATATGGTTAGCGGCTATTGTC GTGC	ptsl 4
X12340	118	TTTTGTCTGCTTAAATTTTGGTATCATTTATAGGATGGCGAATGGCACTCCCTATGT TA	pykA 1
X12341	119	TGACAATCACCAAGTCAACAGACATCCGAATGAAATAACGCCGCGATG	pykA 2
X12342	120	CATCGCGCGTTATTTCAATTCGGATGCTGTGTGACCTGGTGATTGTCA	pykA 3
X12343	121	GATAACAATTTTCACACAGGAAACAGCTATGACCATTTGTTGATGAGATGTTGCCA CCGC	pykA 4
X12344	122	TTTTGTCTGCTTAAATTTTGGTATCATTTATAGGAATGCTGTACGTAATACGCCCTGC GA	pykF 1
X12345	123	TCTTTAAACAAGCTGCGGCACAACGATGGGAGAAAATTGCTTTCTGGGC	pykF 2
X12346	124	GCCAGAAAGCAAGTTTCTCCCATCGTTGTGCGCAGCTTGTAAAGA	pykF 3
X12347	125	GATAACAATTTTCACACAGGAAACAGCTATGACCATATCTTTAGCAGCCTGAACGT CGGA	pykF 4
X13802	126	TTATAGGTTAATGTTCATGATAATAATGGTTTCTTCCGTCAAAGGGCAAATCACCGA AA	fdhF 1
X13803	127	GGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCACCTCGGAATAACCGGTTTCGGG	fdhF 2

X13804	128	AAA CGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGTATACGACAAAAGCGTTCGTCGCTTC	fdhF 3
X13805	129	A ACTGTTGGGAAGGGCGATCGGTGCGGGCCCTCTTCGAATGAAGCCAGTTCGCCCA	fdhF 4
X14576	130	TTT TTATAGGTTAATGTTCATGATAATAATGGTTTCTTTCGGGATGCCGAAGGCTTTTGTGT	Pspc pckA 1
X14577	131	AT TGGGTAGAAAAATAAACGGCTCAGATTCTCTGTACGAAAACGGTTGCT	Pspc pckA 2
X14578	132	AGCAACCGTTTCGTGACAGGAATCTGAGCCGTTTATTTTCTACCCA	Pspc pckA 3
X14579	133	GGTCAAAACCATTTGTTAACGCGCATTTTAGTGCTCCGCTAATGTCAACT	Pspc pckA 4
X14580	134	AGTTGACATTAGCGGAGCACTAAAATGCGCGTTAACAAATGGTTTGACC	Pspc pckA 5
X14581	135	GGTTCGCGCACATTTCCCGGAAAGTGCCACCTAGAACGATACCTTTCAGCG	Pspc pckA 6
X14588	136	GCA TTATAGGTTAATGTTCATGATAATAATGGTTTCTTCTTCTATGTAAACCCAGGAAGCGGC	pta::fdh3 1
X14589	137	AA TGGGTAGAAAAATAAACGGCTCACTTTGCGTTCCATTGCACGGATCA	pta::fdh3 2
X14590	138	TGATCCGTGCAATGGAAACGCAAGTGAGCCGTTTATTTTCTACCCA	pta::fdh3 3
X14591	139	ATAAGAACTAAACACAACTTTCATTTTAGTGCTCCGCTAATGTCAACT	pta::fdh3 4
X14592	140	AGTTGACATTAGCGGAGCACTAAAATGAAGATTGTCTTAGTTCTTTAT	pta::fdh3 5
X14593	141	ACGAGATTACTGCTGCTGTGCAGACTATTCTTATCGTGTTCACCGTA	pta::fdh3 6
X14594	142	TACGGTAAACACGATAAGAAATAGTCTGCACAGCAGCAGTAATCTCGT	pta::fdh3 7
X14595	143	GGTTCCGCGCACATTTCCCGGAAAGTGCCACCAACGCGCTCTTCTCCCATACCAA	pta::fdh3 8
X15570	144	AT TTATAGGTTAATGTTCATGATAATAATGGTTTCTTTTGGCATCCGTAGCAGACACCAT	maeA 1
X15571	145	AA GAATACTGCGCCAGCGTTTCACTTCGTTCCGCTTGTTCTTCGATGGTT	maeA 2
X15572	146	AACCATCGAAGAACAAAGCGGAACGAAGTGAACGCTGGCGCAGTATTC	maeA 3
X15573	147	GGTTCCGCGCACATTTCCCGGAAAGTGCCACCCCATCAATGGCGATCACTTTGGC	maeA 4
		GT	

- 85 -

X15574	148	TTATAGGTTAATGTCATGATAATAATGGTTTCTTAATTGACCGGCAGTTTGTCAACA	maeB 1
		CG	
X15575	149	TCGCCGTGCATTTACCATCAATCGAGCGCGGACAACTTCAATAAA	maeB 2
X15576	150	TTTATTGAAGTTGTCGCCGCTCGATTGATGGTGAATGCACGGCGA	maeB 3
X15577	151	GGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCGCCATAAATCACCCAAATGCACC	
		GCT	maeB 4
X15578	152	TTATAGGTTAATGTCATGATAATAATGGTTTCTTTCAGCTGGCAGGCACGTAAACCAT	
		TT	mdh 1
X15579	153	TCAAAATGCGCTCAGGGTACCGATATTTCTGAACCTGAAGGCAGTTGGGT	mdh 2
X15580	154	ACCCAACGTGCCCTTCAGGTTCAAGATAATCGGTACCCTGAGCGCATTTGA	mdh 3
X15581	155	GGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCACTGGCGGTTTACCTACCATTC	
		CA	mdh 4
X15586	156	TTATAGGTTAATGTCATGATAATAATGGTTTCTTTCGACATCGCTATTGTCAACCAC	
		CA	adhE 1
X15587	157	TTTCGGAAAGTTTGTGCCACAACATAATGCTCTCCTGATAATGTTAAAC	adhE 2
X15588	158	GTTTAAACATTATCAGGAGAGCATTAATGTTGTGGCACAAACTTCCGAAA	adhE 3
X15589	159	GGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCCCAAGTGTGCGCAATTTTCAG	
		CAT	adhE 4

[0269] Gene modifications were confirmed on an agarose gel. See Figure 10 and Table 7. Primers external to each region of interest were used to amplify DNA (Table 8), which was subsequently run on an agarose gel and stained with ethidium bromide to visualize DNA length. The ladder shown in lanes 1 and 22 of Figure 10 is New England Biolabs 1kb. DNA was amplified from strains M2939 (deletion and overexpression strain) and M2162 (wildtype). See Table 9 for a description of the bacterial strains.

Table 7. Predicted Size of Wildtype and Modified Target Genes

Target	WT length (bp)	KO length (bp)
<i>ldhA</i>	2686	1855
<i>edd</i>	4343	1931
<i>pta::fdh</i>	3992	3186
<i>frd</i>	5358	2241
<i>fdhF</i>	3927	2141
<i>ppc</i>	4739	2321
<i>pckA</i> (promoter exchange)	2003	1053/958 *
<i>pykA</i>	3204	1803
<i>pykF</i>	2863	1594
<i>ptsI</i>	4418	1938

* after *psfI* restriction digest of PCR product to distinguish the wildtype (2003 bp) and promoter exchange (2011 bp) alleles.

Table 8. Primers Used to Verify Genome Alterations

Primer #	SEQ ID NO:	Primer sequence	Description
X12354	160	TTGCTGTATTTGACACCGCGTTCC	pta ext 1
X12355	161	TTTCACGAAAGAAGCGGTCGGACT	pta ext 2
X12356	162	GGCAAGTTTAACGTCGCAGTAGCA	ldh ext 1
X12357	163	TTTATGGCGGTGTCGTTTGGCTTG	ldh ext 2
X12358	164	ATATCTGGAAGAAGAGGGCGCGAA	edd ext 1
X12359	165	GATGCATTACGCCGTGTGGTTGAA	edd ext 2
X12360	166	AACAGCAATTGTAGCAGCGTGTCG	frd ext 1
X12361	167	TTGTTTGCCAGCATCACGATACCC	frd ext 2
X12362	168	CTGGGCGTTTATGCTTGCCGTATT	pfl ext 1
X12363	169	AGTCGTCAGTTGTGAGCTCGACTT	pfl ext 2
X12364	170	TATTCACGGTGGCGACGCTTCTAA	ppc ext 1
X12365	171	CGCCTGTTGCAGGATTTCAATGGT	ppc ext 2
X12366	172	AAAGCGTTAGGTGCAAACCTGGTG	pts ext 1
X12367	173	ATTGCCGTGCCTGCTATCAAACAG	pts ext 2

X12368	174	GCTATGGCACTGGAAGCCAATGTT	pykA ext 1
X12369	175	AGAACGTAGTGAAGCTGAACGCGA	pykA ext 2
X12370	176	TGAAGCTTACCGCCTCATCCTGAA	pykF ext 1
X12371	177	AGAATGGTGAACCAGAGCAAGGGA	pykF ext 2
X12801	178	GATTGATTACGCGGTGAAAGCGCA	fdh ext 1
X12802	179	ACACCCGGTATCAAACCCTTCCAT	fdh ext 2
X14574	180	CCGTGGCGATTAACGTGAACAACT	pckA ext 1
X14575	181	AGTCGATAGTGCCATCTTCACGCA	pckA ext 2
X15590	182	ACTGTTCCCTTCCCGCGTTTGATA	maeA ext 1
X15591	183	GCATCAACTGCCGAGTTAAACGCA	maeA ext 2
X15592	184	AGGTCGAAGCCAGCTTGATCAGAA	maeB ext 1
X15593	185	CGCTGACGGTTTGTGATAACGCTT	maeB ext 2
X15594	186	TACCTTCTGCTTTGCCCAGTGAGT	mdh ext 1
X15595	187	TGAAGCATTGCTGGTGGGATCTGA	mdh ext 2
X15596	188	AGTGGCACCACACCAATGCTTTCA	adhE ext 1
X15597	189	TGAACGCCAGCTTCACGGATAGAT	adhE ext 2
X13673	300	ATACGGGATAATACCGCGCCACAT	internal 1
X13674	301	CCATTCGACCACCAAGCGAAACAT	internal 2

Table 9. Lineage of Strains From M2162 to M2939

M number	Genotype	Parent
M2162	strepR	MG1655
M2264	strepR, Δ edd Δ eda	M2162
M2273	strepR, Δ edd Δ eda Δ ldh	M2264
M2348	strepR, Δ edd Δ eda Δ ldh Δ ppc	M2273
M2371	strepR, Δ edd Δ eda Δ ldh Δ ppc Δ fdhF	M2348
M2379	strepR, Δ edd Δ eda Δ ldh Δ ppc Δ fdhF Δ frd	M2371
M2492	strepR, Δ edd Δ eda Δ ldh Δ ppc Δ fdhF Δ frd Δ PpckA::Pspc pckA	M2379
M2590	strepR, Δ edd Δ eda Δ ldh Δ ppc Δ fdhF Δ frd Δ PpckA::Pspc pckA Δ pta::fdh3	M2492
M2645	strepR, Δ edd Δ eda Δ ldh Δ ppc Δ fdhF Δ frd Δ PpckA::Pspc pckA Δ pta::fdh3 Δ pykF	M2590
M2698	strepR, Δ edd Δ eda Δ ldh Δ ppc Δ fdhF Δ frd Δ PpckA::Pspc pckA Δ pta::fdh3 Δ pykF Δ pykA	M2645
M2939	strepR, Δ edd Δ eda Δ ldh Δ ppc Δ fdhF Δ frd Δ PpckA::Pspc pckA Δ pta::fdh3 Δ pykF Δ pykA Δ ptsI	M2909

[0270] 2.2 Creating a balanced reduction/oxidation pathway during anaerobic fatty acid production

[0271] Reduction and oxidation (redox) reactions play a key role in catabolic metabolism, allowing the transfer of electrons from one compound to another, and in the process, creating free energy for use elsewhere in cellular metabolism. To facilitate transfer of electrons from one compound to another, cells use redox co-factors to shuttle electrons. Several compounds and proteins can function as redox co-factors—the most relevant for anaerobic growth on carbohydrates are the nicotinamide adenine dinucleotides NADH and NADPH, and the iron-sulfur protein Ferredoxin (Fd).

[0272] Since NADH, NADPH, and Fd function as electron shuttles, they must discharge as many electrons as they accept, *i.e.*, their net electron accumulation is zero. Catabolic metabolism can be thought of in two parts: carbohydrate deconstruction, where electrons are placed onto redox co-factors, and end-product construction, where electrons are removed from redox co-factors. In order for a metabolic pathway to function efficiently and at high yield, the type of co-factors used in carbohydrate deconstruction must balance those used in end product construction.

[0273] During carbohydrate deconstruction, which in the anaerobic fatty acid pathway ultimately results in acetyl-CoA, electrons are removed at two steps: the conversion of glyceraldehyde-3-phosphate to 1,3-biphosphoglycerate + $2e^-$ and the conversion of pyruvate to acetyl-CoA + CO_2 + $2e^-$. In *E. coli*, NAD⁺ is used as electron acceptor for the first conversion. For the second conversion, *E. coli* employs a NAD⁺ linked pyruvate dehydrogenase during aerobic growth, and pyruvate formate lyase (*pfl*) and a formate dehydrogenase directly linked to hydrogen production (*fdhF*) to produce formate or H_2 from the $2e^-$ removed from pyruvate.

[0274] *E. coli* strains have been engineered to produce ethanol from acetyl-CoA at high yield via anaerobic expression of pyruvate dehydrogenase (PDH) (Kim *et al.*, *AEM* 73: 1766-1771 (2007)) or via heterologous expression of NAD⁺ formate dehydrogenase (Berríos-Rivera *et al.*, *Met Eng* 4:217-229 (2002)). In both wildtype and these engineered *E. coli* strains, NADH is the primary redox co-factor.

[0275] In contrast, the electron accepting reactions of fatty acid elongation require either exclusively NADPH or 1:1 stoichiometric levels of NADPH and NADH, depending on the co-factor specificity (NADPH or NADH) of enoyl-ACP reductase.

[0276] In order to balance the NADPH necessary for fatty acid elongation, the redox enzymes involved in carbohydrate deconstruction should be engineered to produce

NADPH. In Table 10 below, different redox enzyme systems are described that can produce, per $\frac{1}{2}$ glucose molecule, 2 NADH, 1 NADH and 1 NADPH, or 2 NADPH. Use of one of these systems in a host microorganism, or a combination thereof, will allow for an overall balanced co-factor pathway for anaerobic fatty acid production. In addition to, or instead of, using these systems, the enzymes can be modified to have different cofactor specificities.

Table 10. Enzymes for an overall balanced co-factor pathway for anaerobic fatty acid production

Figure 32		Carbohydrate deconstruction reactions	Redox enzymes	NADH	NADPH
A	1/2 glucose $\rightarrow$ acetyl-CoA + CO ₂ + 4e ⁻		NAD+ GAPDH, PDH	2	0
B	1/2 glucose $\rightarrow$ acetyl-CoA + CO ₂ + 4e ⁻		NAD+ GAPDH, PFL, NAD+ FDH	2	0
C	1/2 glucose $\rightarrow$ acetyl-CoA + CO ₂ + 4e ⁻		NAD+ GAPDH, PFL, NADP+ FDH	1	1
D	1/2 glucose $\rightarrow$ acetyl-CoA + CO ₂ + 4e ⁻		NAD+ GAPDH, PNO	1	1
E	1/2 glucose $\rightarrow$ acetyl-CoA + CO ₂ + 4e ⁻		NADP+ GAPDH, PFL, NAD+ FDH	1	1
F	1/2 glucose $\rightarrow$ acetyl-CoA + CO ₂ + 4e ⁻		NADP+ GAPDH, PFL, NADP+ FDH	0	2
G	1/2 glucose $\rightarrow$ acetyl-CoA + CO ₂ + 4e ⁻		NAD+ GAPDH, POR, Fd, NFN	0	2
H	1/2 glucose $\rightarrow$ acetyl-CoA + CH ₂ O ₂ + 2e ⁻		NAD+ GAPDH, PFL	1	0
		End product construction reactions	Redox enzymes	NAD+	NADP+
I	acetyl-CoA + 4e- + acyl _(n) -ACP $\rightarrow$ acyl _(n+2) -ACP		FabG (NADPH), FabI (NADH)	1	1
J	acetyl-CoA + 4e- + acyl _(n) -ACP $\rightarrow$ acyl _(n+2) -ACP		FabG (NADPH), FabI (NADPH)	0	2
K	acyl _(n+2) -ACP + 4e- $\rightarrow$ acyl alcohol (fatty alcohol)		AcDH, ADH	0-2	0-2
L	acyl _(n+2) -ACP + H ₂ O $\rightarrow$ acyl acid (fatty acid) + 2 CH ₂ O ₂		BTE	0	0

Abbreviations: GAPDH – glycerol-3-phosphosphate dehydrogenase, PFL – pyruvate formate lyase, PDH – pyruvate dehydrogenase, PNO – pyruvate:NADP+ oxidoreductase, POR – pyruvate:ferredoxin oxidoreductase, Fd – ferredoxin, NFN – NADH ferredoxin:NADP+ oxidoreductase, FabG – β -ketoacyl-ACP reductase, FabI enoylacyl-ACP reductase, AAR – acyl-ACP reductase, ADH – alcohol dehydrogenase, BTE – acyl-ACP thioesterase, AdhE – bifunctional acetaldehyde/alcohol dehydrogenase.

[0277] Enzymes used in the carbohydrate deconstruction reactions can be cloned into plasmids for expression in a host strain. For example, plasmids FP45, FP47, FP66, FP67, FP68, and FP75 are examples of heterologous redox enzymes designed for expression in *E. coli* to modify the native carbohydrate deconstruction pathway. See Figure 12. Genes or gene operons are cloned under expression of either the constitutively active P_{spc} ribosomal promoter or the inducible Plac promoter. Cloning was performed via PCR amplification, using the primers listed in Table 11, or direct DNA synthesis of the desired gene products, followed by yeast gap-repair cloning with 30-60 bp homologous flanking regions. See Shanks *et al.*, *AEM* 72:5027-5036 (2006). Transformed yeasts were selected via growth on SD-ura minimal medium for the presence of the *ura3* gene. Plasmids were recovered from *ura3*⁺ yeast strains by standard mini-prep (Qiagen) or phenol extract and ethanol precipitation. Crude yeast plasmid preps were then used to transform *E. coli* TOP10 cells (Invitrogen) using selection with the plasmid appropriate antibiotic, either carbenicillin (100 µg/mL), tetracycline (15 µg/mL), or chloroamphenicol (25 µg/mL). *E. coli* mini-prepped plasmids were confirmed by restriction digest and agarose gel analysis.

Table 11. Primers for the Construction of Redox Balancing Plasmids FP45, FP47, FP67, FP68, and FP75 (FP66 was created from direct DNA synthesis (SEQ ID NO:206), so no primers were used during its construction)

Primer #	SEQ ID NO:	Primer sequence	Description
X16072	190	TCTCAGTAGTAGTACATTAGCGGAGCACTAAATGAAGATTGTCTTAGTTCTTTAT	FP45 1
X16073	191	CAGTCTTTCGACTGAGCCTTTCGTTTACGGCCGCTATTCTTATCGTGTTTACCGTA	FP45 2
X16082	192	TCTCAGTAGTAGTACATTAGCGGAGCACTAAATGGCAACCGTTCTGTGTCTTG	FP47 1
X16083	193	CAGTCTTTCGACTGAGCCTTTCGTTTACGGCCGTTAGGTCAGACGATAGCTCTGTGC	FP47 2
X16829	194	TGTGAGCGGATAACAATTTCCACACAGGAAACAGCTATGTGCAAGGTTATGAAAACCA	FP67 1
		TG	FP67 2
X16064	195	GCTTTCCACACCTCCAAGATTTCGTCTAAATTTGTTTCAGCAAGCTTCTT	FP67 3
X16065	196	AAGAAGCTTGCTGAACAAATTAAGACGAAATCTTGGAGGTGTGAAAGC	FP67 4
X16830	197	CCTCGAGGTCGACGGTATCGATAAGCTTGATATCTTATTCAGCCTTAATAGCTCCTGT	FP68 1
		T	FP68 2
X16831	198	TGTGAGCGGATAACAATTTCCACACAGGAAACAGCTATGGGAAAGAAATGATGACG	FP68 3
		ACT	FP68 4
X16075	199	TACACCTCCTTATCTTAATAGGCGTTCTACTTCTTCGTCCGCTTGCTGAG	FP68 5
X16076	200	CTCAGCAAGCGGACGAAAGTAGAACGCCCTATTAAAGATAAGGAGGTGTA	FP68 6
X16077	201	CCCGTCTGATATTATGGTCTACGACTTACTCTTGAACCTGGAGCTCCTAC	FP75 1
X16078	202	GTAGGAGCTCCAGTTCAGAGTAAGTCGTAGAACCATAAATATCAGACGGG	FP75 2
X16832	203	CCCTCGAGGTCGACGGTATCGATAAGCTTGATATCCTATTGGTTCTGCCGGATATA	
		T	
X16981	204	TGTGAGCGGATAACAATTTCCACACAGGAAACAGCTATGCCCCGATATGACAAACGAAT	
		CT	
X16982	205	CCCTCGAGGTCGACGGTATCGATAAGCTTGATATCTTAAACACCAGCTTCGAAGTCCT	
		T	

[0278] Sequence of the PNO gene and flanking regions used to create FP66 (SEQ ID NO:206):

tgagccgtttatTTTTtaccatataccttgaagcgggtgtataatgccgcgccctcgatatggggatttttaacgacctgatttcggg
tctcagtagtagttgacattagcggagcactaaaatgaaacagagcgttcgtccgattattagcaatgttctgcgtaaagaagtgc
cctgtatagcaccattattggtcaggataaaggtaaagaaccgacagggtctacctataccagcgggtccgaaaccggcaagcca
tatigaagttccgcaicatgtttaccgttccggcaaccgatcgtaccccgaaatccggatgcacagtttttcagagcgttgatggtagc
caggcaaccagccatgttgcataatgccctgagcgataccgcatttatctatccgattaccccgagcagcgttatgggtgaactggc
agatgtttggaatggcacaggggtcgtaaaaatgcccttgggtcaggttgtgatgttgcgtgaaatgcagagcgaagccgggtgcagcg
gggtgcactgcatgggtgcactggcagccgggtgcgattgcaaccacctttaccgcaagccagggtctgctgctgatgaticcgaata
tgtataaaatcgaggcgaactgatccgagcgttattcatgttgcagcacgtgagctggcagggtcatgcactgagcatttttggtg
gtcatgcagatgttatggcagttcgtcagaccgggtgggcaatgctgtgtagccataccgttcagcagagccatgatatggcactg
attagccatgtggcaaccctgaaaagcagcattccgtttgttcatTTTTgatggttttcgaccagccacgaagtgaacaaaatca
aaatgctgccgtatgccgaactgaaaaactgggtccgcctggcaccatggaacagcattggggcacgtagcctgaatccgatgc
atccgaccattcgtggcaccatcagagcgcagatatctatttcagaataatggaaagcgccaaccagtattataccgatctggca
gaagtgttcaagaaaccatggatgaagtgcaccgtatattgggtcgtcattacaaaatctttgagtatgttgggtgcaccggatgcag
aagaggtgaccgttctgatgggtagcgggtgccaccaccgttaatgaagcagttgatctgctggttaaacgcggtaaaaaagtgg
tgcagttctgggtcatctgtatcgtccgtgggtcaacaaagcatttgaaaaagtctgcgaaaaaccgtgaaacgtattgcagcact
ggatcgttgc aaagaagtaccgcactggggcgaaccgctgtatctggatgttagcgcaccctgaacctgtttccggaaacgtcag
aatgttaaaagtattgggtgggtcgttatgggtcgggtagcaaaagtattcattccggaacatgcactggccatttatgcaaatctggcaa
gcgaaaaatccgattcagcgtttaccgttgggtattaccgatgatgttaccggcaccagcgtgccgtttgtaataacgtgttgatac
ccgtccggaaaggcaccgcgtcagtggtttttgggggtattggtagtgatggcaccgttggtgcaaatcgtagcgcagttcgtattatt
gggtgataatagcgtatcgaagggtgcagggcgtattttcagtttgatgcattfaaaagcgggtgggttaccagcagccatctgcgtttg
gtcctaaaccgattaccgcacagtatctggtaccaatgcagattatattgctgccactttcaagagtatgtgaaacgttttgataig
ctggatgcaattcgtgaagggtggcacctttgttcgaatagccgttggaccaccgaagataatggaaaaagaaattccggcagatttt
cgtcgtaatgtggcacagaaaaagtgcgcgtttataacgttgatgcccgtaaaattgcgtagctttggtctgggcaaacgcatt
aacatgctgatgcaggcatgtttttcaaactgagcgggtgttcgtccgctggccgaagcacagcgtctgctgaatgaaagcattgtt
catgagtatggcaaaaagggtggtaaagtgggtggaatgaatcaggcagttgtaatgcagtggttgcgggtgatctgcctcaaga
agttcagggttccggcagcatgggcaaatgcagttgataccagcaccgcaccccgaccgggtattgaattgttgataaaatcatgc
gtccgctgatggattcaaagggtgatcagctgccgggtiagcgttatgacaccgggtgggtacatttccgggtggcaccaccagtat
gcaaaacgtgcaattgcggcatttattccgcagtggtatccggcaaaattgtaccagtgtaattattgcagctatgtttgtccgatg
caaccattcgtccgtttgtgctgaccgatcaagaagtgcagctggcaccggaaagcgtttgtaccgtaaaagcaaaagggtgattat
cagggtatgaactttcgtattcaggttgcaccggaaagattgtaccgggtgtcaggtttgtgtgaaacctgtccggatgatgcactgg

aatgaccgatgcgtttaccgccacaccggttcagcgtaccaattgggaatttgcaattaaagtccgaatcgtggtacgatgacc
gatcgctatagcctgaaaggtagccagtttcagcaaccgctgctggaatttagcggtgcatgtgaaggttggtgaaacccgt
atgttaaactgctgaccagctgttgggtgaacgtaccgttattgcaaatgccaccggtgtagcagcatttgggtgttacggcag
gtctggctccgtataccaccaatgcaaaaggtcagggctccggcatggggtaatagcctgttgaagataatgccgaatttggtttg
gtattgcagttgccaatgcacagaaacgtagccgtgttcgtgattgtattctgcaggccgttgaaaaaaagtggccgatgaaggt
ctgaccaccctgctggcacagtggctgcaggattggaataccggtgataaaacactgaaatatcaggaccagattattgccggtc
tggcacagcagcgtagttaaagatcctctgctggaacaaattatggcatgaaagatatgctgccgaatatcagccagtggattatt
ggcgggtgatggttgggccaatgatattggcttgggtggcctggatcatgttctggcgagcggtcagaatctgaatgttctggtgctg
gataccgaaatgtatagcaatacaggtgggtcaggcaagcaaaagcaccatatggcaagcgttgcaaaattgcctgggtggt
aaacgtaccaacaaaaaaacctgaccgaaatggccatgagctatggttaattgttgaaccggttagccatggtaatatggc
ccagtgtgttaaagccttgttgaagcagaaagctatgatgtccgagcctgattgttggttatgcaccgtgcattgaacatggtctg
cgtgcaggatggcacgtatggttcaagaatcagaagcagcaattgcaaccggttattggccactgtatcgtttgatccgcgtctg
gcaaccgaaggtaaaaacccgtttcagctggatagcaaacgtattaaaggtaacctgcaagaatatctggatccgagaatcgtt
atgtgaacctgaaaaaaacaatccgaaaggtgccgatctgctgaaaagccagatggcagataacattacagcacgcttaatcg
ttatcgtcgtatgctggaaggtccgaataccaaagcagcagcaccgagcggtaatcatgtgaccattctgtatggtagtgaacc
ggtaatagcgaaggtctggcaaaagaactggccaccgattttgaacgtcgtgaatatagcgttgacgttcaggccctggatgat
tgatgttgcggatctggaaaatatgggcttgttgttatgtccgtttcaacctgtggtcagggccagtttccgcgtaatagtcagctgtt
ttggcgtgaactgcagcgtgataaacgggaaggttggctgaaaaatctgaaatacaccgttttggcctgggtgatagcacctatt
acttttattgtcataccgccaacaaatcgatgcacgtctggcagcgtgggtgcacagcgtgtgttccgattggttccggtgatga
tggtgatgaagatatgttcataccggcttcaataattggattccgagcgttgggaatgagctgaaaacaaaactccggaagaag
cactgtttaccccgtaattgccgttcagctgacccgaatgcaacaccgcaggattttcatttgcaaaagcacaccggtgctga
gcattaccggtgcagaacgtattacaccggcagatcataccgcaattttgttaccattcgttgaaaaccgatctgagctatcagg
ttggtgatagcctgggtgttttccagaaaatacccgtagcgttgttgaagaattcctgcagtattatggcctgaaccgaaagatgt
tattaccattgaaaataaaggctcacgcgaactgccgcattgtatggccgttgggtgacctgtttacaaaagtcttgatattctgggt
aaaccgaataaccgcttctataaaacccgtgagctatttcgccgttgataaagcagaaaaagaacgcctgctgaaaattgcagaaat
gggtccggaatatagcaacattctgtcagagatgtatcattatgccgacatcttcatatgtttccgagcgcacgtccgacactgca
gtatctgattgaaatgatcccgaacattaaaccggttattatagcattagtagcgcaccgattcatactccgggtgaagtgcatag
cctggttctgattgatacctggattaccctgagcggtaaacatctacgggtctgacctgtaccatgctggaacatctgcaggcag
gtcagggtgggtggtgattgtattcatccgaccgcaatggaatttccgatcatgaaaaaccggttgtatgtgtgcaatgggttcag
gtctggcaccttttgttcatttctcggtgaacgtagcaccttcgtaaacagggttaaaaaaacgggcaatatggcgtgtattttg
gcaatcgttacgaaaaaccgaatttctgatgaaagaggaactgaaaggccatatcaatgatggctgctgacactgcgttgtgca
tttagccgtgatgatccgaaaaaaaagtctatgtgcaggatctgatcaaatggatgaaaaaatgatgtatgattacctgggtgttc

agaaaggcagcatgtattgttggttagccgtagttttatcaaaccgggtgcaagaaagcctgaaacattgtttatgaaagcgggtg
gtctgaccgcagaacaggcagaaaatgaagttattgatatgtttaccacgggtcgcgtataacattgaagcgtggcggccgtaaaa
cgaaaggctcagtcgaaagactg

[0279] NAD⁺ linked *fdh* from *Candida boidinii* and NADP⁺ linked *fdh* from *Burkholderia stabilis* were expressed in *E. coli* TOP10. Biochemical activity measurements were made on cell free extracts, which resulted in the data presented in Table 12 below. The assay was conducted with 50 mM sodium formate and 1.1 mM NAD⁺ or NADP⁺ at pH 7.0 in sodium phosphate buffer, as adapted from Hopner, T. and Knappe, J., *Methods of Enzymatic Analysis*, 3:1551-1555 (1974). In a final volume of 1 mL, 0.55 mL of water, 0.375 mL of 200 mM sodium phosphate, pH 7.0, 0.375 mL of 200 mM sodium formate, 0.15 mL of 10.5 mM β -NAD⁺ or 10.5 mM β -NADP⁺, and 0.05 mL of crude enzyme prep were added to a 1.5 mL plastic cuvette in the order indicated. Absorbance at 340 nm was recorded for 1 minute with a Shimadzu spectrophotometer, and the rate was used to determine specific activity. Protein concentrations were determined by the Bradford method with BSA as the standard. As expected, *fdh* from *C. boidinii* preferred NAD⁺ as a co-factor, while *fdh* from *B. stabilis* preferred NADP⁺.

Table 12. Biochemical activity of Cell-free Extracts

plasmid	description	$\mu\text{mol min}^{-1} \text{mg}^{-1}$	
		NAD ⁺	NADP ⁺
pMU2726	empty vector	0.00 $\pm$ 0.00	0.00 $\pm$ 0.01
FP45	<i>C. boidinii fdh</i>	0.39 $\pm$ 0.04	0.01 $\pm$ 0.01
FP47	<i>B. stabilis fdh</i>	0.06 $\pm$ 0.00	0.31 $\pm$ 0.02

[0280] 2.3 Acyl-ACP chain termination enzymes

[0281] The final step of the anaerobic fatty acid pathway involves cleavage of the acyl carrier protein (ACP) from the acyl chain, and addition of a functional group to the final carbon of the growing chain (Figure 13). The chain termination enzyme(s) determine both the terminal functional group and the overall acyl chain length.

[0282] Plasmids encoding an *E. coli* codon optimized C12 acyl-ACP thioesterase (pMU3061), an *E. coli* codon optimized C16 acyl-ACP thioesterase (pMU3062), an acyl-ACP reductase (pMU3063), and an acyl-ACP reductase homolog (pMU3064) have been

expressed in *E. coli* strain M2933 harboring a deletion in the acyl-CoA dehydrogenase *fadE*, an enzyme involved in fatty acid degradation. The expression plasmids used for these enzymes are shown in Figure 14.

[0283] *E. coli* strains were grown to saturation over 48 hours in 5 mL LB medium at 30°C in aerobic culture tubes containing 100 µg/mL Carbenicillin and 1 mM IPTG. Total fatty acid quantification was performed by lipid extraction followed by methyl ester derivatization and analysis by gas chromatograph with flame ionization detection. Extraction and derivatization was performed by adding 0.5 mL sample to a 13x100mm glass tube with Teflon coated cap, addition of 4 mL 4% sulfuric acid in methanol followed by vortexing. The samples were then incubated at 70°C in a water bath for 30 minutes, cooled to room temperature, followed by addition of 2 mL water and 2 mL hexane with vortexing at each step. The hexane layer was transferred to a new tube and dried under nitrogen. 50 µL hexane was then used to re-constitute the fatty acids for gas chromatograph analysis. Total fatty acids for M2933 strains carrying either plasmid pMU960 (empty vector), pMU3061, pMU3062, pMU3063, or pMU3064 are shown in Figures 22A and 22B. Individual fatty acids are also shown using a standard naming convention of X:Y, where X is the carbon number and Y is the number of unsaturated bonds.

Example 3

[0284] 3.1 Methodology to screen for transcarboxylase activity

[0285] To confirm that putative transcarboxylase genes have *in vivo* oxaloacetate:acetyl-CoA carboxytransferase activity, an *E. coli* strain was constructed that requires recombinant production of malonyl-CoA for growth. Wildtype *E. coli* produces malonyl-CoA, a metabolite essential for growth, exclusively via the enzyme acetyl-CoA carboxylase (ACC). ACC is composed of the four subunit genes *accA*, *accB*, *accC*, and *accD*, which are located at three different loci on the *E. coli* genome.

[0286] Because malonyl-CoA is essential, ACC cannot be disrupted directly in wildtype *E. coli* without resulting in a lethal phenotype. To overcome this, a conditional pathway for malonyl-CoA biosynthesis was first introduced into wildtype *E. coli*. This pathway, encoded by *matBC* from *Rhizobacterium trifolii*, transports exogenous malonate across

the cell membrane, and then uses malonate, ATP, and CoA to produce malonyl-CoA, AMP, and PP_i. See An and Kim, *Eur. J. Biochem.*, 257:395-402 (1998).

[0287] 3.1.1 Construction of strain M2470

[0288] Strain M2470 is a $\Delta accC::matBC$ strain built from *E. coli* K12 strain MG1655 (ATCC Accession No. 700926). To construct M2470, plasmid pMU2737 (Figure 15; SEQ ID NO: 285) was transformed into strain MG1655 with selection on 100 µg/mL ampicillin and 50 µg/mL kanamycin. pMU2737 is a non-replicating plasmid, and confirmation of a single cross over integration was detected via colony PCR. The single cross-over meridioid contains both a functional and a non-functional copy of *accC*, as well as the *matBC* genes, and positive (amp^R, kan^R) and negative selective (*sacB*, *rpsL*) markers. Upon plating on the negative selective condition, 10% w/v sucrose supplemented with 10 mM sodium malonate, the meridioid resolved exclusively to the wildtype, functional *accC* gene copy. This suggested that *matBC* was not able to catalyze the conversion of exogenous malonate to malonyl-CoA at a rate sufficient to allow for observable colony formation. To overcome this, the meridioid strain was grown aerobically in M9 minimal medium supplemented with 1.4 mM glucose and 10 mM malonate. After two transfers, each lasting ~ 48 hours in 50 mL of this medium, the culture was re-plated on 10% w/v sucrose supplemented with 10 mM sodium malonate. Upon screening, most (>90%) of the colony isolates now had the non-functional *accC* copy and *matBC* genes. An isolate was further purified and designated M2470. It is able to grow only when exogenous malonate is present in the medium.

[0289] 3.1.2 Construction and screening of putative transcarboxylase genes

[0290] Plasmids pMU2898 (SEQ ID NO:286), pMU2899 (SEQ ID NO:287), pMU2900 (SEQ ID NO:288), and pMU2901 (SEQ ID NO:289) (Figures 16-19) were constructed via yeast homologous cloning to express putative transcarboxylases from *Clostridium cellulolyticum* H10 ATCC 35319, *Caldicellulosiruptor bescii* DSM 6725, *Thermoanaerobacterium saccharolyticum* JW/SL-YS485, and *Propionibacterium freudenreichii* CIRM-BIA1^T, respectively. Percent consensus and identity positions

across the four subunits of these putative transcarboxylases, as well as from *Corynebacterium kroppenstedtii* DSM 44385, *Geobacter bemidjiensis* Bem^(T), and *Clostridium thermocellum* ATCC 27405, is depicted in Table 13. A phylogenetic tree and alignment of these transcarboxylases, including from *D. propionincus* DSM 2032, is shown in Figures 30-31.

Table 13. Percent Consensus and Identity Positions Across Putative Transcarboxylases

	Consensus positions	Identity positions
5S subunit	75.5%	29.8%
1.3S subunit	58.6%	12.0%
12S subunit	85.0%	35.0%
12S C-terminal subunit*	36.5%	1.9%

* *C. cellulolyticum* does not have a 12S C-terminal subunit

[0291] The four putative transcarboxylases were cloned into pMU2727, a replicating vector with the pBR322 origin, amp^R, Pspc promoter, and T1T2 terminator. Pspc is a moderately high level constitutive ribosomal promoter. See Liang *et al.*, *J. Mol Bio* 292:19-37 (1999).

[0292] These plasmids were then transformed into M2470 and transformants were selected on medium containing, per liter, 10 g glucose, 1.48 g disodium malonate, 100 mg ampicillin, 15 g agar, and the modified M9 base medium: 12.8 g Na₂HPO₄·7H₂O, 3 g KH₂PO₄, 0.5 g NaCl, 1 g NH₄Cl, 0.5 g MgSO₄, 0.015 g CaCl₂, 0.02 g thiamine, 0.02 g CoSO₄, 0.02 g ZnSO₄, 0.02 g MnSO₄, 0.015 g biotin. Transformants were confirmed by plasmid mini-prep, and re-patched onto modified M9 medium plates containing 20 g glucose and 15 g agar per liter ("M9 + 20 glucose"). If growth was observed on M9 + 20 glucose plates, colonies were re-grown in either liquid or solid medium of the same composition, and scored for growth and growth rate (Table 14 and Figure 20A). As a control, transformants were also plated on solid medium comprising M9 base medium, 10 g/L glucose, 10 mM malonate, and 100 µg/mL ampicillin (Table 14 and Figure 20B). The following transformants were isolated and tested for growth: MG1655 – wildtype, M2560 – $\Delta accC::matBC$ + pMU2727 empty vector (amp^R), and M2557, M2558, M2559 – $\Delta accC::matBC$ + pMU2900 *T. saccharolyticum* TC (amp^R).

Table 14. Growth of Transformants Containing Putative Transcarboxylases

Strain	M9 + 20 glu	M9 + malonate + amp
MG1655 WT	++++	-
M2470	-	+++
M2560 (aka. M2470 + pMU2727)	-	+++
M2557 (aka. M2470 + pMU2900) #1	+++	+++
M2558 (aka. M2470 + pMU2900) #2	+++	+++
M2559 (aka. M2470 + pMU2900) #3	+++	+++
M2470 + pMU2898	+	+++
M2470 + pMU2899	++	+++
M2470 + pMU2901	+	+++

++++ = visible growth within 24 hours

+++ = visible growth within 48 hours

++ = visible growth within 96 hours

+ = visible growth within 168 hours

- = no visible growth after 200+ hours

[0293] 3.2 Assays for recombinant transcarboxylase and *in vitro* transcarboxylase activity

[0294] To determine the presence and activity of the *T. saccharolyticum* transcarboxylase enzyme that was engineered into the *E. coli* $\Delta accC::matBC$ strain and screened using the assay above, several biochemical assays were conducted. Initial evaluation of activity in cell lysate was inconclusive. The *T. saccharolyticum* transcarboxylase enzyme was then purified using the biotin binding domain located in the 1.3S protein. Without wishing to be bound by theory, Streptavidin binding of the 1.3S subunit could co-purify both the 5S and 12S proteins which associate with the 1.3S subunit in the native host. *E. coli* $\Delta accC::matBC$ cells were grown in M9+ medium at 37°C in aerobic shake flasks to an OD of 6 in 1.8 L total volume and lysed with Y-PER[®] (Pierce) according to product instructions, in the presence of 100 mM potassium phosphate, pH 6.8, 1 mg/mL reduced glutathione, 1:10,000 dilution of Sigma bacterial protease inhibitors, and 0.5 U/mL DNase I. After 2-3 freeze/thaw cycles, the cells were lysed as determined by microscopic evaluation. The lysate was centrifuged to remove debris and the supernatant was retained for further evaluation of activity. Two constructs were evaluated using this affinity assay, M2557 and M2560, which were either the strain engineered to produce the 12S, 5S, 1.3S,

and 12S C-terminal components of the *T. saccharolyticum* transcarboxylase system or the empty vector control strain, respectively (see above).

[0295] To determine the presence of the biotin-containing enzyme, the lysates were then purified using monomeric avidin resin with a batch binding protocol (Pierce) according to product instructions. After the sample was incubated with the resin, the protein was eluted from the column with 4 mM biotin. The eluted fractions were analyzed on via Western blot with avidin-HRP as the detection. Samples were run on a 4-20% tris glycine gel then transferred to a PVDF membrane. After overnight blocking in TBS/1% BSA, streptavidin HRP was added. The HRP was detected with ECL chemiluminescent and imaged on a chemiluminescent gel doc system. Figure 25 depicts the results of this affinity assay comparing M2557 and M2560.

[0296] The monomeric avidin-purified lysate was purified again with Streptavidin Dynabeads with a batch binding system carried out with 1 mL of lysate mixed with 100 μ L of washed streptavidin Dynabeads. After incubation at room temperature for 30 min, the sample was washed with 100 mM potassium phosphate, 1 mg/mL reduced glutathione, pH 6.8 and eluted from the beads by boiling in SDS-PAGE sample buffer. The re-purified lysate was then analyzed via Western Blot as above. The band on the Western Blot that ran at the same location as the one indicated with an arrow in Figure 25 was sequenced on a Procise N-terminal sequencer. The sequencing data indicated that the N-terminus of the protein was MKKFIVTVNG (SEQ ID NO:299), consistent with the N-terminus of the 1.3S protein.

[0297] The enzymatic activity of the monomeric avidin-purified transcarboxylase complex was then assessed using an LC/MS detection assay. The monomeric avidin purified lysate was mixed with oxaloacetate, acetyl CoA and reduced glutathione and incubated at 40°C for 1.5 hours. The sample was then analyzed by LC/MS using a BioRad 87H column and a Thermo LCQ (HPLC C₁₈ column-formate/methanol eluent) ion trap mass spectrometer. The results are shown in Figure 26A and 26B. In Figure 26A, the negative control sample was analyzed. Using selected ion monitoring, acetyl CoA was detected but no malonyl CoA was detected (Figure 26A, lower two panels). When the transcarboxylase sample was analyzed with selected ion monitoring both acetyl and malonyl CoA (Figure 26B, lower two panels) were detected thereby indicating that the transcarboxylase enzyme complex was functional.

[0298] 3.3 Use of *E. coli* $\Delta accC::matBC$ strain to select for faster-growing transcarboxylase-expressing strains

[0299] The *E. coli* $accC::matBC$ strain M2470 can also be used to select for more efficient malonyl-CoA production by transcarboxylases. This selection is based on the principle that malonyl-CoA generation is the rate-limiting factor for growth of this strain. Thus, more efficient generation of malonyl-CoA will result in a faster growing strain which is able to out-compete the remaining culture and dominate the cell population during serial transfer or other continuous or semi-continuous selection systems. *See, e.g.*, Figure 27A.

[0300] First, strain M2470 was transformed with an integrating plasmid (*e.g.*, pMU2924, pMU2969) carrying a transcarboxylase and spectinomycin antibiotic resistance marker flanked by DNA regions homologous to the *ldh* gene (lactate dehydrogenase). Using kanR, ampR, sacB, and rpsL based selections, the transcarboxylase and specR marker were securely integrated into the genome via two homologous recombination events. During this period, the strain was grown on M9+ base medium with the addition of 2-20 g/L glucose and 1.48 g/L disodium malonate. The medium was prepared at room temperature, adjusted to pH 7.5 with 10 M NaOH or 10 M HCl, and filter sterilized into a pre-sterilized bottle with a 0.22 μ m filter. Subsequently, the strain was grown aerobically at 37°C in 350 mL of M9+ medium with only glucose in a 1 L shake flask. If substantial growth ($OD > 1$) occurred, a 0.1 mL transfer was made to a fresh 350 mL flask, which is repeated 3 times, at which point a small culture volume is plated to isolate a single colony on solid M9+ glucose medium (prepared via addition of 15 g/L melted agar as a 2X stock to 2X liquid media, pre-incubated at 50°C). *See* Figure 27A. This strain is referred to as the 3rd transfer (T3) isolate. Growth rates for the original strain and T3 isolate were then compared in M9+ medium with 20 g/L glucose as the sole carbon and energy source. An increased growth rate indicates an improved ability to generate malonyl-CoA. Plasmids used for this example were pMU2924 (*T. saccharolyticum* TC; Figure 28; SEQ ID NO:207) and pMU2969 (*P. freudenreichii* TC; Figure 29; SEQ ID NO:208), which generated strains M2767 and M2772, respectively. Growth rates for the original strain and T3 strain of each are shown in Figure 27B.

Example 4

[0301] 4.1 High Yield Palmitic Acid Production in *S. cerevisiae*

[0302] The present prophetic example describes the engineering of a recombinant yeast microorganism to convert a native pyruvate decarboxylase (*pdh*) based ethanol pathway (Figure 34A) to an intermediary pyruvate formate lyase and alcohol/aldehyde dehydrogenase (*pfl adhE*) based ethanol pathway (Figure 34B), and finally to a transcarboxylase based palmitic acid pathway (Figure 34C).

[0303] The genetic modifications described below are used to create a strain capable of anaerobic growth in the absence of functional pyruvate decarboxylase and glycerol-3-phosphate dehydrogenase. To accomplish this, constructs were designed to replace GPD1, GPD2, FDH1, and FDH2 with two copies of a bifunctional alcohol/aldehyde dehydrogenase and two copies of a pyruvate formate lyase, both of which were cloned from *B. adolescentis* (Table 15). See, e.g., PCT/US2011/035416, which is incorporated by reference herein in its entirety, for additional details on the construction of such strain. Additionally, constructs were designed to make deletions of PDC5, PDC6, and PDC1. Either a NAD⁺ or NADP⁺ linked formate dehydrogenase is then re-introduced into the strain to create the metabolic pathway shown in (Figure 34B).

Table 15. Coding sequences of *pfl* and *adhE*

GenBank Accession #	Host strain	Gene donor	gene	protein
YP_909854	<i>S. cerevisiae</i>	<i>Bifidobacterium adolescentis</i>	<i>pflA</i>	pyruvate formate lyase activating enzyme
YP_909855	<i>S. cerevisiae</i>	<i>Bifidobacterium adolescentis</i>	<i>pflB</i>	pyruvate formate lyase
YP_909182	<i>S. cerevisiae</i>	<i>Bifidobacterium adolescentis</i>	<i>adhE</i>	alcohol/aldehyde dehydrogenase

[0304] *B. adolescentis adhE* (amino acid sequence) (SEQ ID NO:209)

MADAKKKEPTKPTPEEKLA AA EAEVDALVKKGLKALDEF EKLDQKQVDHIVA
KASVAALNKHVLAKMAVEETHRGLVEDKATKNIFACEHVTNYLAGQKTVGII
REDDVLGIDEIAEPVGVVAGVTPVTNPTSTAIFKSLIALKTRCPIIFGFHPGAQNC
VAAAKIVRDAAIAAGAPENCIQWIEHPSIEATGALMKHGDGVATILATGGPGMVK
AAYSSGKPALGVGAGNAPAYVDKNVDVVRANDLILSKHFDYGMICATEQAIIA

DKDIYAPLVKELKRRKAYFVNADEKAKLEQYMFGCTAYSGQTPKLNSVVPKGK
 PQYIAKAAGFEIPEDATILAAECKEVEGENEPLTMEKLAPVQAVLKSDNKEQAFEM
 CEAMLKHGAGHTAAIHTNDRDLVREYQORMHACRIIWNSSSLGGVVDIYNAIA
 PSLTLGCGSYGGNSVSGNVQAVNLINIKRIARRNNNMQWFKIPAKTYFEPNAIKY
 LRDMYGIEKAVIVCDK'VMEQLGIVDKIIDQLRARSNRVTFRIIDYVEPEPSVETVE
 RGAAMMREEFEPDTIIAVGGGSPMDASKIMWLLYEHPEISFSDVREKFFDIRKRA
 FKIPPLGKKAKLVCIPTSSGTGSEVTPFAVITDHKTGYKYPITDYALTPSVAIVDPV
 LARTQPRKLASDAGFDALTHAFEAYVSVYANDFTDGMALHAAKL VWDNLAES
 VNGEPGEEKTRAQEKMHNAATMAGMAFGSAFLGMCHGMAHTIGALCHVAHG
 RTNSILLPYVIRYNGSVPEEPTSWPKYNKYIAPER YQEIAKNLGVNPGKTPEEGVE
 NLAKAVEDYRDNKLGMNKS FQECGVDEDDYYWSIIDQIGMRAYEDQCAPANPRI
 PQIEDMKDIAIAAAYYGVSSQAEGHKL RVQRQGEAATEEASERA

[0305] *B. adolescentis pflA* (amino acid sequence) (SEQ ID NO:210)

MSEHIFRSTTRHMLRDSKDYVNQTLMGGLSGFESPIGLDRLDRIKALKSGDIGFV
 HSWDINTSV DGPGRMTVFMMSGCPLRCQYCQNPDTWKMRDGKPVYYEAMVKK
 IERYADLFKATGGGITFSGGESMMQPAFVSRVFHAAKQMGVHTCLDTSGFLGAS
 YTDDMVDDIDLCLLDVKS GDEETYHKVTGGILQPTIDFGQRLAKAGKKIWVRFV
 LVPGLTSSEENVENVAKICETFGDALEHIDVLPFHQLGRPKWHMLNIPYPLEDQK
 GPSAAMKQRVVEQFQSHGFTVY

[0306] *B. adolescentis pflB* (amino acid sequence) (SEQ ID NO:211)

MAAVDATAVSQEELEAKAWEGFTEGNWQKDIDVRDFIQKNYTPYEGDESFLAD
 ATDKTKHLWKYLLDDNYLSVERKQRVYDVDTHTPAGIDAF PAGYIDSPEVDNVIV
 GLQTDVPCKRAMMPNGGWRMVEQAIKEAGKEPDPEIKKIFTKYRKTHNDGVFG
 VYTKQIKVARHNKILTGLPDAYGRGRIIGDYRRVALYGVNALIKFKQRDKDSIPY
 RNDLTEPEIEHWIRFREEHDEQIKALKQLINLGNEYGLDLSRPAQTAQEA VQWTY
 MGYLASVKSQDGAAMSFG RVSTFFDVYFERDLKAGKITETDAQEIIDNLVMKLR
 IVRFLRTKDYDAIFSGDPYWATWSDAGFGDDGRTMVTKTSFRLLNTLTLEHLGP
 GPENITIFWDPKLPEAYKRFCARISIDTSAIQYESDKEIRSHWGDDAAIACCVSPM
 RVGKQMQFFAARVNSAKALLYAINGGRDEMTGMQVIDKGVIDPIKPEADGTLD
 YEKVKANYEKALEWLSETYVMALNIIHYMHDKYAYESIEMALHDKEVYRTLGC
 GMSGLSIAADSLSACKYAKVYPIYNKDAKTTTPGHENEYVEGADDDLIVGYRTEG
 DFPLYGNDDDRADDIAKWVVSTVMGQVKRLPVYRDAVPTQSILTITSNVEYGKA
 TGAFPSGHKKGTPYAPGANPENGMDSHGMLPSMFSVGKIDYNDALDGISLTNTI
 TPDGLGRDEEERIGNLVGILDAGNGHGLYHANINVL RKEQLEDAVEHPEKYPHL
 TVRVSGYAVNFVKLTKEQQLDVISRTFHQGA VVD

[0307] To generate a recombinant yeast microorganism as described in this example, individual molecular components are integratively assembled.

[0308] 1) The deletion of the FDH1 gene and replacement with two copies of ADH and two copies of PFL is illustrated in Figure 35. The primers used to generate the molecular components of this integrative assembly are shown in Figure 35, Table 16, and Table 26 (below).

Table 16. Primers used to generate the integrative assembly of Figure 35. Each column indicates a PCR fragment that needs to be generated.

Name	FDH15' Flank rc	pTPI-ADH-FBA _t	ADHpPFK _{rc}	PFL-pADH5	PFL -pENORC	FDH13' Flank rc
Primers	X15559/X15565	X15564/X14843	X14844/X14835	X14836/X14837	X14838/X15567	X15566/X15553
Template	<i>S. cerevisiae</i> gDNA	pMU2746	pMU2745	pMU2770	pMU2606	<i>S. cerevisiae</i> gDNA

[0309] 2) The deletion of the FDH2 gene and replacement with two copies of ADH and two copies of PFL is illustrated in Figure 36. The primers used to generate the molecular components of this integrative assembly are shown in Figure 36, Table 17, and Table 26 (below).

Table 17. Primers used to generate the integrative assembly of Figure 36. Each column indicates a PCR fragment that needs to be generated.

Name	FDH2 5' Flank	FBA(t)-ADH1	PFK-ADH2	PFL-F	PFL-R	FDH2 3' Flank
Primers	X16096/X16097	X16098/X14843	X14844/X14835	X14836/X14837	X14838/X16099	X16100/X11845
Template	<i>S. cerevisiae</i> gDNA	pMU2746	pMU2745	pMU2770	pMU2606	<i>S. cerevisiae</i> gDNA

[0310] 3) The deletion of the GPD2 gene and replacement with two copies of ADH and two copies of PFL is illustrated in Figure 37. The primers used to generate the molecular components of this integrative assembly are shown in Figure 37, Table 18, and Table 26 (below).

Table 18. Primers used to generate the integrative assembly of Figure 37. Each column indicates a PCR fragment that needs to be generated.

Name	GPD2 5' flank	FBA(t)-ADH1	PFK-ADH2	PFL-F	PFL-R	GPD2 3' flank
Primers	X11816 / X14847	X14845 / X14843	X14844 / X14835	X14836 / X14837	X14838 / X14849	X14850 / X11821
Template	<i>S. cerevisiae</i> gDNA	YCL150	YCL149	pMU2770	pMU2760	<i>S. cerevisiae</i> gDNA

[0311] 4) The deletion of the GPD1 gene and replacement with two copies of ADH and two copies of PFL is illustrated in Figure 38. The primers used to generate the molecular components of this integrative assembly are shown in Figure 38, Table 19, and Table 26 (below).

Table 19. Primers used to generate the integrative assembly of Figure 38. Each column indicates a PCR fragment that needs to be generated.

Name	GPD1 5' flank	FBA(t)-ADH1	PFK-ADH2	PFL-F	PFL-R	GPD1 3' flank
------	---------------	-------------	----------	-------	-------	---------------

Primers	X11824/X14776	X14775/X14843	X14844/X14835	X14836/X14837	X14838/X14829	X14778/X11829
Template	<i>S. ce</i> gDNA	pMU2746	pMU2745	pMU2770	pMU2606	<i>S. ce</i> gDNA

[0312] 5) The deletion of the PDC5 gene and replacement with a counter selective gene HSV-TDK and an antibiotic marker (Kan) is illustrated in Figure 39. The primers used to generate the molecular components of this integrative assembly are shown in Figure 39, Table 20, and Table 26 (below).

Table 20. Primers used to generate the integrative assembly of Figure 39. Each column indicates a PCR fragment that needs to be generated.

Name	PDC5 5' Flank	KNT	PDC5 3' Flank
Primers	X16463/X16464	X16467/X16468	X16465/X16466
Template	<i>S. ce</i> gDNA	M2543/TB396	<i>S. ce</i> gDNA

[0313] 6) The removal of the marker shown in Figure 39 resulting in a clean deletion of the PDC5 gene is illustrated in Figure 40. The primers used to generate the molecular components of this integrative assembly are shown in Figure 40, Table 21, and Table 26 (below).

Table 21. Primers used to generate the integrative assembly of Figure 40. Each column indicates a PCR fragment that needs to be generated.

Name	PDC5 5' Flank	PDC5 3' Flank
Primers	X16463/X16495	X16494/X16466
Template	<i>S. ce</i> gDNA	<i>S. ce</i> gDNA

[0314] 7) The deletion of the PDC6 gene and replacement with a counter selective gene HSV-TDK and an antibiotic marker (Kan) is illustrated in Figure 41. The primers used to generate the molecular components of this integrative assembly are shown in Figure 41, Table 22, and Table 26 (below).

Table 22. Primers used to generate the integrative assembly of Figure 41. Each column indicates a PCR fragment that needs to be generated.

Name	PDC6 5' Flank	KNT	PDC6 3' Flank
Primers	X16471/X16472	X16475/X16476	X16473/X16474
Template	<i>S. ce</i> gDNA	M2543/TB396	<i>S. ce</i> gDNA

[0315] 8) The removal of the marker shown in Figure 41 resulting in a clean deletion of the PDC6 gene is illustrated in Figure 42. The primers used to generate the molecular components of this integrative assembly are shown in Figure 42, Table 23, and Table 26 (below).

Table 23. Primers used to generate the integrative assembly of Figure 42. Each column indicates a PCR fragment that needs to be generated.

Name	PDC6 5' Flank	PDC6 3' Flank
Primers	X16471/X16497	X16496/X16474
Template	<i>S.ce</i> gDNA	<i>S.ce</i> gDNA

[0316] 9) The deletion of the PDC1 gene and replacement with a counter selective gene HSV-TDK and an antibiotic marker (Kan) is illustrated in Figure 43. The primers used to generate the molecular components of this integrative assembly are shown in Figure 43, Table 24, and Table 26 (below).

Table 24. Primers used to generate the integrative assembly of Figure 43. Each column indicates a PCR fragment that needs to be generated.

Name	PDC1 5' Flank	KNT	PDC1 3' Flank
Primers	X16951/X16952	X16953/X16954	X16955/X16956
Template	<i>S.ce</i> gDNA	M2543/TB396	<i>S.ce</i> gDNA

[0317] 10) The removal of the marker shown in Figure 43 resulting in a clean deletion of the PDC1 gene is illustrated in Figure 44. The primers used to generate the molecular components of this integrative assembly are shown in Figure 44, Table 25, and Table 26 (below).

Table 25. Primers used to generate the integrative assembly of Figure 44. Each column indicates a PCR fragment that needs to be generated.

Name	PDC1 5' Flank	PDC1 3' Flank
Primers	X16952 / X16953	X16954 / X16955
Template	<i>S.ce</i> gDNA	<i>S.ce</i> gDNA

Table 26. Primer sequences used to create the integrative assemblies illustrated in Figures 35-44.

Primer	SEQ ID	Primer sequence
--------	--------	-----------------

#	NO.	
X11316	212	GTAATACATCACCTCGATGAAAGAGA
X11816	213	GCAGTCATCAGGATCGTAGGAGATAAGCA
X11821	214	TCACAAGAGTGTGCAGAAATAGGAGGTGGA GTTGGGGGAAAAAGAGGCAACAGGAAAGATCAGAGACAGCAA
X11822	215	GCATTGATAAGGAAGGG CCCTTCCTTATCAATGCTTGCTGTCTCTGATCTTTCCTGTTGCCTC
X11823	216	TTTTTCCCCCAAC
X11824	217	AAGCCTACAGGCGCAAGATAACACATCAC
X11829	218	CTCAGCATTGATCTTAGCAGATTCAGGATCTAGGT TATGTTATCTTTCTCCAATAAATCTAATCTTCATGTAGACTATCA
X11830	219	GCAGCAGCAGACAT GATAATATAAAGATGTCTGCTGCTGCTGATAGTCTACATGAAGA
X11831	220	TTAGATTTATTGGAG
X11845	221	TTACTTGTGAAACTGTCTCCGCTATGTCAG CCCCCTCCACAAACACAAATATTGATAATATAAAGATGGCAGAC
X14775	222	GCAAAGAAGAAGGAA
X14778	223	ATTTATTGGAGAAAGATAACATATCATACTTTCC GAAAGTATGATATGTTATCTTTCTCCAATAAATCTAGTCTTCTAG
X14829	224	GCGGGTTATCTACT CAAATTCTAACCAACTTCAAAATGACATAGTACCTCATCTATAA
X14835	225	TTTTTACCCTGATCT AGTTAGATCAGGGTAAAAATTATAGATGAGGTACTATGTCATTT
X14836	226	TGAAGTTGGTTAGAA GGTCCATGTAAAATGATTGCTCCAATGATTGAAATTGATTCAGG
X14837	227	TCAAAATGGATTCAG ACGTCCCTGAATCCATTTTGACCTGAATCAATTTCAATCATTGGA
X14838	228	GCAATCATTTTACA GGTGAACCATTTACTGTATTTTCAATGTAACGCTAGAGAATAA
X14843	229	ATTCAAGTTAAAAGA CATCATCTTTTAACTTGAATTTATTCTCTAGCGTTACATTGAAAA
X14844	230	TACAGTAAATGGTT
X15380	231	TAGGTCTAGAGATCTGTTTAGCTTGC
X15382	232	GAGACTACATGATAGTCCAAAGA GGACGAGGCAAGCTAAACAGATCTCTAGACCTACTTTATATTAT
X15546	233	CAATATTTGTGTTTG CCGTTTCTTTTCTTTGGACTATCATGTAGTCTCATTTATTGGAGA
X15547	234	AAGATAACATATCA GGACGAGGCAAGCTAAACAGATCTCTAGACCTATGATAAGGAA
X15548	235	GGGGAGCGAAGGAAAA CCGTTTCTTTTCTTTGGACTATCATGTAGTCTCCTCTGATCTTTCC
X15549	236	TGTTGCCTCTTTT CCGTTTCTTTTCTTTGGACTATCATGTAGTCTCGAGTGATTATGA
X15552	237	GTATTTGTGAGCAG
X15553	238	ACCAGCGTCTGGTGGACAAACGGCCTTCAAC GGACGAGGCAAGCTAAACAGATCTCTAGACCTAATTAATTTTCA
X15554	239	GCTGTTATTTTCGATT
X15555	240	CCGTTTCTTTTCTTTGGACTATCATGTAGTCTCGAGTGATTATGA

		GTATTTGTGAGCAG
X15559	241	GGAAGGCACCGATACTAGAACTCCG
		CTAATCAAATCAAATAACAGCTGAAAATTAATCTACTTATTCC
X15564	242	CTTCGAGATTATATC
		G TTCCTAGATATAATCTCGAAGGGAATAAGTAGATTAATTTTCA
X15565	243	GCTGTTATTTTGATT
		TCGGATCAGTAGATAACCCGCCTAGAAGACTAGGAGTGATTATG
X15566	244	AGTATTTGTGAGCAG
		AAAAC TTCTGCTCACAAATACTCATAATCACTCCTAGTCTTCTAG
X15567	245	GCGGGTTATCTACT
		CTAATCAAATCAAATAACAGCTGAAAATTAATGAGTGATTATG
X15870	246	AGTATTTGTGAGCAG
		AAAAC TTCTGCTCACAAATACTCATAATCACTCATTAATTTTCAG
X15871	247	CTGTTATTTTGATT
X16096	248	CATGGTGCTTAGCAGCAGATGAAAGTGTCA
		G TTCCTAGATATAATCTCGAAGGGAATAAGTAGATTAATTTTCA
X16097	249	GCTGTTATTTTCGATT
		CTAATCAAATCGAAATAACAGCTGAAAATTAATCTACTTATTCC
X16098	250	CTTCGAGATTATATC
		AAAAC TTCTGCTCACAAATACTCATAATCACTCCTAGTCTTCTAG
X16099	251	GCGGGTTATCTACT
		TCGGATCAGTAGATAACCCGCCTAGAAGACTAGGAGTGATTATG
X16100	252	AGTATTTGTGAGCAG
X16463	253	CAGAGTTTGAAGATATCCAAATGGT
X16464	254	TTTGTTCTTCTTGTTATTGTATTGTGTTG
X16465	255	GCTAATTAACATAAACTCATGATTCAACG
X16466	256	ACATAGGTTTGCAAGCTTTATAATCTG
		AGAACAACACAATAACAATAACAAGAAGAACAATAAGGTCTAGA
X16467	257	GATCTGTTTAGCTTGC
		AAACGTTGAATCATGAGTTTTATGTTAATTAGCGAGACTACATG
X16468	258	ATAGTCCAAAGAAAA
		AGAACAACACAATAACAATAACAAGAAGAACAACACTACTTATTC
X16469	259	CCTTCGAGATTATATC
		AAACGTTGAATCATGAGTTTTATGTTAATTAGCCTAGTCTTCTAG
X16470	260	GCGGGTTATCTACT
X16471	261	AAGAATCTGTTAGTTCGAACTCCAG
X16472	262	TTTGTTGGCAATATGTTTTTGCTATATTAC
X16473	263	GCCATTAGTAGTGTACTCAAACGAA
X16474	264	ACGACTCAACATATGTATGTTGCT
		CACGTAATATAGCAAAAACATATTGCCAACAAATAGGTCTAGA
X16475	265	GATCTGTTTAGCTTGC
		AACAATAATTCGTTTGAGTACACTACTAATGGCGAGACTACATG
X16476	266	ATAGTCCAAAGAAAA
		CACGTAATATAGCAAAAACATATTGCCAACAACTACTTATTCC
X16477	267	CTTCGAGATTATATC
		AACAATAATTCGTTTGAGTACACTACTAATGGCCTAGTCTTCTA
X16478	268	GGCGGGTTATCTACT
X16951	269	ATGTTCCGCTGATGTGATGTGCAAGATAAAC

X16952	270	GAGGCAAGCTAAACAGATCTCTAGACCTATTTGATTGATTGAC TGTGTTATTTTGCCT
X16953	271	ATAACCTCACGCAAAATAACACAGTCAAATCAATCAAATAGGT CTAGAGATCTGTTTAG
X16954	272	AAAACCTTTAACTAATAATTAGAGATTAAATCGCTTAGAGACTAC ATGATAGTCCAAAGA
X16955	273	GTCCCCCGTTTCTTTCTTTGGACTATCATGTAGTCTCTAAGCG ATTTAATCTCTAAT
X16956	274	TCGGTCATTGGGTGAGTTTAAGCATTAGCAGCAATG TAAACTTTAACTAATAATTAGAGATTAAATCGCTTATTTGATT
X16957	275	GATTTGACTGTGTTA CACGCAAAATAACACAGTCAAATCAATCAAATAAGCGATTAA
X16958	276	TCTCTAATTATTAGTT

[0318] Heterologous genes for the production of a transcarboxylase based palmitic acid pathway (Figure 34C) can then be introduced in a yeast microorganism engineered using the above integrative assemblies to replace GPD1, GPD2, FDH1, and FDH2 with two copies of a bifunctional alcohol/aldehyde dehydrogenase and two copies of a pyruvate formate lyase and to delete PDC5, PDC6, and PDC1. Such heterologous genes include, but are not limited to, *S. cerevisiae* NAD⁺ FDH1 to create the metabolic pathway in (Figure 34B) and *B. stabilis* NADP⁺ FDH, *S. cerevisiae* PCK1, *P. freudenreichii* Transcarboxylase (see SEQ ID NOs:6-16), *A. thaliana* FATB1 to create the metabolic pathway in (Figure 34C). Additional enzymes are identified in PCT/US2011/035416, which is incorporated by reference herein in its entirety. The pathways described herein can be engineered for production of a malonyl-CoA derived product in the yeast cytosol.

[0319] >SceNAD⁺_FDH1 (SEQ ID NO:277)

atgtcgaagggaaggttttgctggttcttacgaaggtgtaagcatgctgaagagcaggaaaagttattggggtgtattgaaaat
gaacttggtatcagaaatttcattgaagaacagggtatcgagttggttactaccattgacaaggaccctgagccaacctcaacggt
agacaggaggtgaaagacgctgaaattgtcattactacgcccttttccccgcctacatctcgagaaacaggattgcagaagctc
ctaacctgaagctctgtgtaaccgctggcgtcggttcagaccatgtcgatttagaagctgcaaatgaacggaaaatcacgggtcac
cgaagtactggttctaactgctgttctgtcgagagcacgttatggccacaatttggtttgataagaaactataatggtggtcatc
aacaagcaattaatggtgagtggtgattgcccggcgtggctaaaaatgagtatgatctggaagacaaaataattcaacggtaggt
gccggtagaattggatatagggttctggaaagattggtcgcatftaatccgaagaagtactgtactacgactaccaggaactacct
gcggaagcaatcaatagattgaacgaggccagcaagcttttcaatggcagaggtgatattgttcagagagtagagaaattggag
gatattggtgctcagtcagatgttgttaccatcaactgtccattgcacaaggactcaaggggtttattcaataaaaagcttattccca
catgaaagatggtgcatacttggtgaataccgctagaggtgctatttgtgtcgagaagatgttccgaggcagtcagcttggtta
aattggctggctatggtggtgatgtctgggataagcaaccagcaccaaaagaccatccctggaggactatggacaataaggacc

acgtgggaaacgcaatgactgttcataatcagtggcacatctctggatgctcaaaagaggtagctcagggagtaaagaacatcct
aaatagttacttttccaaaaagtttgattaccgtccacaggatattattgtgcagaatgggttcttatgccaccagagcttatggacaga
agaaa

[0320] >SceNAD₊_FDH1 (SEQ ID NO:278)

MSKGKVLVLYEGGKHAEQEKLKGCIENELGIRNFIEEQGYELVTTIDKDPEPTS
TVDRELKDAEIVITTPFFPAYISRNRIAEAPNLKLCVTAGVGS DHVDLEAANERKI
TVTEVTGSNVVSVAEHVMATILVLIRNYNGGHQQAINGEWDIAGVAKNEYDLED
KIISTVGAGRIGYRVLERL VAFNPKKLLYYDYQELPAEAINRLNEASKLFNGRGDI
VQRVEKLED MVAQSDVVTINCPLHKDSRGLFNKKLISHMKDGAYLVNTARGAIC
VAEDVAEAVKSGKLAGYGGDVWDKQPAPKDHPWRTMDNKHVGNAMTVHIS
GTSLDAQKRYAQGVKNILNSYFSKKFDYRPQDIIVQNGSYATRAYGQKK

[0321] >BstabilisNADP₊_FDH (SEQ ID NO:279)

atggctaccgtttgtgtgtcttgatccagatccagttgatgggtatccaccacattatgtagagataccattccagttattaccagat
acgctgatggtaaactgtccaactccagctgggtccaccaggttttagaccaggtgaattgggtgtctgtttctgggtctttgggt
ttgagagggtatttgaagctcatggctactttgatcggtacctctgataaggatgggtccagattctgaattcgaaagaagattgcc
agacgccgatgtgttatttctcaaccatttggccagcttacttgaccgctgaaagaattgctagagcaccaaaattgagattggctt
tgactgctgggtattgggtctgatcatgttgattggatgctgctgctagagcccatattactgttctgaagtactgggtccaactctatt
tcagttgccgaacacgtgttatgactactttggcttgggtcagaactacttgccatctcatgctattgctcaacaaggtgggtggaat
attgctgattgtgtctctagatcctacgatgtgaaggatgcattttgggtactgttggtgctggtagaattgggtgggtgtttgagaa
gattgaagccatttgggttacacttgactacacccaaagacatagattggatgcagctatcgaaagaattgggtttaacttatca
tgctgatccagcttcattggctgctgctgttgatagtttaacttgcaaatcccattatacccatccaccgaacatttgttgatgctgct
atgattgctagaatgaagagaggtgcatacttgattaacaccgctagagctaaattgggtgatagagatgctgttgtagagctgtta
cttctggctatttggctgggtatgggtggtgatgttgggttccacaaccagctccagctgatccttggagagctatgccttttaag
gtatgactccacatatctccggtacatcttctgtctgctcaagctagatatgctgctggtactttggaaatattgcaatgttgggttgacg
gtagaccaatcagaacgaatatttgattgtcgacgggtggtacttttagctgggtactgggtgctcaatcttacagattaact

[0322] >BstabilisNADP₊_FDH (SEQ ID NO:280)

MATVLCVLYPDPVDGYPPHYVRDTPVITRYADGQTAPTPAGPPGFRPGELVGSV
SGALGLRGYLEAHGHTLIVTSDKDGPDSEFERRLPDADVVISQPFWPAYLTAERI
ARAPKLRLALTAGIGSDHVDLDAAARAHITVAEVTGSNSISVAEHVVMTTLALV

RNYLPSHAIAQQGGWNIADCVSRSYDVEGMHFGTVGAGRIGLAVLRRLKPFGLH
LHYTQRHRLDAAIEQELGLTYHADPASLAAVDIVNLQIPLYPSTEHLFDAAMIA
RMKRGAYLINTARAKLVDRDAVVRAVTSGHLAGYGGDVWFPPQAPADHPWRA
MPFNGMTPHISGTSLSAQARYAAGTLEILQCWFDGRPIRNEYLIVDGGTLAGTGA
QSYRLT

[0323] >ScePCK1 (SEQ ID NO:281)

atgtcccttctaaaatgaatgctacagtaggatctacttccgaagttgaacaaaaatcagacaagaattggctcttagtgacgaa
gtcaccaccatcagacgcaatgctccagctgccgtttgtatgaagatggctctaaaagaaaaataaaactgtcatttcacaaagcggt
gcattgatcgttattccggtgttaaaaccggaagatctccaaaggacaaacgtattgtgaagaacctacctcgaaagacgaaat
ttgggtgggtccggtcaataaaccatgttctgaaagaacatggctctatcaaccgtgaaagagctgcagattacttgagaacaaga
gaccacatttatattgtcgatgcatttcaggatgggatccaaatacagaatcaaagtcgcgtgtttgtgccagggcttaccac
gcttattcatgacaaatatgcttattagacctacagaagaagaattagccattttggagaacctgattttactgtctggaacgctgg
tcagttcccagccaatttacacaccaggatattgtctcaagagtactatagaaattaacttcaaagcaatggaaatgatcatttag
gtaccgaatacgcgggtgaaatgaaaaaggatatttcacagttatgtttacttgatgcctgtgcaccataacgttttaacttgcact
ctccgccaaccagggtattcaaacgggtgacgttactttattcttggcctaagtgtaccgggaaaaccatttatccgcagacc
cacatagattgttgatggcgatgatgaacattgttggtccgacctgggtgtcttcaatatcgaagggtggtgttacgccaagtgtatt
aatttatctgccgaaaaggagcctgaaatttcgacgctatcaagtttggttctgtattagaaaacgttatctatgacgagaagtcgca
ttagtgcactatgacgactcttctattactgaaaatactagatgtgcctacccaattgactacattccaagtccaagattccatgttt
ggcggactctcatcaaagaacattatcctgctaacttgtatgcttcgggtgtttaccaccagtatctaaattgactcctgaacaa
gtcatgtaccatttcctctgtgttacacttctaaaatggctggtagtgagcaaggtgtcactgaacctgaaccaacatttcattgt
ttcggacaacccttctagccttgaccctattagatacgaacatgttagctacaaagatgtctcaacataaagctaatacgctact
taatcaacaccggctggactggttcttctacgtatctgggtgtaaacgttgccattgaagtacacaagggccattctggattctat
tcatgatggttcgttagccaatgaaacgtacgaaactttaccgattttcaatcttcaagtacctaccaaggttaacgggtgtccagctg
agcttttgaatcctgctaaaaactggctcgaaggtgaatccaaatacagaggtgcagttaccaacttgccaactgtttgttcaaaa
tttcaagatttatcaagacagagccacaccagatgtattagccgctggctcctcaattcgag

[0324] >ScePCK1 (SEQ ID NO:282)

MSPSKMNATVGSTSEVEQKIRQELALSDEVTTIRRNAPAAVLYEDGLKENKTVIS
SSGALIAYSKVKTGRSPKDKRIVEEPTSKDEIWWGPVKNKPCERTWSINRERAAD
YLRTDHIYIVDAFAGWDPKYRIKVRVVCARAYHALFMTNMLIRPTEEEELAHFG
EPDFTVWNAGQFPANLHTQDMSSKSTIEINFKAMEMIILGTEYAGEMKKGIFTVM

FYLMPVHHNVLT LHSSANQGIQNGDVT LFFGLSGTGKTTLSADPHRLLIGDDEHC
WSDHGVFNIEGGCYAKCINLSAEKEPEIFDAIKFGSVLENNVIYDEKSHVVDYDDSS
ITENTRCAYPIDYIPSAKIPCLADSHPKNIILLTCDASGVLPPVSKLTPEQVMYHFIS
GYTSKMAGTEQGVTEPEPTFSSCFGQPFLALHPIRYATMLATKMSQHKANAYLIN
TGWTGSSYVSGGKRCPLKYTRAILDSIHGSLANETYETLPIFNLQVPTKVNGVP
AELLNPAKNWSQGESKYRGAVTNLANLFVQNFKIYQDRATPDVLAAGPQFE

[0325] >Ath_FATB1_mature_peptide (SEQ ID NO:283)

atgcttgattggaaacctagggcgttctgacatgctggggatcctttgggtatagggagaattgtcaggatggcctgtgttccgca
gaattttctattagggtcatatgaaataggtgctgacgctctgcacatagaaaccgtcatgaatcatctgcaggaaacggcgctta
atcatgttaagactgctggattgctggagatgggttggctctacacctgagatgtttaagaagaactgatatgggtgtcactcgt
atgcagggtgtggttgataaatcctacttggggagatgttgtgaagtagacacctgggtcagtcacatctggaaagaatggtatg
cgctgtgattggctagtccgggttgaatactggagaaaccttaacacgagcatcaagtgtgtgggtgatgatgaataaactgac
aaggagatgtcaaagattcctgaagaggttcgaggggaaatagagccttatttgtgaattctgactcgtccttggcgaggacag
cagaaagttaacaaaaattgatgacaagactgctgaciatgttcgatctggctcactcctcgatggagtgacctagatgttaacca
gcatgtgaataatgtaaagtacattgggtggatcctggagagtgtccagtggggaataatggagaggcagaagctgaaaagcat
gactctggagtaicggagggaatgeggagagacagtggtcctcagtcctcactgcagttacgggttgcgatatcggtaacctg
gcaacagcgggggatgtggaatgtcagcattgtctccgactccaggatggagcgggaagtggtagaggaagaacagagtggga
gtagtaaaacaccaacaacaacttgggggaactgcacgcg

[0326] >Ath_FATB1_mature_peptide (SEQ ID NO:284)

MLDWKPRRSDMLVDPFGIGRIVQDGLVFRQNFSIRS YEIGADRSAS IETVMNHLQ
ETALNHVKTAGLLGDGFGSTPEMFKKNLIWVVTRMQVVVDKYPTWGDVVEVD
TWVSQSGKNMRRDWLVRDCNTGETLTRASSVWVMMNKLTRRLSKIPEEVRG
EIEPYFVNSDPVLAEDSRKLT KIDDKTADYVRSGLTPRWSDL DVNQHVNNVKYI
GWILESAPVGIMERQKLKSM TLEYRRECGRDSVLQSLTAVTGCDIGNLATAGDV
ECQHLLRLQDGAEVVRGRTEWSSKTPTTTWTAP

Incorporation by Reference

[0327] All of the U.S. patents and U.S. published patent applications cited herein are hereby incorporated by reference.

Equivalents

[0328] Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

WHAT IS CLAIMED IS:

1. A recombinant microorganism comprising one or more native and/or heterologous enzymes that function in one or more engineered metabolic pathways to convert a carbohydrate source to a hydrocarbon, wherein said one or more native and/or heterologous enzymes is activated, upregulated, downregulated, or deleted.
2. The recombinant microorganism of claim 1, wherein the conversion of a carbohydrate source to a hydrocarbon is under anaerobic or microaerophilic conditions.
3. The recombinant microorganism of any one of claims 1 or 2, wherein at least one of said engineered metabolic pathways produces net ATP.
4. The recombinant microorganism of claim 3, wherein said engineered metabolic pathway produces:
 - (a) at least about 0.5 net ATP;
 - (b) at least about 1.0 net ATP;
 - (c) at least about 1.5 net ATP; or
 - (d) at least about 2.0 net ATP.
5. The recombinant microorganism of any one of claims 1-4, wherein said hydrocarbon is selected from the group consisting of:
 - (a) an alkane;
 - (b) an alkene;
 - (c) an alkyne;
 - (d) a hydrocarbon derivative; and
 - (e) combinations of (a)-(d).
6. The recombinant microorganism of claim 5, wherein said hydrocarbon derivative is selected from the group consisting of:

- (a) an aldehyde;
 - (b) an alcohol;
 - (c) an ester;
 - (d) a fatty acid;
 - (e) an unsaturated fatty acid;
 - (f) a branched-chain fatty acid;
 - (g) a branched methoxy fatty acid;
 - (h) a multi-methyl branched acid;
 - (i) a divinyl-ether fatty acid;
 - (j) a w-phenylalkanoic acid;
 - (k) a dicarboxylic acid; and
 - (l) combinations of (a)-(k).
7. The recombinant microorganism of any one of claims 5 or 6, wherein said hydrocarbon or hydrocarbon derivative comprises a carbon backbone of C₄-C₄₀.
8. The recombinant microorganism of claim 7, wherein said hydrocarbon or hydrocarbon derivative comprises a carbon backbone selected from the group consisting of:
- (a) C₆-C₃₆;
 - (b) C₈-C₃₂;
 - (c) C₁₀-C₂₈;
 - (d) C₁₂-C₂₄;
 - (e) C₁₄-C₂₂;

- (f) C₁₆-C₂₀; and
 - (g) combinations of (a)-(f).
9. The recombinant microorganism of claim 7, wherein said hydrocarbon or hydrocarbon derivative comprises a carbon backbone selected from the group consisting of:
- (a) C₁₂;
 - (b) C₁₄;
 - (c) C₁₆;
 - (d) C₁₈;
 - (e) C₂₀;
 - (f) C₂₂;
 - (g) C₂₄; and
 - (h) combinations of (a)-(g).
10. The recombinant microorganism of claim 9, wherein said hydrocarbon or hydrocarbon derivative comprises a carbon backbone of C₁₆.
11. The recombinant microorganism of any one of claims 1-10, wherein said carbohydrate source is a lignocellulosic material.
12. The recombinant microorganism of any one of claims 1-11, wherein one of said engineered metabolic pathways comprises the conversion of oxaloacetate and acetyl-CoA to malonyl-CoA and pyruvate.
13. The recombinant microorganism of any one of claims 1-11, wherein one of said engineered metabolic pathways comprises the conversion of phosphoenolpyruvate to oxaloacetate.
14. The recombinant microorganism of any one of claims 1-11, wherein one of said engineered metabolic pathways comprises the following steps:

- (a) conversion of phosphoenolpyruvate to oxaloacetate; and
 - (b) conversion of oxaloacetate and acetyl-CoA to malonyl-CoA and pyruvate.
15. The recombinant microorganism of any one of claims 12-14, wherein one of said engineered metabolic pathways further comprises the conversion of pyruvate and CoA-SH into acetyl-CoA and CO₂ and NAD(P)H.
16. The recombinant microorganism of any one of claims 12-15, wherein one of said engineered metabolic pathways further comprises at least one of the following steps:
- (a) conversion of malonyl-CoA to malonyl-ACP;
 - (b) conversion of malonyl-ACP to an acyl_n-ACP;
 - (c) conversion of an acyl_n-ACP to a β -keto ester_{n+2}-ACP;
 - (d) conversion of a β -keto ester_{n+2}-ACP to a β -D-hydroxyacyl_{n+2}-ACP;
 - (e) conversion of a β -D-hydroxyacyl_{n+2}-ACP to a trans-2-unsaturated acyl_{n+2}-ACP; or
 - (f) conversion of a trans-2-unsaturated acyl_{n+2}-ACP to an acyl_{n+2}-ACP.
17. The recombinant microorganism of any one of claims 13 or 14, wherein said phosphoenolpyruvate is converted to oxaloacetate by a phosphoenolpyruvate carboxykinase.
18. The recombinant microorganism of claim 17, wherein said phosphoenolpyruvate carboxykinase is encoded by a heterologous phosphoenolpyruvate carboxykinase polynucleotide.
19. The recombinant microorganism of claim 17, wherein said phosphoenolpyruvate carboxykinase is encoded by a polynucleotide from a *Thermoanaerobacter* species, *E. coli*, *S. cerevisiae*, or *C. thermocellum*.
20. The recombinant microorganism of any one of claims 12 or 14, wherein said oxaloacetate and acetyl-CoA is converted to malonyl-CoA and pyruvate by a transcarboxylase.

21. The recombinant microorganism of claim 20, wherein said transcarboxylase is encoded by a heterologous transcarboxylase polynucleotide.
22. The recombinant microorganism of claim 20, wherein said transcarboxylase is encoded by a polynucleotide from a *Thermoanaerobacter* species, *P. freudenreichii*, *P. acnes*, or *C. thermocellum*.
23. The recombinant microorganism of claim 20, wherein said transcarboxylase is encoded by a polynucleotide from *C. bescii*, *C. cellulolyticum*, *C. kroppenstedtii*, *B. fragilis*, *V. parvula*, *V. gazogenes*, *P. thermopropionicum*, *Candidatus Cloacamonas acidaminovorans*, *G. bemidjiensis*, or *D. propionicus*.
24. The recombinant microorganism of any one of claims 20-23, wherein said transcarboxylase is genetically modified.
25. The recombinant microorganism of any one of claims 1-24, wherein one or more of said native enzymes is downregulated or deleted.
26. The recombinant microorganism of claim 25, wherein one or more of said downregulated or deleted native enzymes is an enzyme involved in central metabolism.
27. The recombinant microorganism of claim 26, wherein said downregulated or deleted native enzyme is selected from the group consisting of:
 - (a) a pyruvate kinase;
 - (b) a hydrogenase;
 - (c) a lactate dehydrogenase;
 - (d) a phosphotransacetylase;
 - (e) an acetate kinase;
 - (f) an acetaldehyde dehydrogenase;
 - (g) an alcohol dehydrogenase;

- (h) a pyruvate formate lyase;
 - (i) a pyruvate decarboxylase;
 - (j) an enzyme involved in degradation of fatty acids and their derivatives; and
 - (k) combinations of (a)-(j).
28. The recombinant microorganism of any one of claims 1-27, wherein said microorganism is a thermophilic or a mesophilic bacterium.
29. The recombinant microorganism of claim 28, wherein said thermophilic or mesophilic bacterium is a species of the genera *Escherichia*, *Propionibacterium*, *Thermoanaerobacterium*, *Thermoanaerobacter*, *Clostridium*, *Geobacillus*, *Saccharococcus*, *Paenibacillus*, *Bacillus*, *Caldicellulosiruptor*, *Anaerocellum*, *Anoxybacillus*, *Klebsiella*, *Lactobacillus*, *Lactococcus*, or *Corynebacterium*.
30. The recombinant microorganism of claim 28, wherein said microorganism is a bacterium selected from the group consisting of: *E. coli strain B*, *E. coli strain C*, *E. coli strain K*, *E. coli strain W*, *Shewanella*, *Propionibacterium acnes*, *Propionibacterium freudenreichii*, *Propionibacterium shermanii*, *Propionibacterium pentosaceum*, *Propionibacterium arabinosum*, *Clostridium acetobutylicum*, *Clostridium beijerinckii*, *Thermoanaerobacterium thermosulfurigenes*, *Thermoanaerobacterium aotearoense*, *Thermoanaerobacterium polysaccharolyticum*, *Thermoanaerobacterium zeae*, *Thermoanaerobacterium xylanolyticum*, *Thermoanaerobacterium saccharolyticum*, *Thermoanaerobium brockii*, *Thermoanaerobacterium thermosaccharolyticum*, *Thermoanaerobacter thermohydrosulfuricus*, *Thermoanaerobacter ethanolicus*, *Thermoanaerobacter brocki*, *Clostridium thermocellum*, *Clostridium clariflavum*, *Clostridium cellulolyticum*, *Clostridium phytofermentans*, *Clostridium straminosolvens*, *Geobacillus thermoglucosidasius*, *Geobacillus stearothermophilus*, *Saccharococcus caldxylosilyticus*, *Saccharococcus thermophilus*, *Paenibacillus campinasensis*, *Bacillus flavothermus*, *Anoxybacillus kamchatkensis*, *Anoxybacillus gonensis*, *Caldicellulosiruptor acetigenus*, *Caldicellulosiruptor saccharolyticus*, *Caldicellulosiruptor kristjanssonii*, *Caldicellulosiruptor owensensis*, *Caldicellulosiruptor lactoaceticus*, *Lactobacillus*

thermophilus, *Lactobacillus bulgaricus*, *Lactococcus lactis*, and *Anaerocellum thermophilum*.

31. The recombinant microorganism of claim 25, wherein said downregulated or deleted native enzyme is selected from the group consisting of:
 - (a) a lactate dehydrogenase;
 - (b) a phosphate acetyltransferase;
 - (c) an acetaldehyde dehydrogenase/alcohol dehydrogenase;
 - (d) a pyruvate kinase;
 - (e) a malate dehydrogenase;
 - (f) a PEP-protein phosphotransferase of PTS system; and
 - (g) combinations of (a)-(f).
32. The recombinant microorganism of claim 30, wherein said microorganism is selected from the group consisting of *Clostridium thermocellum* and *Thermoanaerobacterium saccharolyticum*.
33. A process for converting a carbohydrate source to a hydrocarbon comprising contacting the carbohydrate source with a recombinant microorganism according to any one of claims 1-32.
34. The process of claim 33, wherein said carbohydrate source comprises lignocellulosic biomass.
35. The process of claim 34, wherein said lignocellulosic biomass is selected from the group consisting of grass, switch grass, cord grass, rye grass, reed canary grass, mixed prairie grass, miscanthus, sugar-processing residues, sugarcane bagasse, sugarcane straw, agricultural wastes, rice straw, rice hulls, barley straw, corn cobs, cereal straw, wheat straw, canola straw, oat straw, oat hulls, corn fiber, stover, soybean stover, corn stover, forestry wastes, recycled wood pulp fiber, paper sludge, sawdust, hardwood, softwood, agave, and combinations thereof.

36. The process of claim 33, wherein said carbohydrate source comprises a carbohydrate.
37. The process of claim 36, wherein said carbohydrate is a sugar, a sugar alcohol, or a mixture thereof.
38. The process of any one of claims 33-37, wherein said hydrocarbon is secreted from the recombinant microorganism.
39. An engineered metabolic pathway for producing a hydrocarbon in a consolidated bioprocessing (CBP) organism according to any one of claims 1-38.
40. A recombinant microorganism comprising a native and/or heterologous enzyme that converts oxaloacetate and acetyl-CoA to malonyl-CoA and pyruvate, wherein said one or more native and/or heterologous enzymes is activated, upregulated, downregulated, or deleted.
41. The recombinant microorganism of claim 40, wherein said microorganism produces a hydrocarbon.
42. The recombinant microorganism of any one of claims 40 or 41, wherein the enzyme is a transcarboxylase.
43. The recombinant microorganism of claim 42, wherein said transcarboxylase is encoded by a polynucleotide from a *Thermoanaerobacter* species, *P. freudenreichii*, *P. acnes*, or *C. thermocellum*.
44. The recombinant microorganism of claim 42, wherein said transcarboxylase is encoded by a polynucleotide from *C. bescii*, *C. cellulolyticum*, *C. kroppenstedtii*, *B. fragilis*, *V. parvula*, *V. gazogenes*, *P. thermopropionicum*, *Candidatus Cloacamonas acidaminovorans*, *G. bemidjiensis*, or *D. propionicus*.
45. The recombinant microorganism of any one of claim 42-44, wherein the transcarboxylase is genetically modified.
46. A recombinant microorganism comprising an engineered metabolic pathway to direct metabolic flux from phosphoenolpyruvate to oxaloacetate rather than from phosphoenolpyruvate to pyruvate.

47. The recombinant microorganism of claim 46, wherein said microorganism comprises a modification to at least one gene selected from the group consisting of *edd*, *ldhA*, *pta*, *adhE*, *frdABCD*, *fdhF*, *ppc*, *pykA*, *pykF*, *mdh*, *maeA*, *maeB*, *fadE*, *ptsI*, *pflB*, *aceEF*, *poxB*, *mgsA*, and *ppsA* and combinations thereof.
48. The recombinant microorganism of claim 47, wherein said modification is a downregulation of said gene, an upregulation of said gene, a substitution of all or a part of said gene for all or a part of another gene, an exchange of one or more promoters for said gene, a mutation of said gene, or a combination of two or more thereof.
49. The recombinant microorganism of any of claims 46-48, wherein said microorganism is an *E. coli*.
50. A method for identifying a transcarboxylase capable of converting oxaloacetate and acetyl-CoA to pyruvate and malonyl-CoA comprising:
- a) transforming a bacterial strain that requires recombinant production of malonyl-CoA for growth with a vector comprising a transcarboxylase; and
 - b) assaying the transformed bacterial strain for growth on a culture medium lacking malonate,
- wherein growth on a culture medium lacking malonate indicates a transcarboxylase capable of converting oxaloacetate and acetyl-CoA to pyruvate and malonyl-CoA.
51. The method of claim 50, wherein said bacterial strain is *E. coli* $\Delta accC::matBC$.
52. The method of claim 50 or 51, wherein said transcarboxylase is genetically modified.
53. The method of any one of claims 50-52, wherein said transcarboxylase is from a thermophilic or a mesophilic bacterium.
54. The method of claim 53, wherein said thermophilic or mesophilic bacterium is a species of the genera *Escherichia*, *Propionibacterium*, *Thermoanaerobacterium*, *Thermoanaerobacter*, *Clostridium*, *Geobacillus*, *Saccharococcus*, *Paenibacillus*, *Bacillus*, *Caldicellulosiruptor*, *Anaerocellum*, *Anoxybacillus*, *Klebsiella*, *Lactobacillus*, *Lactococcus*, or *Corynebacterium*.

55. The method of claim 54, wherein said bacterium is selected from the group consisting of *E. coli* strain B, *E. coli* strain C, *E. coli* strain K, *E. coli* strain W, *Shewanella*, *Propionibacterium acnes*, *Propionibacterium freudenreichii*, *Propionibacterium shermanii*, *Propionibacterium pentosaceum*, *Propionibacterium arabinosum*, *Clostridium acetobutylicum*, *Clostridium beijerinckii*, *Thermoanaerobacterium thermosulfurigenes*, *Thermoanaerobacterium aotearoense*, *Thermoanaerobacterium polysaccharolyticum*, *Thermoanaerobacterium zae*, *Thermoanaerobacterium xylanolyticum*, *Thermoanaerobacterium saccharolyticum*, *Thermoanaerobium brockii*, *Thermoanaerobacterium thermosaccharolyticum*, *Thermoanaerobacter thermohydrosulfuricus*, *Thermoanaerobacter ethanolicus*, *Thermoanaerobacter brocki*, *Clostridium thermocellum*, *Clostridium clariflavum*, *Clostridium cellulolyticum*, *Clostridium phytofermentans*, *Clostridium straminosolvens*, *Geobacillus thermoglucosidasius*, *Geobacillus stearothermophilus*, *Saccharococcus caldxylosilyticus*, *Saccharococcus thermophilus*, *Paenibacillus campinasensis*, *Bacillus flavothermus*, *Anoxybacillus kamchatkensis*, *Anoxybacillus gonensis*, *Caldicellulosiruptor acetigenus*, *Caldicellulosiruptor bescii*, *Caldicellulosiruptor saccharolyticus*, *Caldicellulosiruptor kristjanssonii*, *Caldicellulosiruptor owensensis*, *Caldicellulosiruptor lactoaceticus*, *Lactobacillus thermophilus*, *Lactobacillus bulgaricus*, *Lactococcus lactis*, *Anaerocellum thermophilum*, and *Corynebacterium kroppenstedtii*.
56. The method of claim 53, wherein said bacterium is selected from the group consisting of a *Thermoanaerobacter* species, *P. freudenreichii*, *P. acnes*, *C. thermocellum*, *C. bescii*, *C. cellulolyticum*, *C. kroppenstedtii*, *B. fragilis*, *V. parvula*, *V. gazogenes*, *P. thermopropionicum*, *Candidatus Cloacamonas acidaminovorans*, *G. bemidjiensis*, and *D. propionicus*.
57. A recombinant microorganism comprising one or more native and/or heterologous enzymes that function in one or more engineered metabolic pathways to convert a carbohydrate source to a malonyl-CoA derived product, wherein said one or more native and/or heterologous enzymes is activated, upregulated, downregulated, or deleted.
58. The recombinant microorganism of claim 57, wherein the conversion of a carbohydrate source to a hydrocarbon is under anaerobic or microaerophilic conditions.

59. The recombinant microorganism of any one of claims 57 or 58, wherein at least one of said engineered metabolic pathways produces net ATP.
60. The recombinant microorganism of claim 59, wherein said engineered metabolic pathway produces:
- (a) at least about 0.5 net ATP;
 - (b) at least about 1.0 net ATP;
 - (c) at least about 1.5 net ATP; or
 - (d) at least about 2.0 net ATP.
61. The recombinant microorganism of any one of claims 57-60, wherein said product is a polyketide or an organic acid.
62. The recombinant microorganism of claim 61, wherein said polyketide is an antibiotic, antitumor, antifungal, and immunosuppressive.
63. The recombinant microorganism of claim 61, wherein said organic acid is 3-hydroxypropionic acid.
64. The recombinant microorganism of claim 63, wherein one of said engineered metabolic pathways comprises the following steps:
- (a) conversion of malonyl-CoA to malonate semialdehyde and coA; and
 - (b) conversion of malonate semialdehyde to 3-hydroxypropanoate.
65. The recombinant microorganism of claim 64, wherein said malonyl-CoA is converted to malonate semialdehyde and coA by a malonyl-CoA reductase
66. The recombinant microorganism of claim 65, wherein said malonyl-CoA reductase is encoded by a polynucleotide from a *C. aurantiacus*.
67. The recombinant microorganism of claim 64, wherein said malonate semialdehyde is converted to 3-hydroxypropanoate by a 3-hydroxypropionate dehydrogenase.

68. The recombinant microorganism of claim 63, wherein said malonyl-CoA is converted to 3-hydroxypropanoate by a bifunctional dehydrogenase.
69. The recombinant microorganism of claim 61, wherein said organic acid is adipic acid.
70. The recombinant microorganism of any one of claims 57-69, wherein said carbohydrate source is a lignocellulosic material.
71. The recombinant microorganism of any one of claims 57-70, wherein one of said engineered metabolic pathways comprises the conversion of oxaloacetate and acetyl-CoA to malonyl-CoA and pyruvate.
72. The recombinant microorganism of any one of claims 57-70, wherein one of said engineered metabolic pathways comprises the conversion of phosphoenolpyruvate to oxaloacetate.
73. The recombinant microorganism of any one of claims 57-70, wherein one of said engineered metabolic pathways comprises the following steps:
 - (a) conversion of phosphoenolpyruvate to oxaloacetate; and
 - (b) conversion of oxaloacetate and acetyl-CoA to malonyl-CoA and pyruvate.
74. The recombinant microorganism of any one of claims 71-73, wherein one of said engineered metabolic pathways further comprises the conversion of pyruvate and CoA-SH into acetyl-CoA and CO₂ and NAD(P)H.
75. The recombinant microorganism of any one of claims 72 or 73, wherein said phosphoenolpyruvate is converted to oxaloacetate by a phosphoenolpyruvate carboxykinase.
76. The recombinant microorganism of claim 75, wherein said phosphoenolpyruvate carboxykinase is encoded by a heterologous phosphoenolpyruvate carboxykinase polynucleotide.

77. The recombinant microorganism of claim 75, wherein said phosphoenolpyruvate carboxykinase is encoded by a polynucleotide from a *Thermoanaerobacter* species, *E. coli*, *S. cerevisiae*, or *C. thermocellum*.
78. The recombinant microorganism of any one of claims 71 or 73, wherein said oxaloacetate and acetyl-CoA is converted to malonyl-CoA and pyruvate by a transcarboxylase.
79. The recombinant microorganism of claim 78, wherein said transcarboxylase is encoded by a heterologous transcarboxylase polynucleotide.
80. The recombinant microorganism of claim 79, wherein said transcarboxylase is encoded by a polynucleotide from a *Thermoanaerobacter* species, *P. freudenreichii*, *P. acnes*, *C. thermocellum*, *C. bescii*, *C. cellulolyticum*, *C. kroppenstedtii*, *B. fragilis*, *V. parvula*, *V. gazogenes*, *P. thermopropionicum*, *Candidatus Cloacamonas acidaminovorans*, *G. bemidjiensis*, or *D. propionicus*.
81. The recombinant microorganism of any one of claims 78-80, wherein said transcarboxylase is genetically modified.
82. The recombinant microorganism of any one of claims 57-81, wherein one or more of said native enzymes is downregulated or deleted.
83. The recombinant microorganism of claim 82, wherein one or more of said downregulated or deleted native enzymes is an enzyme involved in central metabolism.
84. The recombinant microorganism of claim 83, wherein said downregulated or deleted native enzyme is selected from the group consisting of:
 - (a) a pyruvate kinase;
 - (b) a hydrogenase;
 - (c) a lactate dehydrogenase;
 - (d) a phosphotransacetylase;
 - (e) an acetate kinase;

- (f) an acetaldehyde dehydrogenase;
- (g) an alcohol dehydrogenase;
- (h) a pyruvate formate lyase;
- (i) a pyruvate decarboxylase;
- (j) an enzyme involved in degradation of fatty acids and their derivatives; and
- (k) combinations of (a)-(j).

85. The recombinant microorganism of claim 82, wherein said downregulated or deleted native enzyme is selected from the group consisting of:
- (a) a lactate dehydrogenase;
 - (b) a phosphate acetyltransferase;
 - (c) an acetaldehyde dehydrogenase/alcohol dehydrogenase;
 - (d) a pyruvate kinase;
 - (e) a malate dehydrogenase;
 - (f) a PEP-protein phosphotransferase of PTS system; and
 - (g) combinations of (a)-(f).
86. The recombinant microorganism of any one of claims 57-85, wherein said microorganism is a thermophilic or a mesophilic bacterium.
87. The recombinant microorganism of claim 86, wherein said thermophilic or mesophilic bacterium is a species of the genera *Escherichia*, *Propionibacterium*, *Thermoanaerobacterium*, *Thermoanaerobacter*, *Clostridium*, *Geobacillus*, *Saccharococcus*, *Paenibacillus*, *Bacillus*, *Caldicellulosiruptor*, *Anaerocellum*, *Anoxybacillus*, *Klebsiella*, *Lactobacillus*, *Lactococcus*, or *Corynebacterium*.

88. The recombinant microorganism of claim 87, wherein said microorganism is a bacterium selected from the group consisting of: *E. coli strain B*, *E. coli strain C*, *E. coli strain K*, *E. coli strain W*, *Shewanella*, *Propionibacterium acnes*, *Propionibacterium freudenreichii*, *Propionibacterium shermanii*, *Propionibacterium pentosaceum*, *Propionibacterium arabinosum*, *Clostridium acetobutylicum*, *Clostridium beijerinckii*, *Thermoanaerobacterium thermosulfurigenes*, *Thermoanaerobacterium aotearoense*, *Thermoanaerobacterium polysaccharolyticum*, *Thermoanaerobacterium zeae*, *Thermoanaerobacterium xylanolyticum*, *Thermoanaerobacterium saccharolyticum*, *Thermoanaerobium brockii*, *Thermoanaerobacterium thermosaccharolyticum*, *Thermoanaerobacter thermohydrosulfuricus*, *Thermoanaerobacter ethanolicus*, *Thermoanaerobacter brocki*, *Clostridium thermocellum*, *Clostridium clariflavum*, *Clostridium cellulolyticum*, *Clostridium phytofermentans*, *Clostridium straminosolvens*, *Geobacillus thermoglucosidasius*, *Geobacillus stearothermophilus*, *Saccharococcus caldoxylosilyticus*, *Saccharococcus thermophilus*, *Paenibacillus campinasensis*, *Bacillus flavothermus*, *Anoxybacillus kamchatkensis*, *Anoxybacillus gonensis*, *Caldicellulosiruptor acetigenus*, *Caldicellulosiruptor saccharolyticus*, *Caldicellulosiruptor kristjanssonii*, *Caldicellulosiruptor owensensis*, *Caldicellulosiruptor lactoaceticus*, *Lactobacillus thermophilus*, *Lactobacillus bulgaricus*, *Lactococcus lactis*, and *Anaerocellum thermophilum*.
89. A process for converting a carbohydrate source to a malonyl-CoA derived product comprising contacting the carbohydrate source with a recombinant microorganism according to any one of claims 57-88.
90. The process of claim 89, wherein said carbohydrate source comprises lignocellulosic biomass.
91. The process of claim 90, wherein said lignocellulosic biomass is selected from the group consisting of grass, switch grass, cord grass, rye grass, reed canary grass, mixed prairie grass, miscanthus, sugar-processing residues, sugarcane bagasse, sugarcane straw, agricultural wastes, rice straw, rice hulls, barley straw, corn cobs, cereal straw, wheat straw, canola straw, oat straw, oat hulls, corn fiber, stover, soybean stover, corn stover, forestry wastes, recycled wood pulp fiber, paper sludge, sawdust, hardwood, softwood, agave, and combinations thereof.

92. The process of claim 89, wherein said carbohydrate source comprises a carbohydrate.
93. The process of claim 92, wherein said carbohydrate is a sugar, a sugar alcohol, or a mixture thereof.
94. The process of any one of claims 89-93, wherein said malonyl-CoA derived product is secreted from the recombinant microorganism.
95. An engineered metabolic pathway for producing a malonyl-CoA derived product in a consolidated bioprocessing (CBP) organism according to any one of claims 57-94.
96. The recombinant microorganism of any one of claims 1-32 or 40-45, wherein the conversion of a carbohydrate source to a hydrocarbon is redox neutral.
97. The recombinant microorganism of any one of claims 57-88, wherein the conversion of a carbohydrate source to a malonyl-CoA derived product is redox neutral.
98. The recombinant microorganism of claim 96 or 97, wherein the one or more engineered metabolic pathways comprises one or more enzymes in Table 10.
99. The recombinant microorganism of any one of claims 1-32, 40-45, or 57-88, wherein one of said engineered metabolic pathways further comprises the conversion of acyl_{n+2}-ACP to a fatty acid.
100. The recombinant microorganism of claim 99, wherein said acyl_{n+2}-ACP is converted to a fatty acid by a chain termination enzyme.
101. The recombinant microorganism of claim 100, wherein said chain termination enzyme is selected from an *E. coli* codon optimized C12 acyl-ACP thioesterase, an *E. coli* codon optimized C16 acyl-ACP thioesterase, an acyl-ACP reductase, and an acyl-ACP reductase homolog.
102. An isolated bacterial strain *E. coli* $\Delta accC::matBC$.
103. A recombinant yeast microorganism comprising one or more native and/or heterologous enzymes that function in one or more engineered metabolic pathways to convert a

carbohydrate source to a malonyl-CoA derived product, wherein said one or more native and/or heterologous enzymes is activated, upregulated, downregulated, or deleted.

104. The recombinant yeast microorganism of claim 103, wherein the conversion of a carbohydrate source to a malonyl-CoA derived product is under anaerobic or microaerophilic conditions.
105. The recombinant yeast microorganism of any one of claims 103 or 104, wherein at least one of said engineered metabolic pathways produces net ATP.
106. The recombinant yeast microorganism of claim 105, wherein said engineered metabolic pathway produces:
 - (a) at least about 0.5 net ATP;
 - (b) at least about 1.0 net ATP;
 - (c) at least about 1.5 net ATP; or
 - (d) at least about 2.0 net ATP.
107. The recombinant yeast microorganism of any one of claims 103-106, wherein said product is a hydrocarbon, a polyketide, or an organic acid.
108. The recombinant yeast microorganism of any one of claims 103-107, wherein said carbohydrate source is a lignocellulosic material.
109. The recombinant yeast microorganism of any one of claims 103-108, wherein one of said engineered metabolic pathways comprises the conversion of oxaloacetate and acetyl-CoA to malonyl-CoA and pyruvate.
110. The recombinant yeast microorganism of any one of claims 103-108, wherein one of said engineered metabolic pathways comprises the conversion of phosphoenolpyruvate to oxaloacetate.
111. The recombinant yeast microorganism of any one of claims 103-108, wherein one of said engineered metabolic pathways comprises the following steps:

- (a) conversion of phosphoenolpyruvate to oxaloacetate; and
 - (b) conversion of oxaloacetate and acetyl-CoA to malonyl-CoA and pyruvate.
112. The recombinant yeast microorganism of any one of claims 109-111, wherein one of said engineered metabolic pathways further comprises the conversion of pyruvate and CoA-SH into acetyl-CoA and CO₂ and NAD(P)H.
113. The recombinant yeast microorganism of any one of claims 110 or 111, wherein said phosphoenolpyruvate is converted to oxaloacetate by a phosphoenolpyruvate carboxykinase.
114. The recombinant yeast microorganism of claim 113, wherein said phosphoenolpyruvate carboxykinase is encoded by a heterologous phosphoenolpyruvate carboxykinase polynucleotide.
115. The recombinant yeast microorganism of claim 113, wherein said phosphoenolpyruvate carboxykinase is encoded by a polynucleotide from *S. cerevisiae*.
116. The recombinant yeast microorganism of any one of claims 109 or 111, wherein said oxaloacetate and acetyl-CoA is converted to malonyl-CoA and pyruvate by a transcarboxylase.
117. The recombinant yeast microorganism of claim 116, wherein said transcarboxylase is encoded by a heterologous transcarboxylase polynucleotide.
118. The recombinant yeast microorganism of claim 116, wherein said transcarboxylase is encoded by a polynucleotide from a *Thermoanaerobacter* species, *P. freudenreichii*, *P. acnes*, *C. thermocellum*, *C. bescii*, *C. cellulolyticum*, *C. kroppenstedtii*, *B. fragilis*, *V. parvula*, *V. gazogenes*, *P. thermopropionicum*, *Candidatus Cloacamonas acidaminovorans*, *G. bemiidjensis*, or *D. propionicus*.
119. The recombinant yeast microorganism of any one of claims 116-118, wherein said transcarboxylase is genetically modified.

120. The recombinant yeast microorganism of any one of claims 103-119, wherein one of said engineered metabolic pathways further comprises one or more formate dehydrogenases capable of converting formate to CO₂ and NAD(P)H.
121. The recombinant yeast microorganism of claim 120, wherein said formate dehydrogenase is encoded by an *S. cerevisiae* NAD⁺ FDH1, a *B. stabilis* NADP⁺ FDH, or both.
122. The recombinant yeast microorganism of any one of claims 103-119, wherein one of said engineered metabolic pathways further comprises an enzyme encoding a palmitoyl-acyl carrier protein thioesterase (FatB1).
123. The recombinant yeast microorganism of claim 122, wherein said palmitoyl-acyl carrier protein thioesterase (FatB1) is from *A. thaliana*.
124. The recombinant yeast microorganism of any one of claims 103-123, wherein one or more of said native enzymes is downregulated or deleted.
125. The recombinant yeast microorganism of claim 124, wherein said downregulated or deleted native enzyme is selected from a glycerol-3-phosphate dehydrogenase 1 (GPD1), a glycerol-3-phosphate dehydrogenase 2 (GPD2), a formate dehydrogenase 1 (FDH1), a formate dehydrogenase 2 (FDH2), or a combination thereof.
126. The recombinant yeast microorganism of claim 125, wherein said native enzyme is downregulated or deleted by insertion of a heterologous enzyme at the locus of the native enzyme
127. The recombinant yeast microorganism of claim 126, wherein said heterologous enzyme is a pyruvate formate lyase A (PFLA), a pyruvate formate lyase B (PFLB), a bifunctional alcohol/aldehyde dehydrogenase (AdhE), or a combination thereof.
128. The recombinant yeast microorganism of claim 124, wherein said microorganism further comprises a downregulation or deletion in a pyruvate decarboxylase 5 (PDC5), a pyruvate decarboxylase 6 (PDC6), a pyruvate decarboxylase 1 (PDC1), or a combination thereof.
129. The recombinant yeast microorganism of any one of claims 103-129, wherein said yeast microorganism is selected from the group consisting of *Saccharomyces cerevisiae*,

Kluyveromyces lactis, *Kluyveromyces marxianus*, *Pichia pastoris*, *Yarrowia lipolytica*, *Hansenula polymorpha*, *Phaffia rhodozyma*, *Candida utliis*, *Arxula adeninivorans*, *Pichia stipitis*, *Debaryomyces hansenii*, *Debaryomyces polymorphus*, *Schizosaccharomyces pombe*, *Candida albicans*, and *Schwanniomyces occidentalis*.

130. The recombinant microorganism of claim 7, wherein said hydrocarbon or hydrocarbon derivative comprises a carbon backbone of C₁₂-C₁₈.
131. The recombinant microorganism of any one of claims 1, 57, or 103, wherein said carbohydrate source comprises corn mash, corn starch, or sugarcane sucrose.
132. The process of claim 33 or 89, wherein said carbohydrate source comprises corn mash, corn starch, or sugarcane sucrose.

Figure 1

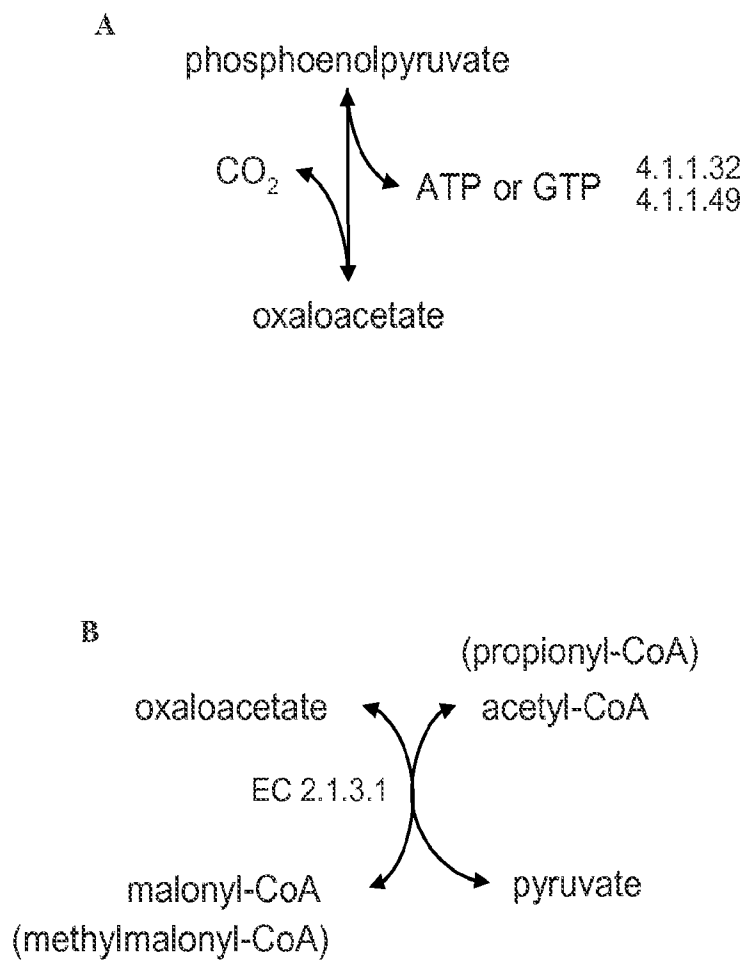


Figure 2

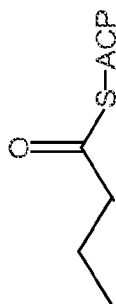
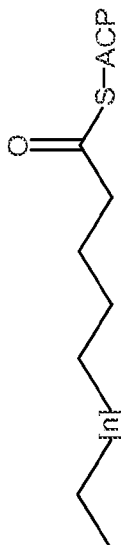
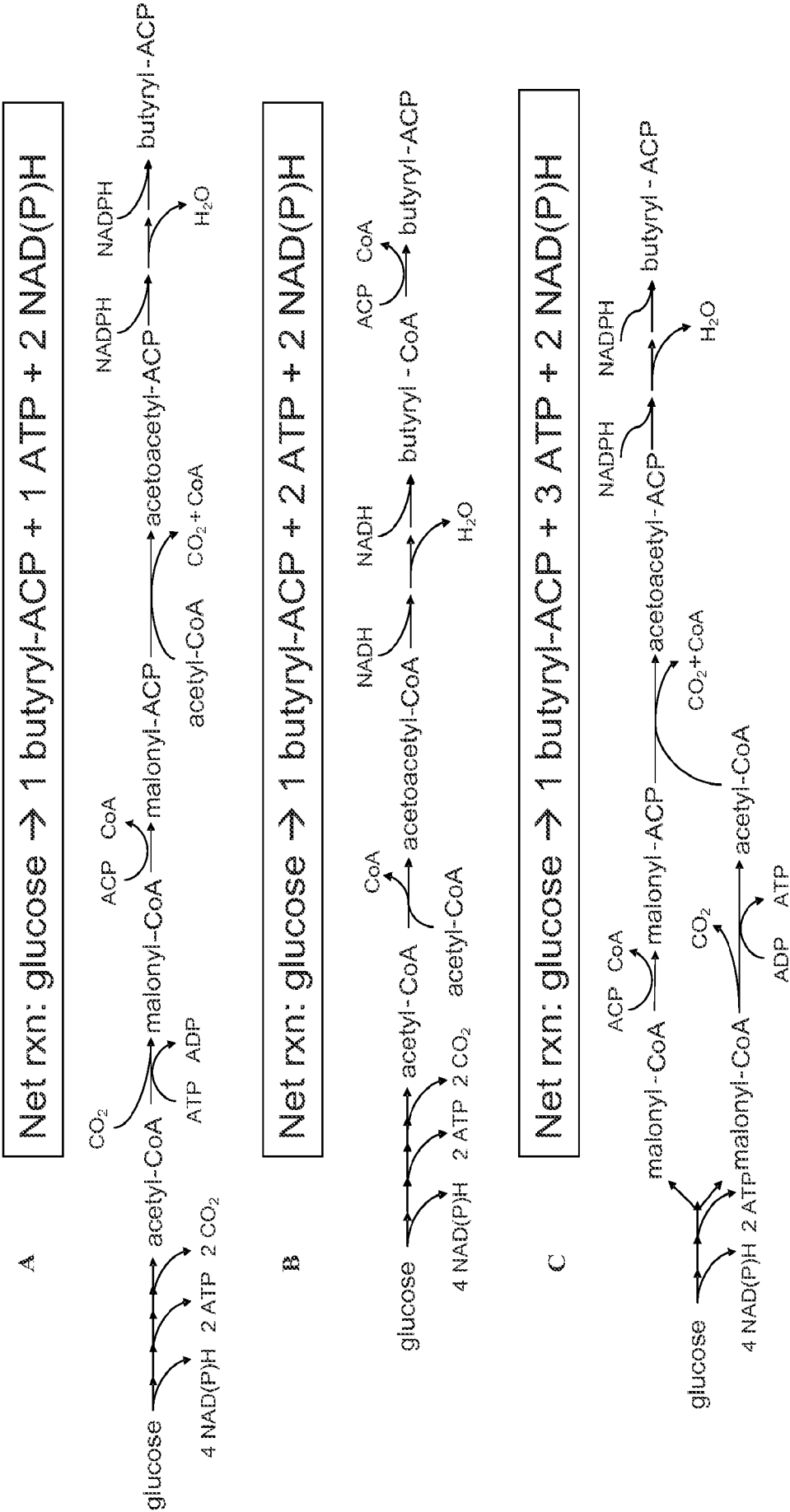
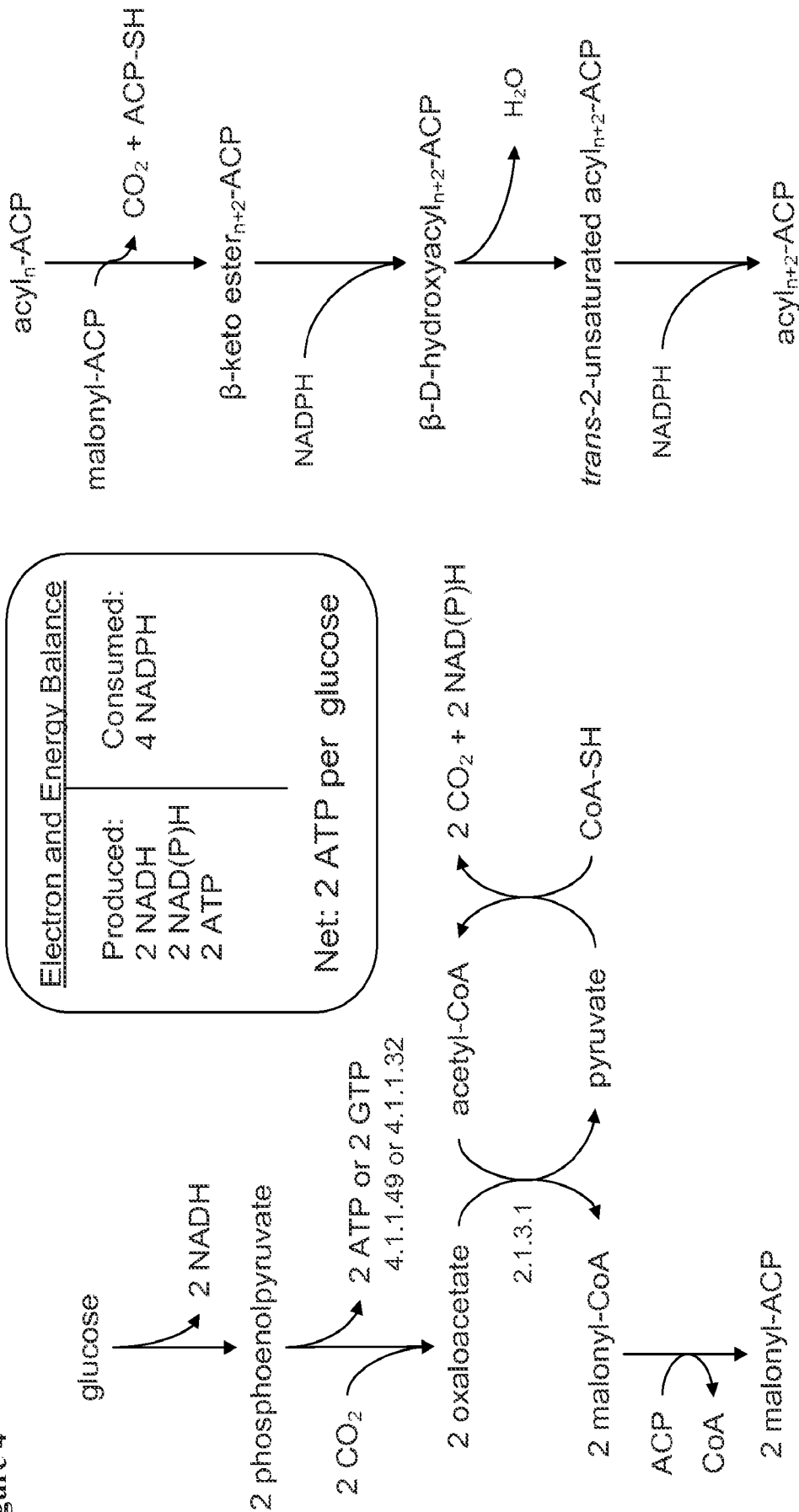
Chain initiation**Chain extension****Chain termination**

Figure 3



4/72

Figure 4



5/72

Figure 5

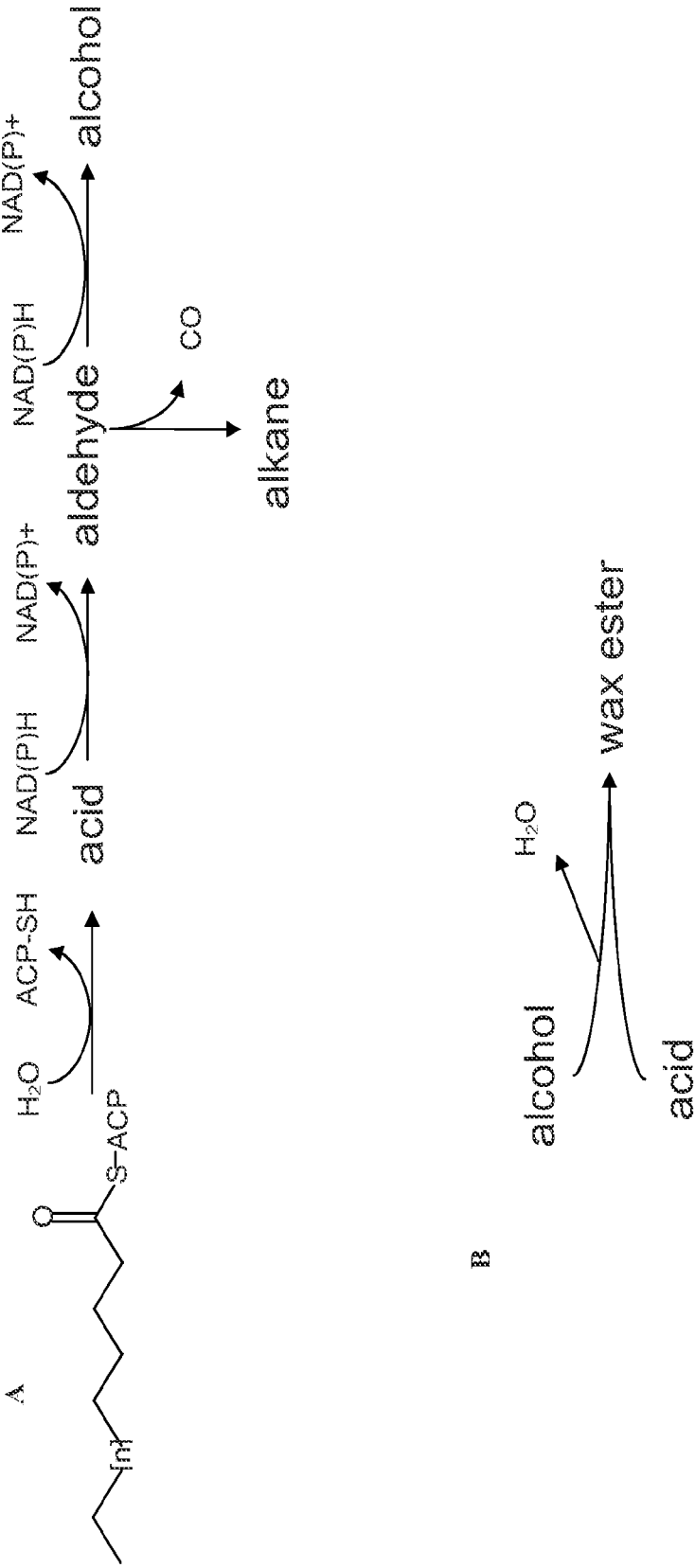


Figure 6

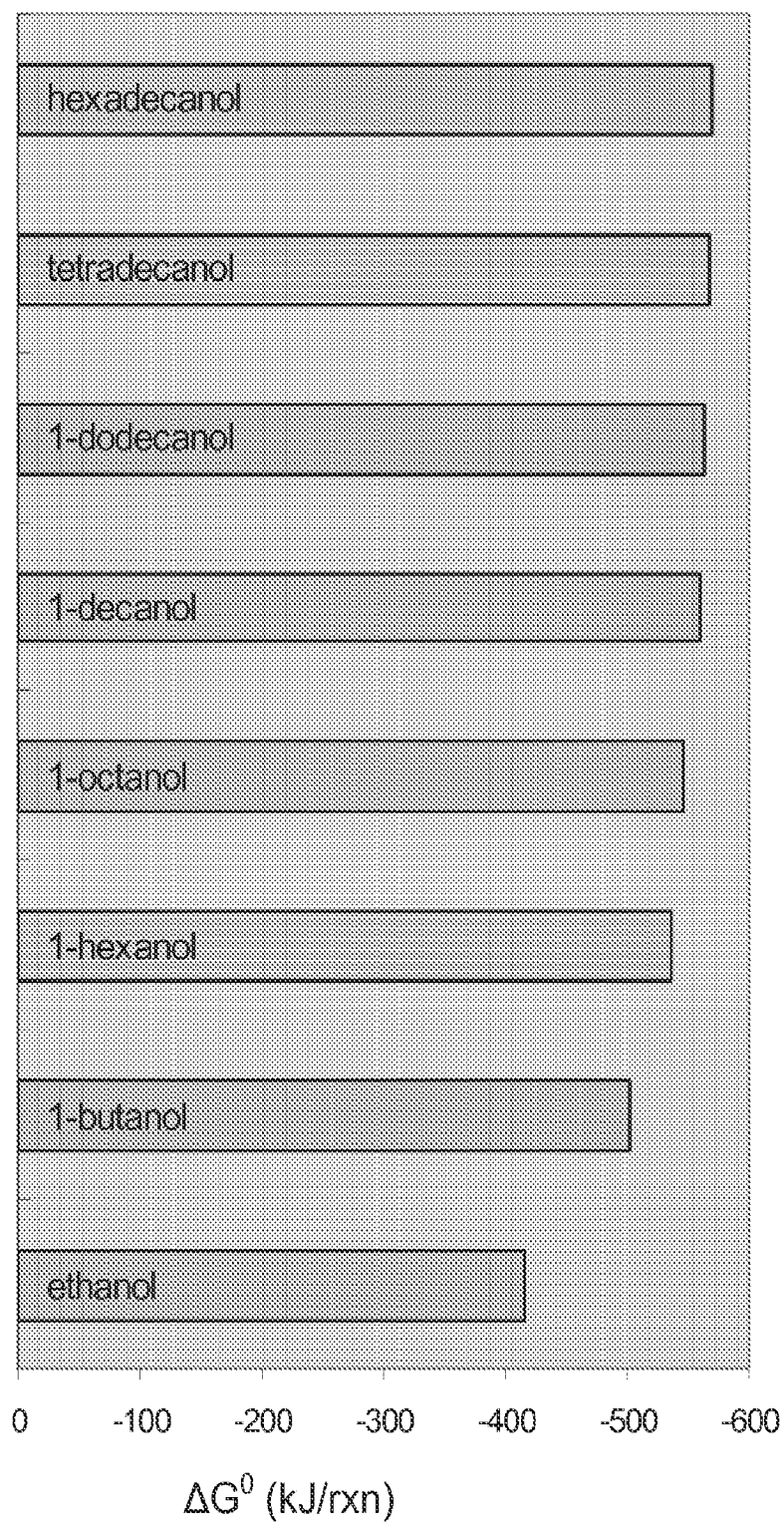


Figure 7A

Transcarboxylase 5S subunit

		1		50
Pfreud 5S	(1)	MSPREIEVSEPRVGTTELRLDAHOSIATRMAIEDV	GACATDAAG	
Pacnes 5S	(1)	-----MSPRVGTTELRLDAHOSIATRMAIEDV	DACAIDAG	
Cthe_0701	(1)	-----MAVKITETRLDAHOSIATRRIEELPIIDKIL	DEISY	
Tsacch orl888	(1)	-----MSKIKITETVRLDAHOSIATRTTDELEPI	EKLIECT	
		51		100
Pfreud 5S	(51)	WSVECWGGATIDSCIREFINEDPWERLERFRKLMPNSF	LQMLLRGONILGY	
Pacnes 5S	(43)	WSVECWGGATIDSCIREFINEDPWERLERFRKLLNSF	LQMLLRGONILGY	
Cthe_0701	(41)	HSLEVWGGATIDACIREFINEDPWERLERLKHCKTFL	LQMLLRGONILGY	
Tsacch orl888	(41)	HSLEVWGGATIDACIREFINEDPWERLERLKKAIRTFL	LQMLLRGONILGY	
		101		150
Pfreud 5S	(101)	RYMDEVVDKFEVKSENGMDVRFEDAMDPRLMAHAMA	AVRKAGHAG	
Pacnes 5S	(93)	RYMDEVVDKFEVKSENGMDVRFEDAMDPRLMAHAMA	AVRKAGHAG	
Cthe_0701	(91)	RYMADVWEYDVKSVANGLIIRFEDAMDPRLMTAKAC	KEGCHAG	
Tsacch orl888	(91)	RYMADVWEYDVKSVANGLIIRFEDAMDPRLMTAKAC	KEGCHAG	
		151		200
Pfreud 5S	(151)	GTICYTISPEHVEGYVRLAGHFLMGADSI	LKEDAAITPEQPAYEITK	
Pacnes 5S	(143)	GTICYTISPEHPESEVKKAFRLIMGADSI	LKEDAAITPEQPAYEITK	
Cthe_0701	(141)	GTVCYTISPEHLELVKDAKTIVMGADSI	CVKDMAGDILEYVAYEITK	
Tsacch orl888	(141)	ATVITVSEVHSDHYLVKASLQMGADSI	CKDMAGDILEYVAYEITK	
		201		250
Pfreud 5S	(201)	ATLKYGQKATLHCHSTTGTVSLMAIEAGD	PTALISSMSLG	
Pacnes 5S	(193)	CTLENH-PVQDCHSTTGTVILVLOMAIEAGD	PTALISSMSLG	
Cthe_0701	(191)	ALLENH--FVPLCHSTTGTSGVSMTYLMAIEAGD	PTALISSMSLS	
Tsacch orl888	(191)	SLLENAL--YFPLCHSTTGTGLSMTYLMAIEAGD	PTALISSLSLATS	
		251		300
Pfreud 5S	(251)	NPTESVEMLCGTGTTLDYDRTHKPDHEKAI	RFXKFFESKTLVDI	
Pacnes 5S	(242)	NPTESLVEMLCGEMITGLDMDRLEKRDHKK	VPRKXKFFESKTLVNT	
Cthe_0701	(239)	QPFIEILVATIKCPNDIGLBDLPSADYERPL	NEXYIEGLLDVKVM	
Tsacch orl888	(239)	QPAIESIYAALNDEIDIGLDELILPAIAHT	VVKQHFDDMSLLMS	
		301		350
Pfreud 5S	(301)	---S-LKSKPGGMISNMESOLRAQAEEDMDEM	MAEVPVRKAAGPP	
Pacnes 5S	(292)	---N-LKSKPGGMISNMESOLRAQAGDRMDEM	MAEVPVRKBAGNPP	
Cthe_0701	(289)	GVDVNTLKYQPGGMISNLVSOLKONAVDK	EEVLEVPVRVDFGYP	
Tsacch orl888	(289)	-VDVKALSKPGGMISNLVSOLKQQLKYDVL	KEVPVRVDLGNPP	
		351		400
Pfreud 5S	(347)	LVTSSQINGTOAVNVMTC--HPRMTGSPADIM	SGVGAPADRFPV	
Pacnes 5S	(338)	LVTSSQINGTOAVNVLNMGNSYINLTAFADLM	LGKGPIGELNPEI	
Cthe_0701	(339)	LVTETSQINGTOAVNVTIC-EFHYVNSKALIK	EGCAPAFVNPEV	
Tsacch orl888	(338)	LVTETSQINGTOAVNVTIC-EFHYVNSKALIK	EGCAPAFVNPEV	
		401		450
Pfreud 5S	(395)	VKLAEVSGKPTTPPADILPEEWEEQNEAAAL	KGNGTDEDVITNAL	
Pacnes 5S	(388)	VMAKKOTGKEPTDPPADILPEWDLVQAQSL	GFDSDEDVITNAL	
Cthe_0701	(388)	QKKILK--EFPTVPADILPEELDNRN--EM	KYLEQDEDVITSYAL	
Tsacch orl888	(387)	RKKIIG--EVSKEPPADILPELDEKN--E	KREFQDEDVITSYAL	
		451		500
Pfreud 5S	(445)	EEQVAVVTEFESPHSVALTDAOLKAGDEKSL	AVAGFVTYNNVE	
Pacnes 5S	(438)	EEQVAPKTEESQGEKSVAMTEAOLKAKG	TGAAGIAGFVNNTVG	
Cthe_0701	(433)	EEQVAKTEERKAKQYIEPDMYIRVHPV	-----	
Tsacch orl888	(432)	EEQVARETEERKAKKYRIDSTILNT	ERVHEI-----	
		501		511
Pfreud 5S	(495)	CTVRENTVQQA		
Pacnes 5S	(487)	GNSHQVTVEPA		
Cthe_0701	(466)	-----		
Tsacch orl888	(465)	-----		

Figure 7B

Transcarboxylase 1.3S subunit

		1			50
Pfreud 1.3S	(1)	-MELXVTYVETATDDEVDVDSHENPMCTILFGGGGGAPAPRAAGAG-			
Pacnes 1.3S	(1)	-MELXVTYVETATDDEVDVDTANAPMAPILFGGGAG---PMKASGGG-			
Cthe_0700	(1)	MKRLIYVLENQDEEEIEIRGASAPQVTLSLPAPAPSPAPAEEKT			
Tsacch or0947	(1)	MKRLIYVLENQDEEEIEIRGASAPQVTLSLPAPAPSPAPAEEKT			
		51			100
Pfreud 1.3S	(49)	-----AGKAGEGEIAPLAGTVSKILVAGDTVKAGGVVVEEAMKME			
Pacnes 1.3S	(46)	-----AGKAGEGEVPAPLASTVAKILVAGDVKAGGVVVEEAMKME			
Cthe_0700	(51)	AAPKEDSVPAATAIKAPMPGTLIPVAGDTVKAGGVVVEEAMKME			
Tsacch or0947	(48)	QVEVKN--EVKSGFINAPMPGTLIPVAGDTVKAGGVVVEEAMKME			
		101			132
Pfreud 1.3S	(92)	ELNAETDSVEVLAQRDAQGGGCHVIG			
Pacnes 1.3S	(89)	ELNAPADGTWKILAVGDAQGGGCHVALG			
Cthe_0700	(101)	ELVAPMDGTVASIRVSKGASVNVGVVSLK			
Tsacch or0947	(96)	ELVSHYDGTISIRVSKGASVNTGVVLYLK			

Figure 7C

Transcarboxylase 12S subunit (N-terminus)

		1	50
Pfreud 12S-N	(1)	MAENNNLKLASTNMGKVECHAE PQVTEAGGCE KVTFC CKQIARER	
Pacnes 12S	(1)	MAEKKPIKLADTMAGRIECLAD RHVELGGGEDRLKQDCKQIARER	
Cthe_0699	(1)	-----MDKV KIGLIREKLAQVEQGGGAEKIARNDACKMIARER	
Tsacch or0945	(1)	-----MSIDRIELILPREMVLEGGGLDKVTKCKKGLIARER	
		51	100
Pfreud 12S-N	(51)	LNNILLPHSDEVGAEKRRHTTLGGDRAVVEADGVATCPGTILGRPVH	
Pacnes 12S	(51)	LSNLDAYSDEVGAEKRRHTTLGGDRAVVEADGVATCPGTILGRPVH	
Cthe_0699	(41)	LGALDESEVELTFVTRSIDELNKKVEGGGVVTCYGSIDGRLNV	
Tsacch or0945	(41)	LYILDESEVELAYVHPCIDELNKKVEGGGVVTCYGTILGRNV	
		101	150
Pfreud 12S-N	(101)	ASQDFTVMGCSAGTQSTKVVVTEQSLITCTEFLFFDSGGARLOEGID	
Pacnes 12S	(101)	ASQDFTVMGCSAGTQSTKVVVTEQSLITCTEFLFFDSGGARLOEGID	
Cthe_0699	(91)	ASQDFTVMGCSAGTQSTKVVVTEQSLITCTEFLFFDSGGARLOEGID	
Tsacch or0945	(91)	YASQDFTVMGCSAGTQSTKVVVTEQSLITCTEFLFFDSGGARLOEGID	
		151	200
Pfreud 12S-N	(151)	SPSGYCKMFTANVKISGVPEQATFAGPCAGGASYSFATDFTIM--KKA	
Pacnes 12S	(151)	SPSGYCKMFTANVKISGVPEQATFAGPCAGGASYSFATDFTIM--KKA	
Cthe_0699	(141)	ALSGYCKMFTANVKISGVPEQATFAGPCAGGASYSFATDFTIM--KKA	
Tsacch or0945	(141)	ALSGYCKMFTANVKISGVPEQATFAGPCAGGASYSFATDFTIM--KKA	
		201	250
Pfreud 12S-N	(200)	QMFITGPOVILVTGEDVTSAEELGCSITHTSKSGVAHTRAEDEECLKM	
Pacnes 12S	(200)	QMFITGPOVILVTGEDVTSAEELGCSITHTSKSGVAHTRAEDEECLKM	
Cthe_0699	(191)	QMFITGPOVILVTGEDVTSAEELGCSITHTSKSGVAHTRAEDEECLKM	
Tsacch or0945	(191)	QMFITGPOVILVTGEDVTSAEELGCSITHTSKSGVAHTRAEDEECLKM	
		251	300
Pfreud 12S-N	(250)	KKLSLPLPNNTEEA VNNNEVSPNTE--IRDIVEIDGKGDVDRDIA	
Pacnes 12S	(250)	KKLSLPLPNNTEEA VNNNEVSPNTE--IRDIVEIDGKGDVDRDIA	
Cthe_0699	(241)	KKLSLPLPNNTEEA VNNNEVSPNTE--IRDIVEIDGKGDVDRDIA	
Tsacch or0945	(241)	KKLSLPLPNNTEEA VNNNEVSPNTE--IRDIVEIDGKGDVDRDIA	
		301	350
Pfreud 12S-N	(299)	KIVDMGDIYEVAGQNTNIVTAEFRNGRSVGIANORVMSGCIIDINAS	
Pacnes 12S	(299)	KIVDMGDIYEVAGQNTNIVTAEFRNGRTVGIANORVMSGCIIDINAS	
Cthe_0699	(291)	KIVDMGDIYEVAGQNTNIVTAEFRNGRTVGIANORVMSGCIIDINAS	
Tsacch or0945	(291)	KIVDMGDIYEVAGQNTNIVTAEFRNGRTVGIANORVMSGCIIDINAS	
		351	400
Pfreud 12S-N	(349)	DKAAEVFCDSNIELVQLDVPGLPGVGOEYGSIRHGAKLYAISE	
Pacnes 12S	(349)	DKAAEVFCDSNIELVQLDVPGLPGVGOEYGSIRHGAKLYAISE	
Cthe_0699	(341)	DKAAEVFCDSNIELVQLDVPGLPGVGOEYGSIRHGAKLYAISE	
Tsacch or0945	(341)	DKAAEVFCDSNIELVQLDVPGLPGVGOEYGSIRHGAKLYAISE	
		401	450
Pfreud 12S-N	(399)	ATVEPKITVVLPRAYGGSYLAMCRLGADAVYAWESAEIAVMGAEGAANV	
Pacnes 12S	(399)	ATVEPKITVVLPRAYGGSYLAMCRLGADAVYAWESAEIAVMGAEGAANV	
Cthe_0699	(391)	ATVEPKITVVLPRAYGGSYLAMCRLGADAVYAWESAEIAVMGAEGAANV	
Tsacch or0945	(391)	ATVEPKITVVLPRAYGGSYLAMCRLGADAVYAWESAEIAVMGAEGAANV	
		451	500
Pfreud 12S-N	(449)	ERKEIKAADDEDMBAENIEEYNAENTPYVAAARGQDDVIDPADTRK	
Pacnes 12S	(449)	ERKEIKAADDEDMBAENIEEYNAENTPYVAAARGQDDVIDPADTRK	
Cthe_0699	(441)	ERKEIKAADDEDMBAENIEEYNAENTPYVAAARGQDDVIDPADTRK	
Tsacch or0945	(441)	ERKEIKAADDEDMBAENIEEYNAENTPYVAAARGQDDVIDPADTRK	
		501	526
Pfreud 12S-N	(499)	RLISALEMYATKPOIRBAKKHGVPL	
Pacnes 12S	(499)	RLISALEMYATKPOIRBAKKHGVPL	
Cthe_0699	(491)	RLISALEMYATKPOIRBAKKHGVPL	
Tsacch or0945	(491)	RLISALEMYATKPOIRBAKKHGVPL	

10/72

Figure 8

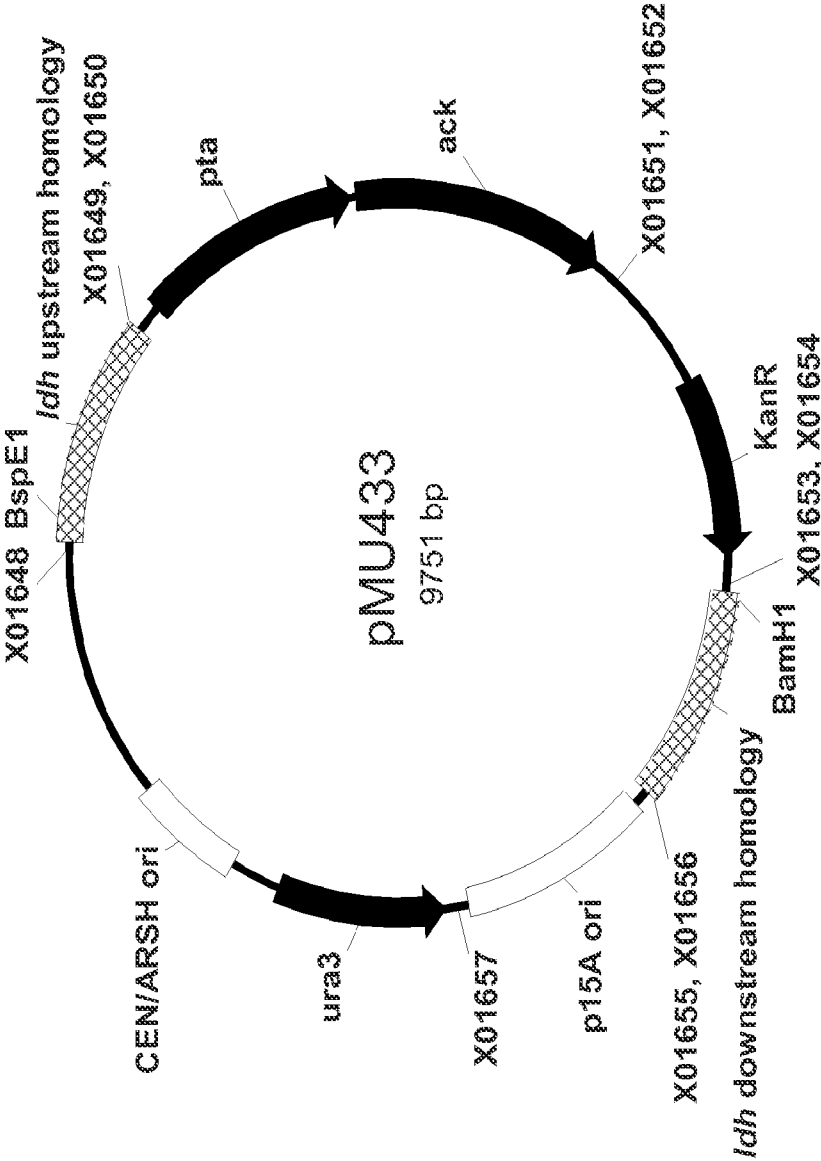
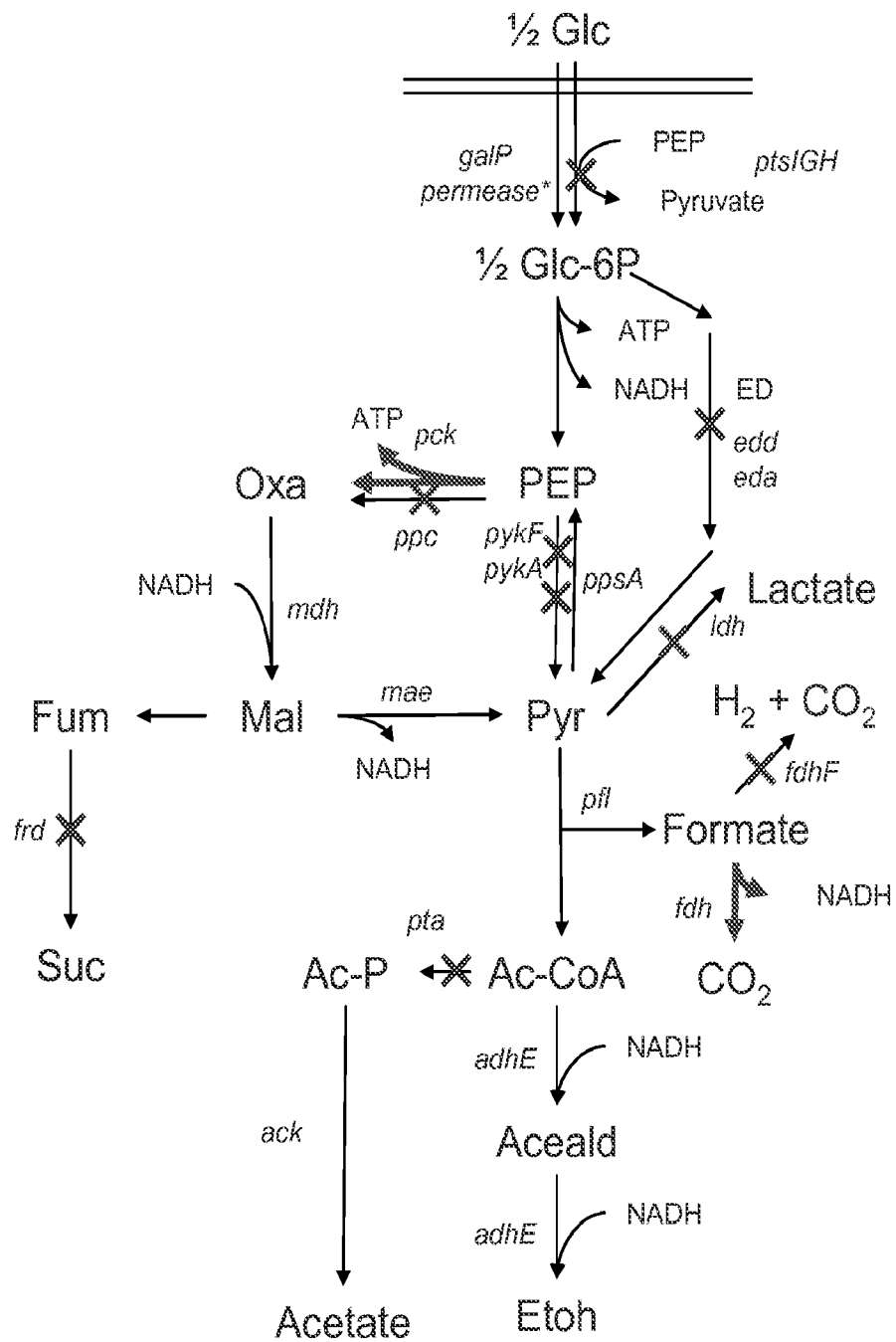


Figure 9



12/72

Figure 10

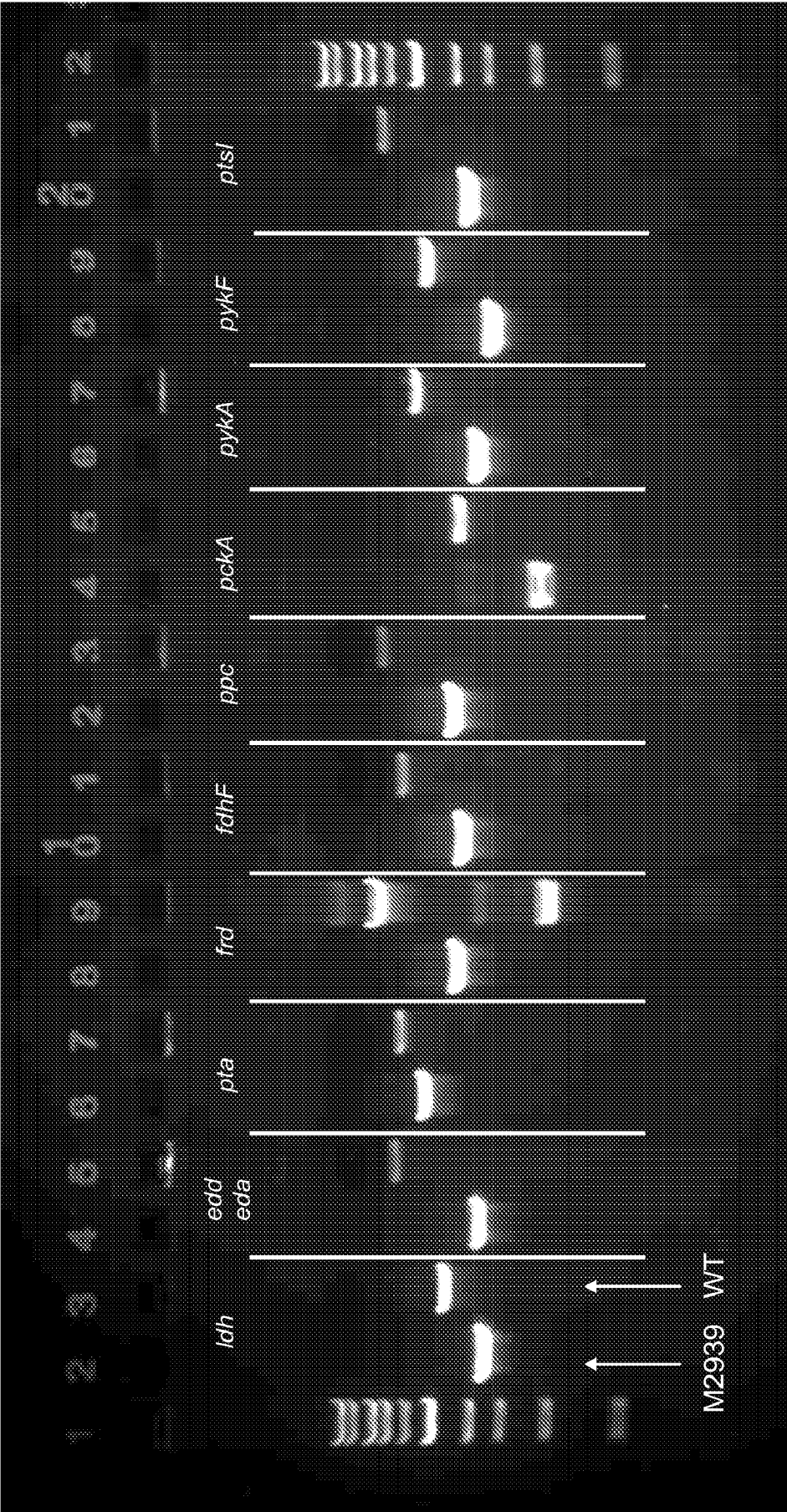


Figure 11

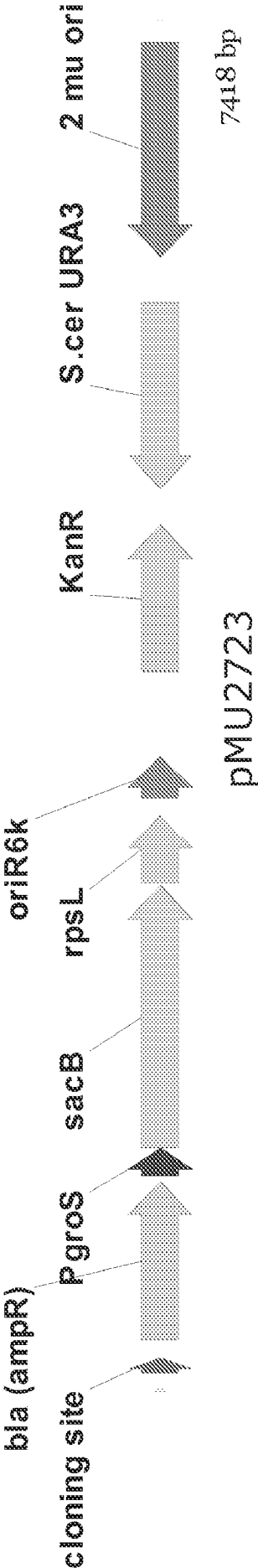
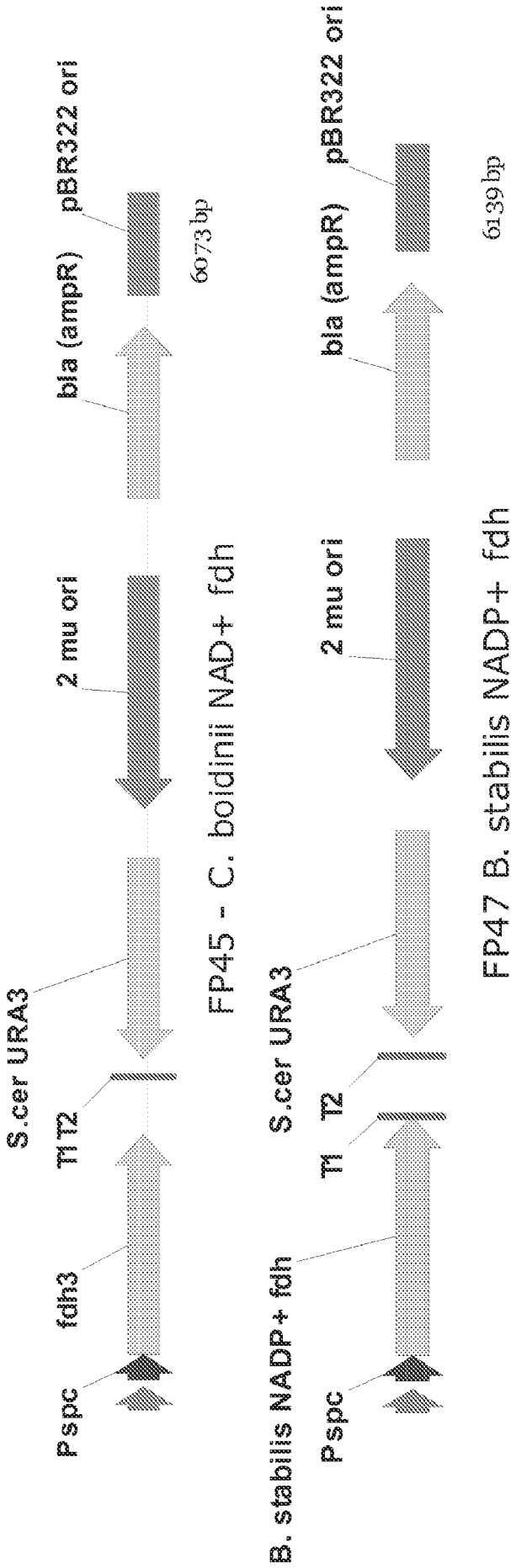


Figure 12

14/72



15/72

Figure 12 cont.

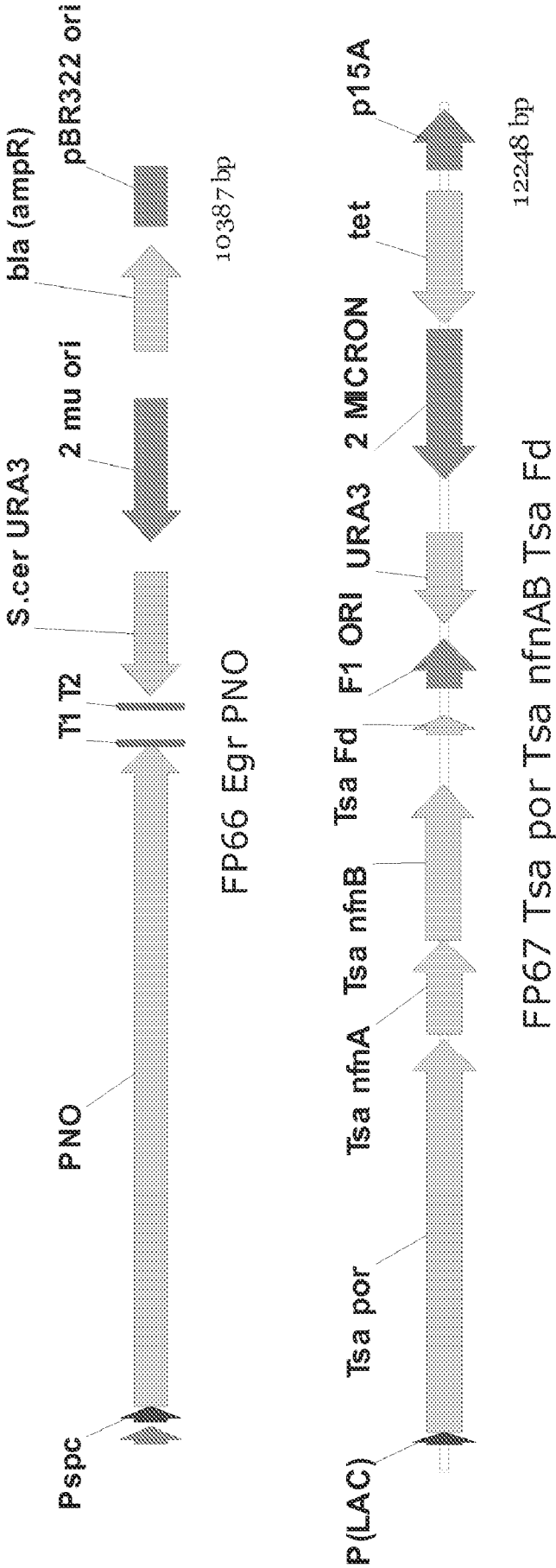
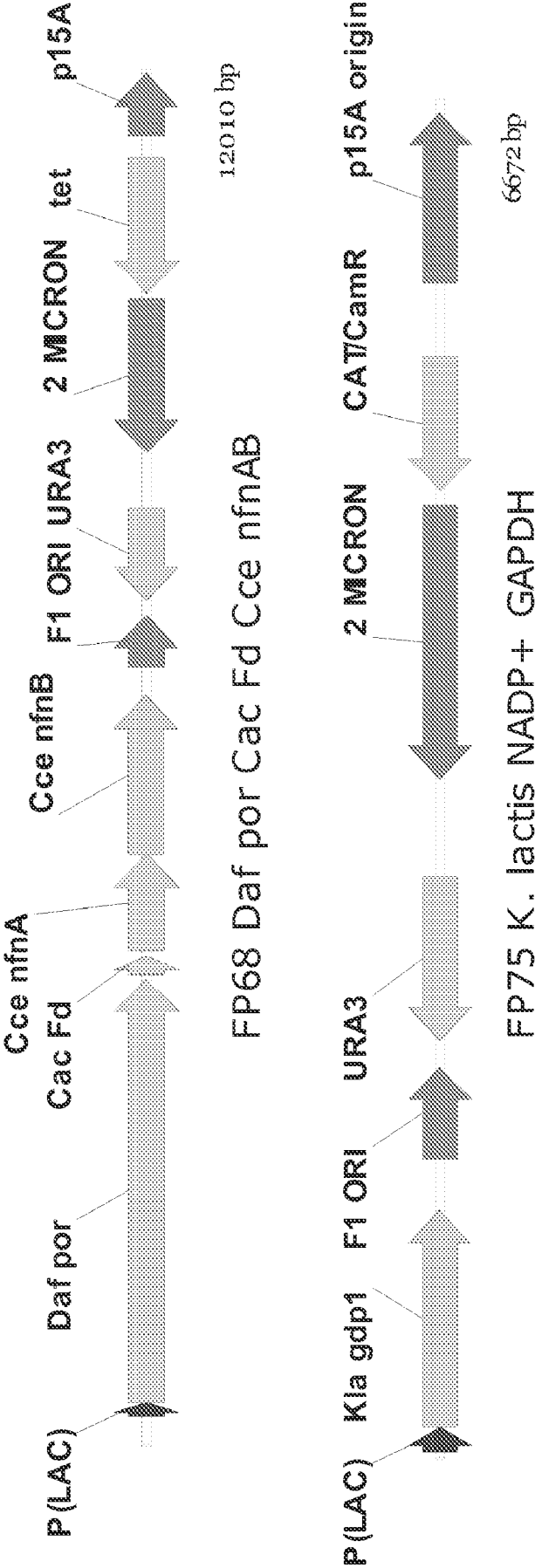


Figure 12 cont.



17/72

Figure 13

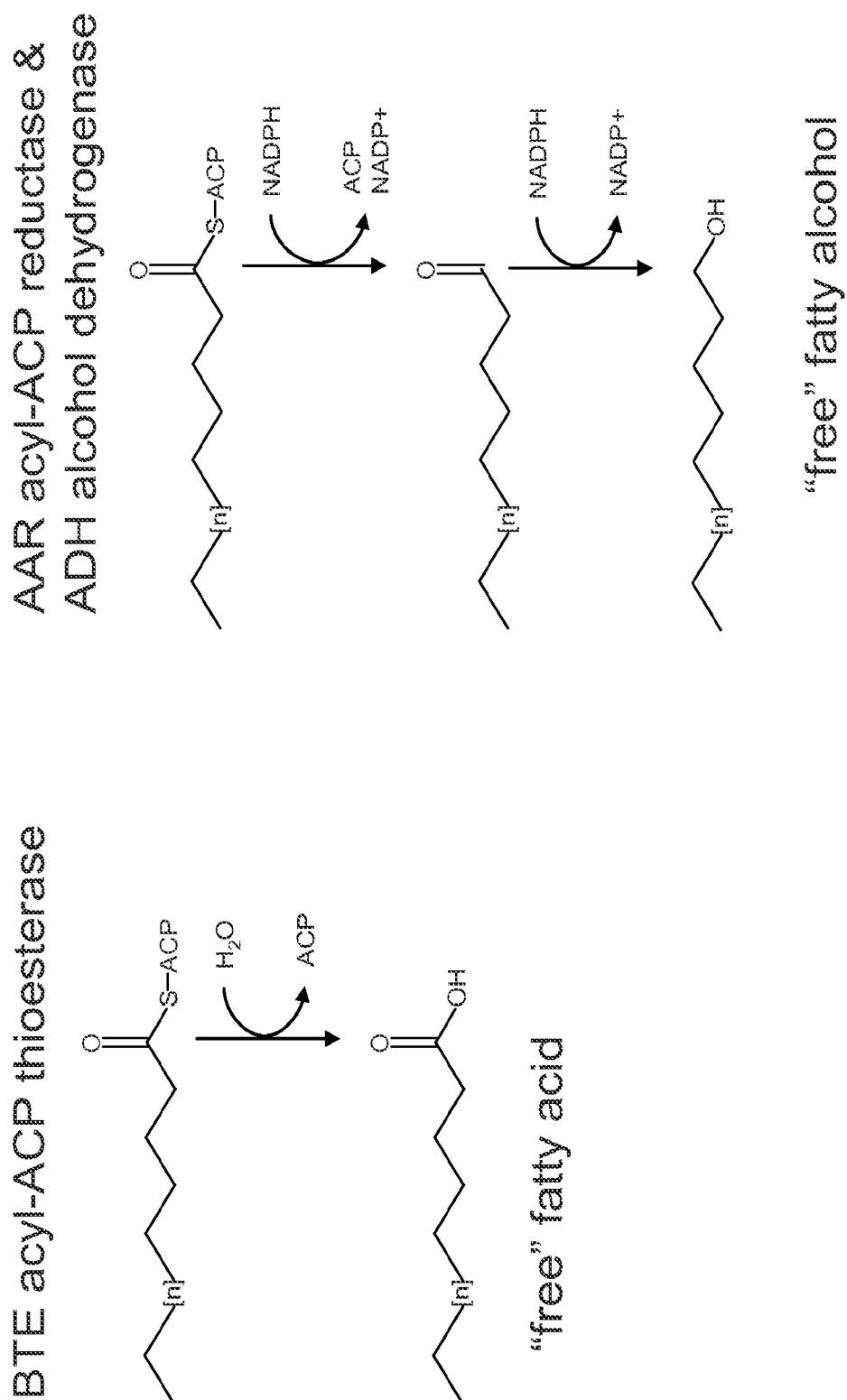
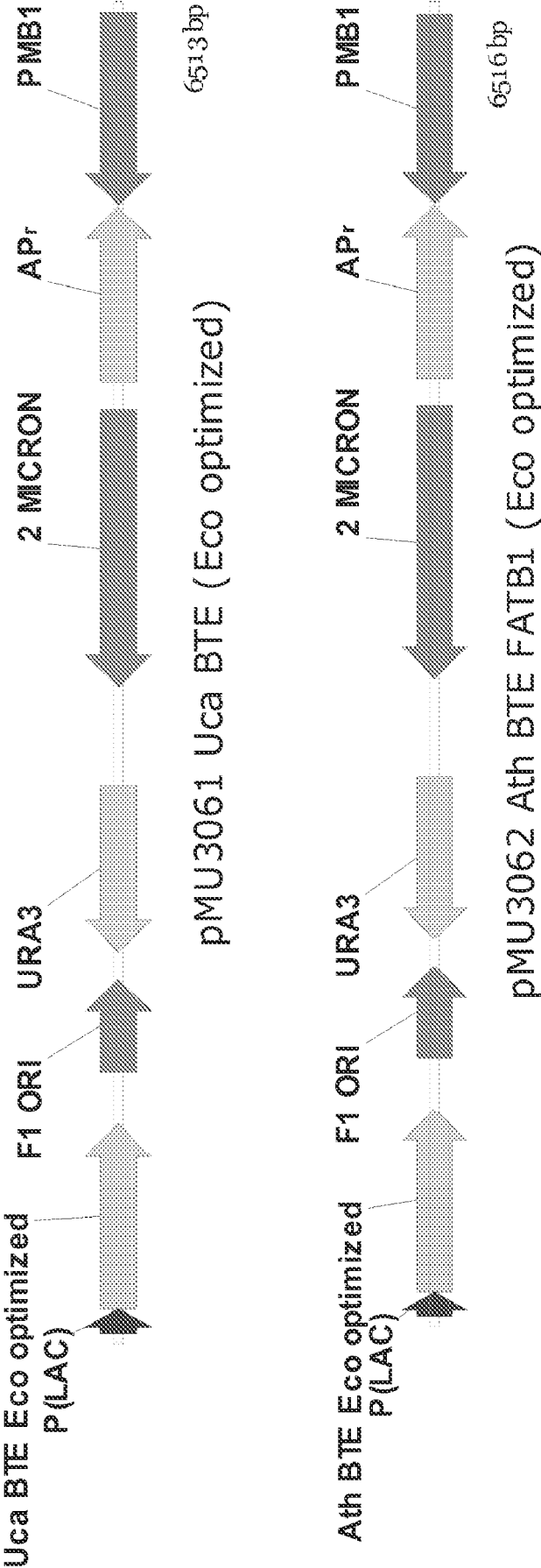


Figure 14



19/72

Figure 14 cont.

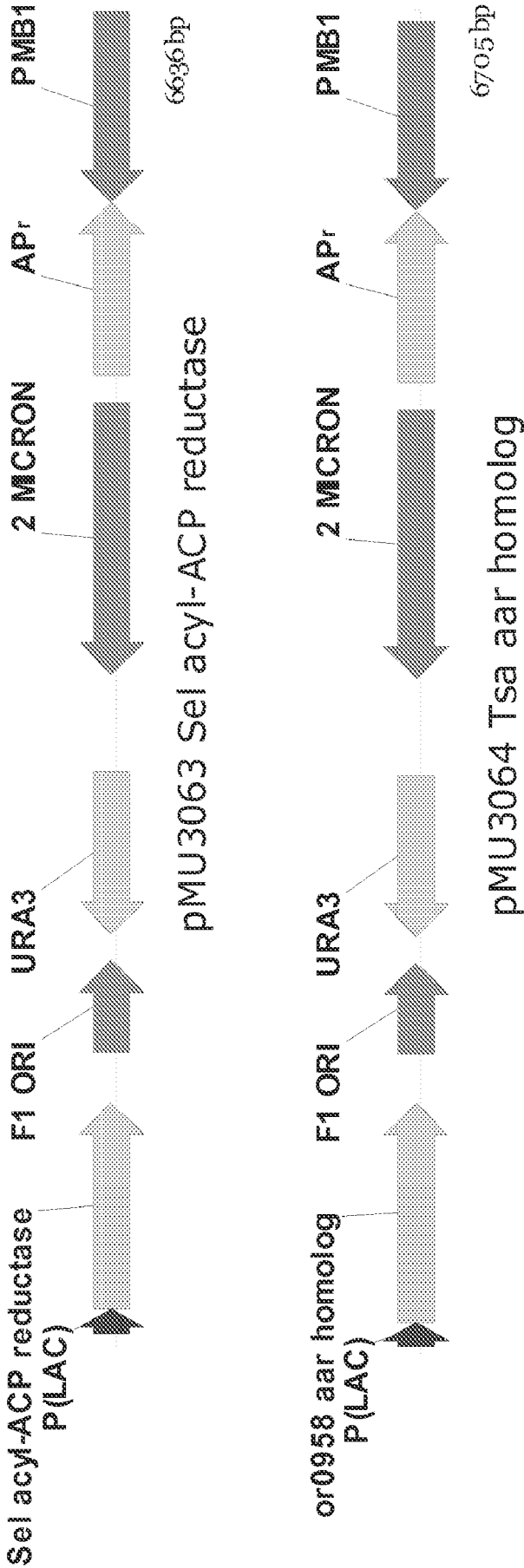


Figure 15

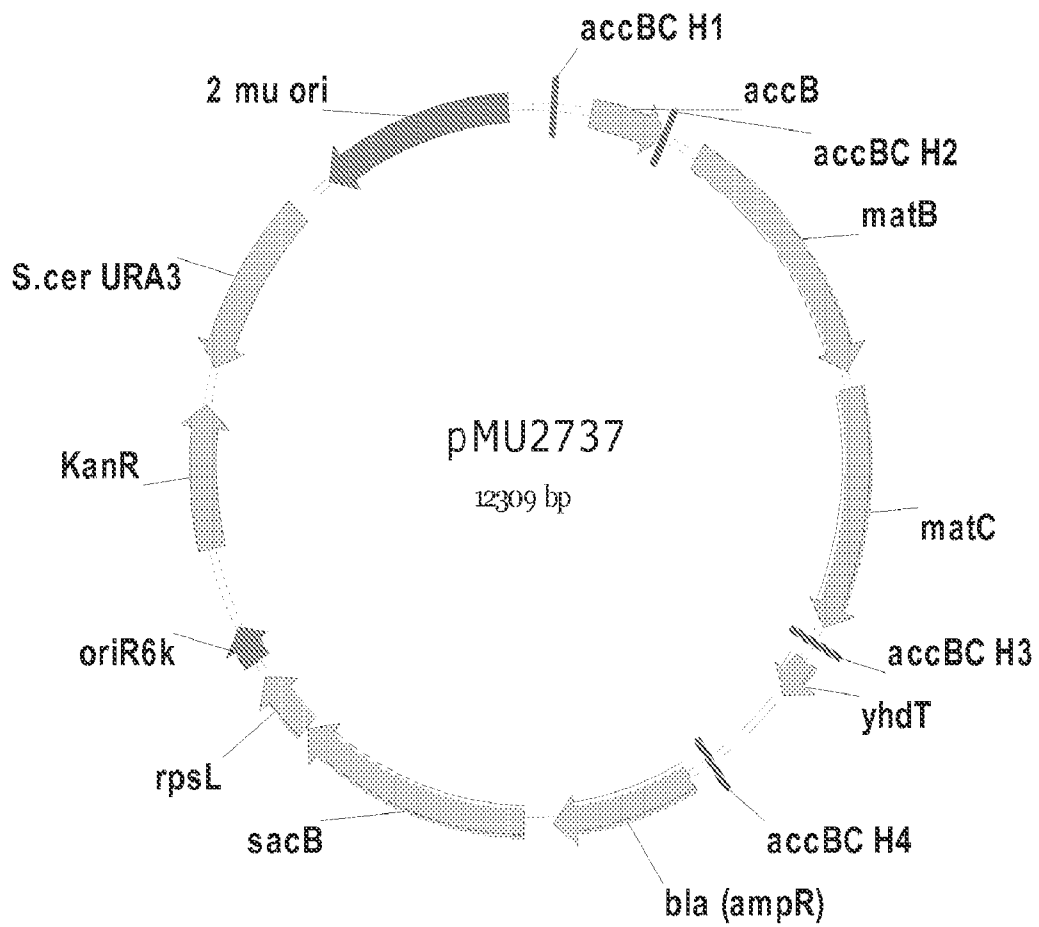


Figure 16

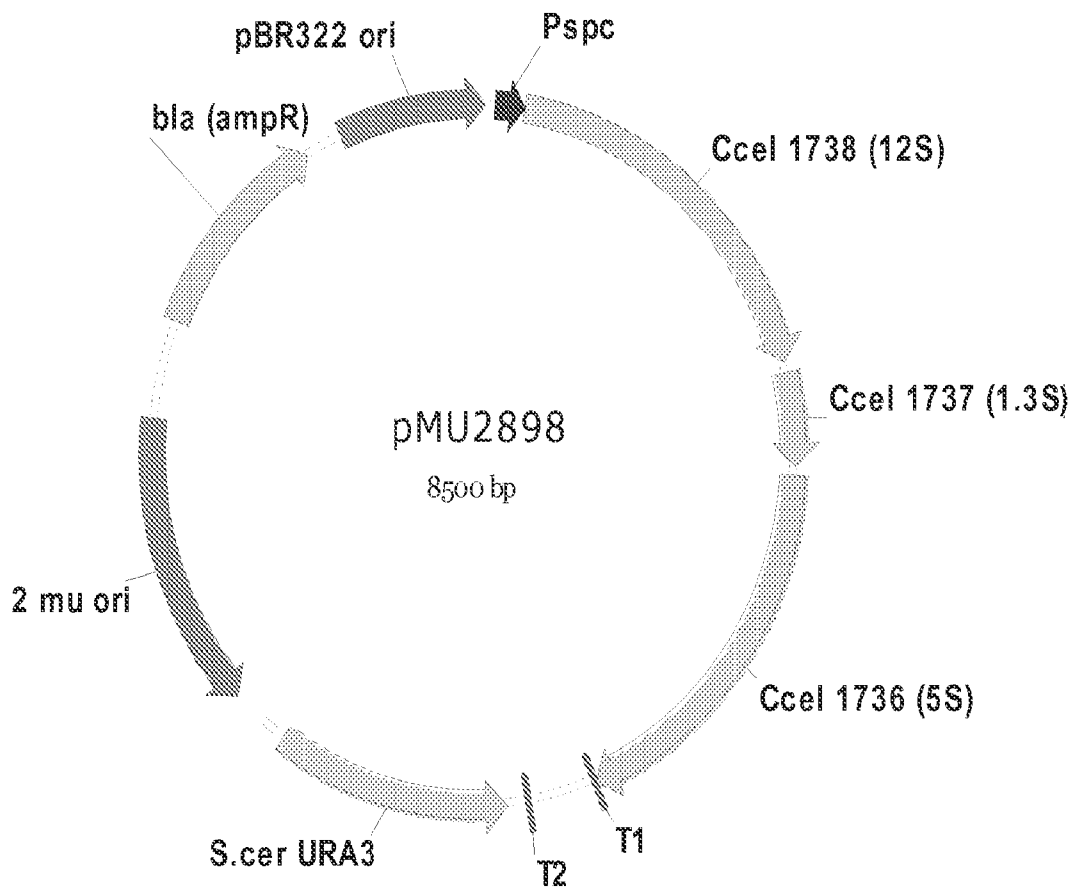


Figure 17

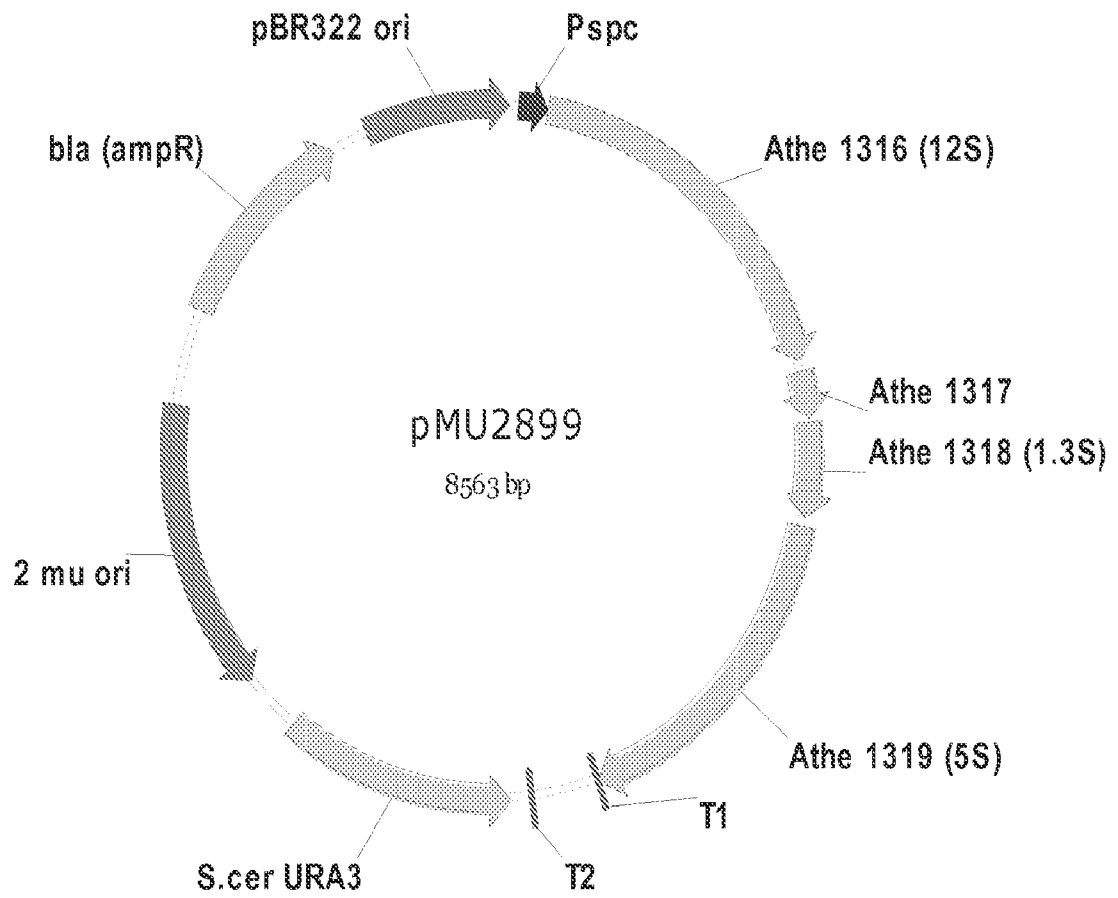


Figure 18

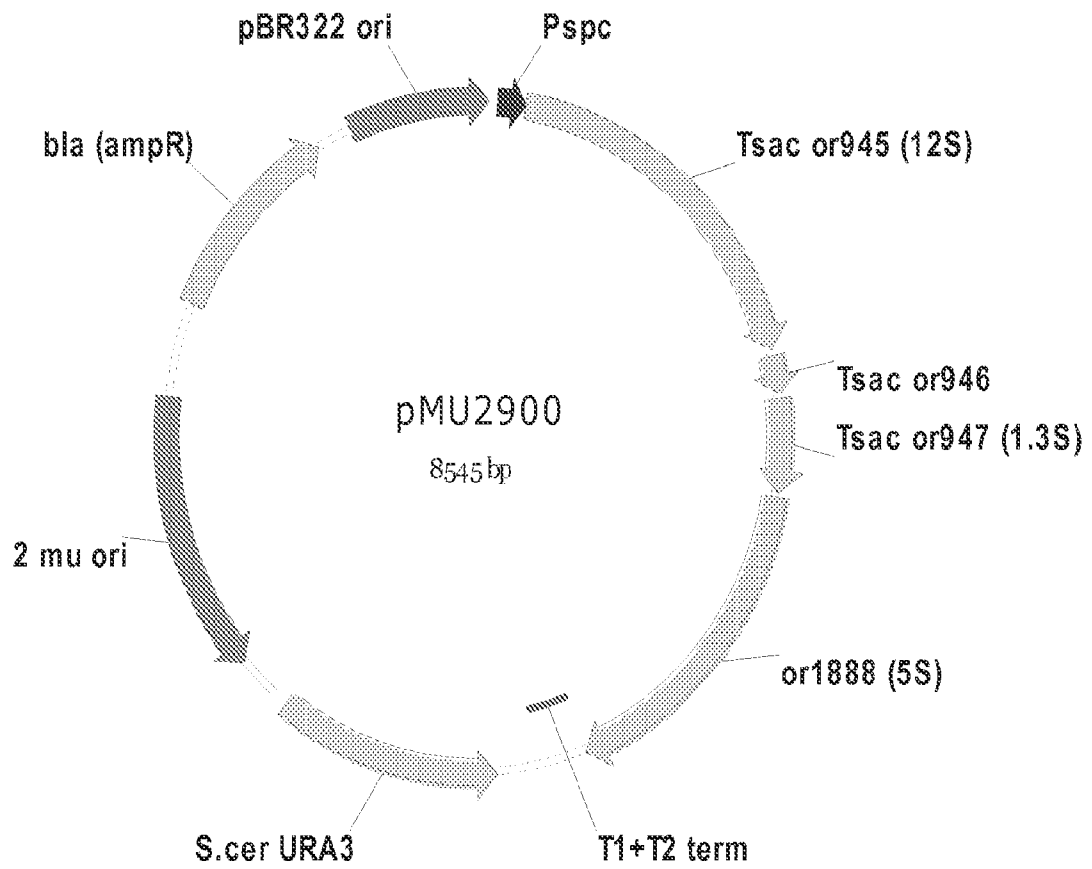
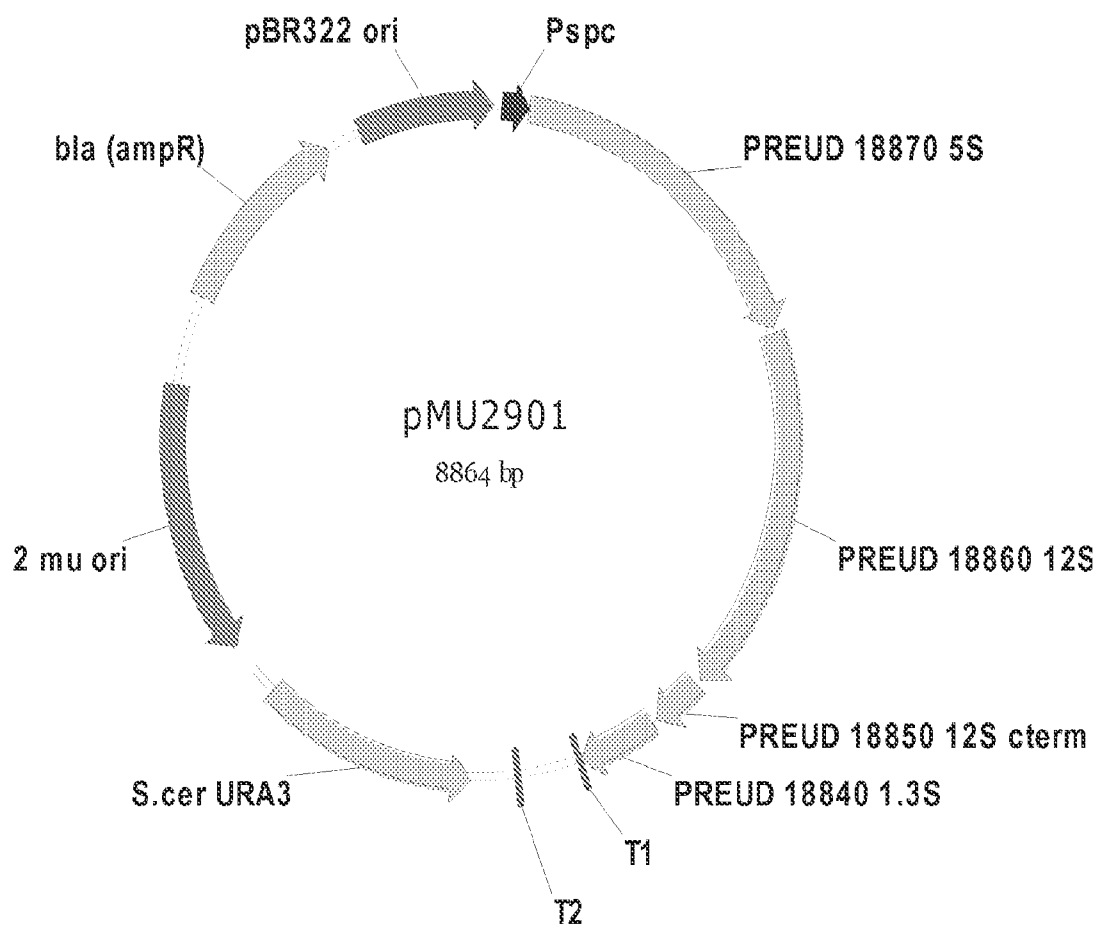


Figure 19



25/72

Figure 20

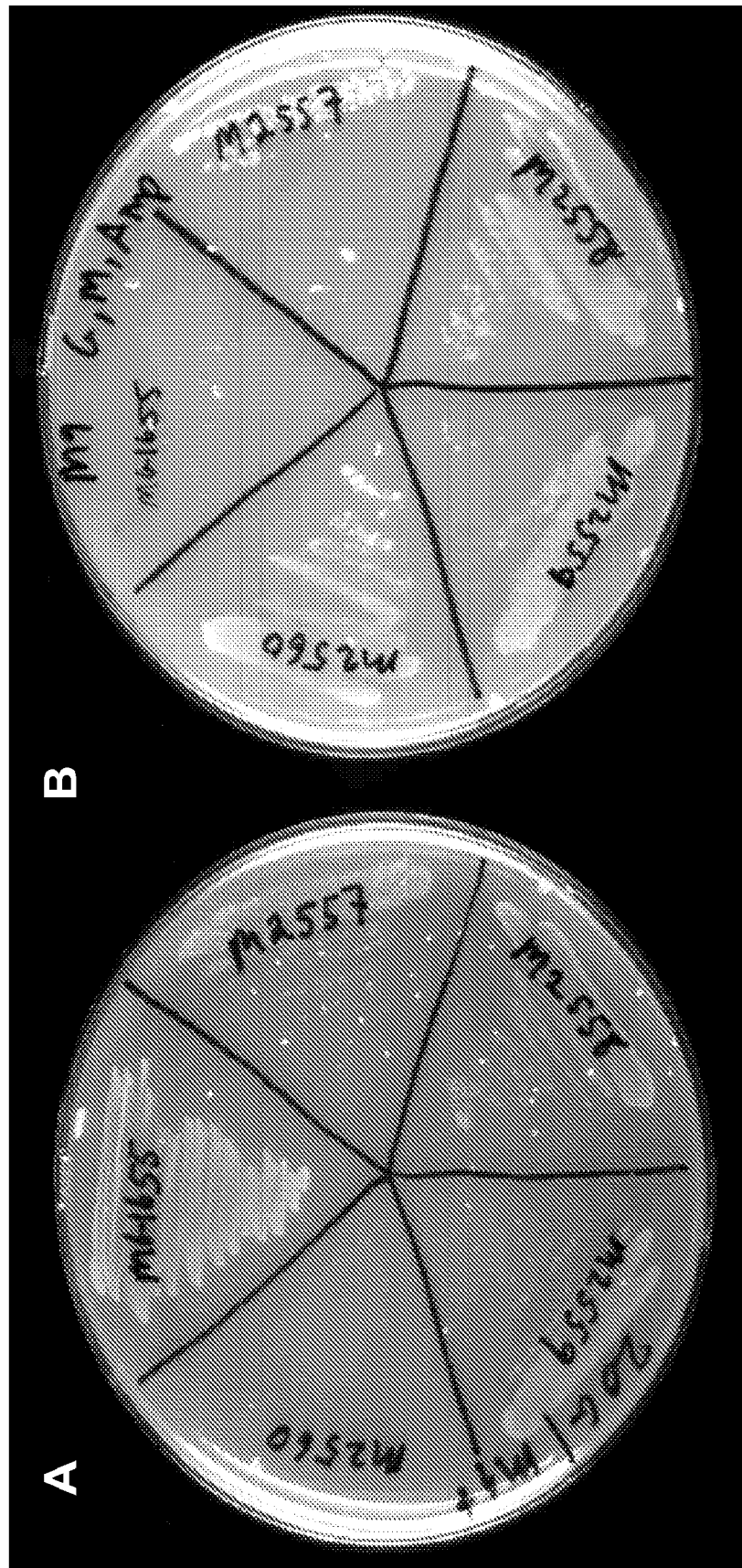


Figure 21

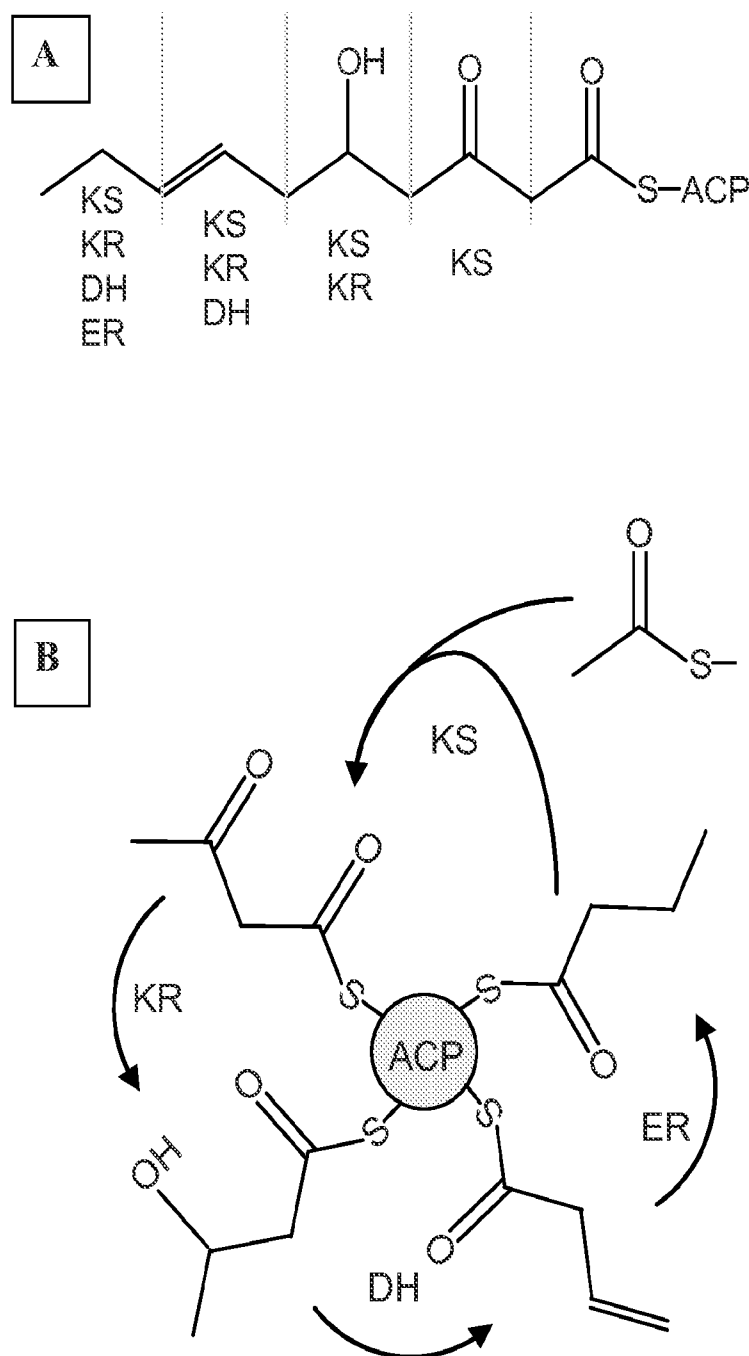


Figure 22A

	pMU690	pMU3061	pMU3062	pMU3063	pMU3064
Sum	45	154	201	39	54
12:0	0.00	49.53	2.82	0.25	0.25
14:0	1.94	1.88	70.37	3.57	1.76
14:1	0.00	15.92	2.56	0.16	0.14
15:0	0.00	0.00	0.43	0.14	0.00
15:1	0.00	0.00	0.00	0.00	0.00
16:0	21.87	5.39	19.30	12.41	20.21
16:1w5	0.00	0.00	0.00	0.00	0.00
16:1w7	5.74	4.01	75.74	6.28	5.57
17:0	0.49	0.38	0.52	0.37	0.34
17:1	5.76	1.85	3.21	4.93	5.72
18:0	1.76	1.94	1.72	1.37	2.19
18:1w9	0.37	0.54	0.16	0.14	0.17
18:1w7	5.73	10.16	7.53	5.36	12.13
18:2w6	0.00	0.36	7.65	0.34	0.35
20:0	0.00	0.00	0.00	0.00	0.00
20:1w9	0.00	0.00	0.49	0.00	0.27
20:2w6	0.00	0.00	3.64	0.00	0.00
20:3w6	0.00	0.00	1.33	0.00	0.00
20:5w3	0.00	0.00	1.05	0.00	0.00
22:0	0.00	0.00	0.00	0.00	0.00
22:1w9	0.30	0.13	0.29	0.23	0.22
24:0	0.00	0.00	0.00	0.00	0.00
other	0.57	61.53	2.05	3.72	4.42

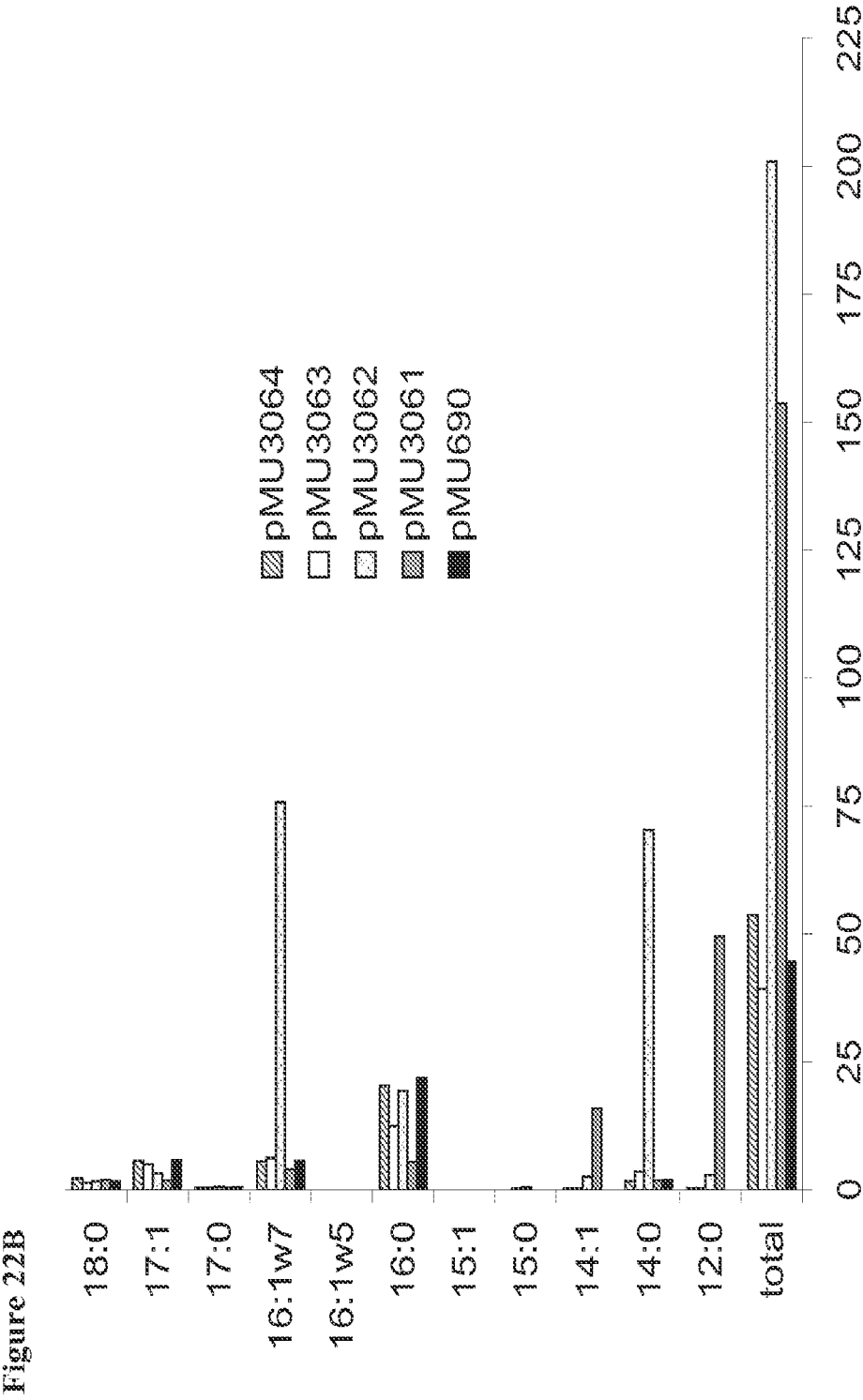


Figure 23

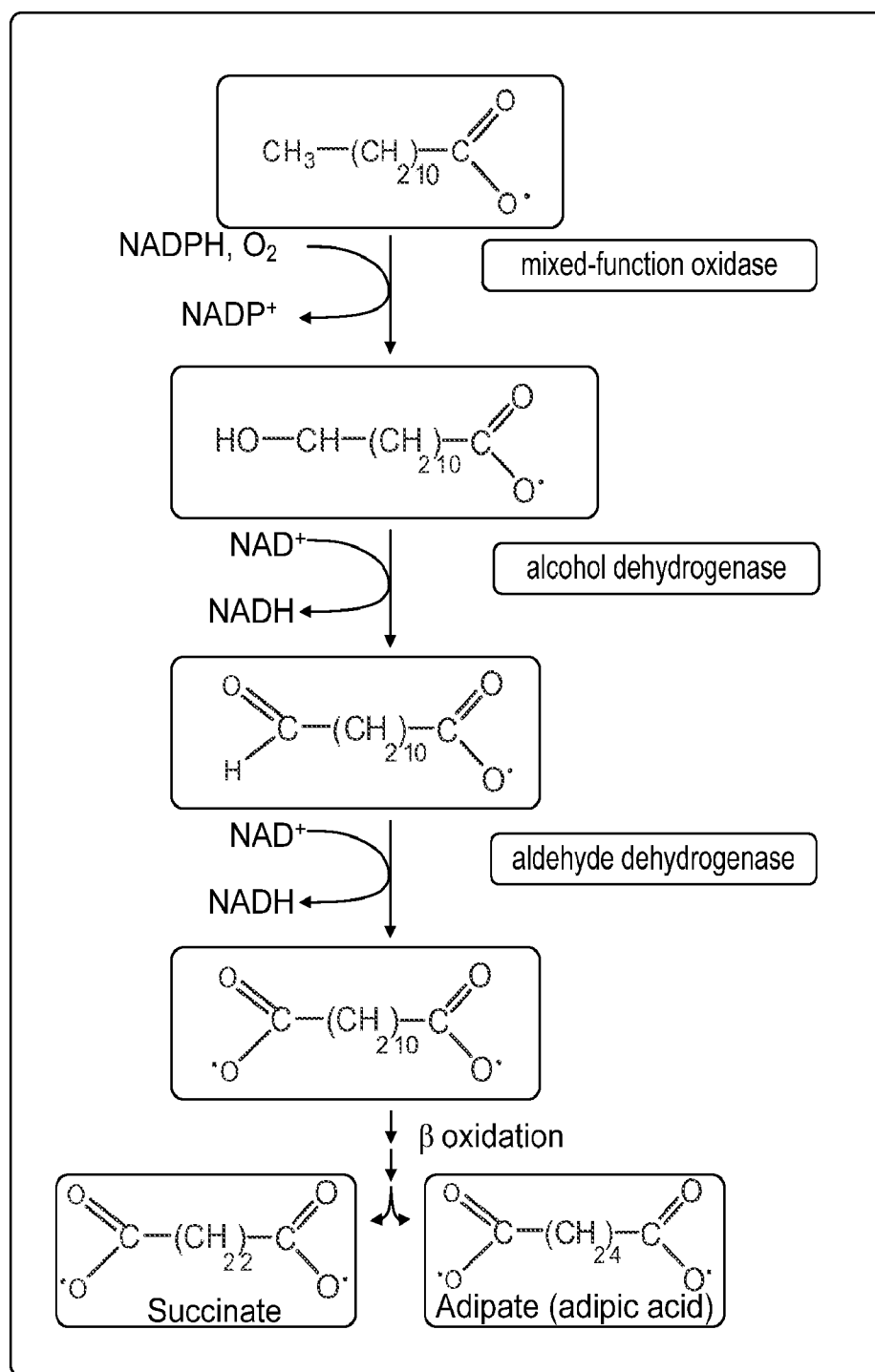


Figure 24

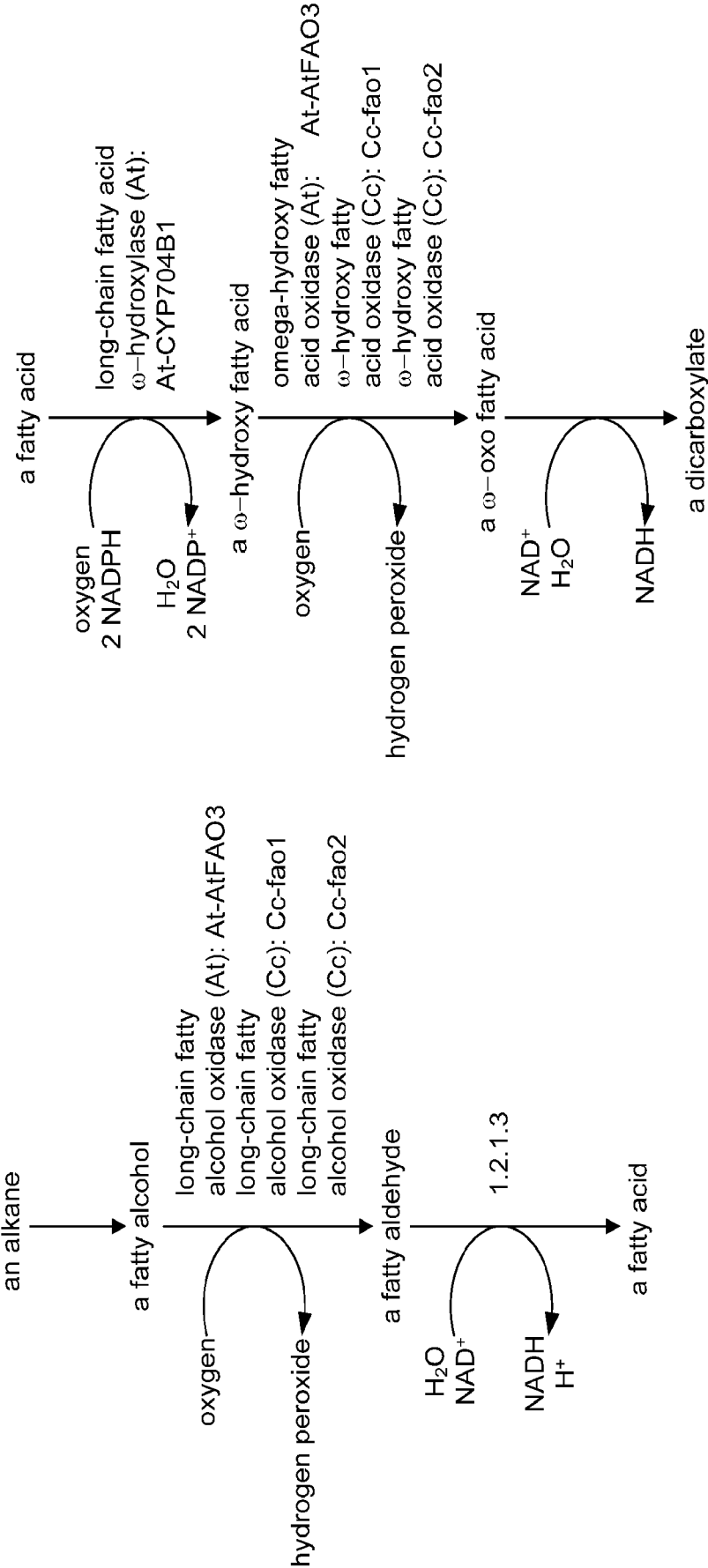


Figure 25

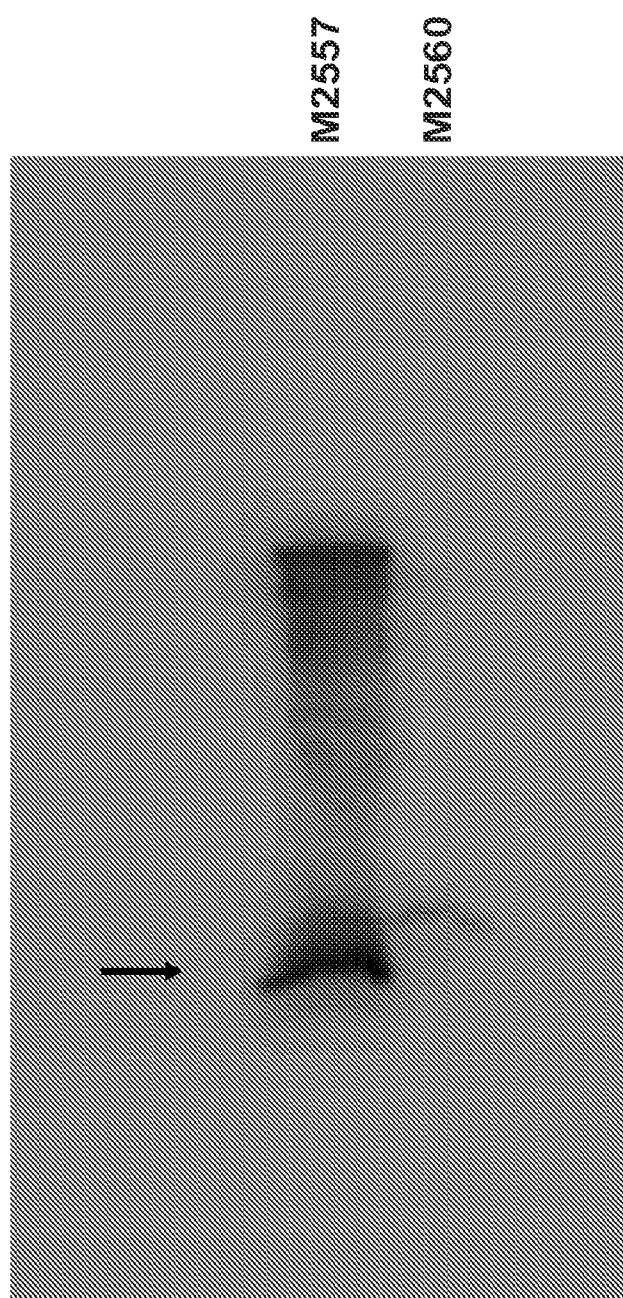


Figure 26A

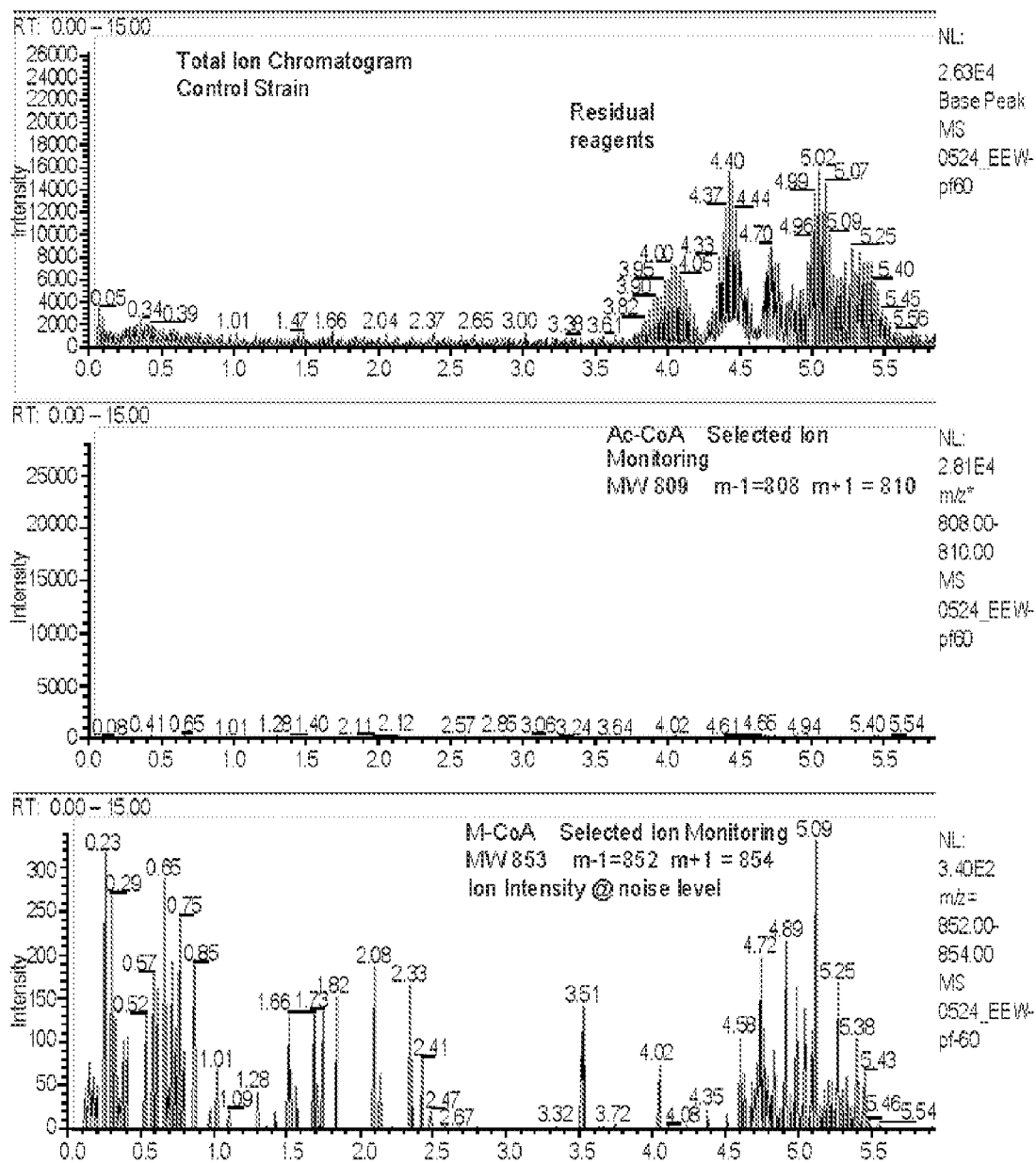


Figure 26A cont.

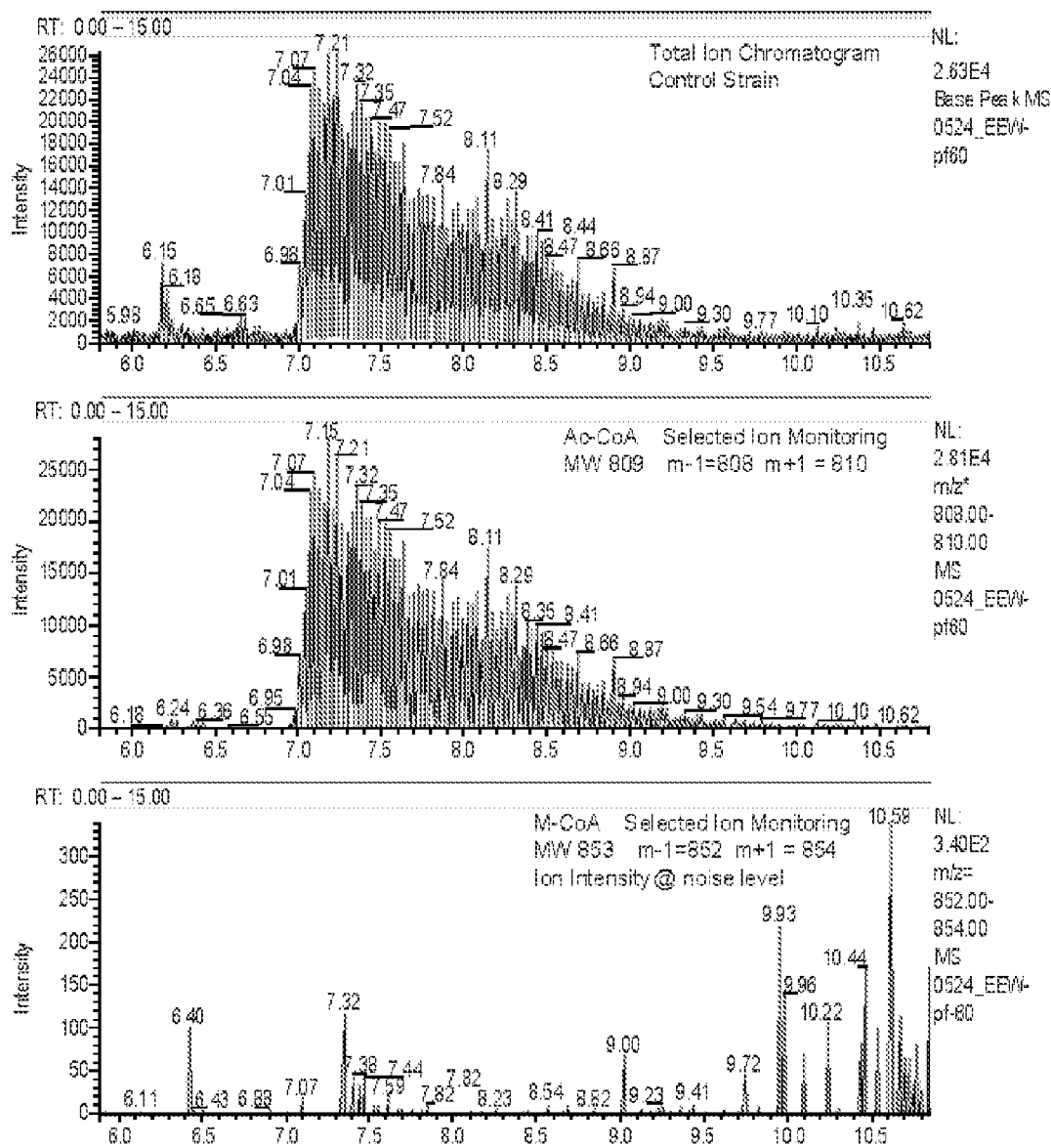


Figure 26A cont.

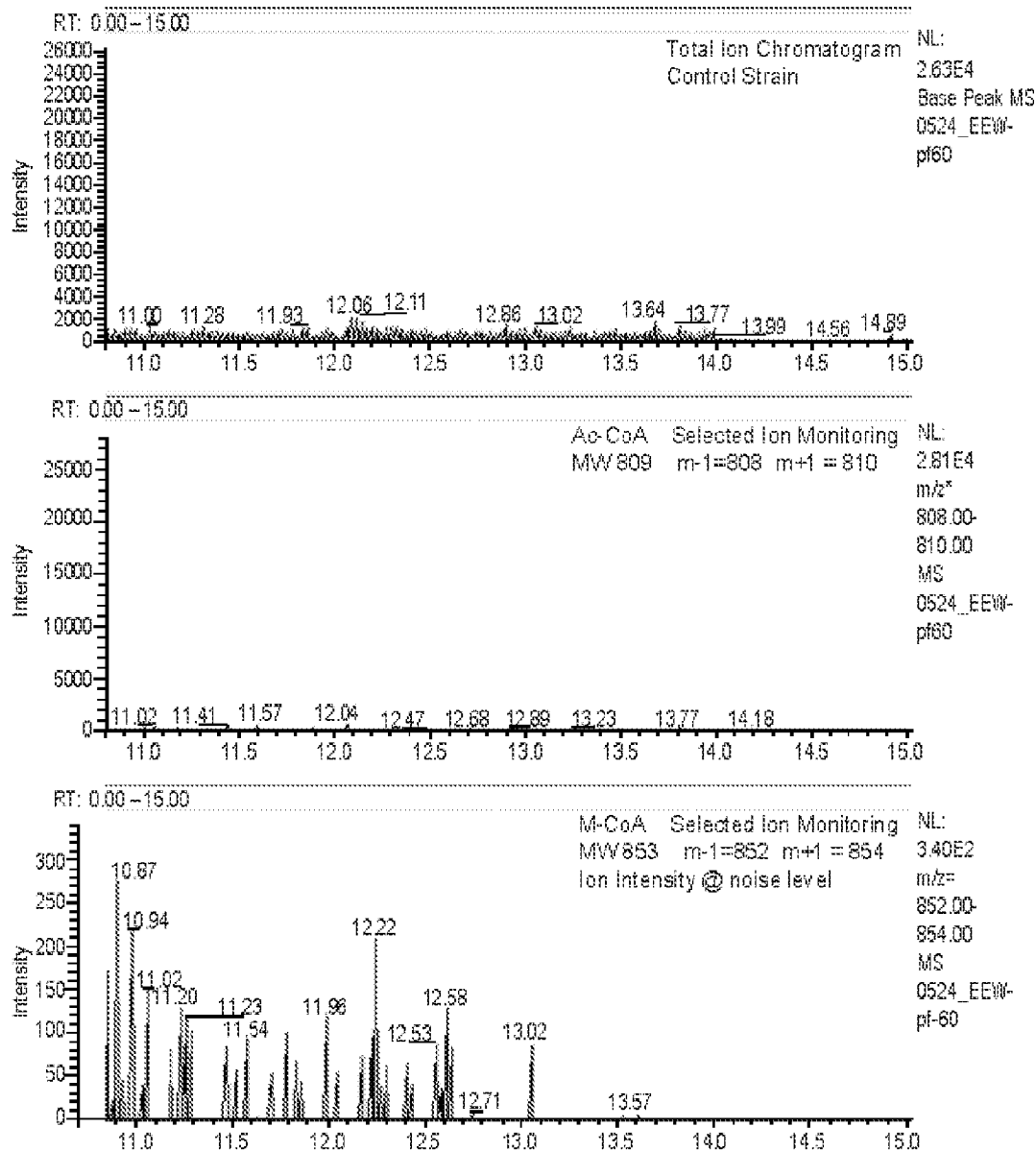


Figure 26B

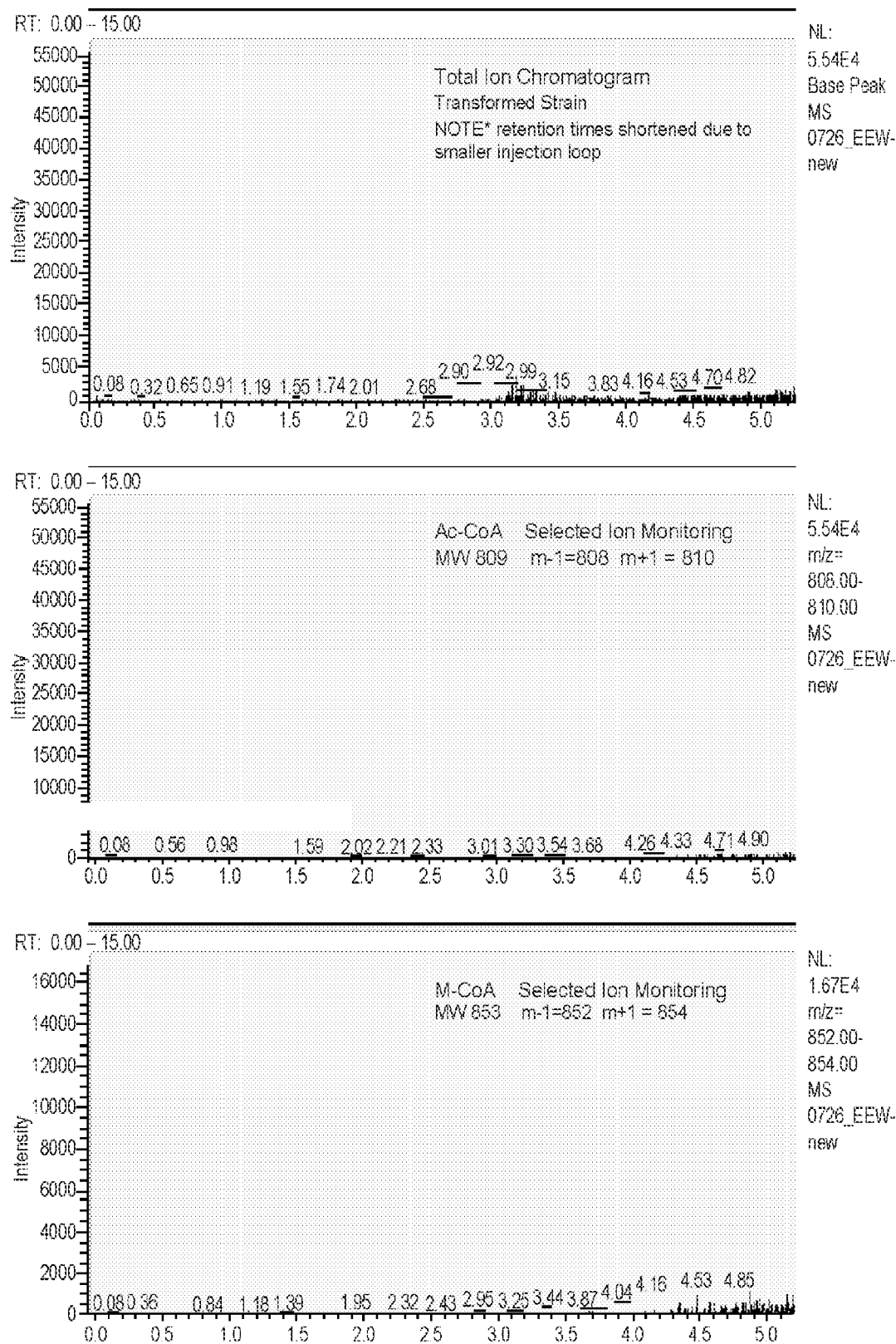


Figure 26B cont.

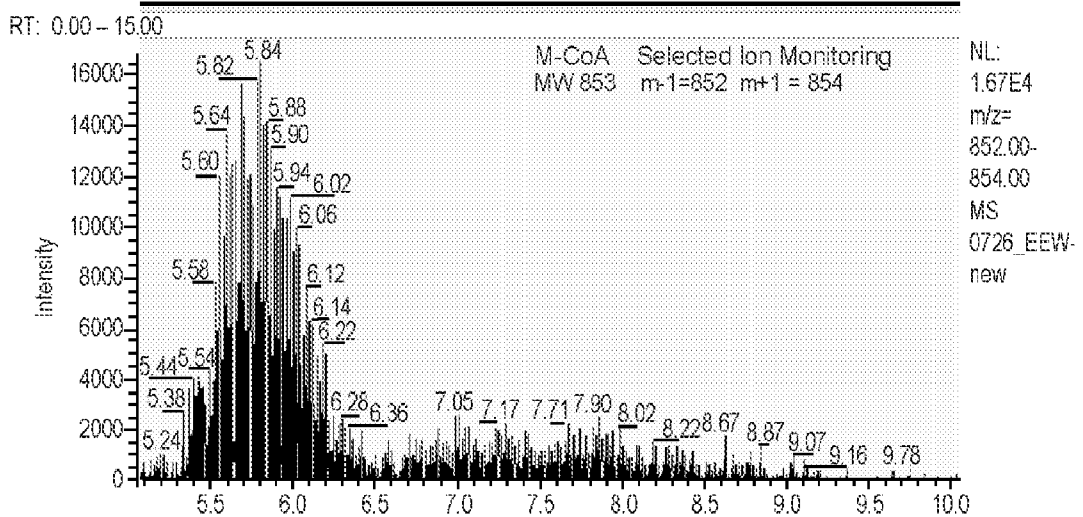
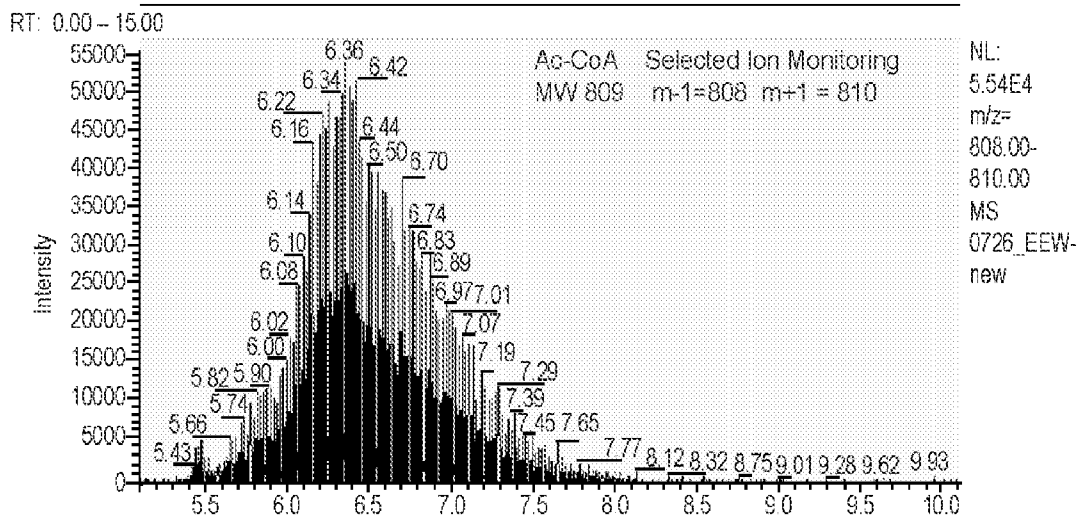
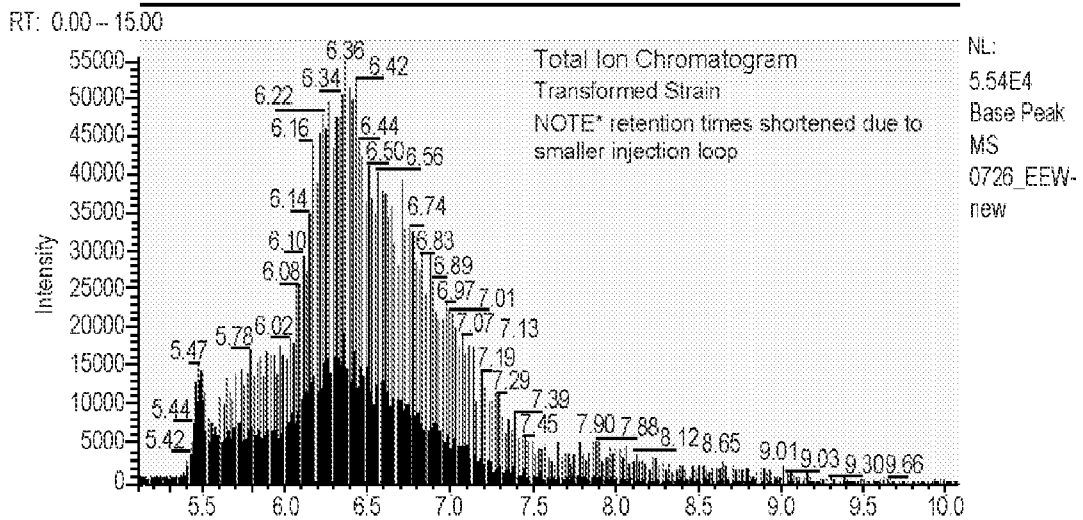


Figure 26B cont.

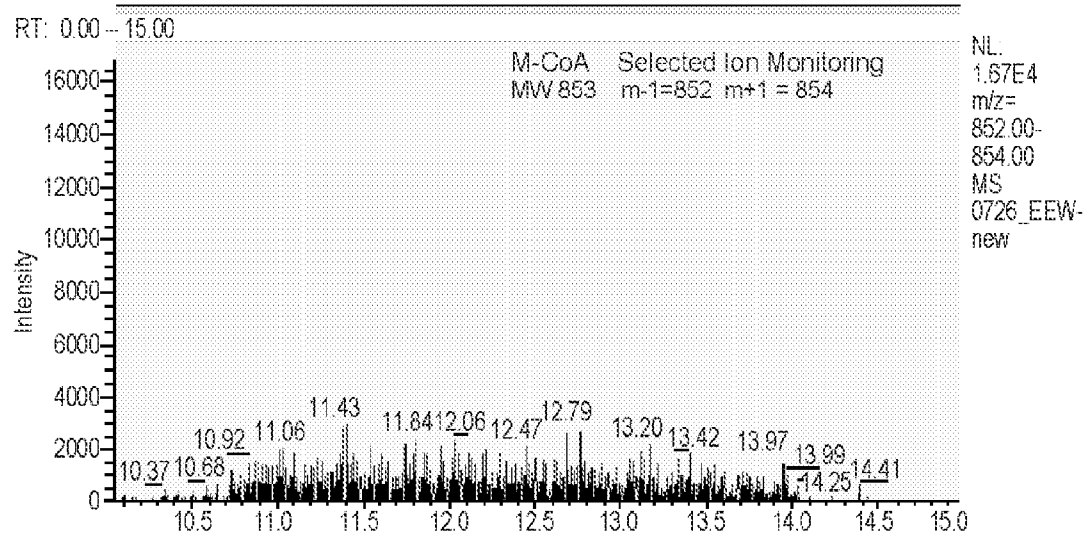
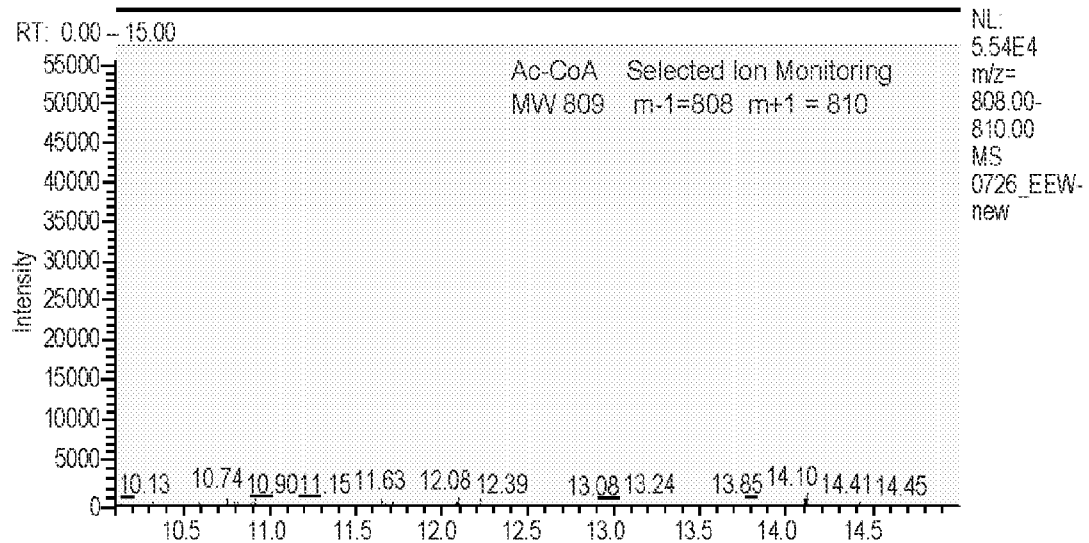
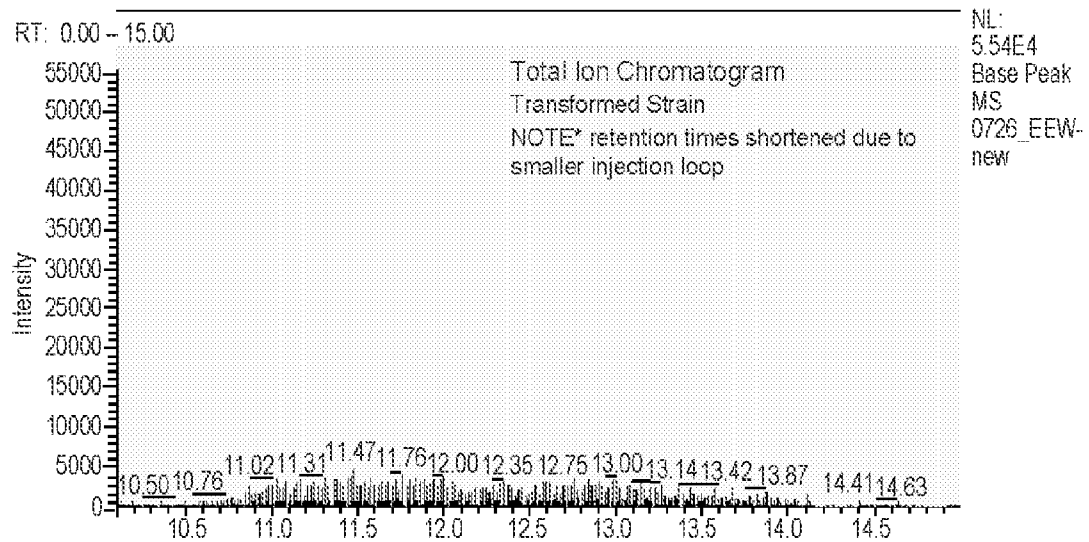
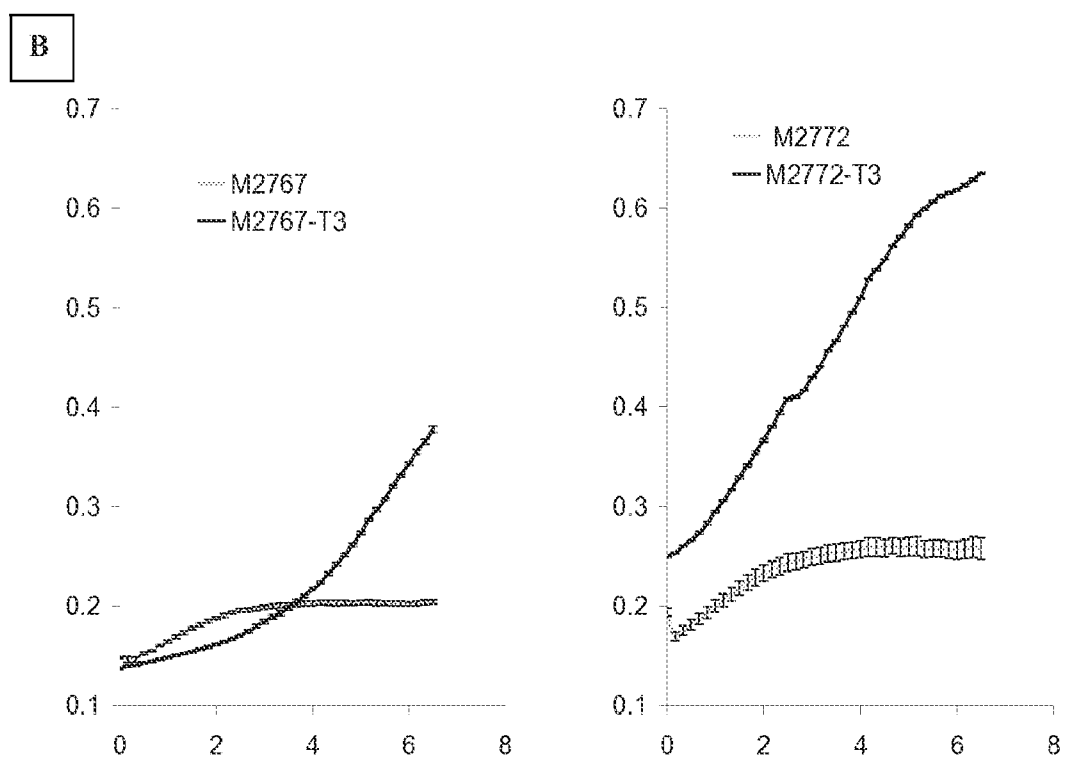
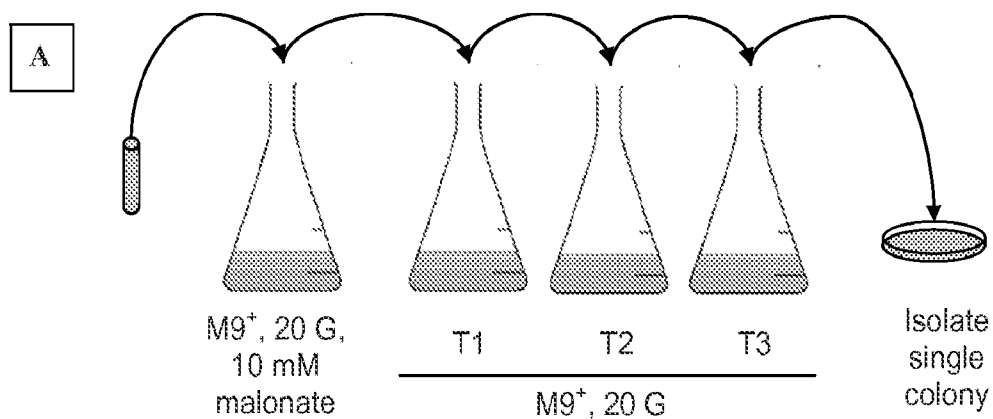


Figure 27



	$\mu_{\max}$	SD
M2767	0.005	0.00
M2767-T3	0.24	0.00

	$\mu_{\max}$	SD
M2772	0.01	0.00
M2772-T3	0.18	0.00

39/72

Figure 28

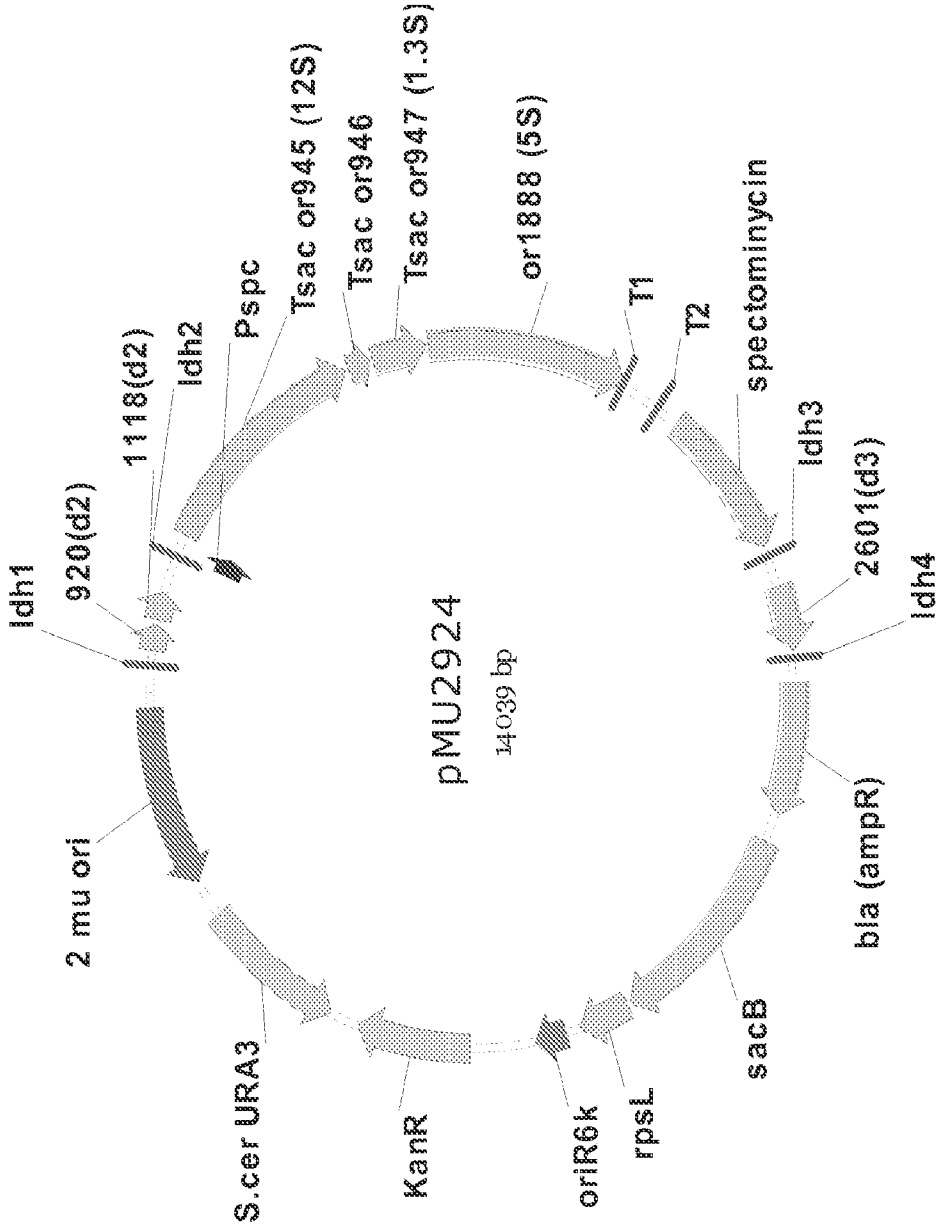


Figure 29

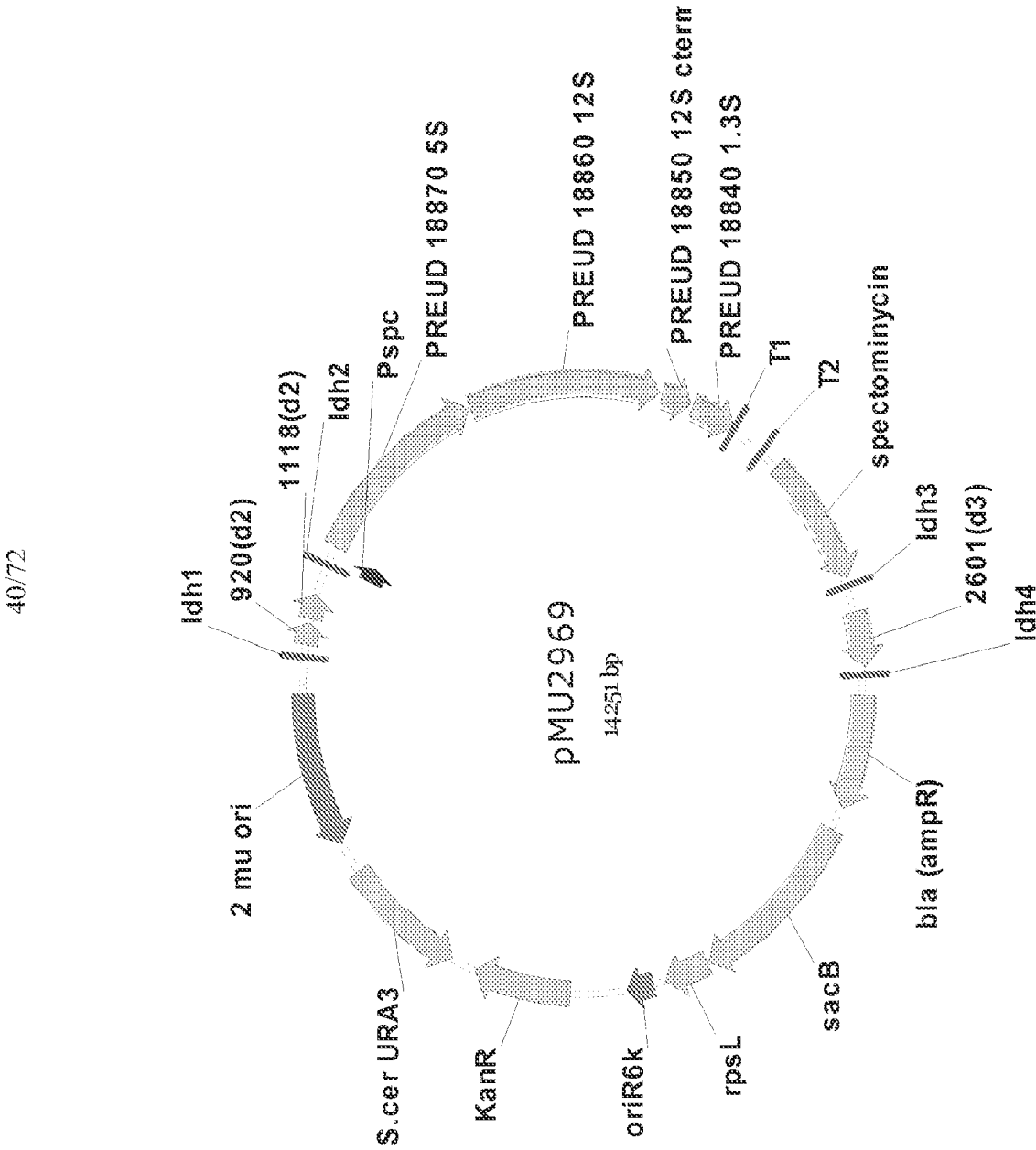


Figure 30

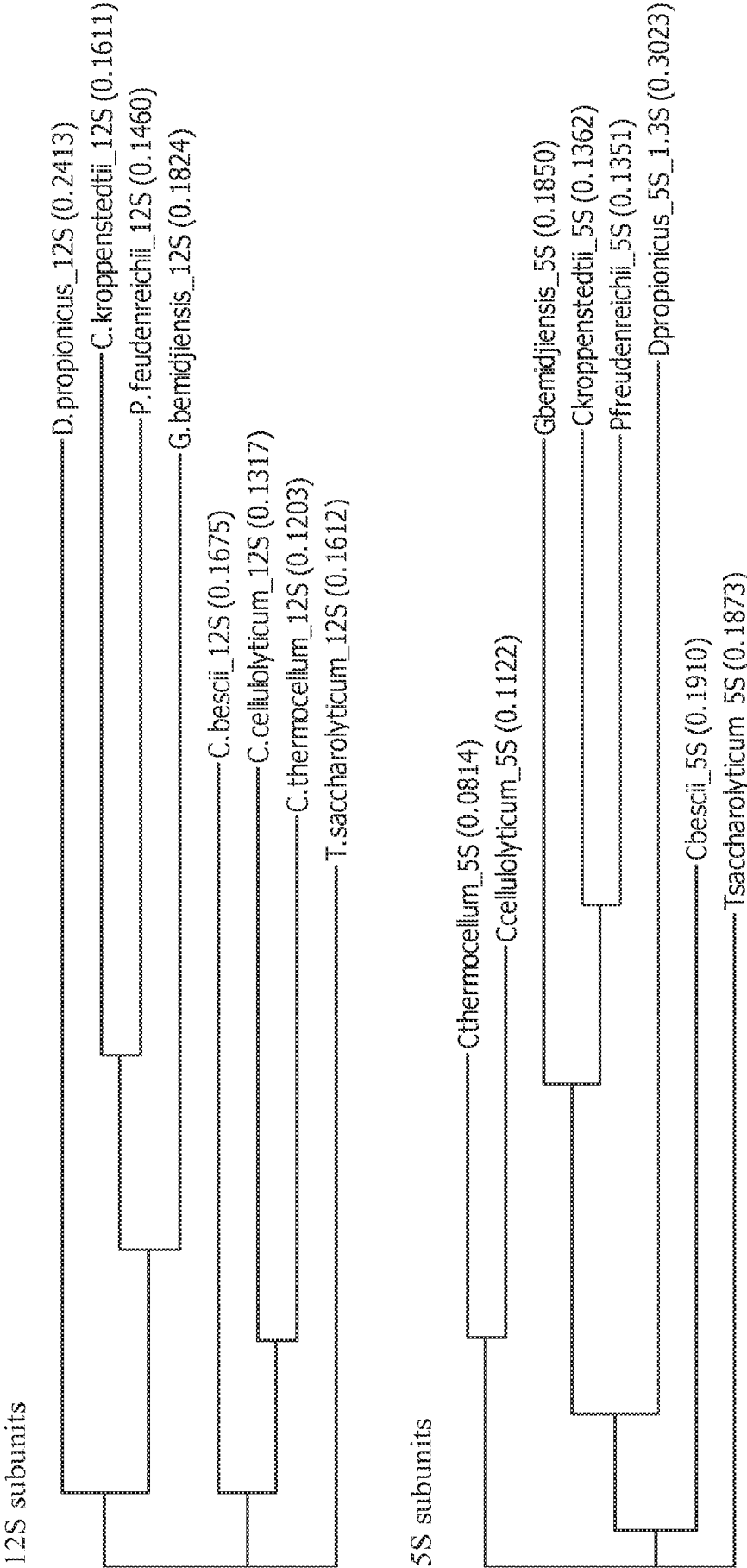


Figure 30 cont.

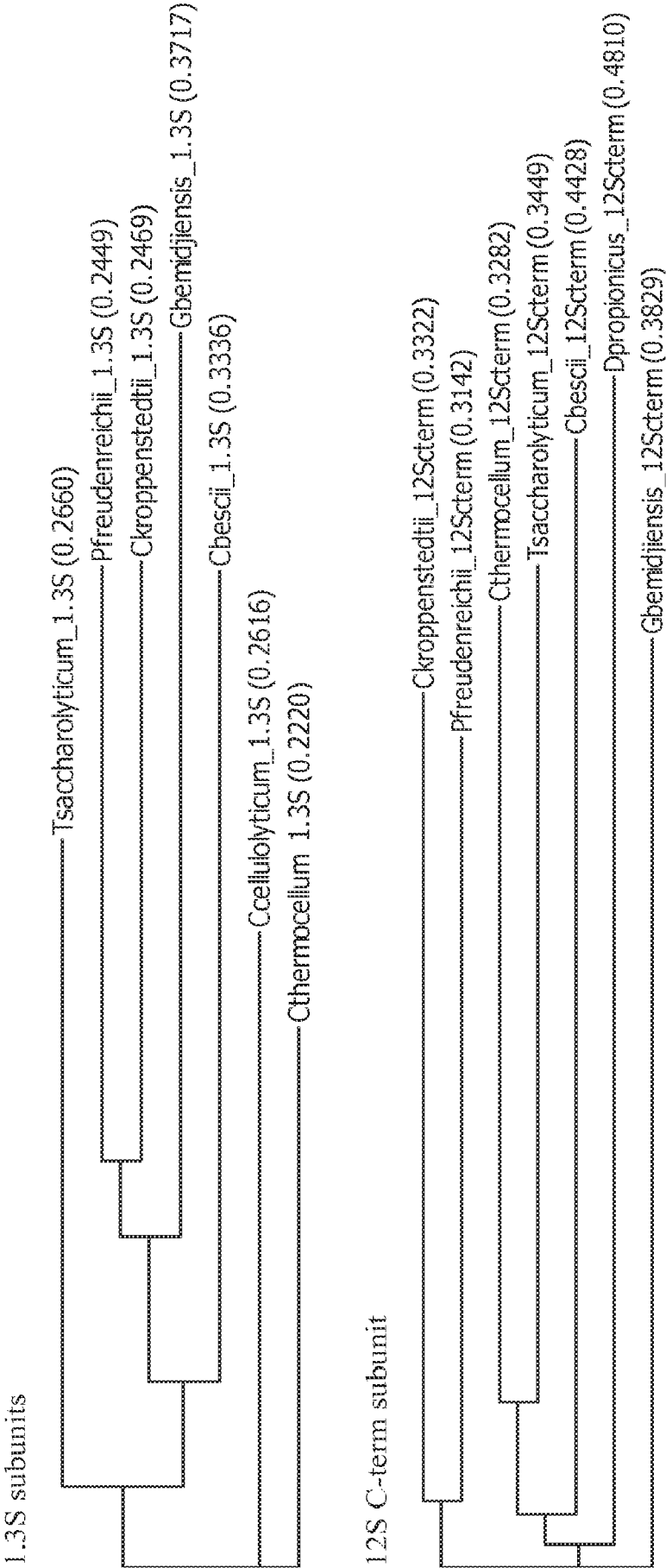


Figure 31

12S subunits

		1	50
D.propionicus_12S	(1)	-----TKK E E C A A L L I Q A I I D Y	
C.kroppenstedtii_12S	(1)	MSEQ---PHDP F P G G E E E N E R L S Q Q A S P A R	
P.feudenreichii_12S	(1)	MAENNNKLAS E E G E E A G V E A E E F A A C C O	
G.bemidjiensis_12S	(1)	-----E E K K A A A A S S K L E R E A G C E E A	
C.bescii_12S	(1)	-----T T E P R E C E E L K I E E E A A E S K E E	
C.cellulolyticum_12S	(1)	-----Q I S Q C A M M K A K E E E A A S P D	
C.thermocellum_12S	(1)	-----D K V S G L E E L A G E Q C A A A A A D A	
T.saccharolyticum_12S	(1)	-----D D E E L E E E M L E E L A S E E E K	
Consensus	(1)	MSM EKI QLKEKR KI GGG DKDKQHA GKLTARER	
		51	100
D.propionicus_12S	(41)	Q L E P C E F Y A K I S Y V A I I P F A I A E	
C.kroppenstedtii_12S	(48)	RTK E E E L I G M A M T T T A A A A A A A A A A	
P.feudenreichii_12S	(51)	NNN P P H E E E E E K I T T I A A V A A A A A	
G.bemidjiensis_12S	(41)	SEA E E E L E G I N A S T T A A A K E E E E E	
C.bescii_12S	(39)	E E E P G E E E M M A E E E E E T T F E C E A	
C.cellulolyticum_12S	(41)	SH E A A A A A A A A A A A A A A A A A A	
C.thermocellum_12S	(41)	Q A A A A A A A A A A A A A A A A A A A	
T.saccharolyticum_12S	(41)	Y Y K E E E E E E E E E E E E E E E E E	
Consensus	(51)	I LLDE SP EIDAEVEHRC DFGMDK KVGDCGVVCGYGTI GR VHV	
		101	150
D.propionicus_12S	(91)	Y A A A A A E G T I E E C A A A A A A A A A	
C.kroppenstedtii_12S	(98)	A A A A A A A A A A A A A A A A A A A A	
P.feudenreichii_12S	(101)	A A A A A A A A A A A A A A A A A A A A	
G.bemidjiensis_12S	(91)	A A A A A A A A A A A A A A A A A A A A	
C.bescii_12S	(89)	Y A A A A A A A A A A A A A A A A A A A	
C.cellulolyticum_12S	(91)	A A A A A A A A A A A A A A A A A A A A	
C.thermocellum_12S	(91)	A A A A A A A A A A A A A A A A A A A A	
T.saccharolyticum_12S	(91)	Y A A A A A A A A A A A A A A A A A A A	
Consensus	(101)	AAQDFTVIGGSLGEMHA KI KVMDYALK G PFIGINDSGGARIQEGID	
		151	200
D.propionicus_12S	(141)	E T A T A A A A A A A A A A A A A A A A	
C.kroppenstedtii_12S	(148)	A A A A A A A A A A A A A A A A A A A A	
P.feudenreichii_12S	(151)	A A A A A A A A A A A A A A A A A A A A	
G.bemidjiensis_12S	(141)	A A A A A A A A A A A A A A A A A A A A	
C.bescii_12S	(139)	A A A A A A A A A A A A A A A A A A A A	
C.cellulolyticum_12S	(141)	A A A A A A A A A A A A A A A A A A A A	
C.thermocellum_12S	(141)	A A A A A A A A A A A A A A A A A A A A	
T.saccharolyticum_12S	(141)	A A A A A A A A A A A A A A A A A A A A	
Consensus	(151)	ALSGYGDIFYRNVLASGVIPQISVINGECAGAVYSPALTDFTIMVDKTS	
		201	250
D.propionicus_12S	(191)	Y A A A A A A A A A A A A A A A A A A A	
C.kroppenstedtii_12S	(197)	A A A A A A A A A A A A A A A A A A A A	
P.feudenreichii_12S	(200)	A A A A A A A A A A A A A A A A A A A A	
G.bemidjiensis_12S	(190)	A A A A A A A A A A A A A A A A A A A A	
C.bescii_12S	(189)	A A A A A A A A A A A A A A A A A A A A	
C.cellulolyticum_12S	(191)	A A A A A A A A A A A A A A A A A A A A	
C.thermocellum_12S	(191)	A A A A A A A A A A A A A A A A A A A A	
T.saccharolyticum_12S	(191)	A A A A A A A A A A A A A A A A A A A A	
Consensus	(201)	QMFITGPQVIKSVTGEDVSAEELCGADTHSS SGVAHFVAE DEEA I I	
		251	300
D.propionicus_12S	(241)	E E E E E E E E E E E E E E E E E E E E	
C.kroppenstedtii_12S	(247)	A A A A A A A A A A A A A A A A A A A A	
P.feudenreichii_12S	(250)	A A A A A A A A A A A A A A A A A A A A	
G.bemidjiensis_12S	(240)	A A A A A A A A A A A A A A A A A A A A	
C.bescii_12S	(239)	Y A A A A A A A A A A A A A A A A A A A	
C.cellulolyticum_12S	(241)	A A A A A A A A A A A A A A A A A A A A	
C.thermocellum_12S	(241)	A A A A A A A A A A A A A A A A A A A A	
T.saccharolyticum_12S	(241)	A A A A A A A A A A A A A A A A A A A A	
Consensus	(251)	KKLLSFLF NNLEDPE V SDD I F D L D I I F D NKAYDMKDVT	
		301	350
D.propionicus_12S	(291)	E E E E E E E E E E E E E E E E E E E E	
C.kroppenstedtii_12S	(296)	A A A A A A A A A A A A A A A A A A A A	
P.feudenreichii_12S	(299)	A A A A A A A A A A A A A A A A A A A A	
G.bemidjiensis_12S	(289)	A A A A A A A A A A A A A A A A A A A A	
C.bescii_12S	(289)	A A A A A A A A A A A A A A A A A A A A	
C.cellulolyticum_12S	(291)	A A A A A A A A A A A A A A A A A A A A	
C.thermocellum_12S	(291)	A A A A A A A A A A A A A A A A A A A A	
T.saccharolyticum_12S	(291)	A A A A A A A A A A A A A A A A A A A A	
Consensus	(301)	KIVD GDFLEVQA FAQIVIGFARINGRSVGIVANQPKVMAGVLDINAS	

Figure 31 cont.

D.propionicus_12S	(341)	351	400
C.kroppenstedtii_12S	(346)		
P.feudenreichii_12S	(349)		
G.bemidjiensis_12S	(339)		
C.bescii_12S	(339)		
C.cellulolyticum_12S	(341)		
C.thermocellum_12S	(341)		
T.saccharolyticum_12S	(341)		
Consensus	(351)	OKAAREFVRFCDAFNIPLLTLVDVFGFLPGVNDQEHGGIIRHGAKMLYAYSE	
D.propionicus_12S	(391)	401	450
C.kroppenstedtii_12S	(396)		
P.feudenreichii_12S	(399)		
G.bemidjiensis_12S	(389)		
C.bescii_12S	(389)		
C.cellulolyticum_12S	(391)		
C.thermocellum_12S	(391)		
T.saccharolyticum_12S	(391)		
Consensus	(401)	ATVPKIDVILRKAYGGAYLAMCSKDLGAD VFAWPSASIAVMGPDGAANI	
D.propionicus_12S	(441)	451	500
C.kroppenstedtii_12S	(446)		
P.feudenreichii_12S	(449)		
G.bemidjiensis_12S	(439)		
C.bescii_12S	(439)		
C.cellulolyticum_12S	(441)		
C.thermocellum_12S	(441)		
T.saccharolyticum_12S	(441)		
Consensus	(451)	IFRKEIKAADDP R EKI EYRE PSNPFVAAARGYVDVVIE	
D.propionicus_12S	(485)	501	532
C.kroppenstedtii_12S	(496)		
P.feudenreichii_12S	(493)		
G.bemidjiensis_12S	(483)		
C.bescii_12S	(483)		
C.cellulolyticum_12S	(485)		
C.thermocellum_12S	(485)		
T.saccharolyticum_12S	(485)		
Consensus	(501)	FADTR KIISALEMLATKRENRPSSKKHCNIFL	

5S subunits

Cthermocellum_5S	(1)	1	50
Ccellulolyticum_5S	(1)		
Gbemidjiensis_5S	(1)		
Ckroppenstedtii_5S	(1)		
Pfreudenreichii_5S	(1)		
Dpropionicus_5S_1.3S	(1)		
Cbescii_5S	(1)		
Tsaccharolyticum_5S	(1)		
Consensus	(1)	VKITETVLROAHQSIIATRM TEDMLFI	
Cthermocellum_5S	(32)	51	100
Ccellulolyticum_5S	(32)		
Gbemidjiensis_5S	(33)		
Ckroppenstedtii_5S	(46)		
Pfreudenreichii_5S	(42)		
Dpropionicus_5S_1.3S	(51)		
Cbescii_5S	(31)		
Tsaccharolyticum_5S	(32)		
Consensus	(51)	E LD IGYWSVECWGGATFDACIRFLNEDPWERLRLTLKLMKKTPLQML	

Figure 31 cont.

		101	150
Cthermocellum_5S	(82)	...Y...C...A...E...E...T...	
Ccellulolyticum_5S	(82)	...Y...C...A...E...E...T...	
Gbemidjiensis_5S	(83)	...Y...C...A...E...E...T...	
Ckroppenstedtii_5S	(96)	...Y...C...A...E...E...T...	
Pfreudenreichii_5S	(92)	...Y...C...A...E...E...T...	
Dpropionicus_5S_1.3S	(101)	...Y...C...A...E...E...T...	
Cbescii_5S	(81)	...Y...C...A...E...E...T...	
Tsaccharolyticum_5S	(82)	...Y...C...A...E...E...T...	
Consensus	(101)	LRGQNLGKYRYADDVVD EV KSAENGIDIFRIFDALND RNIE AKA	
		151	200
Cthermocellum_5S	(132)	...K...L...E...S...C...E...E...	
Ccellulolyticum_5S	(132)	...K...L...E...S...C...E...E...	
Gbemidjiensis_5S	(133)	...K...L...E...S...C...E...E...	
Ckroppenstedtii_5S	(146)	...K...L...E...S...C...E...E...	
Pfreudenreichii_5S	(142)	...K...L...E...S...C...E...E...	
Dpropionicus_5S_1.3S	(151)	...K...L...E...S...C...E...E...	
Cbescii_5S	(131)	...K...L...E...S...C...E...E...	
Tsaccharolyticum_5S	(132)	...K...L...E...S...C...E...E...	
Consensus	(151)	VKK GRHAQG ICYTISF VH LE YVKLAK LLDMGADSIKID	
		201	250
Cthermocellum_5S	(176)	...M...G...L...P...Y...A...Y...D...L...I...K...A...I...K...E...V...	
Ccellulolyticum_5S	(176)	...M...G...L...P...Y...A...Y...D...L...I...K...A...I...K...E...V...	
Gbemidjiensis_5S	(177)	...M...G...L...P...Y...A...Y...D...L...I...K...A...I...K...E...V...	
Ckroppenstedtii_5S	(190)	...M...G...L...P...Y...A...Y...D...L...I...K...A...I...K...E...V...	
Pfreudenreichii_5S	(186)	...M...G...L...P...Y...A...Y...D...L...I...K...A...I...K...E...V...	
Dpropionicus_5S_1.3S	(201)	...M...G...L...P...Y...A...Y...D...L...I...K...A...I...K...E...V...	
Cbescii_5S	(175)	...M...G...L...P...Y...A...Y...D...L...I...K...A...I...K...E...V...	
Tsaccharolyticum_5S	(176)	...M...G...L...P...Y...A...Y...D...L...I...K...A...I...K...E...V...	
Consensus	(201)	MAGLL PY AYDLIKAIKE V KTEIQLHHTYTSGVASMTYLKAIKAEV	
		251	300
Cthermocellum_5S	(224)	...D...V...D...T...A...I...S...S...M...S...L...G...T...S...Q...P...P...T...E...S...L...V...	
Ccellulolyticum_5S	(224)	...D...V...D...T...A...I...S...S...M...S...L...G...T...S...Q...P...P...T...E...S...L...V...	
Gbemidjiensis_5S	(227)	...D...V...D...T...A...I...S...S...M...S...L...G...T...S...Q...P...P...T...E...S...L...V...	
Ckroppenstedtii_5S	(240)	...D...V...D...T...A...I...S...S...M...S...L...G...T...S...Q...P...P...T...E...S...L...V...	
Pfreudenreichii_5S	(236)	...D...V...D...T...A...I...S...S...M...S...L...G...T...S...Q...P...P...T...E...S...L...V...	
Dpropionicus_5S_1.3S	(249)	...D...V...D...T...A...I...S...S...M...S...L...G...T...S...Q...P...P...T...E...S...L...V...	
Cbescii_5S	(223)	...D...V...D...T...A...I...S...S...M...S...L...G...T...S...Q...P...P...T...E...S...L...V...	
Tsaccharolyticum_5S	(224)	...D...V...D...T...A...I...S...S...M...S...L...G...T...S...Q...P...P...T...E...S...L...V...	
Consensus	(251)	DVVDTAISSMSLGTSPPTESLV LGTDYDTGLDLKLAETADHFK L	
		301	350
Cthermocellum_5S	(274)	...K...E...K...I...E...S...S...M...S...L...G...T...S...Q...P...P...T...E...S...L...V...	
Ccellulolyticum_5S	(274)	...K...E...K...I...E...S...S...M...S...L...G...T...S...Q...P...P...T...E...S...L...V...	
Gbemidjiensis_5S	(277)	...K...E...K...I...E...S...S...M...S...L...G...T...S...Q...P...P...T...E...S...L...V...	
Ckroppenstedtii_5S	(290)	...K...E...K...I...E...S...S...M...S...L...G...T...S...Q...P...P...T...E...S...L...V...	
Pfreudenreichii_5S	(286)	...K...E...K...I...E...S...S...M...S...L...G...T...S...Q...P...P...T...E...S...L...V...	
Dpropionicus_5S_1.3S	(299)	...K...E...K...I...E...S...S...M...S...L...G...T...S...Q...P...P...T...E...S...L...V...	
Cbescii_5S	(273)	...K...E...K...I...E...S...S...M...S...L...G...T...S...Q...P...P...T...E...S...L...V...	
Tsaccharolyticum_5S	(274)	...K...E...K...I...E...S...S...M...S...L...G...T...S...Q...P...P...T...E...S...L...V...	
Consensus	(301)	K Y KYGLLDISKVT VDVNLL SQIPGCMLSNLVSQLKQGGALDKIDEV	
		351	400
Cthermocellum_5S	(324)	...L...E...K...E...V...P...R...V...R...K...D...G...Y...P...P...L...V...T...P...T...S...Q...I...V...G...T...Q...A...V...L...N...V...L...	
Ccellulolyticum_5S	(324)	...L...E...K...E...V...P...R...V...R...K...D...G...Y...P...P...L...V...T...P...T...S...Q...I...V...G...T...Q...A...V...L...N...V...L...	
Gbemidjiensis_5S	(324)	...L...E...K...E...V...P...R...V...R...K...D...G...Y...P...P...L...V...T...P...T...S...Q...I...V...G...T...Q...A...V...L...N...V...L...	
Ckroppenstedtii_5S	(336)	...L...E...K...E...V...P...R...V...R...K...D...G...Y...P...P...L...V...T...P...T...S...Q...I...V...G...T...Q...A...V...L...N...V...L...	
Pfreudenreichii_5S	(332)	...L...E...K...E...V...P...R...V...R...K...D...G...Y...P...P...L...V...T...P...T...S...Q...I...V...G...T...Q...A...V...L...N...V...L...	
Dpropionicus_5S_1.3S	(349)	...L...E...K...E...V...P...R...V...R...K...D...G...Y...P...P...L...V...T...P...T...S...Q...I...V...G...T...Q...A...V...L...N...V...L...	
Cbescii_5S	(323)	...L...E...K...E...V...P...R...V...R...K...D...G...Y...P...P...L...V...T...P...T...S...Q...I...V...G...T...Q...A...V...L...N...V...L...	
Tsaccharolyticum_5S	(323)	...L...E...K...E...V...P...R...V...R...K...D...G...Y...P...P...L...V...T...P...T...S...Q...I...V...G...T...Q...A...V...L...N...V...L...	
Consensus	(351)	LKEVPRVRKD GYPPLVTPTSQIVGTQAVLNVL GER YKMTKE KD	
		401	450
Cthermocellum_5S	(371)	...L...M...K...G...Y...G...K...T...P...A...P...I...N...P...E...V...K...K...I...L...K...D...E...E...P...I...T...C...R...P...A...D...L...L...E...P...E...L...E...K...I...K...	
Ccellulolyticum_5S	(371)	...L...M...K...G...Y...G...K...T...P...A...P...I...N...P...E...V...K...K...I...L...K...D...E...E...P...I...T...C...R...P...A...D...L...L...E...P...E...L...E...K...I...K...	
Gbemidjiensis_5S	(370)	...L...M...K...G...Y...G...K...T...P...A...P...I...N...P...E...V...K...K...I...L...K...D...E...E...P...I...T...C...R...P...A...D...L...L...E...P...E...L...E...K...I...K...	
Ckroppenstedtii_5S	(382)	...L...M...K...G...Y...G...K...T...P...A...P...I...N...P...E...V...K...K...I...L...K...D...E...E...P...I...T...C...R...P...A...D...L...L...E...P...E...L...E...K...I...K...	
Pfreudenreichii_5S	(378)	...L...M...K...G...Y...G...K...T...P...A...P...I...N...P...E...V...K...K...I...L...K...D...E...E...P...I...T...C...R...P...A...D...L...L...E...P...E...L...E...K...I...K...	
Dpropionicus_5S_1.3S	(399)	...L...M...K...G...Y...G...K...T...P...A...P...I...N...P...E...V...K...K...I...L...K...D...E...E...P...I...T...C...R...P...A...D...L...L...E...P...E...L...E...K...I...K...	
Cbescii_5S	(370)	...L...M...K...G...Y...G...K...T...P...A...P...I...N...P...E...V...K...K...I...L...K...D...E...E...P...I...T...C...R...P...A...D...L...L...E...P...E...L...E...K...I...K...	
Tsaccharolyticum_5S	(370)	...L...M...K...G...Y...G...K...T...P...A...P...I...N...P...E...V...K...K...I...L...K...D...E...E...P...I...T...C...R...P...A...D...L...L...E...P...E...L...E...K...I...K...	
Consensus	(401)	LMKG YGKTPAPINPEV KKILK DEEPITCRPADLLEPELEKIK	

	Cthermocellum_5S	(415)	--N-----E-----QYK-----AQ	500
	Ccellulolyticum_5S	(415)	-EA-----P-----PKSI-----D	
	Gbemidjiensis_5S	(418)	ALFETGCDGGELT-----TFATSS-----SG	
	Ckroppenstedtii_5S	(430)	ALFETGTGGELT-----NPFCKTP-----G	
	Pfreudenreichii_5S	(426)	AAAELVNGELT-----EFVEHFA-----G	
D	propionicus_5S_1.3S	(447)	-PFGGLAKHIELLAAITGVTELEWYGITPAPEVKPLTLEDVK	
	Chesicii_5S	(414)	-EK-----T-----LKLP-----AK	
T	saccharolyticum_5S	(414)	-N-----E-----EYEQ-----A	
	Consensus	(451)	EIKEYLEQDEDLVTYALFPQVA KFF R	B
		501		550
	Cthermocellum_5S	(448)	---EPDMHYENRVHPV-----	
	Ccellulolyticum_5S	(448)	---ETAPASDEIPEVVAAISAVVNEMGRDCQYRIGNISKLNQNKNW	
	Gbemidjiensis_5S	(452)	FEDMGRDFEGAGETSIPESHGKITCFVTITLTLSQPHKTAPYGD	
	Ckroppenstedtii_5S	(464)	PKNGCKTKELLESAAKASGDATAIREPIMKITIGARSHTSSSPA	
	Pfreudenreichii_5S	(460)	PSGALTDDNLK-ABAEGDEKSLAVACPVTNINVCIVRETSQA---	
D	propionicus_5S_1.3S	(496)	KBELVAKKAKGLIEPKRAAPEKTANVRTNNFVDSEYFNELDPTGDF	
	Chesicii_5S	(447)	---DADLGSMVYFV-----	
T	saccharolyticum_5S	(447)	---DSTLLIEERVHI-----	
	Consensus	(501)	KYKV VQ K Y V G V V	
		551		600
	Cthermocellum_5S	(466)	-----	
	Ccellulolyticum_5S	(498)	SLYGMLDRFR TKI-----	
	Gbemidjiensis_5S	(502)	-----	
	Ckroppenstedtii_5S	(511)	-----	
	Pfreudenreichii_5S	(506)	-----	
D	propionicus_5S_1.3S	(546)	QMNVAAAPRPAAPAAAEKAAAPAAAAAPAAAPKAAAPAAAAAPAAVEGGT	
	Chesicii_5S	(464)	-----	
T	saccharolyticum_5S	(465)	-----	
	Consensus	(551)		601
		601		650
	Cthermocellum_5S	(466)	-----	
	Ccellulolyticum_5S	(511)	-----	
	Gbemidjiensis_5S	(502)	-----	
	Ckroppenstedtii_5S	(511)	-----	
	Pfreudenreichii_5S	(506)	-----	
D	propionicus_5S_1.3S	(596)	ELLAPMGIMVKNLVNGBDAVKAGDPILVLEAMKMENNIGSPCDGTVKAL	
	Chesicii_5S	(464)	-----	
T	saccharolyticum_5S	(465)	-----	
	Consensus	(601)		
		651	668	
	Cthermocellum_5S	(466)	-----	
	Ccellulolyticum_5S	(511)	-----	
	Gbemidjiensis_5S	(502)	-----	
	Ckroppenstedtii_5S	(511)	-----	
	Pfreudenreichii_5S	(506)	-----	
D	propionicus_5S_1.3S	(646)	NFGSGDSVAKDTVLAIIG	
	Chesicii_5S	(464)	-----	
T	saccharolyticum_5S	(465)	-----	
	Consensus	(651)		

1.3S subunits

12S C-term subunitSUBSTITUTE SHEET (RULE 26)

Figure 32

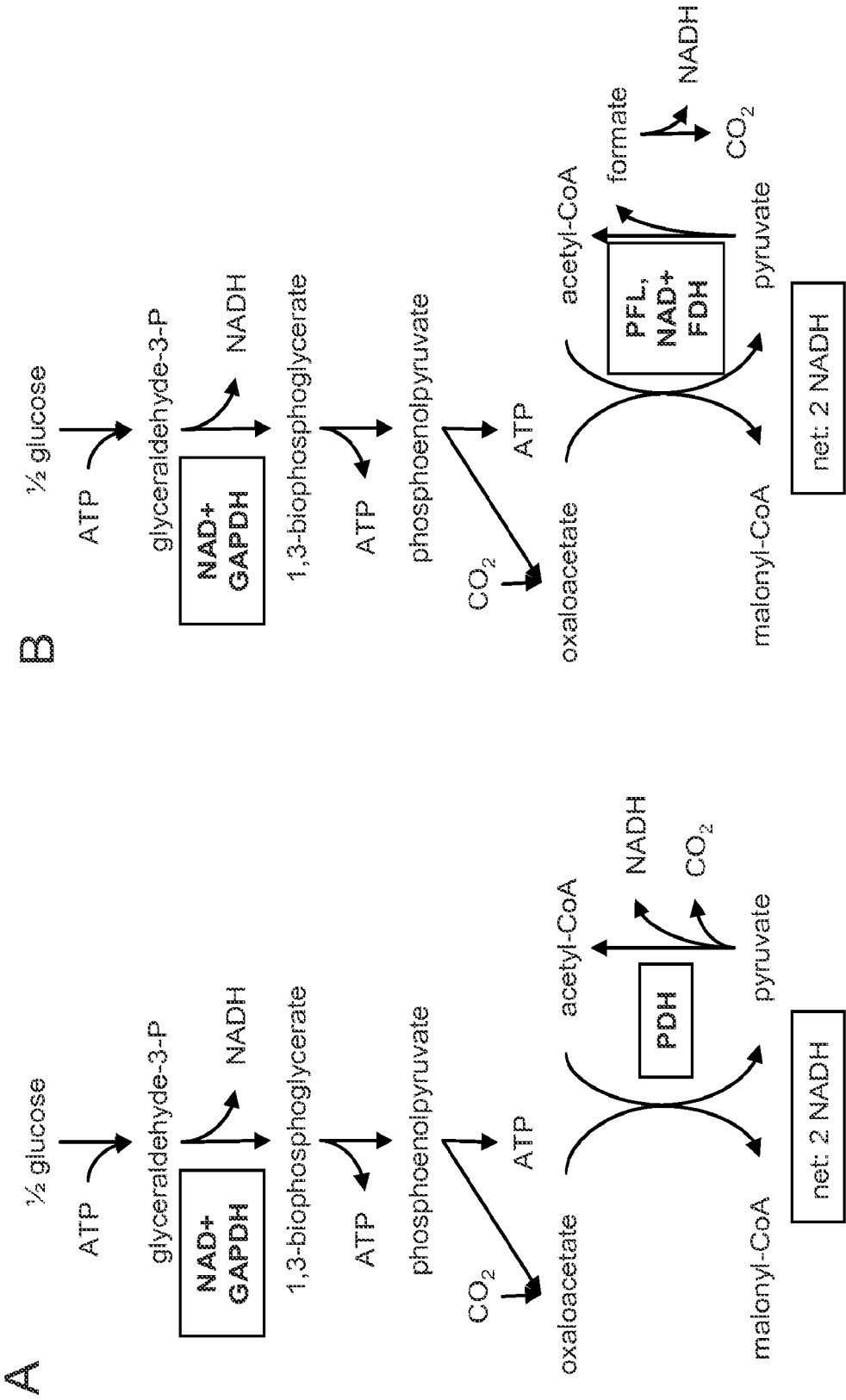


Figure 32 cont.

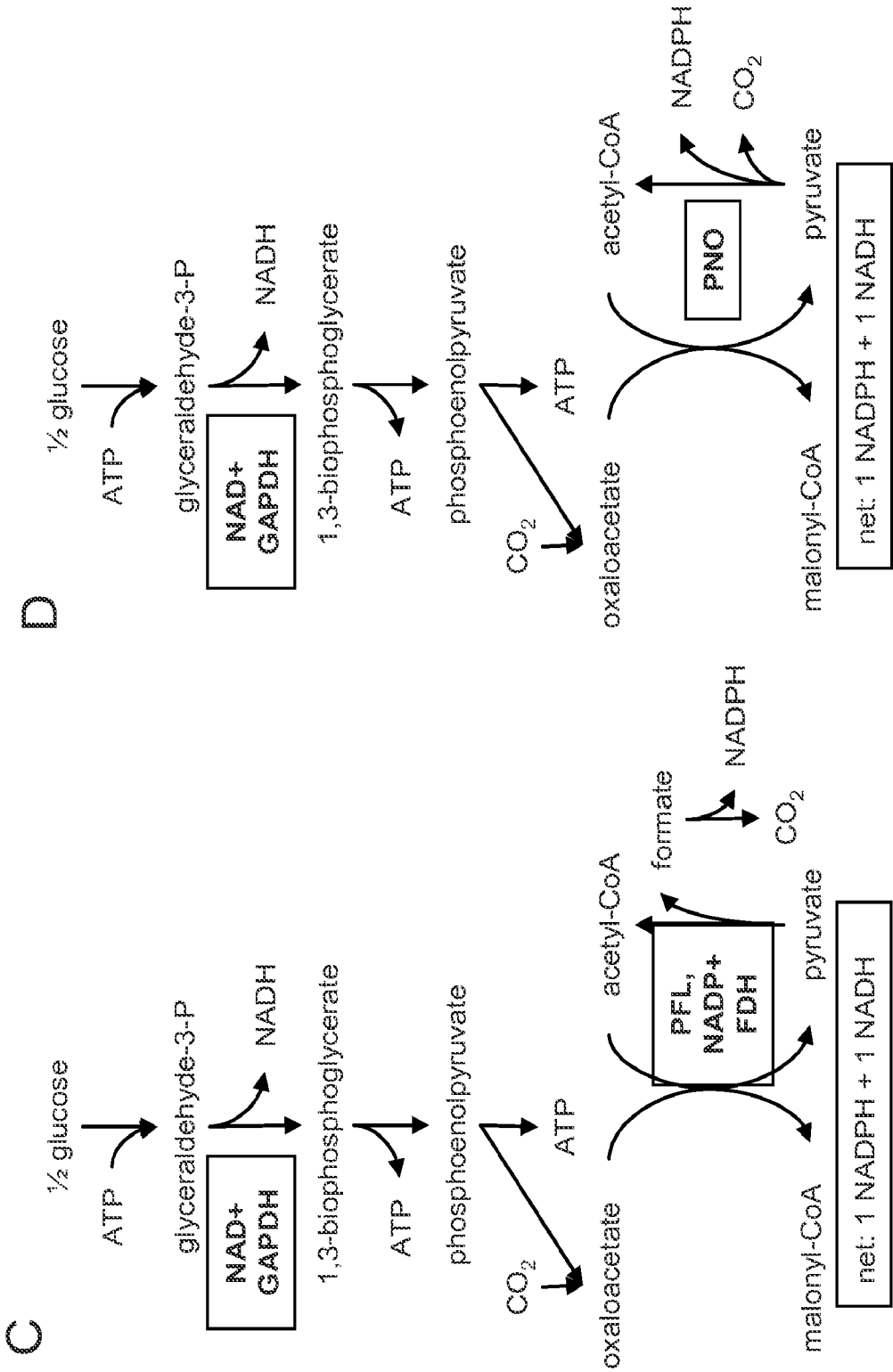
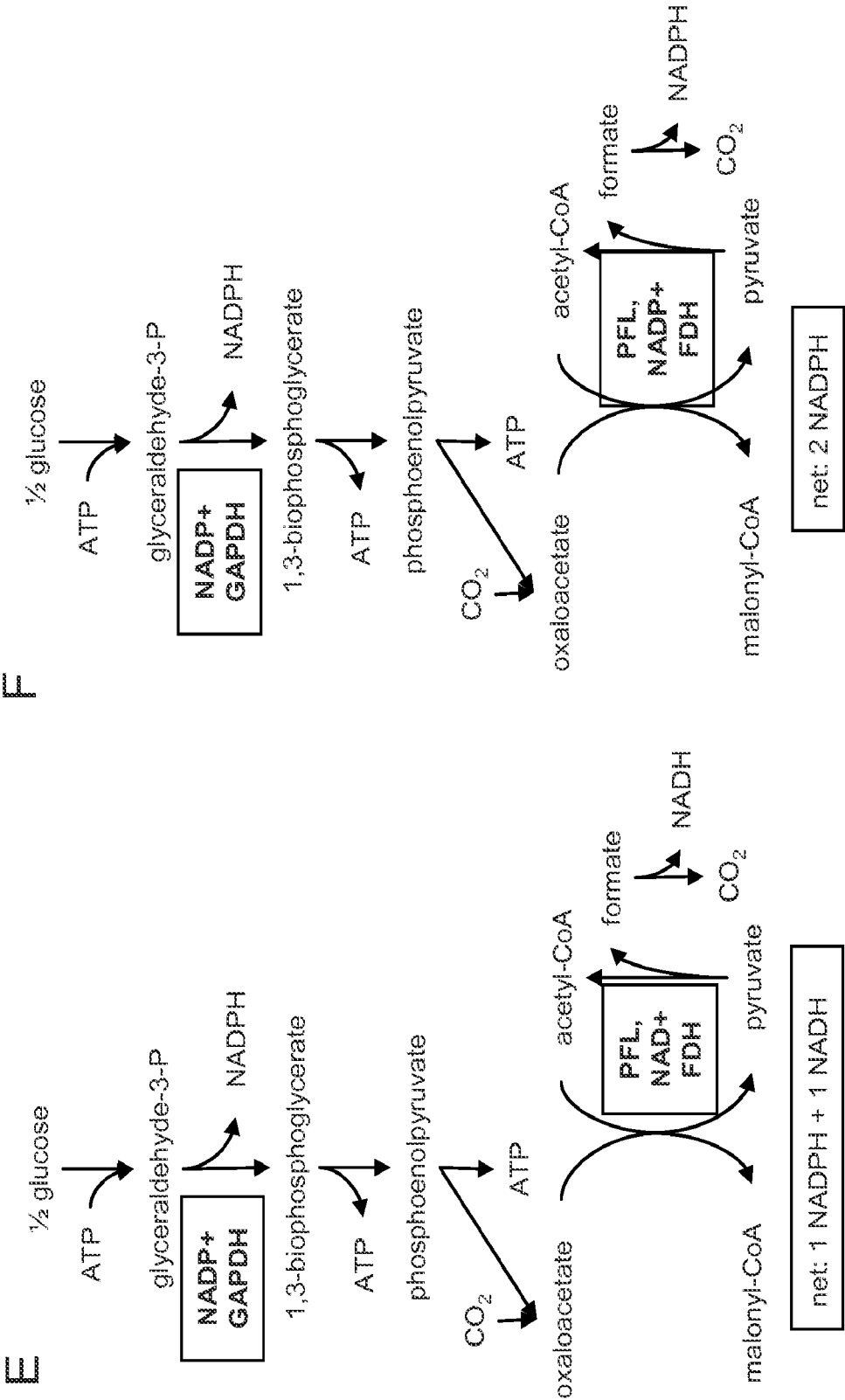
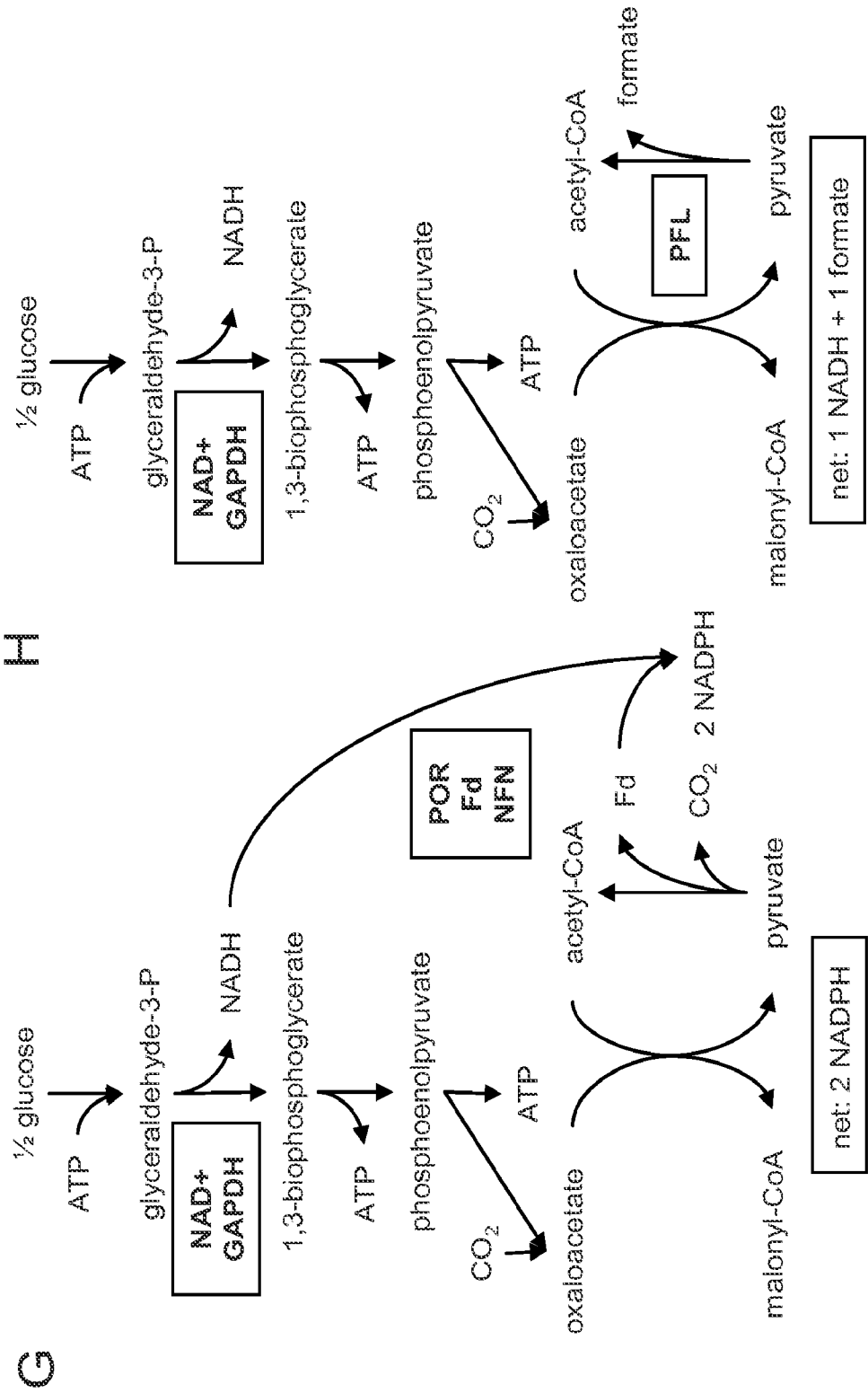


Figure 32 cont.



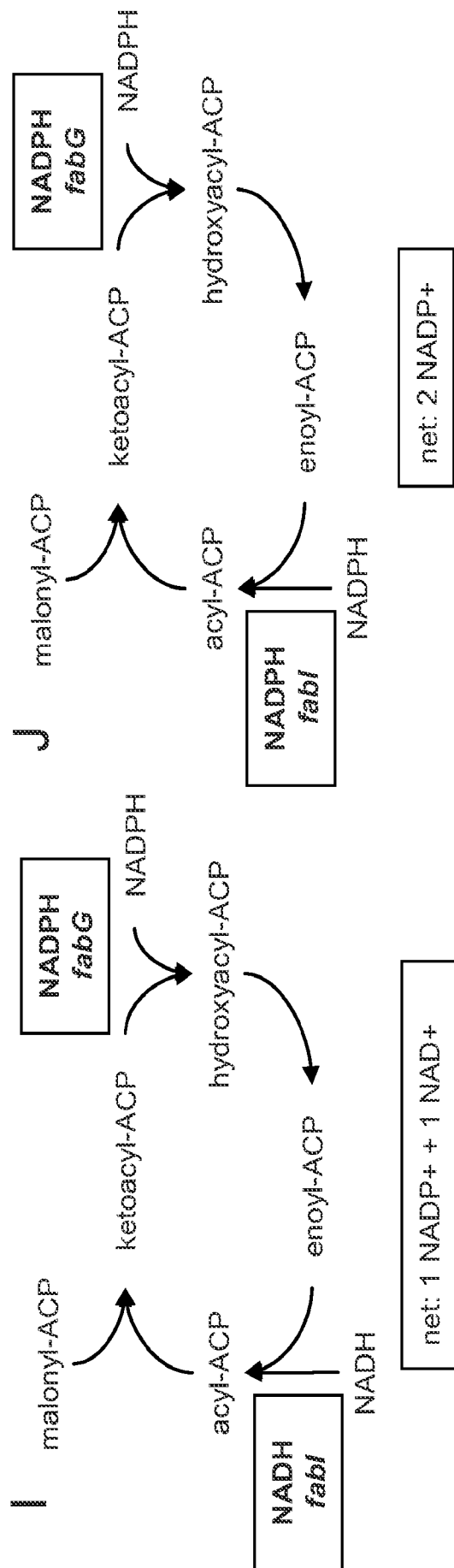
51/72

Figure 32 cont.



52/72

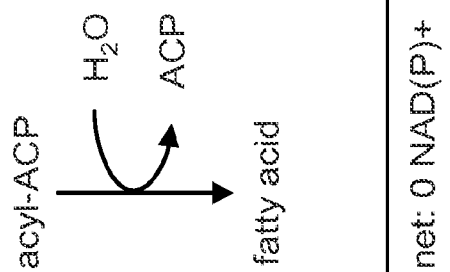
Figure 32 cont.



53/72

Figure 32 cont.

K



L

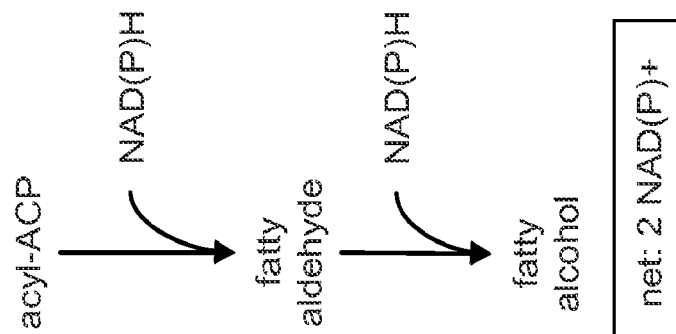


Figure 33

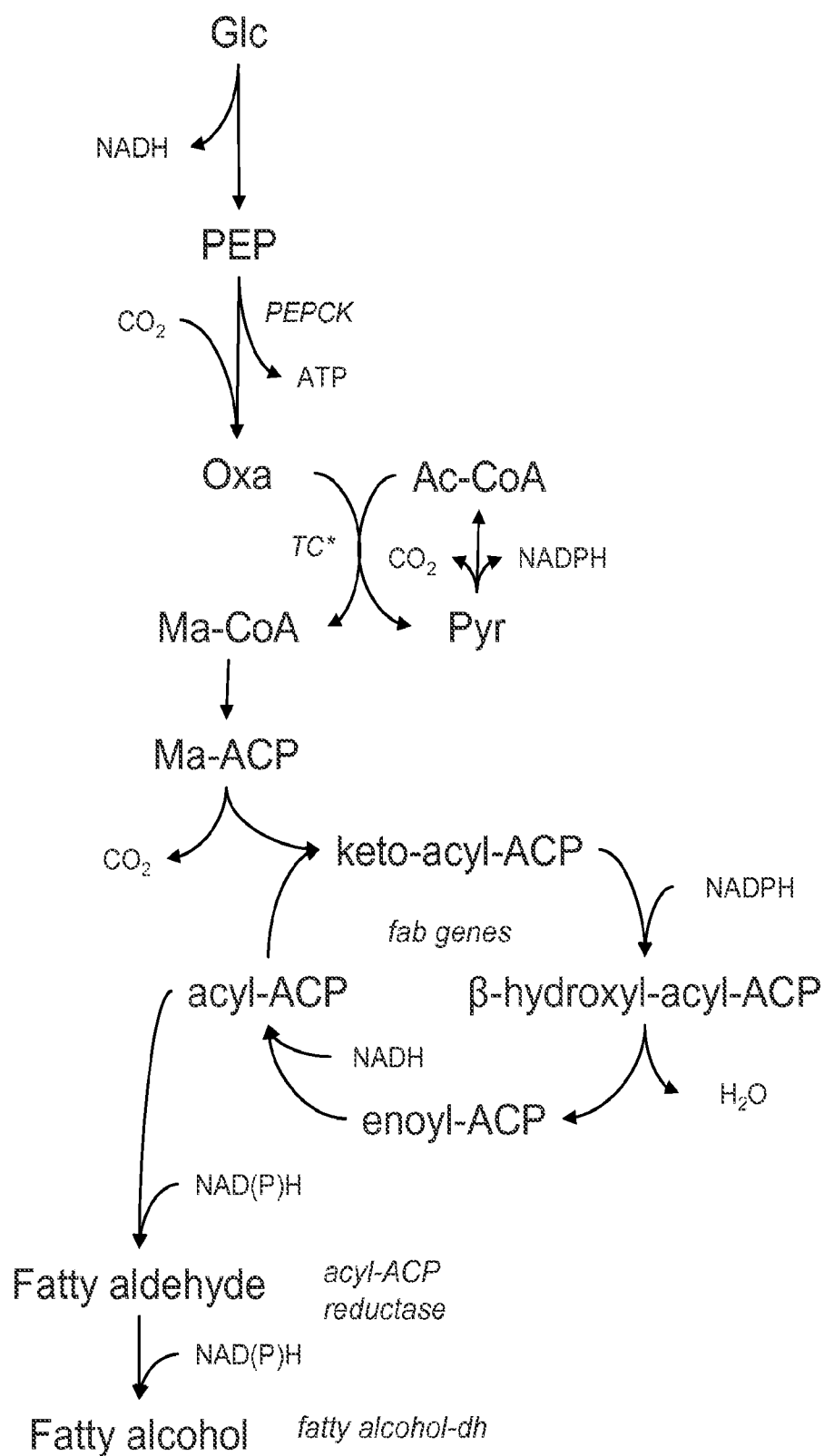
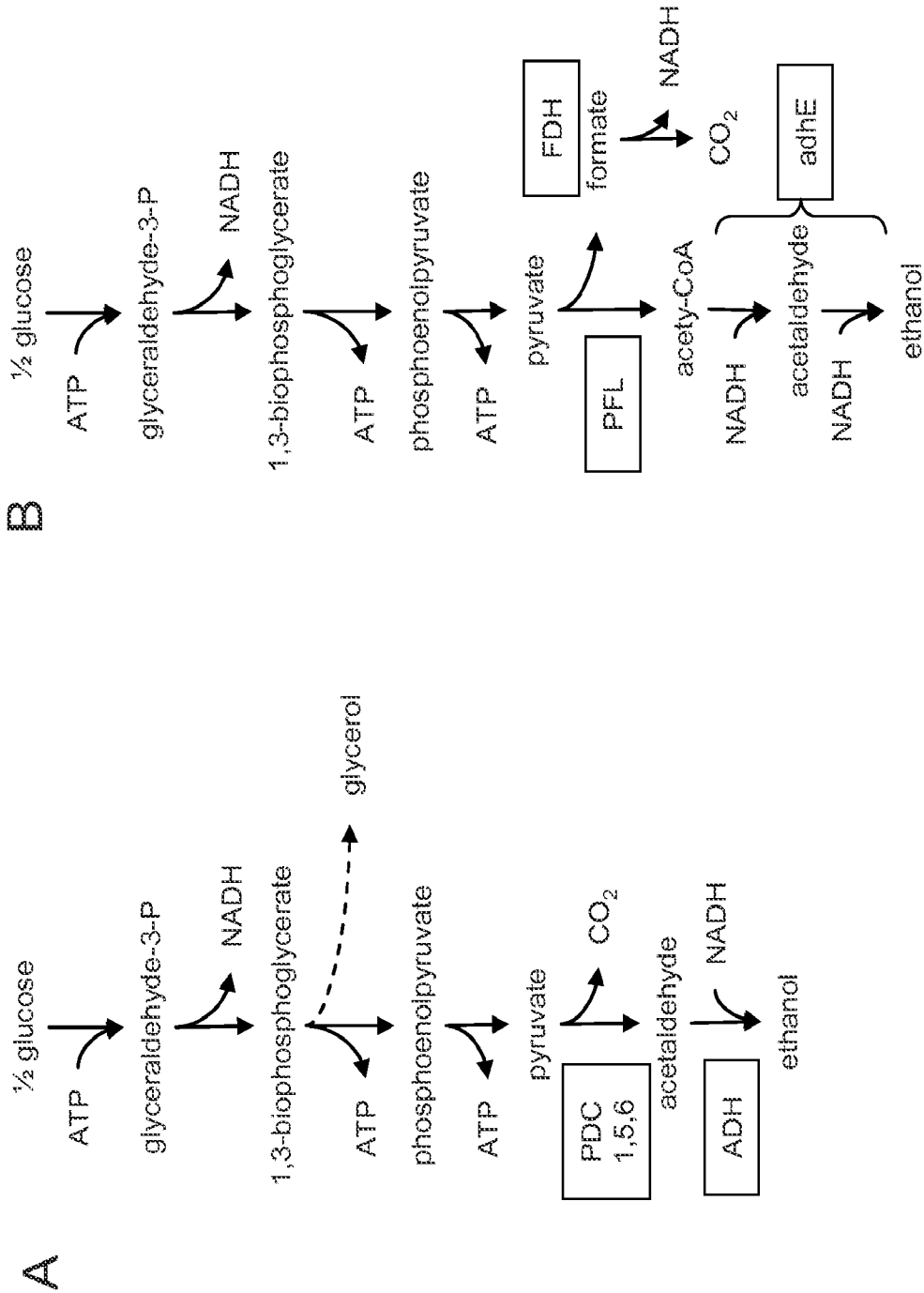


Figure 34



56/72

Figure 34 cont.

C

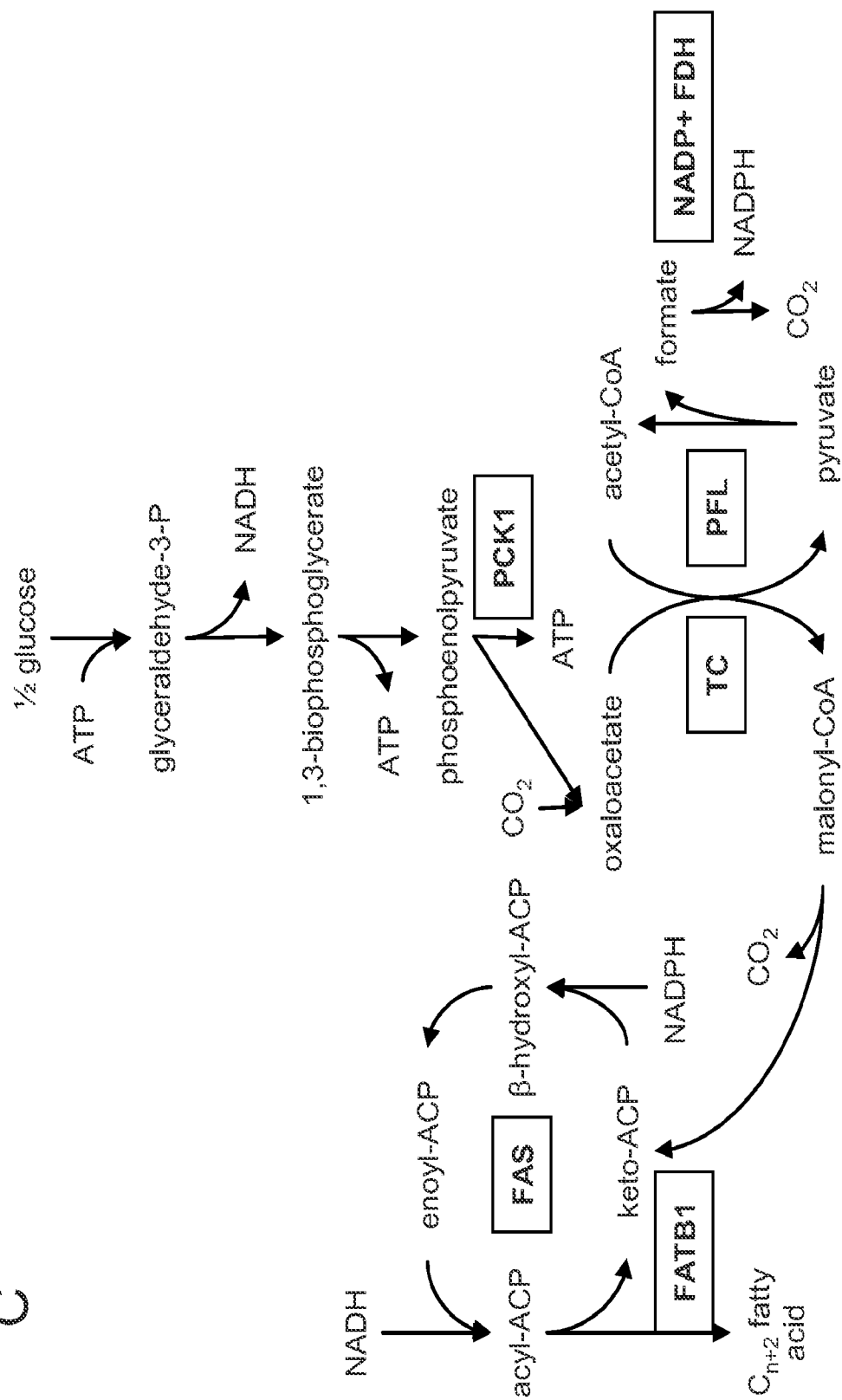


Figure 35

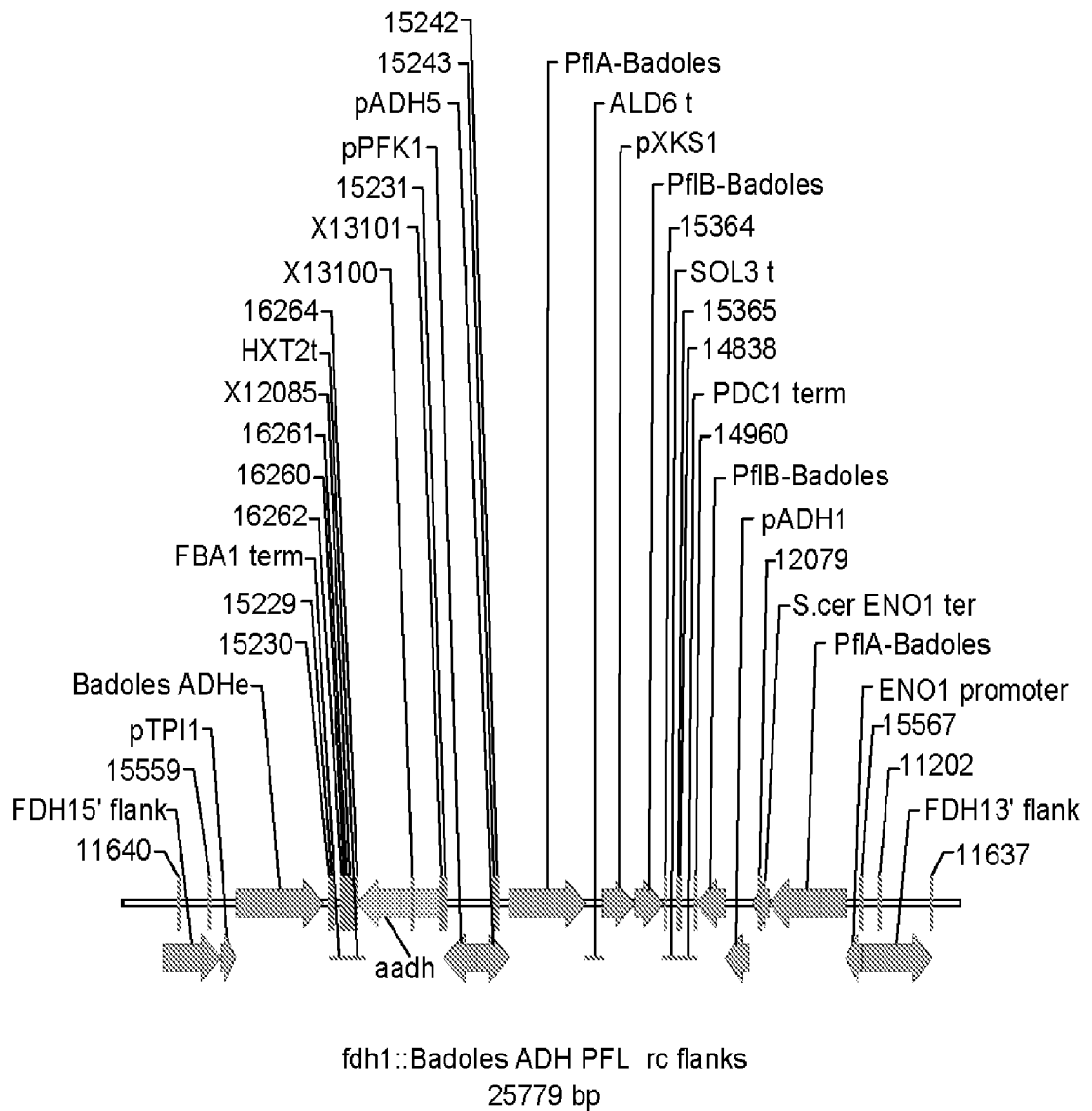


Figure 36

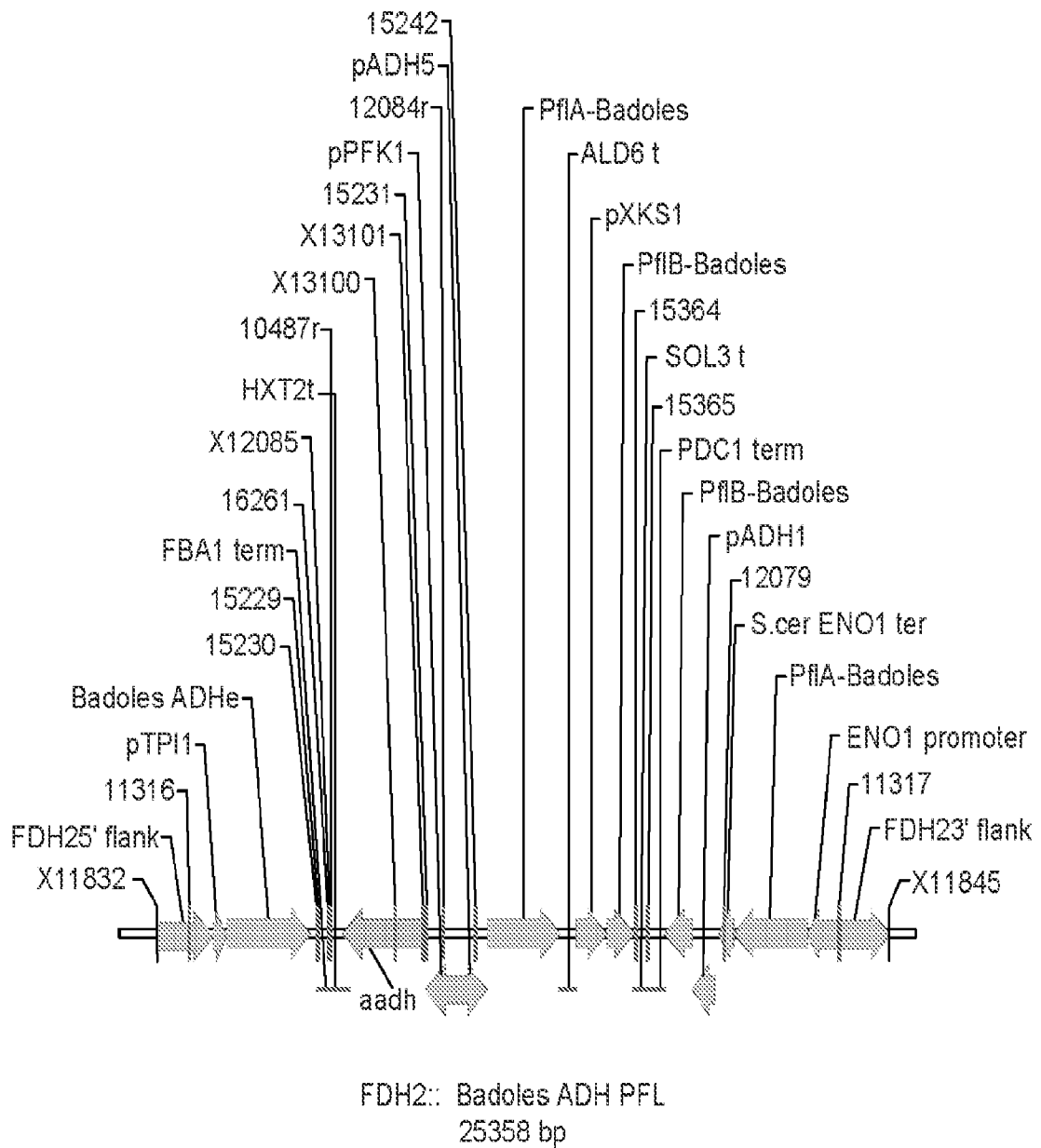


Figure 37

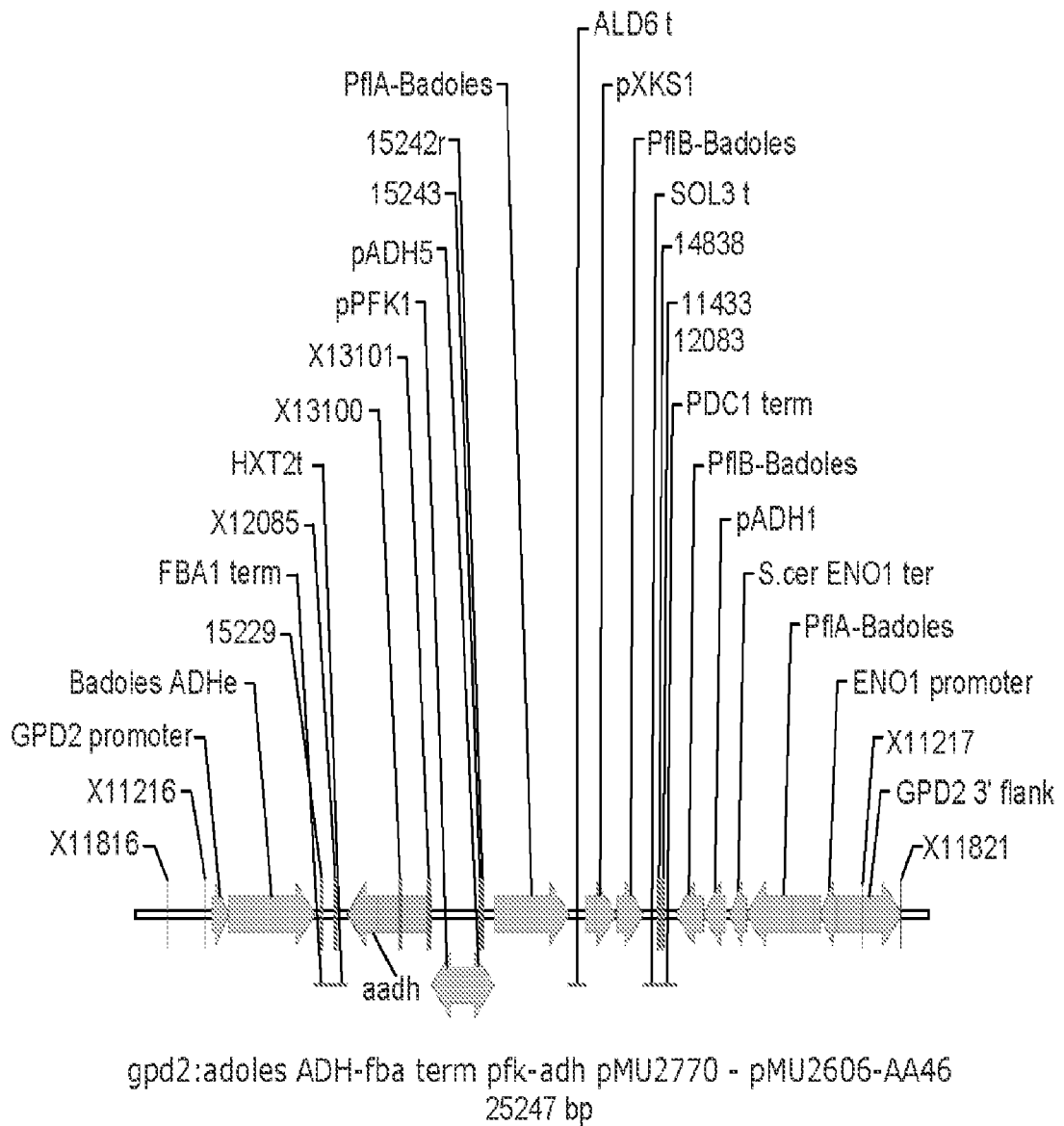


Figure 38

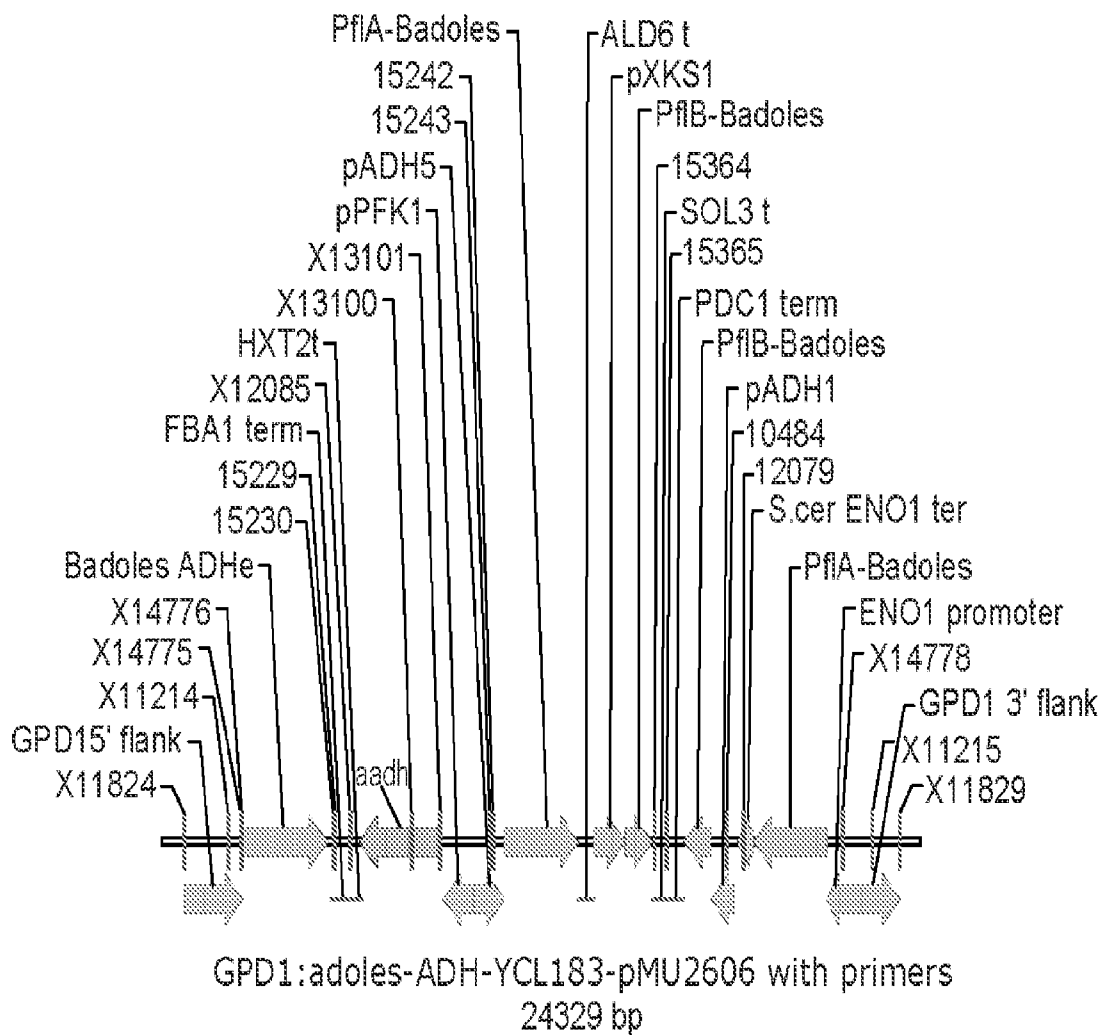
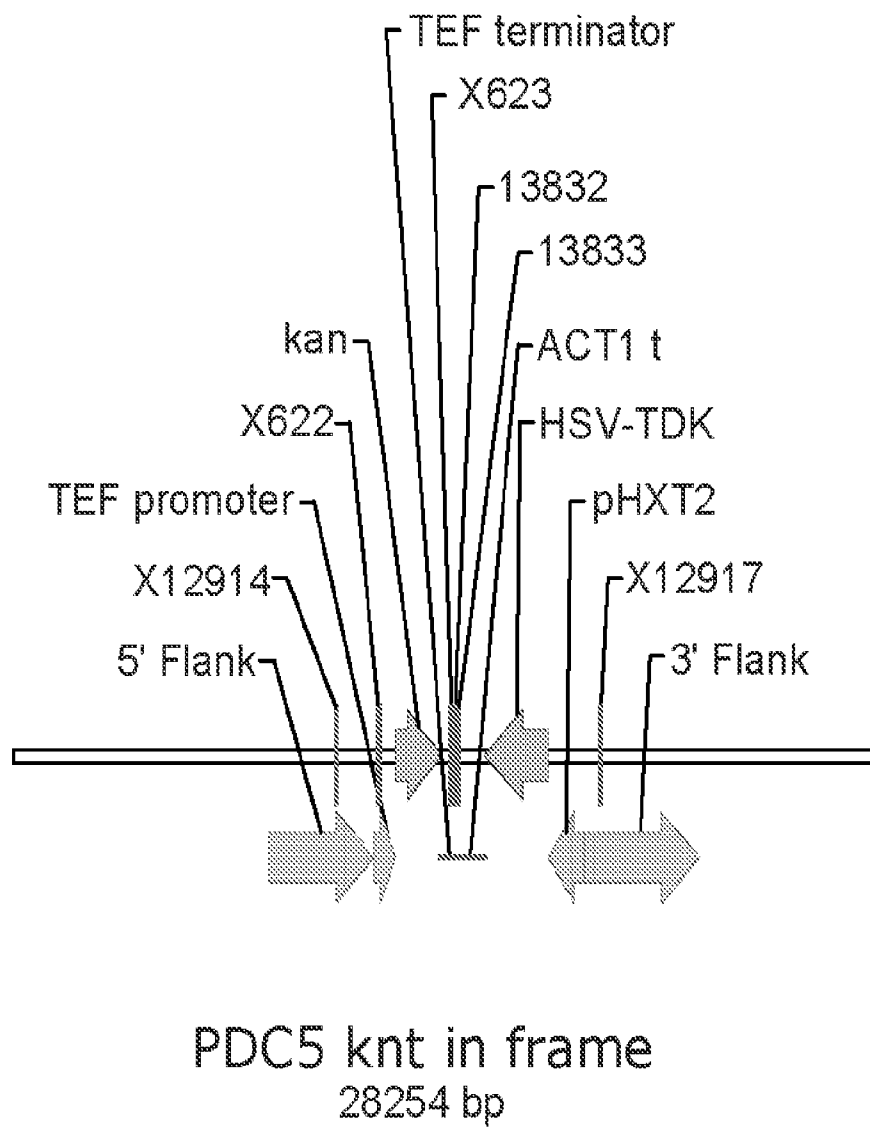


Figure 39



62/72

Figure 40

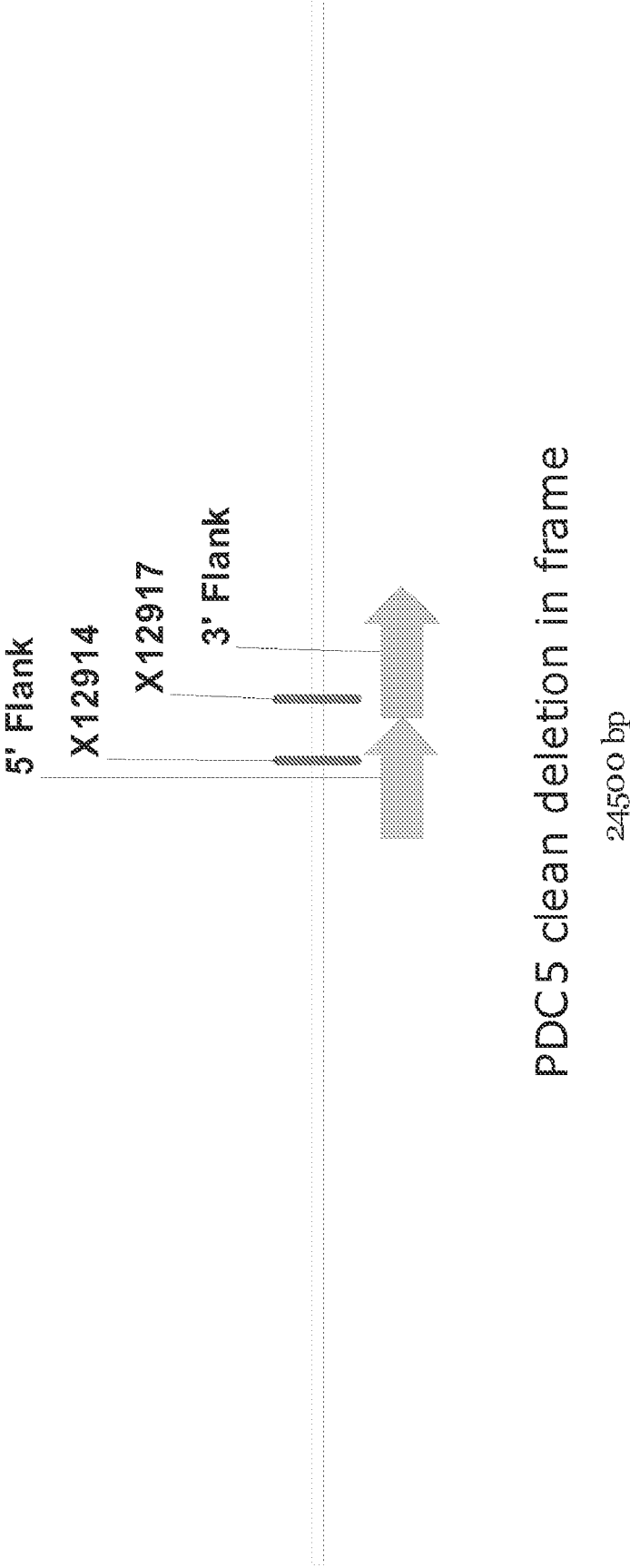
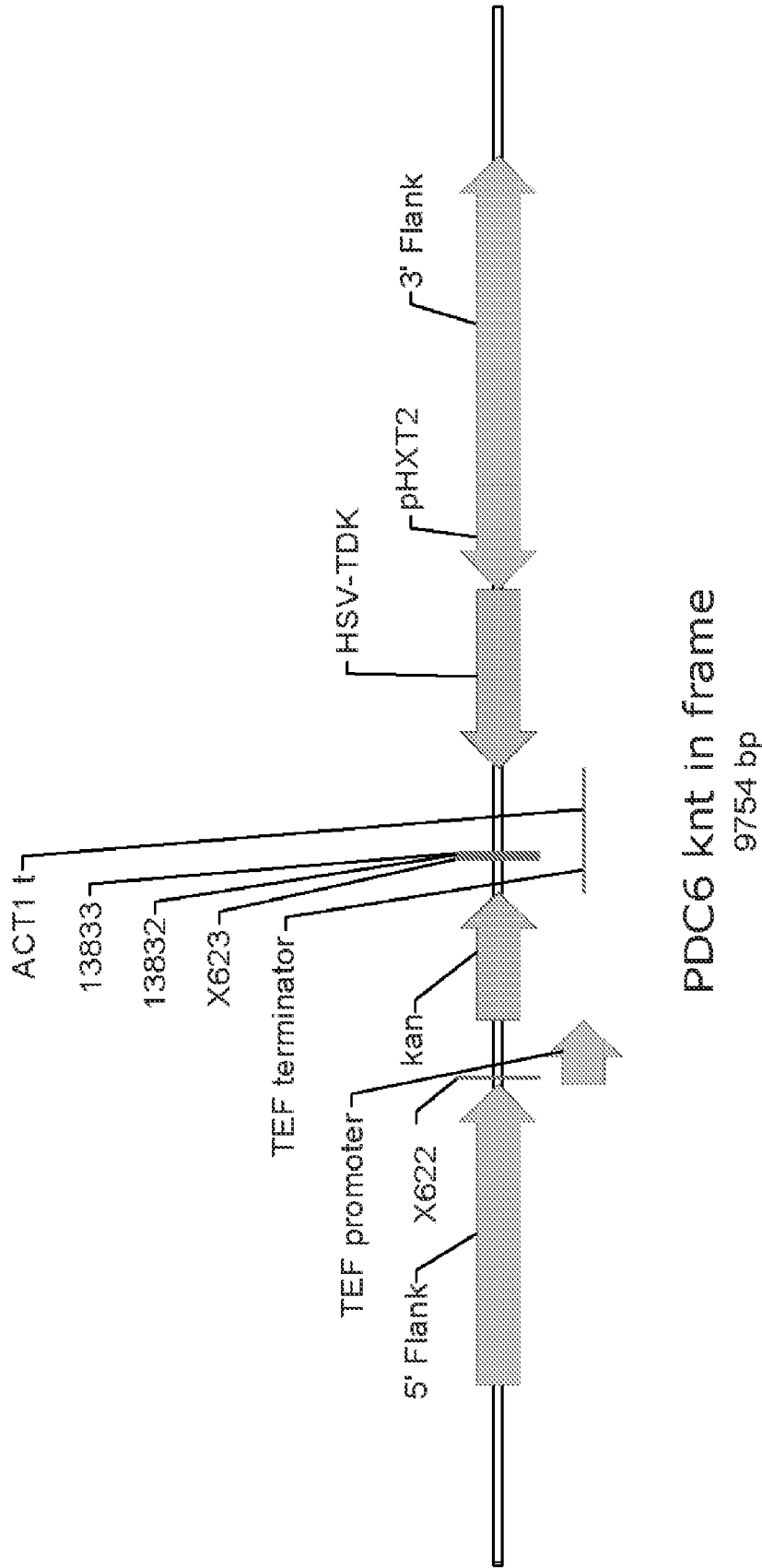
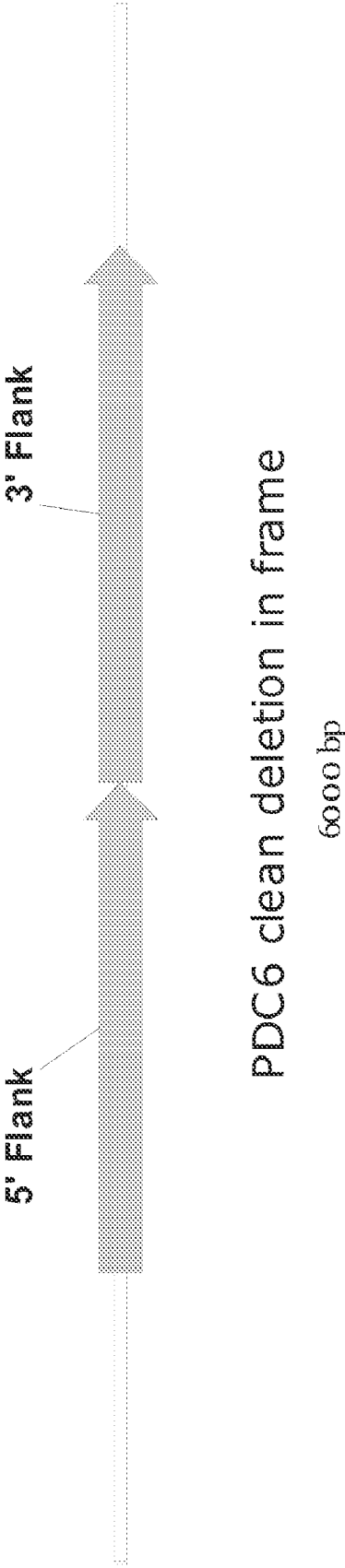


Figure 41



64/72

Figure 42



65/72

Figure 43

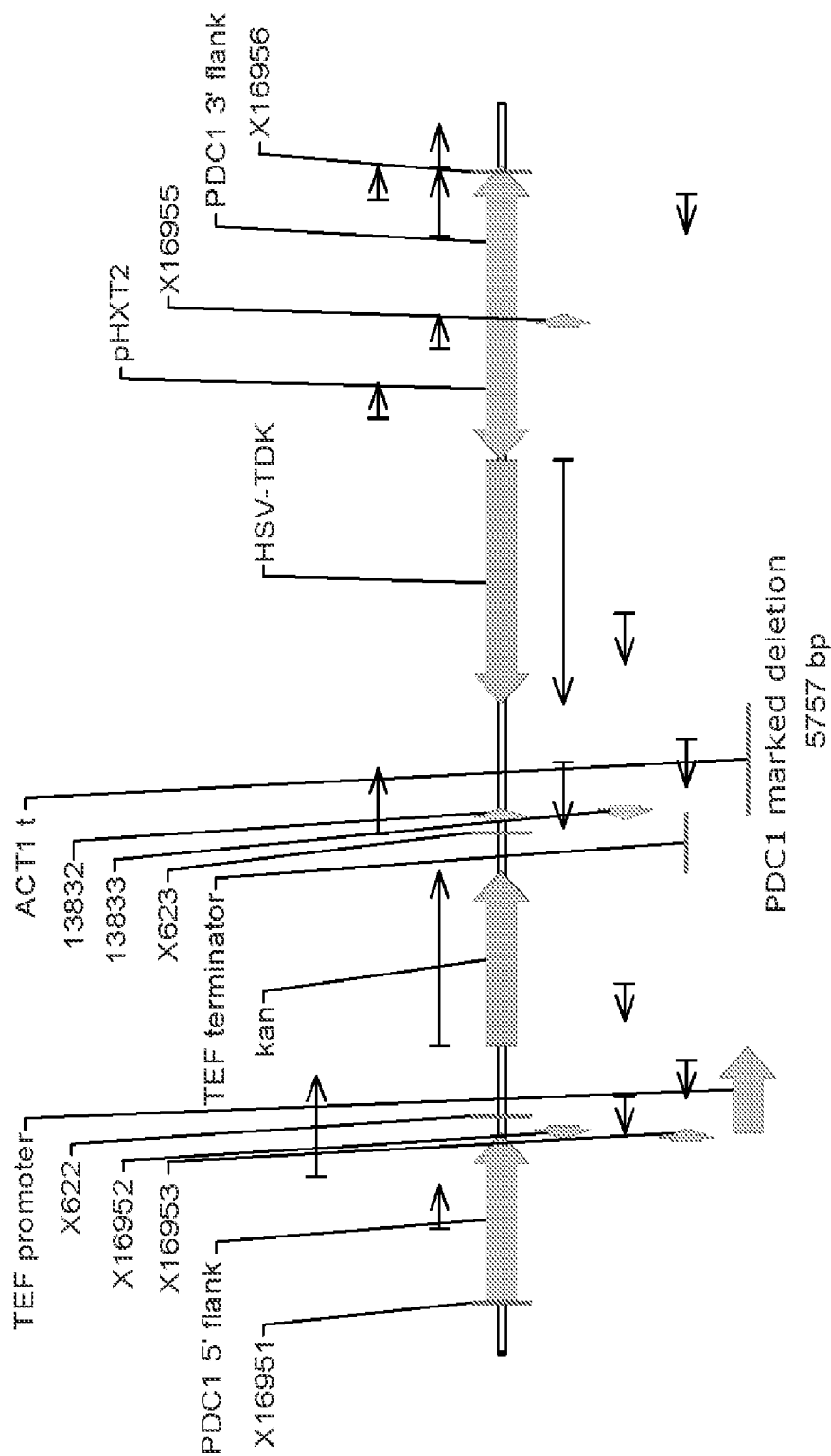
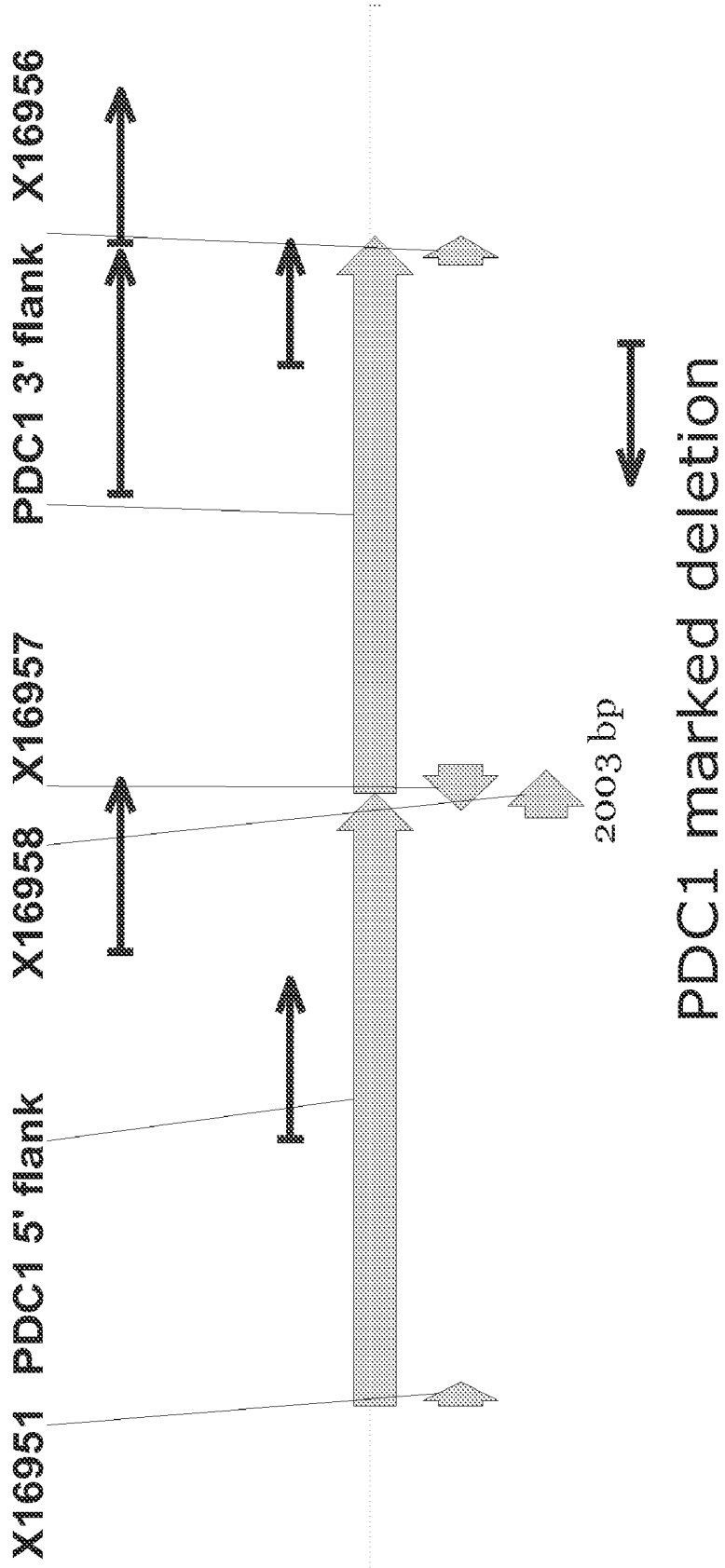


Figure 44



67/72

Figure 45

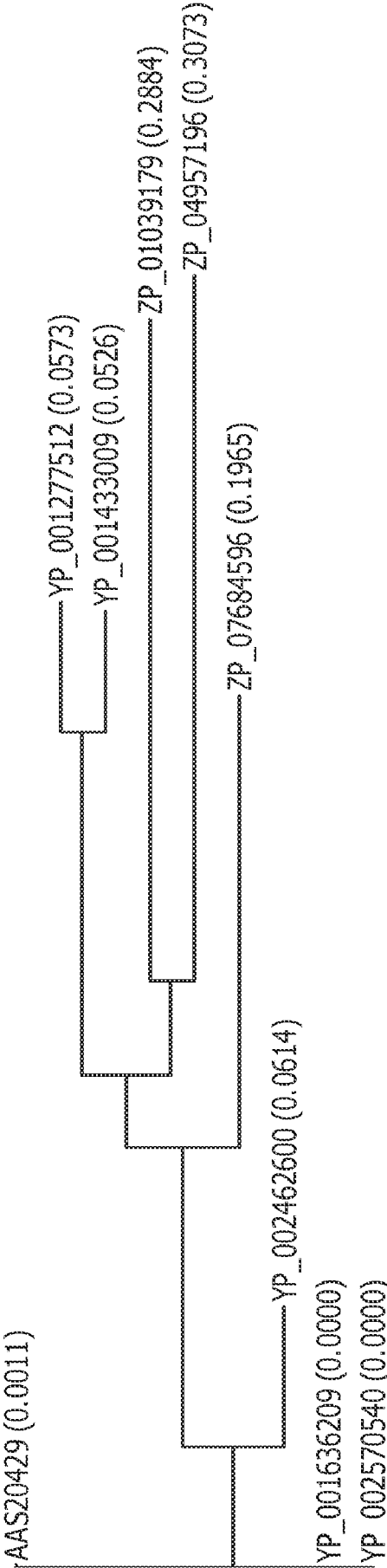


Figure 46

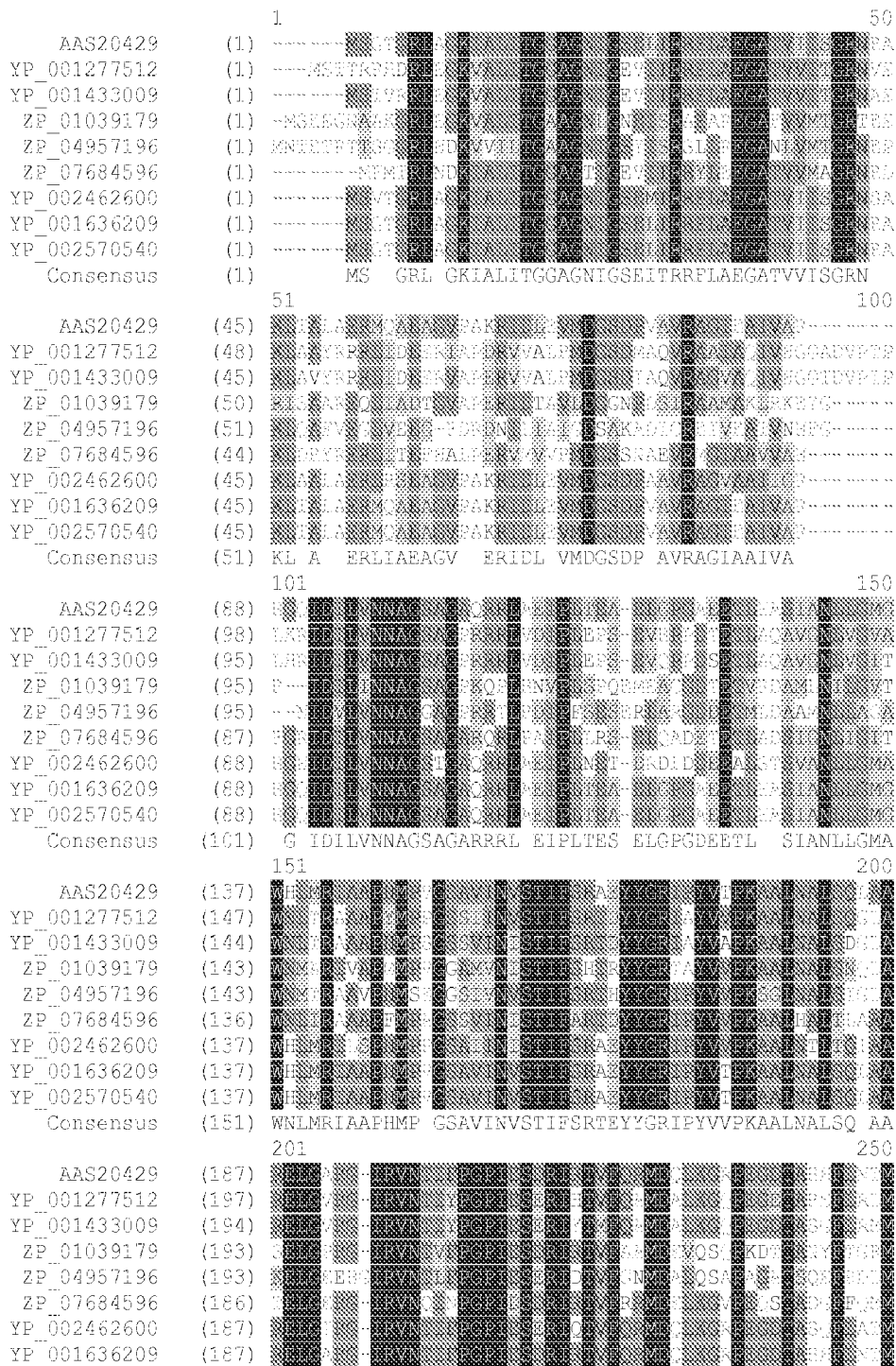


Figure 46 cont.

YP_002570540	(187)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
Consensus	(201)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
AAS20429	(236)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_001277512	(246)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_001433009	(243)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
ZP_01039179	(242)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
ZP_04957196	(243)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
ZP_07684596	(235)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_002462600	(236)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_001636209	(236)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_002570540	(236)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
Consensus	(251)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
AAS20429	(286)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_001277512	(296)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_001433009	(293)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
ZP_01039179	(292)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
ZP_04957196	(293)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
ZP_07684596	(285)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_002462600	(286)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_001636209	(286)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_002570540	(286)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
Consensus	(301)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
AAS20429	(336)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_001277512	(346)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_001433009	(343)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
ZP_01039179	(342)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
ZP_04957196	(343)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
ZP_07684596	(335)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_002462600	(336)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_001636209	(336)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_002570540	(336)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
Consensus	(351)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
AAS20429	(379)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_001277512	(396)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_001433009	(393)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
ZP_01039179	(380)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
ZP_04957196	(380)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
ZP_07684596	(378)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_002462600	(379)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_001636209	(379)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_002570540	(379)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
Consensus	(401)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
AAS20429	(428)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_001277512	(445)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_001433009	(442)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
ZP_01039179	(430)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
ZP_04957196	(429)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
ZP_07684596	(426)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_002462600	(427)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_001636209	(427)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_002570540	(427)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
Consensus	(451)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
AAS20429	(478)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_001277512	(491)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
YP_001433009	(488)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
ZP_01039179	(476)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300
ZP_04957196	(475)	RELG RG IRVNTIFPGPIESDRITVFQMD LAG FEGDIA FL M	251	300

SUBSTITUTE SHEET (RULE 26)

SUBSTITUTE SHEET (RULE 26)

Figure 46 cont.

		1101		1150
AAS20429	(1065)			
YP_001277512	(1078)			
YP_001433009	(1075)			
ZP_01039179	(1068)			
ZP_04957196	(1055)			
ZP_07684596	(1066)			
YP_002462600	(1064)			
YP_001636209	(1064)			
YP_002570540	(1064)			
Consensus	(1101)	FRPLPS PLVGRKDSWSTVLS AEFALCE QLTHHFRVARKIAL DGA		
		1151		1200
AAS20429	(1115)			
YP_001277512	(1127)			
YP_001433009	(1124)			
ZP_01039179	(1112)			
ZP_04957196	(1105)			
ZP_07684596	(1116)			
YP_002462600	(1114)			
YP_001636209	(1114)			
YP_002570540	(1114)			
Consensus	(1151)	SL LVIPETTA STEQFALANFVKITLHAFTATIGVESERIA RILVNQ		
		1201		1250
AAS20429	(1165)			
YP_001277512	(1177)			
YP_001433009	(1174)			
ZP_01039179	(1162)			
ZP_04957196	(1155)			
ZP_07684596	(1166)			
YP_002462600	(1164)			
YP_001636209	(1164)			
YP_002570540	(1164)			
Consensus	(1201)	VDLTRRARAEPRDP E QQELERFIEAVLLVTAPLP EADSRY GRIHR		
		1251		
AAS20429	(1215)			
YP_001277512	(1227)			
YP_001433009	(1224)			
ZP_01039179	(1212)			
ZP_04957196	(1205)			
ZP_07684596	(1216)			
YP_002462600	(1214)			
YP_001636209	(1214)			
YP_002570540	(1214)			
Consensus	(1251)	GRAITV		

1421572_1.DOC