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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 October 2012 (26.10.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/145179 A1

(51) **International Patent Classification**:
C12N 9/04 (2006.01) *C12P 7/06* (2006.01)
C12N 9/68 (2006.01) *C12N 9/02* (2006.01)
C12N 1/19 (2006.01)

(21) **International Application Number**:
PCT/US2012/032558

(22) **International Filing Date**:
6 April 2012 (06.04.2012)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
61/472,664 7 April 2011 (07.04.2011) US

(71) **Applicant** *(for all designated States except US)*: **E. I. DU PONT DE NEMOURS AND COMPANY** [US/US]; 1007 Market Street, Wilmington, Delaware 19898 (US).

(72) **Inventors; and**
(75) **Inventors/Applicants** *(for US only)*: **TAO, Luan** [US/US]; 404 Karen Lane, Wallingford, Pennsylvania 19086 (US). **TOMB, Jean-Francois** [US/US]; 627 Haverhill Road, Wilmington, Delaware 19803 (US). **VIITANEN, Paul V.** [US/US]; 1138 Country Club Road, West Chester, Pennsylvania 19382 (US).

(74) **Agent**: **ODELL, Joan T.**; E. I. du Pont de Nemours and Company, Legal Patent Records Center, 4417 Lancaster Pike, Unionville, Pennsylvania 19375 (US).

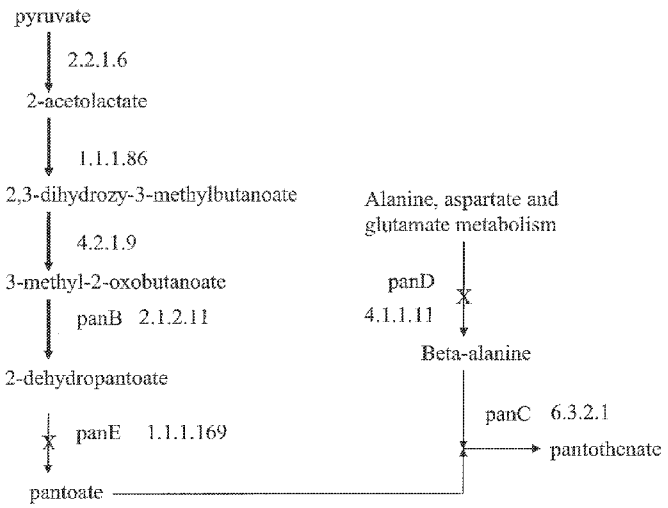
(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, 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, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, 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:
— *with international search report (Art. 21(3))*
— *with (an) indication(s) in relation to deposited biological material furnished under Rule 13bis separately from the description (Rules 13bis.4(d)(i) and 48.2(a)(viii))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) **Title**: PANTOTHENIC ACID BIOSYNTHESIS IN ZYMOMONAS

Figure 1



(57) **Abstract**: *Zymomonas* is unable to synthesize pantothenic acid and requires this essential vitamin in growth medium. *Zymomonas* strains transformed with an operon for expression of 2-dehydropantoate reductase and aspartate 1-decarboxylase were able to grow in medium lacking pantothenic acid. These strains may be used for ethanol production without pantothenic acid supplementation in seed culture and fermentation media.

WO 2012/145179 A1

TITLE

PANTOTHENIC ACID BIOSYNTHESIS IN *ZYMOMONAS*

This application claims the benefit of United States Provisional Application 61/472664, filed April 7, 2011, and is incorporated by reference in its entirety.

STATEMENT OF GOVERNMENT RIGHTS

This invention was made with United States government support under Contract No. DE-FC36-07GO17056 awarded by the Department of Energy. The government has certain rights in this invention.

FIELD OF THE INVENTION

The invention relates to the fields of microbiology and genetic engineering. More specifically, *Zymomonas* was engineered for expression of enzymes to provide a pathway for pantothenate biosynthesis.

BACKGROUND OF THE INVENTION

Production of ethanol by microorganisms provides an alternative energy source to fossil fuels and is therefore an important area of current research. The bacteria *Zymomonas* naturally produces ethanol, and has been genetically engineered for improved ethanol production. Improvements include elimination of competing pathways, utilization of xylose, and better performance in medium containing biomass hydrolysate (for example: US 7,741,119, US 5,514,583, US 5,712,133, WO 95/28476, Feldmann et al. (1992) Appl. Microbiol. Biotechnol. 38: 354-361, Zhang et al. (1995) Science 267:240-243, and US 2009-0203099 A1). The hydrolysate produced from lignocellulosic and cellulosic biomass can provide an abundantly available, low cost source of carbon substrates for biocatalyst fermentation to produce desired products. Biomass hydrolysate typically includes xylose, as well as inhibitors of fermentation.

For economical fermentative production, it is desired that a biocatalyst does not require addition of any costly nutrients to growth and production media. In particlar, it is desired that no vitamin supplements be required for seed or production biocatalyst cultures. *Zymomonas* requires supplementation of pantothenic acid (PA; also pantothenate, vitamin B_5, 3-[(2,4-dihydroxy-3,3-dimethylbutanoyl)amino] propanoic acid) in growth medium, being unable to synthesize this nutrient (Seo et al. (2005) Nat. Biotechnol. 23:63-68; Nipkow et al. (1984) Appl. Microbiol. Biotechnol. 19:237-240 and references therein). PA is an important cellular component as it is required for the synthesis of coenzyme-A (CoA), a compound with many important cellular functions. For many animals it is an essential nutrient, while many plants express enzymes for the synthesis of PA.

E. coli is able to synthesize PA and the biosynthetic pathway is known. *E. coli* genes encoding enzymes of the pathway have been identified. Increased production of pantothenate has been achieved by overexpressing genes in the biosynthetic pathway of microorganisms that naturally produce pantothenate. Disclosed in WO 2003006664 is increasing expression of coding regions in a *Bacillus* that naturally produces D-pantothenic acid, such as ybbT, ywkA, yjmC, ytsJ, mdh, cysK, iolJ, pdhD, yuiE, dhas, adk, yusH, yqhJ, yqhK, and/or yqh-I for increased pantothenic acid production. In addition, panE, ylbQ, panB, panD, panC, ilvB, ilvN, alsS, ilvC, ilvD, serA, serC, ywpJ, and/or glyA may be increased in expression. US 6,171,845 discloses amplification of nucleotide sequences encoding keptopantoate reductase, in particular panE, in pantothenic acid producing microorganisms. It was shown that the *Saccharomyces cerevisiae* YRH063c ORF encodes a protein having ketopantoate reductase activity by complementation of a panE-ilvC mutant in *E. coli*. US 20050089973 discloses producing panto-compounds in microorganisms where existing biosynthetic pathways are manipulated, such as by overexpressing ketopantoate reductase and aspartate alpha-decarboxylase.

US 2005221466 discloses the use of cells with alanine 2,3-aminomutase activity, which converts alpha-alanine to beta-alanine, for production of pantothenate.

There remains a need for creating *Zymomonas* strains that are able to grow and produce ethanol in the absence of externally supplied PA. These *Zymomonas* strains may be used to improve and reduce the cost of ethanol production using this biocatalyst.

SUMMARY OF THE INVENTION

The invention provides recombinant *Zymomonas* cells that express heterologous enzymes to provide a PA biosynthetic pathway.

Accordingly, the invention provides a bacterial strain of the genus *Zymomonas* comprising a heterologous nucleic acid molecule encoding a polypeptide having 2-dehydropantoate reductase activity and a heterologous nucleic acid molecule encoding a polypeptide having aspartate 1-decarboxylase activity.

In another embodiment the invention provides a process for producing a *Zymomonas* strain that synthesizes pantothenic acid comprising:

a) providing a bacterial strain of the genus *Zymomonas*;

b) introducing a heterologous nucleic acid molecule encoding a polypeptide having 2-dehydropantoate reductase activity; and

c) introducing a heterologous nucleic acid molecule encoding a polypeptide having aspartate 1-decarboxylase activity;

wherein steps b) and c) may be in either order or simultaneous and wherein 2-dehydropantoate reductase and aspartate 1-decarboxylase activities are expressed in the strain.

In yet another embodiment the invention provides a method for producing ethanol comprising:

a) providing the recombinant bacterial strain of the genus *Zymomonas* comprising a heterologous nucleic acid molecule encoding a polypeptide having 2-dehydropantoate reductase activity and a

heterologous nucleic acid molecule encoding a polypeptide having aspartate 1-decarboxylase activity; and

b) contacting the strain of (a) with fermentation medium under conditions whereby the strain produces ethanol.

BRIEF DESCRIPTION OF THE FIGURES, BIOLOGICAL DEPOSITS AND SEQUENCE DESCRIPTIONS

Applicants have made the following biological deposits under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure:

INFORMATION ON DEPOSITED STRAINS

Depositor Identification Reference	International Depository Designation	Date of Deposit
Zymomonas ZW658	ATCC No PTA-7858	Sept. 12, 2006

Figure 1 is a diagram of a pantothenic acid biosynthetic pathway with bold arrows marking activities that may be present in *Zymomonas* and arrows with" X" marking absent activities. The numbers with the arrows are EC numbers of enzymes that perform the shown reaction. Gene names associated with the EC niumbers are given in some cases.

Figure 2 shows conserved amino acid positions, using one letter abbreviations, of aspartate 1-decarboxylase polypeptides in a general structure diagram (A), in a representative sequence (B), and in the *E. coli* aspartate 1-decarboxylase amino acid sequence (C).

Figure 3 shows conserved amino acid positions, using one letter abbreviations, of 2-dehydropantoate reductase polypeptides in a general structure diagram (A), in a representative sequence (B), in the *E. coli* 2-dehydropantoate reductase amino acid sequence based on a ten sequence alignment (C), and in the *E. coli* 2-dehydropantoate reductase amino acid sequence based on a 648 sequence alignment (D).

Figure 4 shows a graph of growth curves of wild type *Zymomonas* strain ZW1 grown in minimal medium supplemented with different concentrations of pantothenic acid (PA), after a PA-depletion step.

Figure 5 shows a graph of growth curves of wild type *Zymomonas* strain ZW1(A) and strain ZW1/PanED#1 (also referred to as ZED#1) (B) grown in minimal medium with or without 2.5 mg/L pantothenic acid supplementation, after a PA-depletion step..

Figure 6 shows a graph of growth curves of *Zymomonas* strain ZW801-4 in (A) minimal medium with and without 2.5 mg/L pantothenic acid supplementation, after a PA-depletion step, and in (B) minimal medium containing 15 mg/L p-aminobenzoic acid and different concentrations of pantothenic acid, after a PA-depletion step.

Figure 7 shows a graph of growth curves of *Zymomonas* strains ZW801-4 (A) and ZW801-4/PanED#1(B) grown in minimal medium supplemented with 15 mg/L p-aminobenzoic acid (PABA), 2.5 mg/L pantothenic acid, both compounds, or neither, after a PA-depletion step.

Appendix 1, which is incorporated herein by refernece, is a listing of Accession numbers and annotated identities of 648 2-dehydropantoate reductases of 250-350 amino acids that have an E-value of 0.00001 or smaller to the *E. coli* 2-dehydropantoate reductase of SEQ ID NO:4, with 95% identity and 95% overlap redundancy cutoffs.

Appendix 2, which is incorporated herein by reference, is a listing of Accession numbers and annotated identities of 693 aspartate 1-decarboxylases of 120-150 amino acids that have an E-value of 0.00001 or smaller to the *E. coli* aspartate 1-decarboxylase of SEQ ID NO:7, with 95% identity and 95% overlap redundancy cutoffs.

The invention can be more fully understood from the following detailed description and the accompanying sequence descriptions which form a part of this application.

The following sequences conform with 37 C.F.R. 1.821-1.825 ("Requirements for Patent Applications Containing Nucleotide Sequences and/or Amino Acid Sequence Disclosures - the Sequence Rules") and are consistent with World Intellectual Property Organization (WIPO) Standard

ST.25 (2009) and the sequence listing requirements of the EPO and PCT (Rules 5.2 and 49.5(a-bis), and Section 208 and Annex C of the Administrative Instructions). The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

SEQ ID NO:1 is the nucleotide sequence of a synthetic chimeric *E. coli* panE and panD operon.

SEQ ID NO:2 is the nucleotide sequence of the GI promoter from the *Actinoplanes missouriensis* xylose isomerase gene.

SEQ ID NO:3 is the nucleotide sequence of the *E. coli* panE open reading frame encoding 2-dehydropantoate reductase.

SEQ ID NO:4 is the amino acid sequence of the *E. coli* panE encoded 2-dehydropantoate reductase (strain K-12 substr. MG1655; gi|16128410|ref|NP_414959.1|).

SEQ ID NO:5 is the nucleotide sequence of a stretch of DNA that is upstream from the start codon for the *Z. mobilis* glyceraldehyde 3-phosphate dehydrogenase gene that includes the Shine-Delgarno sequence

SEQ ID NO:6 is the nucleotide sequence of the *E. coli* panD open reading frame encoding aspartate 1-decarboxylase.

SEQ ID NO:7 is the amino acid sequence of the *E. coli* panD encoded aspartate 1-decarboxylase.

SEQ ID NO:8 is the nucleotide sequence of a stretch of DNA that corresponds to the small, stabilizing stem-loop sequence that immediately follows the xylose isomerase (xylA) stop codon in the *E. coli* XylA/B operon.

Table 1. 2-dehydropantoate reductases used in first alignment

Accession number	Organism	SEQ ID NO
gi|53803269|ref|YP_114934.1|	*Methylococcus capsulatus*	9
gi|78223840|ref|YP_385587.1|	*Geobacter metallireducens* GS-15	10
gi|19113647|ref|NP_596855.1|	*Schizosaccharomyces pombe*	11
gi|6321854|ref|NP_011930.1|	*Saccharomyces cerevisiae* S288c	12

gi|73538792|ref|YP_299159.1|	*Ralstonia eutropha* JMP134	13
gi|207722086|ref|YP_002252524.	*Ralstonia solanacearum* MolK2	14
gi|194367655|ref|YP_002030265.	*Stenotrophomonas maltophilia* R551-3	15
gi|29376939|ref|NP_816093.1|	*Enterococcus faecalis* V583	16
gi|16078575|ref|NP_389394.1|	*Bacillus subtilis* subsp. subtilis str. 168]	17

SEQ ID NO:18 is a representation of a conserved amino acid sequence for aspartate 1-decarboxylase showing the highly conserved amino acid positions, without notation of less conserved amino acid positions and with insertion positions omitted.

SEQ ID NO:19 is a representation of a conserved amino acid sequence for 2-dehydropantoate reductase showing the highly conserved amino acid positions, without notation of less conserved amino acid positions and with insertion positions omitted.

DETAILED DESCRIPTION

The following definitions may be used for the interpretation of the claims and specification:

As used herein, the terms “comprises,” “comprising,” “includes,” “including,” “has,” “having,” “contains” or “containing,” or any other variation thereof, are intended to cover a non-exclusive inclusion. For example, a composition, a mixture, process, method, article, or apparatus that comprises a list of elements is not necessarily limited to only those elements but may include other elements not expressly listed or inherent to such composition, mixture, process, method, article, or apparatus. Further, unless expressly stated to the contrary, “or” refers to an inclusive or and not to an exclusive or. For example, a condition A or B is satisfied by any one of the following: A is true (or present) and B is false (or not present), A is false (or not present) and B is true (or present), and both A and B are true (or present).

Also, the indefinite articles “a” and “an” preceding an element or component of the invention are intended to be nonrestrictive regarding the number of instances (i.e. occurrences) of the element or component. Therefore “a” or “an” should be read to include one or at least one, and the singular word form of the element or component also includes the plural unless the number is obviously meant to be singular.

The term “invention” or “present invention” as used herein is a non-limiting term and is not intended to refer to any single embodiment of the particular invention but encompasses all possible embodiments as described in the specification and the claims.

As used herein, the term "about" modifying the quantity of an ingredient or reactant of the invention employed refers to variation in the numerical quantity that can occur, for example, through typical measuring and liquid handling procedures used for making concentrates or use solutions in the real world; through inadvertent error in these procedures; through differences in the manufacture, source, or purity of the ingredients employed to make the compositions or carry out the methods; and the like. The term “about” also encompasses amounts that differ due to different equilibrium conditions for a composition resulting from a particular initial mixture. Whether or not modified by the term "about", the claims include equivalents to the quantities. In one embodiment, the term “about” means within 10% of the reported numerical value, preferably within 5% of the reported numerical value.

"Gene" refers to a nucleic acid fragment that expresses a specific protein or functional RNA molecule, which may optionally include regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" or "wild type gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers to any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same

source, but arranged in a manner different than that found in nature. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes.

"Promoter" or "Initiation control regions" refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters".

The term "expression", as used herein, refers to the transcription and stable accumulation of coding (mRNA) or functional RNA derived from a gene. Expression may also refer to translation of mRNA into a polypeptide. “Overexpression” refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms.

The term "transformation" as used herein, refers to the transfer of a nucleic acid fragment into a host organism, resulting in genetically stable inheritance. The transferred nucleic acid may be in the form of a plasmid maintained in the host cell, or some transferred nucleic acid may be integrated into the genome of the host cell. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" or "recombinant" or "transformed" organisms.

The terms "plasmid" and "vector" as used herein, refer to an extra chromosomal element often carrying genes which are not part of the central metabolism of the cell, and usually in the form of circular double-

stranded DNA molecules. Such elements may be autonomously replicating sequences, genome integrating sequences, phage or nucleotide sequences, linear or circular, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of introducing a promoter fragment and DNA sequence for a selected gene product along with appropriate 3' untranslated sequence into a cell.

The term "operably linked" refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "selectable marker" means an identifying factor, usually an antibiotic or chemical resistance gene, that is able to be selected for based upon the marker gene's effect, i.e., resistance to an antibiotic, wherein the effect is used to track the inheritance of a nucleic acid of interest and/or to identify a cell or organism that has inherited the nucleic acid of interest.

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

The term "codon-optimized" as it refers to genes or coding regions of nucleic acid molecules for transformation of various hosts, refers to the alteration of codons in the gene or coding regions of the nucleic acid molecules to reflect the typical codon usage of the host organism without

altering the polypeptide encoded by the DNA.

The term “carbon substrate” or “fermentable carbon substrate” refers to a carbon source capable of being metabolized by microorganisms. A type of carbon substrate is “fermentable sugars” which refers to oligosaccharides and monosaccharides that can be used as a carbon source by a microorganism in a fermentation process.

The term “lignocellulosic” refers to a composition comprising both lignin and cellulose. Lignocellulosic material may also comprise hemicellulose.

The term “cellulosic” refers to a composition comprising cellulose and additional components, which may include hemicellulose and lignin.

The term “saccharification” refers to the production of fermentable sugars from polysaccharides.

The term “pretreated biomass” means biomass that has been subjected to thermal, physical and/or chemical pretreatment to increase the availability of polysaccharides in the biomass to saccharification enzymes.

“Biomass” refers to any cellulosic or lignocellulosic material and includes materials comprising cellulose, and optionally further comprising hemicellulose, lignin, starch, oligosaccharides and/or monosaccharides. Biomass may also comprise additional components, such as protein and/or lipid. Biomass may be derived from a single source, or biomass can comprise a mixture derived from more than one source; for example, biomass could comprise a mixture of corn cobs and corn stover, or a mixture of grass and leaves. Biomass includes, but is not limited to, bioenergy crops, agricultural residues, municipal solid waste, industrial solid waste, sludge from paper manufacture, yard waste, wood and forestry waste. Examples of biomass include, but are not limited to, corn cobs, crop residues such as corn husks, corn stover, grasses, wheat, wheat straw, barley straw, hay, rice straw, switchgrass, waste paper, sugar cane bagasse, sorghum, components obtained from milling of grains, trees, branches, roots, leaves, wood chips, sawdust, shrubs and bushes, vegetables, fruits, flowers and animal manure.

"Biomass hydrolysate" refers to the product resulting from saccharification of biomass. The biomass may also be pretreated or pre-processed prior to saccharification.

The term "heterologous" means not naturally found in the location of interest. For example, a heterologous gene refers to a gene that is not naturally found in the host organism, but that is introduced into the host organism by gene transfer. For example, a heterologous nucleic acid molecule that is present in a chimeric gene is a nucleic acid molecule that is not naturally found associated with the other segments of the chimeric gene, such as the nucleic acid molecules having the coding region and promoter segments not naturally being associated with each other.

As used herein, an "isolated nucleic acid molecule" is a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid molecule in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA.

A nucleic acid fragment is "hybridizable" to another nucleic acid fragment, such as a cDNA, genomic DNA, or RNA molecule, when a single-stranded form of the nucleic acid fragment can anneal to the other nucleic acid fragment under the appropriate conditions of temperature and solution ionic strength. Hybridization and washing conditions are well known and exemplified in Sambrook, J., Fritsch, E. F. and Maniatis, T. Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory: Cold Spring Harbor, NY (1989), particularly Chapter 11 and Table 11.1 therein (entirely incorporated herein by reference). The conditions of temperature and ionic strength determine the "stringency" of the hybridization. Stringency conditions can be adjusted to screen for moderately similar fragments (such as homologous sequences from distantly related organisms), to highly similar fragments (such as genes that duplicate functional enzymes from closely related organisms). Post-hybridization washes determine stringency conditions. One set of preferred conditions uses a series of washes starting with 6X SSC, 0.5% SDS at room temperature for 15 min, then repeated with 2X SSC, 0.5%

SDS at 45 °C for 30 min, and then repeated twice with 0.2X SSC, 0.5% SDS at 50 °C for 30 min. A more preferred set of stringent conditions uses higher temperatures in which the washes are identical to those above except for the temperature of the final two 30 min washes in 0.2X SSC, 0.5% SDS was increased to 60 °C. Another preferred set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65 °C. An additional set of stringent conditions include hybridization at 0.1X SSC, 0.1% SDS, 65 °C and washes with 2X SSC, 0.1% SDS followed by 0.1X SSC, 0.1% SDS, for example.

Hybridization requires that the two nucleic acids contain complementary sequences, although depending on the stringency of the hybridization, mismatches between bases are possible. The appropriate stringency for hybridizing nucleic acids depends on the length of the nucleic acids and the degree of complementation, variables well known in the art. The greater the degree of similarity or homology between two nucleotide sequences, the greater the value of Tm for hybrids of nucleic acids having those sequences. The relative stability (corresponding to higher Tm) of nucleic acid hybridizations decreases in the following order: RNA:RNA, DNA:RNA, DNA:DNA. For hybrids of greater than 100 nucleotides in length, equations for calculating Tm have been derived (see Sambrook et al., supra, 9.50-9.51). For hybridizations with shorter nucleic acids, i.e., oligonucleotides, the position of mismatches becomes more important, and the length of the oligonucleotide determines its specificity (see Sambrook et al., supra, 11.7-11.8). In one embodiment the length for a hybridizable nucleic acid is at least about 10 nucleotides. Preferably a minimum length for a hybridizable nucleic acid is at least about 15 nucleotides; more preferably at least about 20 nucleotides; and most preferably the length is at least about 30 nucleotides. Furthermore, the skilled artisan will recognize that the temperature and wash solution salt concentration may be adjusted as necessary according to factors such as length of the probe.

The term “complementary” is used to describe the relationship between nucleotide bases that are capable of hybridizing to one another. For

example, with respect to DNA, adenosine is complementary to thymine and cytosine is complementary to guanine.

The term "percent identity", as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in: 1.) Computational Molecular Biology (Lesk, A. M., Ed.) Oxford University: NY (1988); 2.) Biocomputing: Informatics and Genome Projects (Smith, D. W., Ed.) Academic: NY (1993); 3.) Computer Analysis of Sequence Data, Part I (Griffin, A. M., and Griffin, H. G., Eds.) Humania: NJ (1994); 4.) Sequence Analysis in Molecular Biology (von Heinje, G., Ed.) Academic (1987); and 5.) Sequence Analysis Primer (Gribskov, M. and Devereux, J., Eds.) Stockton: NY (1991).

Preferred methods to determine identity are designed to give the best match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Sequence alignments and percent identity calculations may be performed using the MegAlign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI).

Multiple alignment of the sequences is performed using the "Clustal method of alignment" which encompasses several varieties of the algorithm including the "Clustal V method of alignment" corresponding to the alignment method labeled Clustal V (described by Higgins and Sharp, *CABIOS*. 5:151-153 (1989); Higgins, D.G. et al., *Comput. Appl. Biosci.,* 8:189-191 (1992)) and found in the MegAlign v8.0 program of the LASERGENE bioinformatics computing suite (DNASTAR Inc.). For multiple alignments, the default values correspond to GAP PENALTY=10 and GAP LENGTH PENALTY=10. Default parameters for pairwise alignments and calculation of percent identity of protein sequences using

the Clustal method are KTUPLE=1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. For nucleic acids these parameters are KTUPLE=2, GAP PENALTY=5, WINDOW=4 and DIAGONALS SAVED=4. After alignment of the sequences using the Clustal V program, it is possible to obtain a "percent identity" by viewing the "sequence distances" table in the same program.

Additionally the "Clustal W method of alignment" is available and corresponds to the alignment method labeled Clustal W (described by Higgins and Sharp, *CABIOS*. 5:151-153 (1989); Higgins, D.G. et al., *Comput. Appl. Biosci.* 8:189-191(1992); Thompson, J.D. et al, Nucleic Acid Research, 22 (22): 4673-4680, 1994) and found in the MegAlign v8.0 program of the LASERGENE bioinformatics computing suite (DNASTAR Inc.). Default parameters for multiple alignment (stated as protein/nucleic acid (GAP PENALTY=10/15, GAP LENGTH PENALTY=0.2/6.66, Delay Divergen Seqs(%)=30/30, DNA Transition Weight=0.5, Protein Weight Matrix=Gonnet Series, DNA Weight Matrix=IUB). After alignment of the sequences using the Clustal W program, it is possible to obtain a "percent identity" by viewing the "sequence distances" table in the same program.

It is well understood by one skilled in the art that many levels of sequence identity are useful in identifying polypeptides, from other species, wherein such polypeptides have the same or similar function or activity. Useful examples of percent identities include, but are not limited to: 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%, or any integer percentage from 50% to 100% may be useful in identifying polypeptides of interest, such as 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99%. Suitable nucleic acid fragments not only have the above identities but typically encode a polypeptide having at least 50 amino acids, preferably at least 100 amino acids, and more preferably at least 125 amino acids.

The term "sequence analysis software" refers to any computer

algorithm or software program that is useful for the analysis of nucleotide or amino acid sequences. "Sequence analysis software" may be commercially available or independently developed. Typical sequence analysis software will include, but is not limited to: 1.) the GCG suite of programs (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI); 2.) BLASTP, BLASTN, BLASTX (Altschul et al., J. Mol. Biol., 215:403-410 (1990)); 3.) DNASTAR (DNASTAR, Inc. Madison, WI); 4.) Sequencher (Gene Codes Corporation, Ann Arbor, MI); and 5.) the FASTA program incorporating the Smith-Waterman algorithm (W. R. Pearson, Comput. Methods Genome Res., [Proc. Int. Symp.] (1994), Meeting Date 1992, 111-20. Editor(s): Suhai, Sandor. Plenum: New York, NY). Within the context of this application it will be understood that where sequence analysis software is used for analysis, that the results of the analysis will be based on the "default values" of the program referenced, unless otherwise specified. As used herein "default values" will mean any set of values or parameters that originally load with the software when first initialized.

The term "E-value", as known in the art of bioinformatics, is "Expect-value" which provides the probability that a match will occur by chance. It provides the statistical significance of the match to a sequence. The lower the E-value, the more significant the hit.

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described by Sambrook, J. and Russell, D., Molecular Cloning: A Laboratory Manual, Third Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (2001); and by Silhavy, T. J., Bennan, M. L. and Enquist, L. W., Experiments with Gene Fusions, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1984); and by Ausubel, F. M. et. al., Short Protocols in Molecular Biology, 5th Ed. Current Protocols, John Wiley and Sons, Inc., N.Y., 2002.

The present invention relates to engineered strains of *Zymomonas* that have the ability to grow without pantothenic acid (PA; also called pantothenate and vitamin B_5) supplementation in growth and production medium. A challenge for providing an economical process for fermentation

by *Zymomonas* to produce ethanol is to reduce the requirement for vitamin supplementation in the medium, specifically of PA, thereby reducing the cost of growth and/or production medium. *Zymomonas* strains disclosed herein are genetically engineered to synthesize pantothenic acid.

Pantothenic acid biosynthesis

Zymomonas is known to lack the natural ability to synthesize pantothenic acid and therefore requires the presence of this vitamin in medium used for growth of this bacteria. PA is require for CoA (Coenzyme A) production, and is therefore critical for carbohydrate, protein and fatty acid metabolism.

Applicants analyzed the *Zymomonas* genome for the potential to encode enzymes of a pantothenic acid biosynthetic pathway. The complete sequence of the *Zymomonas* genome is known (Seo et al. (2005) Nat. Biotechnol. 23:63-68; NCBI Reference: NC_006526.2) and open reading frames (ORFs) have been annotated as encoding proteins with defined function where possible, based on sequence analysis. The presence of ORFs potentially encoding enzymes that could function in a pantothenic acid biosynthetic pathway was analyzed using KEGG analysis (Kyoto Encyclopedia of Genes and Genomes; Kanehisa et al. (2002) Nucleic Acids Res. 30:42-46; Kanehisa and Goto (2000) Nucleic Acids Res. 28:27-30; Kanehisa et al. (2006) Nucleic Acids Res. 34:D354-357). KEGG provides knowledge-based methods for uncovering higher-order systemic behaviors of the cell and the organism from genomic and molecular information, as stated by KEGG.

The KEGG analysis showed that the *Zymomonas* genome has the potential for encoding proteins with activities for some steps in a pantothenic acid biosynthetic pathway (see Figure 1, bold arrows). The genome includes ORFs predicted to encode proteins with activities that may produce 2-dehydropantoate and L-aspartate. The EC group to which each enzyme activity in the pathway belongs, and in some cases the name of the gene encoding the enzyme, is shown in the pathway diagram of Figure 1.

No ORF was found with the potential to encode a protein with activity that would convert 2-dehydropantoate to pantoate, and no ORF was found with the potential to encode a protein with activity that would convert L-aspartate to beta-alanine. These two steps are marked with an **X** in the Figure 1 diagram. Pantoate and beta-alanine are ligated together to produce pantothenic acid in a known pantothenic acid biosynthetic pathway, such as from *E. coli*. Pantoate-beta-alanine ligase is encoded by the panC gene in many organisms, including in *E. coli*. The *Zymomonas* genome does have the potential for encoding a pantoate-beta-alanine ligase. However, the protein encoded by the ORF annotated as panC has only 46% amino acid sequence identity with the *E. coli* panC gene encoded pantoate-beta-alanine ligase. With presumably no pantoate and beta-alanine substrates available in the cell, the native function of the protein encoded by the ORF annotated as panC is unknown.

Conversion of 2-dehydropantoate to pantoate in *E. coli* is catalyzed by 2-dehydropantoate reductase, which is encoded by the panE ORF in many organisms including in *E. coli* (SEQ ID NO:3). Conversion of L-aspartate to beta-alanine is catalyzed by aspartate 1-decarboxylase, which is encoded by the panD ORF in many organisms including in *E. coli* (SEQ ID NO:6). Whether expression of these two activities in *Zymomonas* cells would confer the ability to synthesize PA was unknown, since that outcome necessitates the assumption that the presumed existing native enzymes, that sequence analysis speculates can participate in PA biosynthesis, actually do encode functional enzymes that do catalyze reactions of a portion of a PA biosynthetic pathway in *Zymomonas*.

Upon experimental analysis as disclosed herein, Applicants have discovered that *Zymomonas* cells engineered for expression of the *E. coli* panE and panD coding regions are able to grow in medium that does not contain PA. This result suggests that said *Zymomonas* cells have a complete functional PA biosynthetic pathway and synthesize PA. Further, the engineered wild type *Zymomonas* cells grow as well in medium that does not contain PA as wild type *Zymomonas* cells grow in the same medium supplemented with a non-limiting amount of PA .

In the present recombinant *Zymomonas* bacterial strains, a nucleic acid molecule encoding a polypeptide having 2-dehydropantoate reductase activity and a nucleic acid molecule encoding a polypeptide having aspartate 1-decarboxylase activity are introduced. These polypeptides are encoded by heterologous nucleic acid molecules that are introduced into the *Zymomonas* cell.

Host Zymomonas cells

Heterologous nucleic acid molecules encoding polypeptides with 2-dehydropantoate reductase activity and aspartate 1-decarboxylase activity may be introduced into any strain of *Zymomonas,* such as *Zymomonas mobilis,* to create a pantothenic acid biosynthesis pathway. Wild type *Zymomonas* strains naturally produce ethanol and may be used as a host for introduction of said nucleic acid molecules. In other embodiments the *Zymomonas* host strains are recombinant strains engineered to be improved biocatalysts for ethanol production and comprise a number of genetic modifications that enhance the production of ethanol. Host strains may be strains engineered in one or more of the following ways, in any combination. *Z. mobilis* strains have been engineered to utilize xylose, a sugar found in biomass hydrolysate, for ethanol production (US 5,514,583, US 5,712,133, WO 95/28476, Feldmann et al. (1992) App.l Microbiol. Biotechnol. 38: 354-361, Zhang et al. (1995) Science 267:240-243). Ethanol has been produced by genetically modified *Zymomonas* in lignocellulosic biomass hydrolysate fermentation media (US 7,932,063). Genetically modified strains of *Z. mobilis* with improved xylose utilization and/or production of ethanol are disclosed in US 7,223,575, US 7,741,119, US 7,897,396, US 7,998,722, and WO2010/075241 (US 2011/0014670), which are herein incorporated by reference. Any of the disclosed strains, including for example ATCC31821/pZB5, ZW658 (ATCC #PTA-7858), ZW800, ZW801-4, ZW801-4:: ΔhimA, AcR#3, ZW705, or other ethanol-producing strains of *Zymomonas,* may be used as host cells for expression of heterologous nucleic acid molecules encoding a

polypeptide having 2-dehydropantoate reductase activity and encoding a polypeptide having aspartate 1-decarboxylase activity, which enables pantothenic acid biosynthesis.

Enzyme activities

Any nucleic acid molecule encoding a polypeptide having 2-dehydropantoate reductase activity may be used in the present strains. Enzymes with 2-dehydropantoate reductase activity are also called 2-dehydropantoate 2-reductase, 2-oxopantoate reductase, ketopantoate reductase, ketopantoic acid reductase, KPA reductase, and KPR. The reaction catalyzed by this enzyme activity is:

(R)-pantoate + $NADP^+$ ←→ 2-dehydropantoate + NADPH + H^+

The 2-dehydropantoate reductase enzyme is classified as EC 1.1.1.169. A nucleic acid molecule encoding any enzyme belonging to this EC group having 2-dehydropantoate reductase activity may be used in the present strains.

Any nucleic acid molecule encoding a polypeptide having aspartate 1-decarboxylase activity may be used in the present strains. Enzymes with aspartate 1-decarboxylase activity are also called aspartate alpha-decarboxylase, L-aspartate alpha-decarboxylase, aspartic alpha-decarboxylase, L-aspartate 1-carboxy-lyase, ADC, AspDC, and Dgad2, . The reaction catalyzed by this enzyme activity is:

L-aspartate ←→ beta-alanine + CO_2

The aspartate 1-decarboxylase enzyme is classified as EC 4.1.1.11. A nucleic acid molecule encoding any enzyme belonging to this EC group having aspartate 1-decarboxylase activity may be used in the present strains. The protein translated from the *E. coli* panD gene is an inactive protein called the pi-protein (Ramjee et al (1997) Biochem. J. 323:661-669). This protein is autocatalytically self-processed into two subunits (alpha and beta) that form the active enzyme.

Polypeptides with 2-dehydropantoate reductase activity or aspartate 1-decarboxylase activity may be identified using bioinformatics and/or experimental methods. Amino acid sequences of these polypeptides can be readily found by EC number, gene name, and/or

enzyme name using databases that are well known to one of skill in the art including NCBI (National Center for Biotechnology Information; Bethesda, MD), BRENDA (The Comprehensive Enzyme Information System; Technical University of Braunschweig Dept. of Bioinformatics), and Swiss-Prot (Swiss Institute of Bioinformatics; Lausanne, Switzerland). In addition, amino acid sequences of these polypeptides can be readily found based on a known sequence using bioinformatics, including sequence analysis software such as BLAST sequence analysis using for example the *E coli* sequences (2-dehydropantoate reductase: SEQ ID NO:4; aspartate 1-decarboxylase: SEQ ID NO:7).

The following analysis of polypeptide sequences identified a structure that is common to 2-dehydropantoate reductases belonging to EC 1.1.1.169. First the amino acid sequences of ten 2-dehydropantoate reductases (SEQ ID NOs: 9-17 and 4) with experimentally verified function and/or characterized structure as identified in the BRENDA database (BRaunschweig ENzyme Database; Cologne University BioInformatics Center; Scheer et al. (2011) Nucleic Acids Res. 39:670-676) and the Protein Data Bank database (RCSB PDB; Berman et al. (2000) Nucleic Acids Res. 28:235-242) were aligned using Clustal W with the following parameters: Slow/Accurate Pairwise Parameters: Gap Opening = 10, Gap Extend = 0.1, Protein weight matrix Gonnet 250; Multiple Parameters: Gap Opening = 10, Gap Extension = 0.2, Protein Weight Matrix= Gonnet series. From this multiple sequence alignment, the amino acid positions with 90% to 100% conservation as a single amino acid among the ten sequences were identified and used to provide the conserved structure diagram shown in Figure 3A. In this figure the conserved amino acids are indicated as G (glycine), L (leucine), K (lysine), N (asparagine), E (glutamic acid), S (serine), and D (aspartic acid). All of the amino acids shown are 100% conserved except the two asparagines (N) which are each 90% conserved. The dashed lines represent positions in the multiple sequence alignment where insertions and deletions, including N-terminal and C-terminal extensions, occur in one or more of the ten analyzed 2-dehydropantoate reductase amino acid sequences belonging to EC

1.1.1.169. The conserved structure of Figure 3A is represented as a sequence in Figure 3B (SEQ ID NO:19). In this sequence the dashed positions are omitted, and notation of other conserved amino acids, which are not as highly conserved, is omitted.

The amino acids at the conserved positions are highlighted in the *E. coli* 2-dehydropantoate reductase amino acid sequence (SEQ ID NO:4) in Figure 3C and are G at position 7, G at position 9, G at position 12, L at position 19, K at position 72, N at position 98, G at position 99, K at position 176, N at position 180, N at position 184, E at position 210, S at position 244, D at position 248, and E at position 256. One of skill in the art will be readily able to align a candidate sequence with SEQ ID NO:4, allowing for extensions, insertions and deletions such as in positions indicated in the structure diagram of Figure 3A, such that the presence of the conserved amino acids highlighted in Figure 3C can be determined. Any polypeptide having at least 10, at least 11, at least 12, at least 13, or all 14 of these 14 conserved amino acids when compared to SEQ ID NO:4 and having 2-dehydropantoate reductase activity may be used in the present strains.

A BLAST search was performed using the *E. coli* 2-dehydropantoate reductase (SEQ ID NO:4) against publicly available sequences, and protein sequence matches with an E-value of 0.00001 or smaller were extracted. Matched protein sequences in the range of 250-350 amino acids were retained. Sequence redundancy was reduced to 95% identity and 95% overlap. This filtering resulted in 648 sequences, which are listed in Appendix 1 by their accession numbers. A multiple sequence alignment was performed using Clustal W with the same parameters used above. All of the amino acid positions identified above in the ten sequence alignment were also highly conserved among the 648 sequences except the L at position 19. In addition D at position 248 is replaced with S in about 4% of the sequences and the Ns at positions 180 and 184 have some variation. Thus characterization of the broader group of 2-dehydropantoate reductases provided a structure with conserved amino acids highlighted in the *E. coli* 2-dehydropantoate reductase amino

acid sequence (SEQ ID NO:4) shown in Figure 3D. The Ns at positions 180 and 184 are underlined but not bolded to represent some amino acid variation at those positions. At position 248, the presence of S instead of D is included in the conserved structure. The conserved amino acids are G at position 7, G at position 9, G at position 12, K at position 72, N at position 98, G at position 99, K at position 176, N at position 180, N at position 184, E at position 210, S at position 244, D or S at position 248, and E at position 256. One of skill in the art will be readily able to align a candidate sequence with SEQ ID NO:4, allowing for extensions, insertions and deletions such as in positions indicated in the structure diagram in Figure 3A, such that the presence of the conserved amino acids of Figure 3D can be determined. Any polypeptide having at least 10, at least 11, at least 12, or all 13 of these 13 conserved amino acids when compared to SEQ ID NO:4 and having 2-dehydropantoate reductase activity may be used in the present strains.

Nucleic acid molecules that may be used in the present strains include those encoding any protein having 2-dehydropantoate reductase activity, including for example: 1) those belonging to EC 1.1.1.169; 2) those with experimentally verified function and/or characterized structure (SEQ ID NOs: 9-17 and 4); 3) those having conserved structure of Figure 3A, represented as a sequence in Figure 3B (SEQ ID NO:19); 4) those having ten or more of the conserved amino acids highlighted in SEQ ID NO:4 in Figure 3C; 5) those with at least about 95% sequence identity to any of the 648 proteins listed in Appendix 1; and 6) those having ten or more of the conserved amino acids highlighted in SEQ ID NO:4 in Figure 3D.

The following analysis of polypeptide sequences identified a structure that is common to aspartate 1-decarboxylases belonging to EC 4.1.1.11. A BLAST search was performed using the *E. coli* aspartate 1-decarboxylase (SEQ ID NO:7) against publicly available sequences, and protein sequence matches with an E-value of 0.00001 or smaller were extracted. Matched protein sequences in the range of 120-150 amino acids were retained. Sequence redundancy was reduced to 95% identity

and 95% overlap. This filtering resulted in 493 sequences, which are listed in Appendix 2 by their accession numbers. A multiple sequence alignment was performed using Clustal W with the following parameters: Slow/Accurate Pairwise Parameters: Gap Opening = 10, Gap Extend = 0.1, Protein weight matrix Gonnet 250; Multiple Parameters: Gap Opening = 10, Gap Extension = 0.2, Protein Weight Matrix= Gonnet series. A sequence logo was generated by LOGO extraction using Weblogo, a publicly available web based application (Crooks et al (2004) Genome Research 14:1188-1190); Schneider and Stephens (1990) Nucleic Acids Res. 18:6097-6100). According to the provided information, each logo consists of stacks of symbols, one stack for each position in the sequence. The overall height of the stack indicates the sequence conservation at that position, while the height of symbols within the stack indicates the relative frequency of each amino or nucleic acid at that position. The percent frequency of each amino acid at each position was calculated for the set of 493 sequences.

From this analysis, the most highly conserved amino acid positions were identified as those having a single amino acid occurring in at least 99% of the 493 sequences analyzed, and were used to provide the conserved structure diagram shown in Figure 2A. In this figure the conserved amino acids are indicated as K (lysine), H (histidine), Y (tyrosine), G (glycine), S (serine), R (arginine), T (threonine), N (asparagine), and I (isoleucine). The dashed lines represent positions in the multiple sequence alignment where insertions and deletions, including N-terminal and C-terminal extensions, occur in one or more of the aligned 493 aspartate 1-decarboxylase amino acid sequences belonging to EC 4.1.1.11. The conserved structure of Figure 2A is represented as a sequence in Figure 2B (SEQ ID NO:18). In this sequence the dashed positions are omitted, and notation of other conserved amino acids, which are not as highly conserved, is omitted.

The amino acids at the most highly conserved positions are highlighted in the *E. coli* aspartate 1-decarboxylase amino acid sequence (SEQ ID NO:7) in Figure 2C and are K at position 9, H at position 11, Y at

position 22, G at position 24, S at position 25, G at position 52, R at position 54, T at position 57, Y at position 58, N at position 72, G at position 73, and I at position 86. One of skill in the art will be readily able to align a candidate sequence with SEQ ID NO:7, allowing for extensions, insertions and deletions such as in positions indicated in the structure diagram of Figure 2A, such that the presence of the conserved amino acids highlighted in Figure 2C can be determined. Any polypeptide having at least 8, at least 9, at least 10, at least 11, or all 12 of these 12 conserved amino acids when compared to SEQ ID NO:7 and having aspartate 1-decarboxylase activity may be used in the present strains.

Of the 12 most highly conserved amino acid positions shown in Figure 2B, five of these occurred in 100% of the 493 sequences aligned. These100% conserved positions are K at position 9, Y at position 22, G at position 24, T at position 57, and Y at position 58. In one embodiment a polypeptide that may be used has all five of these 100% highly conserved amino acid positions when compared to the *E. coli* aspartate 1-decarboxylase amino acid sequence (SEQ ID NO:7).

Nucleic acid molecules that may be used in the present strains include those encoding any protein having aspartate 1-decarboxylase activity, including for example: 1) those belonging to EC 4.1.1.11; 2); 2) those with 95% sequence identity to the *E. coli* aspartate 1-decarboxylase (SEQ ID NO:7); 3) those having conserved structure of Figure 2A, represented as a sequence in Figure 2B (SEQ ID NO:18); 4) those having eight or more of the conserved amino acids highlighted in SEQ ID NO:7 in Figure 2C; 5) those having the five conserved amino acid positions K at position 9, Y at position 22, G at position 24, T at position 57, and Y at position 58, as compared to the *E. coli* aspartate 1-decarboxylase of SEQ ID NO:7; and 6) those with at least about 95% sequence identity to any of the 493 proteins listed in Appendix 2.

DNA sequences encoding polypeptides with 2-dehydropantoate reductase activity or aspartate 1-decarboxylase activity may also be identified using bioinformatics and/or experimental methods. Coding sequences can be found in databases including NCBI (*ibid.*) using gene

name and/or enzyme name as is well know to one of skill in the art. Genes encoding 2-dehydropantoate reductase have multiple names, including for example panE or ApbA in *E. coli*, and PAN5 in *Saccharomyces cerevisiae.* In addition, nucleic acid sequences encoding these polypeptides can be readily found based on a known sequence using bioinformatics, including sequence analysis software such as BLAST sequence analysis using for example the *E coli* sequences (2-dehydropantoate reductase: SEQ ID NO:3; aspartate 1-decarboxylase: SEQ ID NO:6). Experimental methods include those based on nucleic acid hybridization.

Nucleic acid molecules encoding 2-dehydropantoate reductase and aspartate 1-decarboxylase are found in numerous organisms including, for example, in some bacteria (excluding *Zymomonas*), yeast, and plants. A coding region sequence from one of these sources, which is heterologous to *Zymomonas*, may be used directly or it may be optimized for expression in *Zymomonas*. For example, it may be codon optimized for optimal protein expression in *Zymomonas*, and/or introns may be removed if present in a eukaryotic coding region, both of which are well known to one skilled in the art.

Expression of enzyme activities

For expression, a nucleic acid molecule encoding a polypeptide having 2-dehydropantoate reductase activity and a nucleic acid molecule encoding aspartate 1-decarboxylase are each constructed in a chimeric gene with operably linked promoter and typically a termination sequence. Alternatively the coding regions are constructed as part of an operon that is operably linked to a promoter and a termination sequence. In an operon, typically a ribosome binding site is located upstream of the start codons for all open reading frames in the operon. Promoters that may be used in chimeric genes and operons are promoters that are expressed in *Zymomonas* cells such as the promoters of *Z. mobilis* glyceraldehyde-3-phosphate dehydrogenase gene (GAP promoter), *Z. mobilis* enolase gene (ENO promoter), and the *Actinoplanes missouriensis* xylose isomerase gene (GI promoter). Termination signals are also those that are expressed in the target cell.

Chimeric genes or an operon for 2-dehydropantoate reductase and aspartate 1-decarboxylase expression are typically constructed in or transferred to a vector for further manipulations. Vectors are well known in the art. Particularly useful for expression in *Zymomonas* are vectors that can replicate in both *E. coli* and *Zymomonas*, such as pZB188 which is described in U.S. Pat. No. 5,514,583. Vectors may include plasmids for autonomous replication in a cell, and plasmids for carrying constructs to be integrated into bacterial genomes. Plasmids for DNA integration may include transposons, regions of nucleic acid sequence homologous to the target bacterial genome, or other sequences supporting integration. An additional type of vector may be a transposome produced using, for example, a system that is commercially available from EPICENTRE®. It is well known how to choose an appropriate vector for the desired target host and the desired function.

Vectors carrying the desired coding regions are introduced into *Zymomonas* cells using known methods such as electroporation, freeze-thaw transformation, calcium-mediated transformation, or conjugation. The coding regions may be maintained on a plasmid in the cell, or integrated into the genome. Integration methods may be used that are well known in the art such as homologous recombination, transposon insertion, or transposome insertion. In homologous recombination, DNA sequences flanking a target integration site are placed bounding a spectinomycin-resistance gene, or other selectable marker, and the chimeric genes or operon for expression, leading to insertion of the selectable marker and the expression sequences into the target genomic site. In addition, the selectable marker may be bounded by site-specific recombination sites, so that after expression of the corresponding site-specific recombinase, the resistance gene is excised from the genome.

Transformed *Zymomonas* strains expressing 2-dehydropantoate reductase and aspartate 1-decarboxylase may be readily identified by their ability to grow in medium lacking PA. A wild type strain of *Zymomonas mobilis* engineered as described in examples herein was able to grow in minimal medium lacking PA as well as the wild type strain grew with

supplementation of 2.5 mg/L PA (i.e. a saturating concentration of this vitamin). In a strain of *Zymomonas mobilis* previously engineered for expression of xylose utilization enzyme activities and adapted to growth on xylose (US 7,629,156), and engineered for improved ethanol production through disruption of the endogenous glucose-fructose oxidoreductase (US 7,741,119) gene, expressing 2-dehydropantoate reductase and aspartate 1-decarboxylase also conferred the ability to grow in minimal medium lacking PA. However supplementation with p-aminobenzoic acid (PABA) was required for growth of this strain due to disruption of the *pabB* gene encoding p-aminobenzoate synthase subunit I, which occurred in previous engineering steps.

Ethanol production by pantothenic acid producing strain

The present engineered *Zymomonas* strain expressing 2-dehydropantoate reductase and aspartate 1-decarboxylase may be used as a biocatalyst in fermentation to produce ethanol. The *Zymomonas* strain is brought in contact with medium containing a carbon substrate. Typically one or more sugars provide the carbon substate. In one embodiment the medium may be a minimal medium with no addition of a complex ingredient that contains PA such as yeast extract, or PA itself, such that the medium lacks PA. Alternatively, the medium may contain an amount of PA that is suboptimal for growth and/or production of *Zymomonas* strains not engineered for pantothenic acid production. In one embodiment a seed culture is grown in minimal medium lacking PA or in medium having a sub-optimal amount of PA. The seed culture is then used to inoculate a larger fermentation culture. The fermentation medium may lack PA or have a sub-optimal amount of PA. Alternatively, the fermentation medium may contain an adequate amount of PA for growth and/or production of *Zymomonas* strains not engineered for pantothenic acid production.

Seed culture medium and/or fermentation medium may contain biomass hydrolysate which provides mixed sugars as a carbon source, typically including glucose, xylose, and arabinose. It is desirable that the present engineered *Zymomonas* strain also expresses enzyme activities

for utilization of xylose, or of xylose and arabinose. When the mixed sugars concentration is high such that growth is inhibited, the medium may include sorbitol, mannitol, or a mixture thereof as disclosed in US 7,629,156. Galactitol or ribitol may replace or be combined with sorbitol or mannitol. The present *Zymomonas* strain grows in the medium where fermentation occurs and ethanol is produced. The fermentation is run without supplemented air, oxygen, or other gases (which may include conditions such as anaerobic, microaerobic, or microaerophilic fermentation), for at least about 24 hours, and may be run for 30 or more hours. The timing to reach maximal ethanol production is variable, depending on the fermentation conditions. Typically, if inhibitors are present in the medium, as may be present in hydrolysate medium, a longer fermentation period is required. The fermentations may be run at temperatures that are between about 30 ^{o} C and about 37 ^{o} C, at a pH of about 4.5 to about 7.5.

The present *Zymomonas* strains may be grown in medium without PA supplementation in laboratory scale fermenters, and in scaled up fermentation where commercial quantities of ethanol are produced. Where commercial production of ethanol is desired, a variety of culture methodologies may be applied. For example, large-scale production from the present *Zymomonas* strains may be produced by both batch and continuous culture methodologies. A classical batch culturing method is a closed system where the composition of the medium is set at the beginning of the culture and not subjected to artificial alterations during the culturing process. Thus, at the beginning of the culturing process the medium is inoculated with the desired organism and growth or metabolic activity is permitted to occur adding nothing to the system. Typically, however, a “batch” culture is batch with respect to the addition of carbon source and attempts are often made at controlling factors such as pH and oxygen concentration.

A variation on the standard batch system is the Fed-Batch system. Fed-Batch culture processes are also suitable for growth of the present *Zymomonas* strains and comprise a typical batch system with the

exception that the substrate is added in increments as the culture progresses. Batch and Fed-Batch culturing methods are common and well known in the art and examples may be found in Biotechnology: A Textbook of Industrial Microbiology, Crueger, Crueger, and Brock, Second Edition (1989) Sinauer Associates, Inc., Sunderland, MA, or Deshpande, Mukund V., *Appl. Biochem. Biotechnol.*, 36, 227, (1992), herein incorporated by reference.

Commercial production of ethanol may also be accomplished with a continuous culture. Continuous cultures are open systems where a culture medium is added continuously to a bioreactor and an equal amount of conditioned medium is removed simultaneously for processing. Continuous cultures generally maintain the cells at a constant high liquid phase density where cells are primarily in log phase growth.

Continuous or semi-continuous culture allows for the modulation of one factor or any number of factors that affect cell growth or end product concentration. For example, one method will maintain a limiting nutrient such as the carbon source or nitrogen level at a fixed rate and allow all other parameters to moderate. In other systems a number of factors affecting growth can be altered continuously while the cell concentration, measured by medium turbidity, is kept constant. Continuous systems strive to maintain steady state growth conditions and thus the cell loss due to medium being drawn off must be balanced against the cell growth rate in the culture. Methods of modulating nutrients and growth factors for continuous culture processes as well as techniques for maximizing the rate of product formation are well known in the art of industrial microbiology and a variety of methods are detailed by Brock, *supra*.

Ethanol may be produce in simultaneous saccharification and fermentation (SSF) where pretreated biomass is saccharified producing hydrolysate containing fermentable sugars concurrently with ethanol production by the present *Zymomonas* strain.

In one embodiment the present *Zymomonas* strain is grown in shake flasks in minimal medium lacking PA at about 30 ^{o}C to about 37 ^{o}C with shaking at about 150 rpm in orbital shakers and then transferred to a

10 L seed fermentor containing a similar medium. The seed culture is grown in the seed fermentor anaerobically until OD_{600} is between 3 and 6, when it is transferred to the production fermentor where the fermentation parameters are optimized for ethanol production. Typical inoculum volumes transferred from the seed tank to the production tank range from about 2% to about 20% v/v. Typical fermentation medium contains biomass hydrolysate. A final concentration of about 5 mM sorbitol or mannitol is present in the medium. The fermentation is controlled at pH 5.0 – 6.0 using caustic solution (such as ammonium hydroxide, potassium hydroxide, or sodium hydroxide) and either sulfuric or phosphoric acid. The temperature of the fermentor is controlled at 30 ^{o}C - 35 ^{o}C. In order to minimize foaming, antifoam agents (any class- silicone based, organic based etc) are added to the vessel as needed. An antimicrobial, to which the present *Zymomonas* strain has tolernace, may be used optionally to minimize contamination.

Any set of conditions described above, and additionally variations in these conditions that are well known in the art, are suitable conditions for production of ethanol by a pantothenic acid producing *Zymomonas* strain.

EXAMPLES

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

GENERAL METHODS

The meaning of abbreviations is as follows: “kb” means kilobase(s), “bp” means base pairs, “nts” means nucleotides, “hr” means hour(s), “min” means minute(s), “sec” means second(s), “d” means day(s), “L” means liter(s), “ml” means milliliter(s), “μL” means microliter(s), “μg”

means microgram(s), “ng” means nanogram(s), "g" means gram(s), “mM” means millimolar, “μM” means micromolar, “nm” means nanometer(s), “μmol” means micromole(s), “pmol” means picomole(s), "OD" or "OD600" means optical density at 600 nm, "rpm" is revolutions per minute, "~ " means approximately.

Shake flask experiments with minimal media

Unless otherwise noted, all experiments described below were conducted in shake flasks (15-ml loosely-capped, conical shaped test tubes) using PA-depleted cells and a synthetic growth medium, MM-G5, that does not contain pantothenic acid. MM-G5 is a modified version of a minimal medium that is described in Goodman et al. ((1982) Applied and Environmental Microbiology 44:496-498). It contains 50 g/L glucose, 2 g/L KH_2PO_4, 1 g/L $MgSO_4$ (7H20) 2.5 g/L $(NH_4)_2SO_4$, 0.5 g/L NaCl, 50 mg/L $CaCl_2$ ($2H_2O$), 1 mg/L Na_2MoO_4 ($2H_20$), 5 mg/L $FeSO_4$ ($7H_2O$), and 1 mg/L each of pyridoxine, nicotinic acid, biotin and thiamine; with the final pH brought to 5.9 with KOH, and the solution was filtered through a 0.2 μm membrane. It is also important to adjust the pH to ~5.9 with KOH after dissolving the first five ingredients in close to the final volume of deionized water to avoid precipitation of the other components. To deplete intracellular and carryover PA, cells from agar plates or glycerol stocks were inoculated into MM-G5 medium to an OD of 0.1-0.3 and the cultures were incubated at 33 °C (~150 rpm) until the cells stopped growing (~14-20 hrs). However, depending on the history of the cells, the volume and density of the initial inoculum and the extent of growth that occurred during the incubation period, complete depletion of PA may require a second growth period in fresh MM-G5 medium and/or a longer incubation period. Unless stated otherwise, spectinomycin (200 μg/ml) was included in the growth media for all experiments that were performed with the plasmid-bearing strains ZW1/PanED#1 and 801/PanED#1.

Example 1

Construction of the synthetic GipanEpanD operon

To complete a putative pathway for pantothenic acid (PA)

biosynthesis in *Z. mobilis* we designed a synthetic 1620 bp DNA fragment (SEQ ID NO:1) that codes for an artificial, chimeric *E. coli* panE and panD operon (referred to below as either the "GipanEpanD operon" or the "GI-PanED operon"). The 5' end of the operon contains the *A. missouriensis* (ATCC 14538) GI promoter and there is a stretch of DNA at the 3' end that corresponds to the small, stabilizing stem-loop sequence that immediately follows the xylose isomerase (xylA) stop codon in the *E. coli* XylA/B operon. The synthetic DNA fragment also has NcoI and SpeI sites at its 5' end and NotI and EcoRI sites at its 3' end that can be used for cloning purposes. With reference to the DNA sequence of SEQ ID NO:1, nts 23-209 (SEQ ID NO:2) correspond to the GI promoter; nts 210-1121 (SEQ ID NO:3) correspond to the *E. coli* panE open reading frame (GenBank accession number AAC73528) that codes for 2-dehydropantoate reductase; nts 1122-1139 (SEQ ID NO:5) correspond to a stretch of DNA that is upstream from the start codon for the *Z. mobilis* glyceraldehyde 3-phosphate dehydrogenase gene that includes the Shine-Delgarno sequence; nts 1140-1520 (SEQ ID NO:6) correspond to the *E. coli* panD open reading frame (GenBank accession no. AAC73242) that codes for aspartate 1-decarboxylase; and nts 1543-1577 (SEQ ID NO:8) correspond to the stabilizing xylA stem-loop structure described above. The GI-PanED operon DNA fragment was synthesized by Genescript (Piscataway, NJ).

Example 2

Construction of the shuttle vector used for GI-PanED operon expression in *Z. mobilis*, and generation of *PanED strains*

To introduce the GI-PanED operon into *Z. mobilis*, the synthetic DNA molecule described above was digested with NcoI and NotI, and the resulting fragment was ligated into the unique NcoI and NotI sites of the plasmid shuttle vector pZB188/aadA. As described in US 2009-0246876 A1, which is herein incorporated by reference, pZB188/aadA is vector pZB188 described in US 5,514,583, which is herein incorporated by reference, which is able to replicate in *Z. mobilis* and *E. coli since* it has

origins of replication for both bacterial species, with an added spectinomycin resistance DNA fragment. To generate non-methylated plasmid DNA for transformation of *Z. mobilis*, pZB188/aadA-GlpanEpanD was introduced into chemically competent *E. coli* SCS110 cells (Stratagene, San Diego, CA), and transformants were selected on LB medium that contained spectinomycin (100 μg /ml). Isolated non-methylated plasmid DNA was then electroporated into ZW1 (ATCC #31821) and ZW801-4. A detailed description of the construction of the xylose-utilizing recombinant strain, ZW801-4, starting from the wild type parent strain, ZW1, is provided in US 7,741,084, which is herein incorporated by reference. Strain ZW801-4 was derived from strain ZW800, which was derived from strain ZW658, all as described in US 7,741,084. ZW658 was constructed by integrating two operons, P_{gap}*xylAB* and P_{gap}*taltkt*, containing four xylose-utilizing genes encoding xylose isomerase (xylA), xylulokinase (xylB), transaldolase (tal), and transketolase (tkt), into the genome of ZW1 (rename of strain ZM4; ATCC #31821) via sequential transposition events to produce strain X13L3, which was renamed ZW641, and followed by adaptation on selective media containing xylose. ZW658 was deposited under the Budapest Treaty as ATCC #PTA-7858. In ZW658, the gene encoding glucose-fructose oxidoreductase was insertionally-inactivated using host-mediated, double-crossover, homologous recombination and spectinomycin resistance as a selectable marker to create strain ZW800. The spectinomycin resistance marker, which was bounded by loxP sites, was removed by site specific recombination using Cre recombinase to create strain ZW801-4. As disclosed in commonly owned and co-pending US Patent Application Publication #US 20090246846, which is herein incorporated by reference, ZW648 has much more xylose isomerase activity (about 4-fold higher) than ZW641 (represented by X13bC strain) due to a point mutation in the promoter (Pgap) that drives expression of the xylA coding region.

Transformants were selected on agar plates that contained mRM3-G5 media (50 g/L glucose, 10 g/L yeast extract (contains PA), 2 g/L

KH_2PO_4, 1 g/L $MgSO_4$) and 200 μg/ml of spectinomycin. The resulting ZW1 and ZW801-4 strains that harbor the pZB188/aadA-GlpanEpanD shuttle vector were named ZW1/PanED #1 and 801/PanED #1, respectively. It should be noted that two primary transformants for each strain were evaluated in the shake flask experiments described below. Since both transformants behaved essentially the same in both cases, only the results that were obtained with ZW1/PanED#1 and 801/PanED#1 are presented below.

Example 3

Growth of ZW1 in MM-G5 medium requires PA supplementation

The ZW1 strain from an mRM3-G5 plate that contained 50 g/L glucose, 10 g/L yeast extract, 2 g/L KH_2PO4, 1 g/L $MgSO_2$ and 1.5% agar was inoculated into 20 ml of MM-G5 (described in General Methods) and the culture was incubated for ~19 hours at 33 ^oC (150 rpm) to deplete carryover pantothenic acid. The OD600 increased from 0.178 to 0.408 during the incubation period. An aliquot of the PA-depleted cells was then diluted with fresh MM-G5 medium to an OD600 of 0.035, and 10-ml aliquots of the resulting culture were distributed to eight 15-ml conical test tubes that contained various amounts of pantothenic acid (0, 0.025, 0.063, 0.125, 0.25, 0.63, 2.5, or 5 mg/L, final concentrations). After this step the eight cultures were incubated at 33 ^oC at 150 rpm, and growth was monitored by following changes in optical density (OD) at 600 nm as a function of time. As shown in Figure 4, both the exponential growth rate and maximum cell density increased in a dose related manner with increasing concentrations of pantothenic acid until saturation was achieved and growth was no longer limited by this vitamin. It was clear from this experiment that a concentration of ~2.5 mg/L PA or higher is able to support maximum growth of ZW1 in MM-G5 medium under the conditions employed. Growth was not observed unless the cells were supplemented with pantothenic acid.

Example 4

Effect of the synthetic GI-PanED operon in ZW1

Growth curves for ZW1 and ZW1/PanED #1 in the presence and absence of supplemented pantothenic acid were assayed. The protocol for this experiment was as follows. Two 10-ml MM-G5 cultures were started for each of strains ZW1 and ZW1/PanED #1. One was supplemented with PA (2.5 mg/L) while the other received an equivalent volume of sterile water. The initial ODs for all four cultures were ~0.1. After a 15-hr incubation period at 33 °C (150 rpm), aliquots of these cultures were used to start new 10-ml cultures that contained the same growth media as the original cultures; the initial OD was ~0.05 in all cases. The new cultures were incubated at 33 °C (150 rpm) and growth was monitored by OD600. The resulting exponential growth curves are shown in Figure 5.

Consistent with previous results, when ZW1 was depleted of pantothenic acid in MM-G5 medium and then transferred to fresh medium that had the same composition, it failed to grow (Fig. 5A). In contrast, when ZW1 was transferred to medium that contained 2.5 mg/L of pantothenic acid, the cells grew exponentially with a doubling time of about 2 hours to a final OD of ~1.2 (Fig. 5A). Very different results were obtained with the ZW1 derivative that contains the synthetic GI-PanED operon. As shown in Figure 5B, ZW1/PanED#1 (ZED#1) grew with the same kinetics in MM-G5 medium in the presence or absence of supplemented PA. Indeed, both growth curves for this strain were virtually identical to the growth curve for ZW1 when a saturating concentration of PA was present. These results clearly demonstrate that with introduction of panE and panD genes *Zymomonas* was able to synthesize PA, and PA was made in an amount sufficient to support the maximum growth rate in minimal medium lacking PA.

Example 5

Effect of the synthetic GI-PanED operon in ZW801-4

ZW801-4 has two vitamin requirements for growth in minimal

medium. Growth experiments using ZW801-4 were carried out as described in General Methods and the resulting growth curves are shown in Figure 6. As seen in Figure 6A, when ZW801-4 was depleted of pantothenic acid in MM-G5 and transferred to the same medium it failed to grow, similar to the results that were obtained with ZW1. However, very little growth was also observed when pantothenic acid (2.5 mg/L) was added to the growth medium during the second incubation period (Figure 6A). There is a genetic basis for this observation that is related to strain construction. Like other ZW641 derivatives, ZW801-4 *cannot* synthesize p-aminobenzoic acid (PABA), which is a vitamin that is required for folic acid biosynthesis and hence is essential. As described in US 7,741,084, the first step in the construction of ZW641 was the integration of a synthetic P_{gap}*taltkt* operon (encoding *E. coli* transaldolase and transketolase under the control of the *Z. mobilis* P_{gap} promoter) into the ZW1 chromosome. The operon was introduced by a transposon that randomly integrates into DNA, and the transposon insertion site for the strain that was selected for further metabolic engineering is in the open reading frame of the *Z. mobilis* *pab*B gene that codes for p-aminobenzoate synthase, subunit I, which is required for biosynthesis of PABA. The P_{gap}*taltkt* transposon insert is located between nts 102021 and 102022 of GenBank accession number AE008692, as determined by whole genome DNA sequence analysis. Since the disrupted *pab*B gene does not appear to be functional, ZW641 and all strains that were derived from it require two vitamins for growth in minimal media, namely PA and PABA.

A titration experiment was conducted with ZW801-4 to determine the optimal concentration of PA for growth of in MM-G5 medium that contains a saturating concentration of PABA (15 mg/L; Figure 6B). Strain ZW801-4 cells were inoculated into 20 ml of MM-G5 medium and the culture was incubated for ~19 hours at 33 ^{o}C (150 rpm) to deplete intracellular and carryover pantothenic acid. During the incubation period the OD increased from 0.143 to 0.364. The PA-depleted culture was then diluted with MM-G5 medium to an OD of ~0.035 and PABA was added to a final concentration of 15 mg/L. Aliquots (10 ml) of the cell suspension

were distributed to eight 15-ml conical tubes that contained various concentrations of PA (ranging from 0-5 mg/L), and the resulting cultures were incubated at 33 °C (150 rpm) to monitor growth. As shown in Figure 6B, the optimal concentration of PA for growth of ZW801-4 in MM-G5 medium that contains a saturating concentration of PABA was ~2.5 mg/L, similar to the requirement for ZW1 (Figure 4).

GI-PanED operon in ZW801-4

Each of strains ZW801-4 and ZW801-4/PanED#1 was inoculated into 10 ml of MM-G5 and the cultures were incubated at 33 °C for 15 hrs to deplete pantothenic acid and partially deplete PABA; the initial OD was ~0.1 in both cases. Following this step the cultures were diluted with the same growth medium to an OD of ~0.04, and quadruplicate 10-ml aliquots of each cell suspension were distributed to eight 15-ml conical tubes. The tubes were then supplemented with PA, PABA both vitamins, or neither, and the resulting cultures were incubated at 33 °C to monitor growth at 600 nm. The final concentrations of PA and PABA when present were 2.5 mg/L and 15 mg/L, respectively, and the no vitamin control cultures received an equivalent volume of sterile water. The resulting growth curves are shown in Figure 7.

Consistent with previous results, ZW801-4 only grew when PA and PABA were both added to the growth medium (Figure 7A). The small amount of growth that occurred in the culture that was only supplemented with PA is the result of carryover PABA, since there is always a small amount of residual PABA after the PA-depletion step during the first incubation period in MM-G5 medium (i.e. the cells use up all the PA before they run out of PABA). In contrast to the above results, the ZW801-4 strain that contained the synthetic GI-PanED operon only required PABA for growth (Figure 7B)) since it was able to synthesize pantothenic acid. `

Note that the ZW801-4 culture that was supplemented with both vitamins grew slightly better in MM-G5 medium than the corresponding culture of ZW801-4/PanED#1 (Figure 7A versus 7B). The most likely explanation for this result is “plasmid burden”, which is often observed with *Z. mobilis* and other bacterial strains (Kim et al, (2000) Applied and

Environmental Microbiology 66:186-193 and references therein). This phenomenon, whereby energy that would otherwise be available for growth is diverted to plasmid replication and maintenance, would likely be far more pronounced in minimal medium compared to rich medium.

The key finding in this experiment is that ZW801-4/PanED#1 grew with the same kinetics and to the same cell density in the presence and absence of added pantothenic acid when the growth medium contained a saturating concentration of PABA. Taken together the above results clearly indicate that co-expression of the *E. coli* panE and panD coding regions in wild type and recombinant strains of *Z. mobilis* allowed growth under conditions where pantothenic acid was limiting.

CLAIMS

What is claimed is:

1. A recombinant bacterial strain of the genus *Zymomonas* comprising a heterologous nucleic acid molecule encoding a polypeptide having 2-dehydropantoate reductase activity and a heterologous nucleic acid molecule encoding a polypeptide having aspartate 1-decarboxylase activity.

2. A recombinant bacterial strain of claim 1 wherein the polypeptide having 2-dehydropantoate reductase activity is an enzyme belonging to the EC 1.1.1.169 class.

3. A recombinant bacterial strain of claim 2 wherein the polypeptide having 2-dehydropantoate reductase activity has at least ten conserved amino acid positions selected from the group consisting of G at position 7, G at position 9, G at position 12, K at position 72, N at position 98, G at position 99, K at position 176, N at position 180, N at position 184, E at position 210, S at position 244, D or S at position 248, and E at position 256, as compared to the *E. coli* 2-dehydropantoate reductase of SEQ ID NO:4.

4. A recombinant bacterial strain of claim 1 wherein the polypeptide having aspartate 1-decarboxylase activity is an enzyme belonging to the EC 4.1.1.11 class.

5. A recombinant bacterial strain of claim 4 wherein the polypeptide having aspartate 1-decarboxylase activity has at least eight conserved amino acid positions selected from the group consisting of K at position 9, H at position 11, Y at position 22, G at position 24, S at position 25, G at position 52, R at position 54, T at position 57, Y at position 58, N at position 72, G at position 73, and I at position 86, as compared to the *E. coli* aspartate 1-decarboxylase of SEQ ID NO:7.

6. A recombinant bacterial strain of claim 4 wherein the polypeptide having aspartate 1-decarboxylase activity has the five conserved amino acid positions K at position 9, Y at position 22, G at position 24, T at position 57, and Y at position 58, as compared to the *E. coli* aspartate 1-decarboxylase of SEQ ID NO:7.

7. A recombinant bacterial strain of claim 1 wherein the strain grows in media that lacks pantothenic acid.

8. The recombinant bacterial strain of Claim 1 wherein the strain produces ethanol.

9. The recombinant bacterial strain of Claim 8 wherein the strain comprises genetic modifications which enhance the production of ethanol.

10. A process for producing a *Zymomonas* strain that synthesizes pantothenic acid comprising:

a) providing a bacterial strain of the genus *Zymomonas*;

b) introducing a heterologous nucleic acid molecule encoding a polypeptide having 2-dehydropantoate reductase activity; and

c) introducing a heterologous nucleic acid molecule encoding a polypeptide having aspartate 1-decarboxylase activity;

wherein steps b) and c) may be in either order or simultaneous, and wherein 2-dehydropantoate reductase activity and aspartate 1-decarboxylase activity are both expressed in the strain produced by steps (a), (b), and (c).

11. A method for the producing ethanol comprising:

a) providing a recombinant bacterial strain of the genus *Zymomonas* comprising a heterologous nucleic acid molecule encoding a polypeptide having 2-dehydropantoate reductase activity and a heterologous nucleic acid molecule encoding a polypeptide having aspartate 1-decarboxylase activity; and

b) contacting the strain of (a) with fermentation medium under conditions whereby the strain produces ethanol.

12. A method according to claim 11 wherein the polypeptide having 2-dehydropantoate reductase activity is an enzyme belonging to the EC 1.1.1.169 class.

13. A method according to claim 12 wherein the polypeptide having 2-dehydropantoate reductase activity has at least ten conserved amino acid positions selected from the group consisting of G at position 7, G at position 9, G at position 12, K at position 72, N at position 98, G at position 99, K at position 176, N at position 180, N at position 184, E at position 210, S at position 244, D or S at position 248, and E at position 256, as compared to the *E. coli* 2-dehydropantoate reductase of SEQ ID NO:4.

14. A method according to claim 11 wherein the polypeptide having aspartate 1-decarboxylase activity is an enzyme belonging to the EC 4.1.1.11 class.

15. A method according to claim 14 wherein the polypeptide having aspartate 1-decarboxylase activity has at least eight conserved amino acid positions selected from the group consisting of K at position 9, H at position 11, Y at position 22, G at position 24, S at position 25, G at position 52, R at position 54, T at position 57, Y at position 58, N at position 72, G at position 73, and I at position 86, as compared to the *E. coli* aspartate 1-decarboxylase of SEQ ID NO:7.

16. A method according to claim 14 wherein the polypeptide having aspartate 1-decarboxylase activity has the five conserved amino acid positions K at position 9, Y at position 22, G at position 24, T at position 57, and Y at position 58, as compared to the *E. coli* aspartate 1-decarboxylase of SEQ ID NO:7.

17. A method according to claim 11 wherein prior to step (b) the bacterial strain of (a) is contacted with medium that lacks or has a sub-optimal amount of pantothenic acid wherein a seed culture is produced to inoculate the fermentation medium of (b).

18. A method according to claim 11 or 17 wherein the fermentation medium lacks or has a sub-optimal amount of pantothenic acid.

Figure 1

pyruvate
2.2.1.6
2-acetolactate
1.1.1.86
2,3-dihydrozy-3-methylbutanoate
4.2.1.9
3-methyl-2-oxobutanoate
panB 2.1.2.11
2-dehydropantoate
panE 1.1.1.169
pantoate
Alanine, aspartate and glutamate metabolism
panD
4.1.1.11
Beta-alanine
panC 6.3.2.1
pantothenate

Figure 2

A

--------------------------xxxxxxxxKxHxxxxxxxxxxYxG-Sxxxxxxxxxxx

xxxxxxxVxxxxxxxGxRxxTYxxxx---xxxxxxxxxxGxxxxxxxxxxxxxxxxxxxxxx--

xxxxxxxxxxxxxx-----------------xxxxxxxxxx--------------------xx

xxxx-------------------------

B

SEQ ID NO:18

xxxxxxxxKxHxxxxxxxxxxxYxGSxxxxxxxxxxxxxxxxxxxxxxxxxxxGxRxxTYxxxxxxxxxxxxxN
GxxxxxxxxxxxxIxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

B

SEQ ID NO:7

MIRTMLQG**KLH**RVKVTHADLII**Y**E**GS**CAIDQDFLDAAGILENEA
IDIWNVTN**GKR**FS**TY**AIAAERGSRIISV**NG**AAAHCASVGDIV**I**IA
SFVTMPDEEARTWRPNVAYFEGDNEMKRTAKAIPVQVA

A

-------xxxxxxGxGxxGxxxxxxLxxxx---xxxxxxxxx-----xxxxxxxxxxxxxxxxxxxxxx-----------------
xxxxxxxxKxxxxxxxxxxxxxxxxxxxxxxxxN---------------Gxxxxxxxxx--xxxx-------
xxxxxxxxxxxxxxxxxxxxx-xxxxxxxxxxx------------------xxxxxxxxxxx-------
xxxxxxxxxxxxxxKxxxNxxxNxxxxxxxxxxxxxx---xxxxxxxxxxxEXxxxxxxxxx---------------------------
xxxxxxxxxxxxxx-xxxxSxxxDxxxxx-xxExxxxxxxxxxxxxxxxxx---xxxxxxxxxxx-----------------

B

SEQ ID NO:19

xxxxxxGxGxxGxxxxxxLxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxKxxxxxxxxxxxxx
xxxxxxxxxxxNGxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
xxxxxKxxxNxxxNxxxxxxxxxxxxxxxxxxxxxxxxExxxxxxxxxxxxxxxxxxxxxxxxxxxxSxxxDxxx
xxxxExxxxxxxxxxxxxxxxxxxxxxxxxxxxx

C

SEQ ID NO:4

MKITVLGCGALGQLWLTALCKQGHEVQGWLRVPQPYCSVNLVETDGSIFNESLTANDPD
FLATSDLLLVTLKAWQVSDAVKSLASTLPVTTPILLIHNGMGTIEELQNIQQPLLMGTTTHA
ARRDGNVIIHVANGITHIGPARQQDGDYSYLADILQTVLPDVAWHNNIRAELWRKLAVNC
VINPLTAIWNCPNGELRHHPQEIMQICEEVAAVIEREGHHTSAEDLRDYVMQVIDATAENIS
SMLQDIRALRHTEIDYINGFLLRRARAHGIAVPENTRLFEMVKRKESEYERIGTGLPRPW

D

SEQ ID NO:4

MKITVLGCGALGQLWLTALCKQGHEVQGWLRVPQPYCSVNLVETDGSIFNESLTANDPDF
LATSDLLLVTLKAWQVSDAVKSLASTLPVTTPILLIHNGMGTIEELQNIQQPLLMGTTTHAA
RRDGNVIIHVANGITHIGPARQQDGDYSYLADILQTVLPDVAWHNNIRAELWRKLAVNCVI
NPLTAIWNCPNGELRHHPQEIMQICEEVAAVIEREGHHTSAEDLRDYVMQVIDATAENISS
MLQDIRALRHTEIDYINGFLLRRARAHGIAVPENTRLFEMVKRKESEYERIGTGLPPRPW

Figure 3

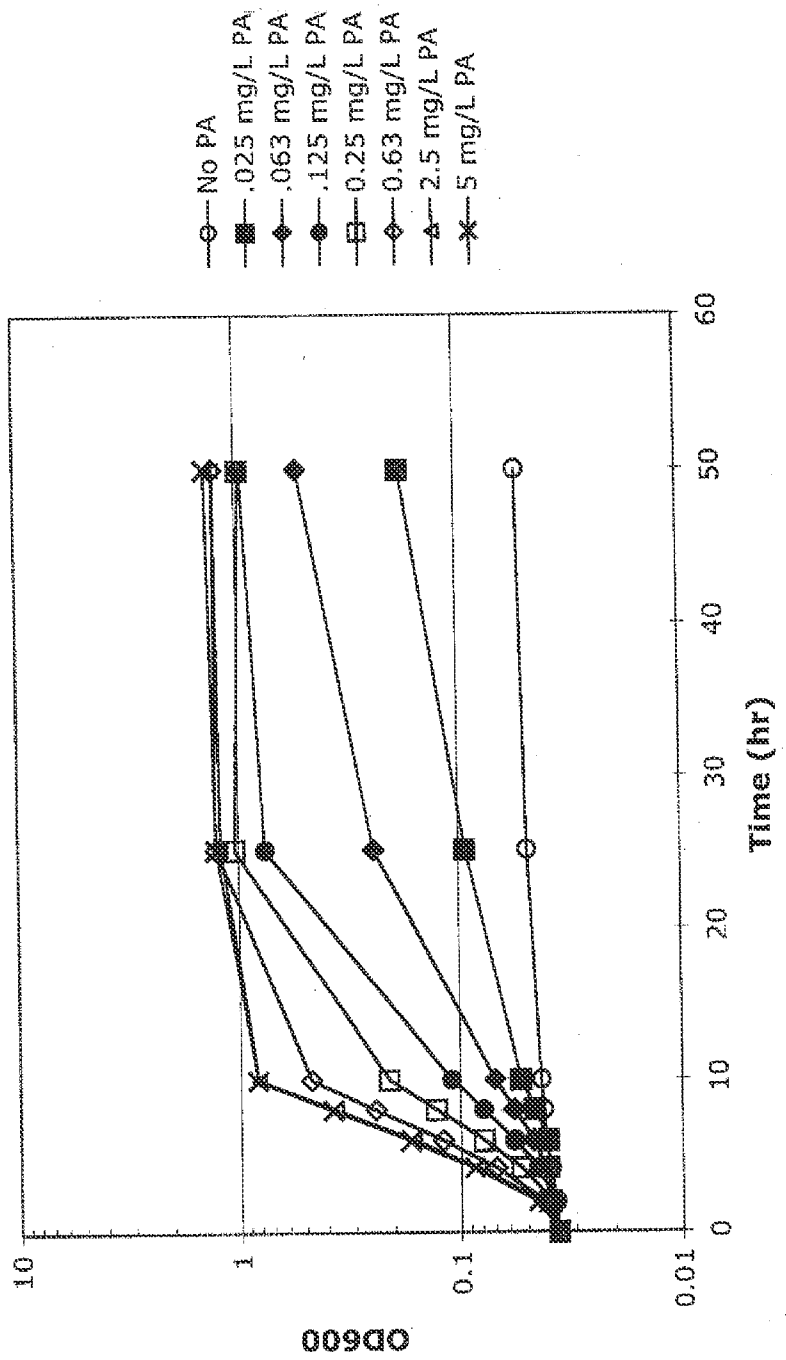

No PA
.025 mg/L PA
.063 mg/L PA
.125 mg/L PA
0.25 mg/L PA
0.63 mg/L PA
2.5 mg/L PA
5 mg/L PA
OD600
10
1
0.1
0.01
0
10
20
30
40
50
60
Time (hr)

FIG. 4

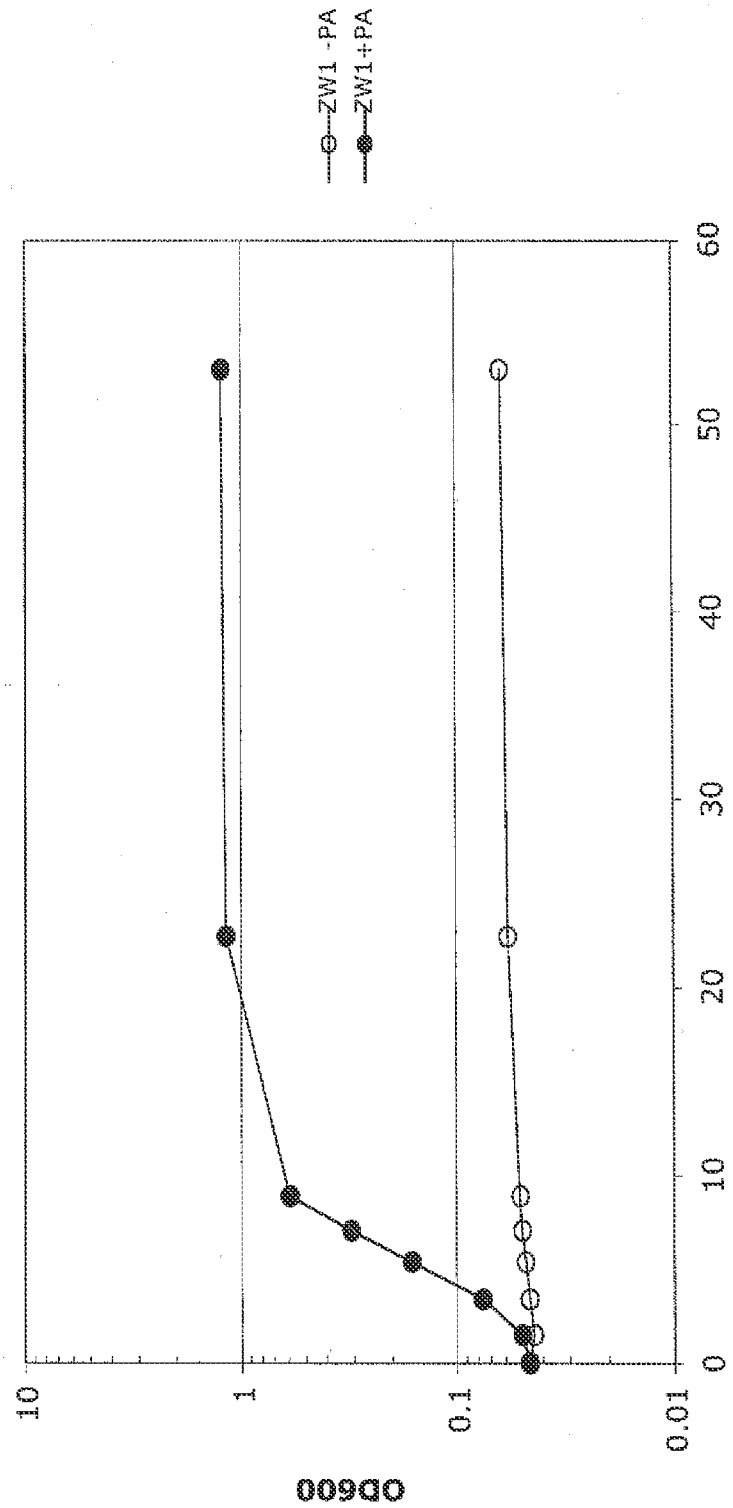


FIG. 5A

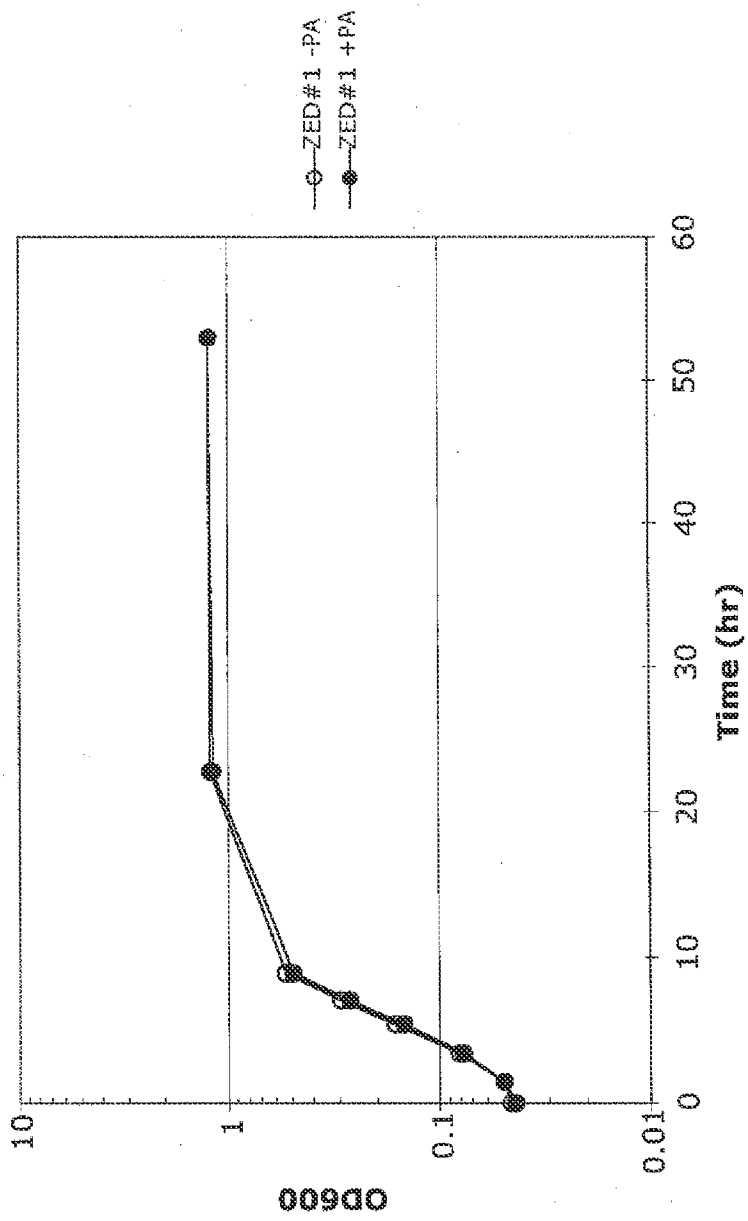


FIG. 5B

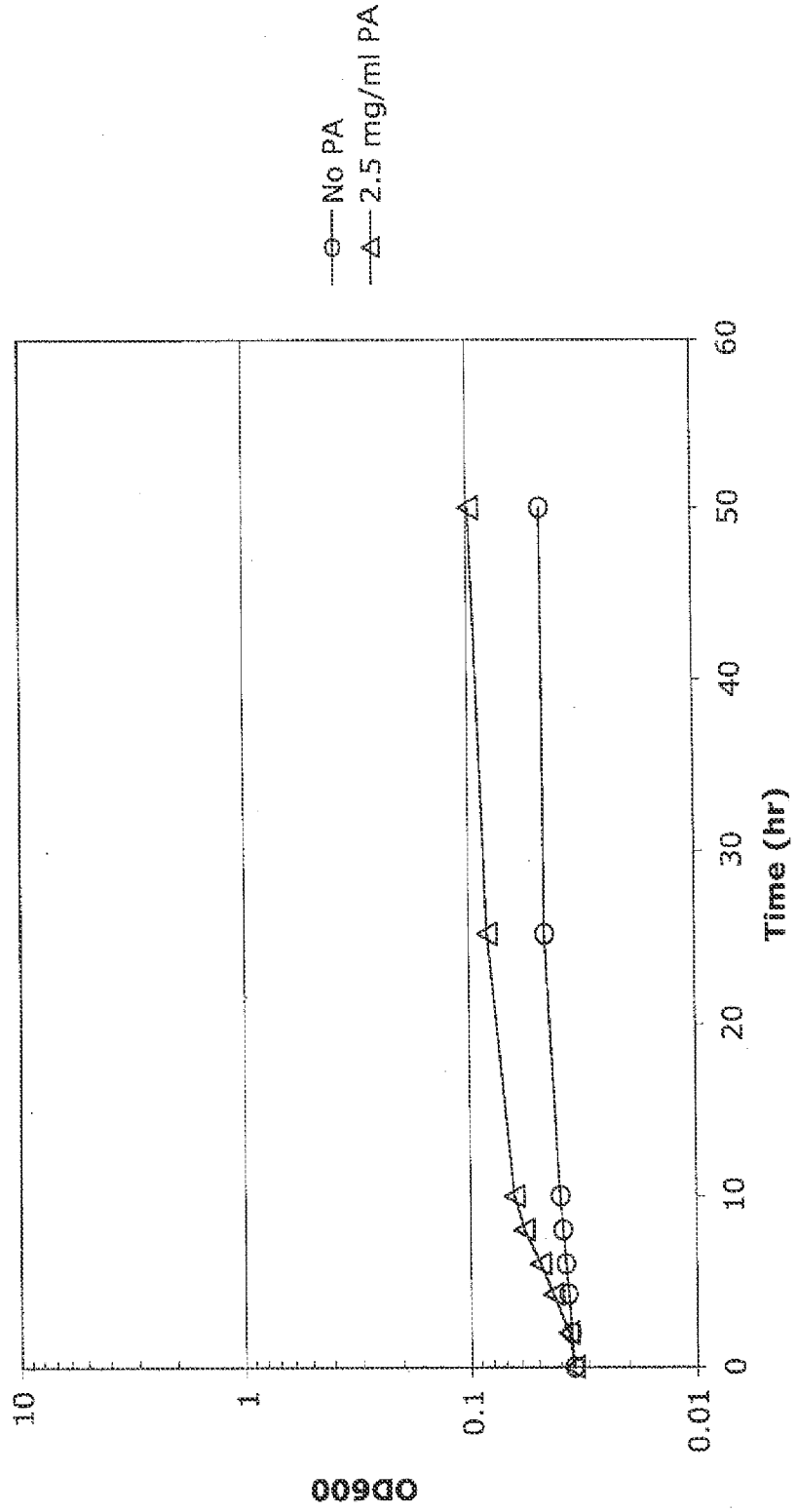


FIG. 6A

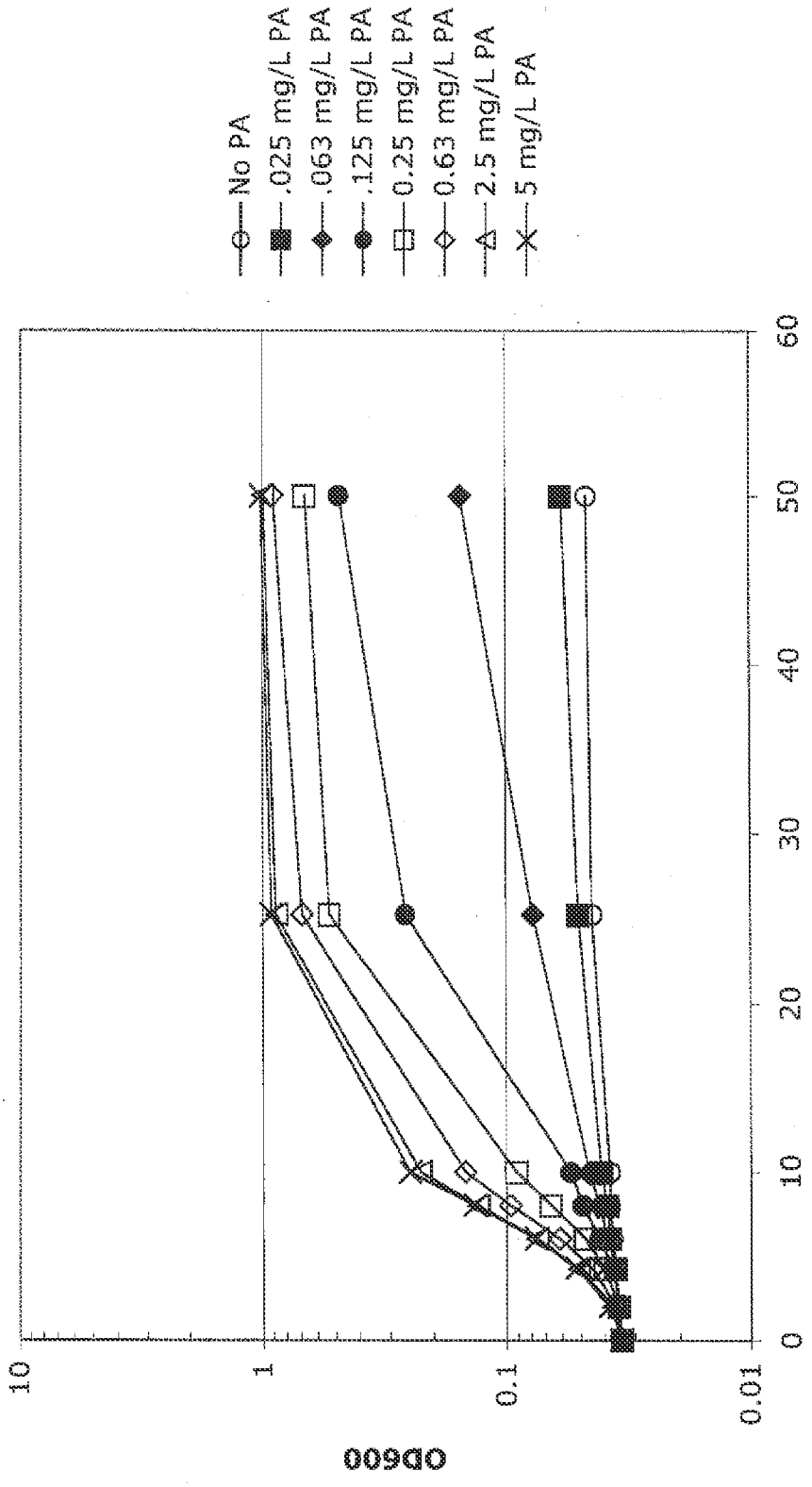

No PA
.025 mg/L PA
.063 mg/L PA
.125 mg/L PA
0.25 mg/L PA
0.63 mg/L PA
2.5 mg/L PA
5 mg/L PA
OD600
10
1
0.1
0.01
0
10
20
30
40
50
60
Time (hr)

FIG. 6B

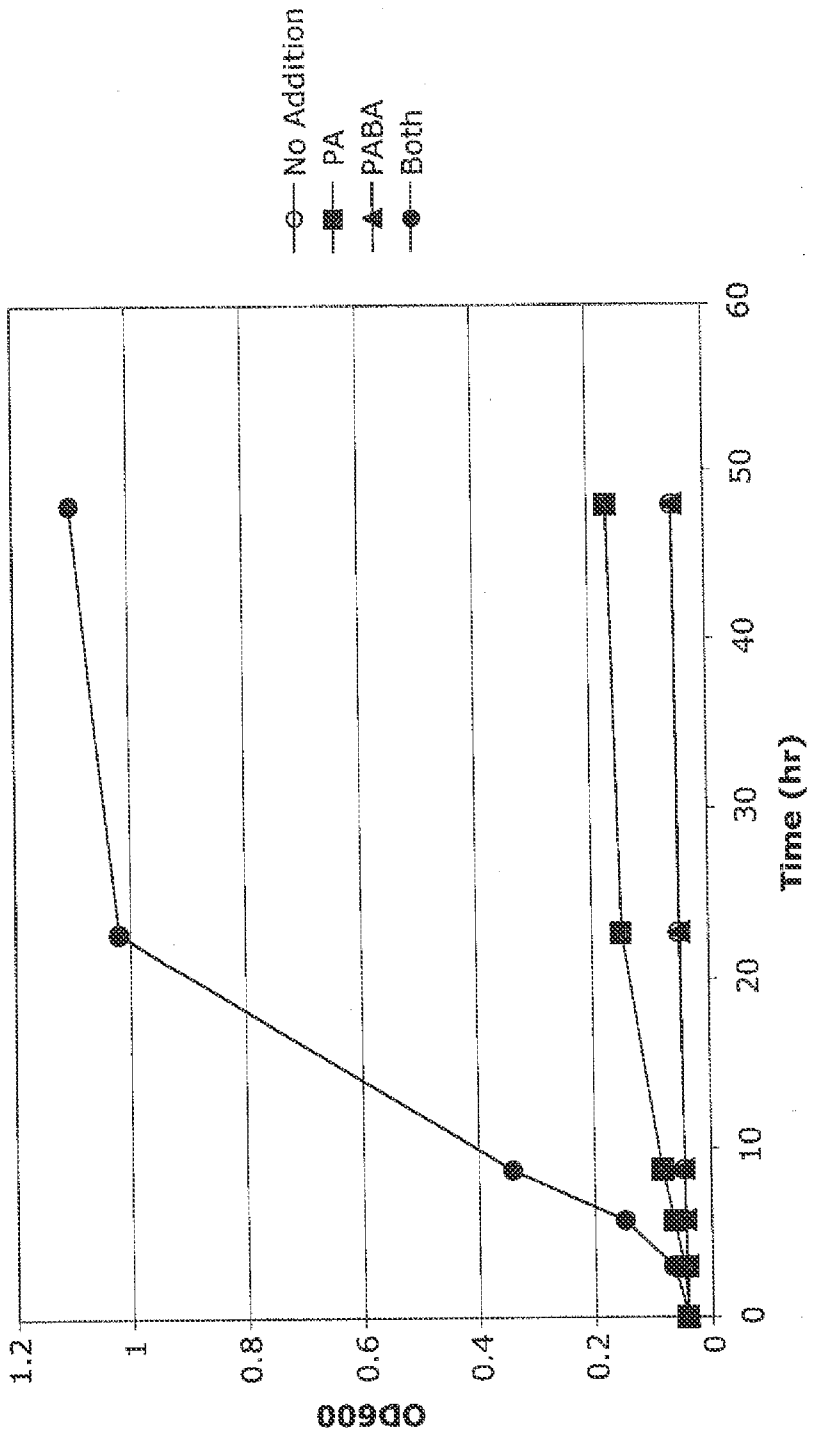

No Addition
PA
PABA
Both
OD600
1.2
1
0.8
0.6
0.4
0.2
0
0
10
20
30
40
50
60
Time (hr)

FIG. 7A

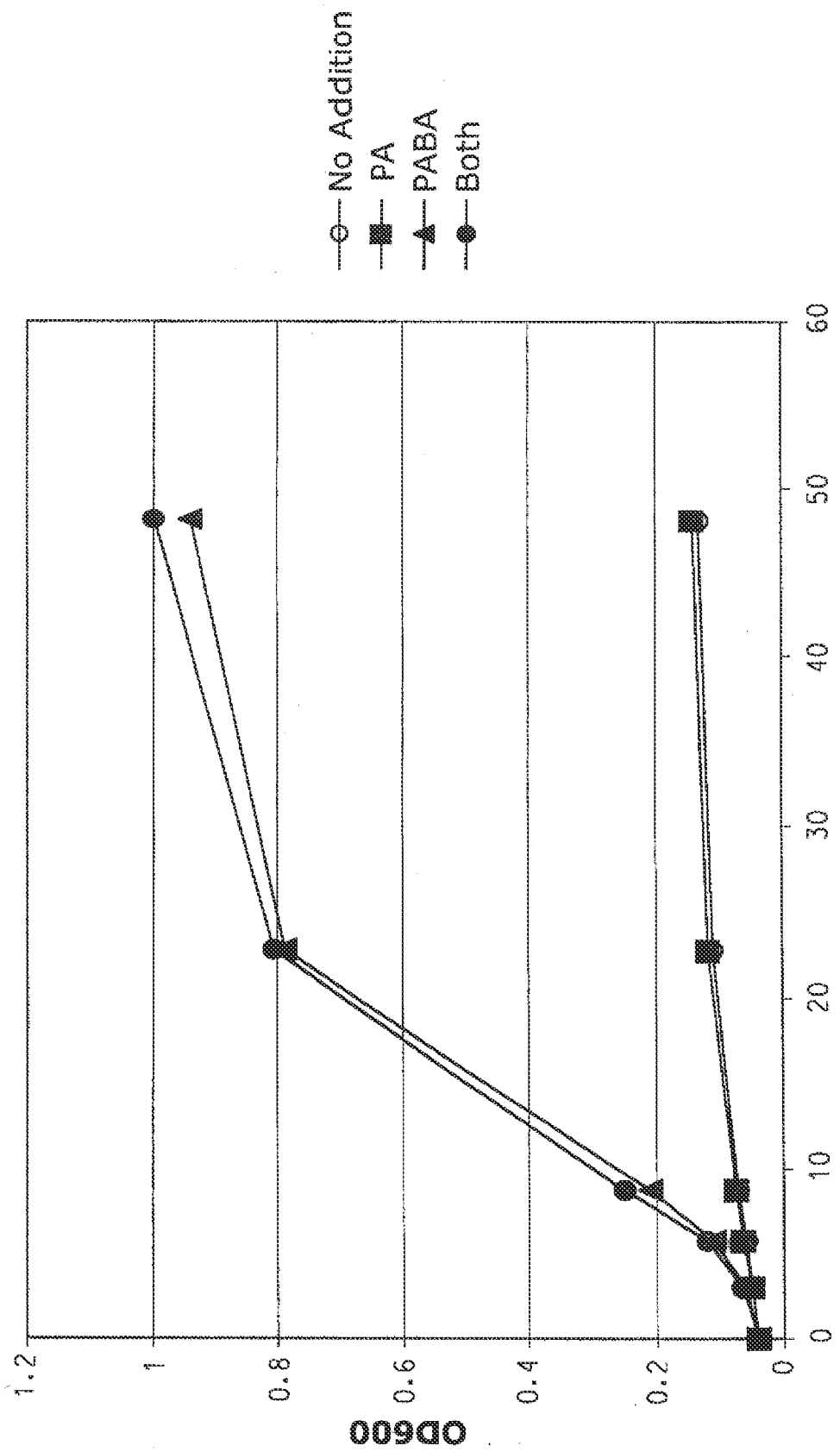

No Addition
PA
PABA
Both
OD600
1.2
1
0.8
0.6
0.4
0.2
0
0
10
20
30
40
50
60
Time (hr)

FIG. 7B

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2012/032558

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N9/04 C12N9/88 C12N1/19 C12P7/06 C12N9/02
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N C12P

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, EMBASE, WPI Data, Sequence Search

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 6 171 845 B1 (ELISCHWESKI FRANK [DE] ET AL) 9 January 2001 (2001-01-09) cited in the application the whole document	1-10
A	US 2005/089973 A1 (YOCUM R R [US] ET AL) 28 April 2005 (2005-04-28) cited in the application the whole document	1-10
A	WO 03/006664 A2 (DEGUSSA [DE]; KRUSE DANIELA [DE]; THIERBACH GEORG [DE]) 23 January 2003 (2003-01-23) cited in the application the whole document	1-10

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
26 June 2012	06/07/2012

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Sonnerat, Isabelle

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2012/032558

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	DATABASE UniProt [Online] 5 October 2010 (2010-10-05), Weinstock G. et al.: "RecName: Full=2-dehydropantoate 2-reductase; EC=1.1.1.169", XP002678555, Database accession no. D8BYP4 the whole document	1-10

A	WO 2010/085731 A2 (MICROBIA INC [US]; TATE & LYLE INGREDIENTS [US]; MADDEN KEVIN T [US];) 29 July 2010 (2010-07-29) abstract sequence 61 page 6, line 9 - line 28 page 9, line 7 - line 18 page 42, line 1 - line 10	1-10

A	WO 2009/058927 A1 (DU PONT [US]; VIITANEN PAUL V [US]; TAO LUAN [US]; KNOKE KYLE [US]; ZH) 7 May 2009 (2009-05-07) abstract	11-18

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2012/032558

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 6171845 B1	09-01-2001	AT 266098 T	15-05-2004
		AU 5357099 A	13-04-2000
		BR 9904449 A	14-11-2000
		CA 2284117 A1	09-04-2000
		CN 1254758 A	31-05-2000
		DE 19846499 A1	20-04-2000
		EP 1001027 A2	17-05-2000
		ES 2221280 T3	16-12-2004
		HU 9903458 A2	28-12-2000
		JP 2000116387 A	25-04-2000
		KR 20000028952 A	25-05-2000
		PL 335926 A1	10-04-2000
		SK 137799 A3	12-09-2000
		US 6171845 B1	09-01-2001
		ZA 9906367 A	13-04-2000
US 2005089973 A1	28-04-2005	NONE	
WO 03006664 A2	23-01-2003	AU 2002316940 A1	29-01-2003
		EP 1404840 A2	07-04-2004
		US 2004213874 A1	28-10-2004
		WO 03006664 A2	23-01-2003
WO 2010085731 A2	29-07-2010	US 2011014669 A1	20-01-2011
		WO 2010085731 A2	29-07-2010
WO 2009058927 A1	07-05-2009	AT 554175 T	15-05-2012
		AU 2008318709 A1	07-05-2009
		CA 2704532 A1	07-05-2009
		CN 101970671 A	09-02-2011
		EP 2209899 A1	28-07-2010
		JP 2011515066 A	19-05-2011
		US 2009221078 A1	03-09-2009
		WO 2009058927 A1	07-05-2009