



(51) International Patent Classification:

C12N 1/21 (2006.01) *C12P 5/00* (2006.01)
C12N 9/04 (2006.01) *C12P 7/16* (2006.01)
C12P 1/00 (2006.01) *C12P 7/24* (2006.01)

(21) International Application Number:

PCT/US2017/013211

(22) International Filing Date:

12 January 2017 (12.01.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/277,849 12 January 2016 (12.01.2016) US

(71) Applicant: THE REGENTS OF THE UNIVERSITY
OF CALIFORNIA [US/US]; 1111 Franklin Street, 12th
Floor, Oakland, CA 94607-5200 (US).(72) Inventors: LIAO, James, C.; 10573 Wellworth Avenue,
Los Angeles, CA 90024 (US). WU, Tung-Yun; 1363
Kelton Avenue, #3, Los Angeles, CA 90024 (US). CHEN,
Chang-Ting; 20854 Quail Run Drive, Diamond Bar, CA
91789 (US).(74) Agent: BAKER, Joseph, R., Jr.; Gavrilovich, Dodd &
Lindsey LLP, 4370 La Jolla Village Drive, Suite 303, San
Diego, CA 92122 (US).

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,
BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN,
KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA,
MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG,
NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS,
RU, RW, SA, 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, ST, SZ,
TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU,
TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE,
DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU,
LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,
SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: METHANOL DEHYDROGENASES

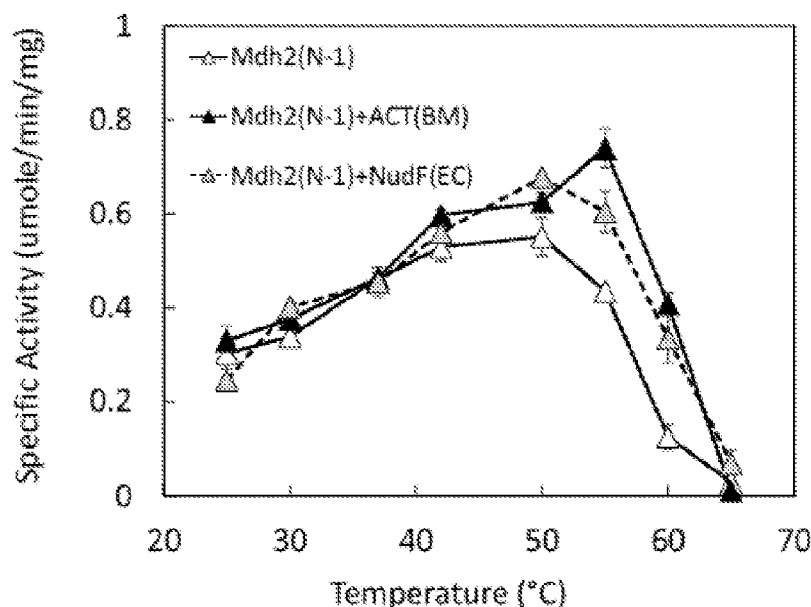


FIG. 3A

(57) Abstract: This disclosure relates to modified alcohol dehydrogenases that metabolize methanol. The disclosure further relates to cells expressing such modified dehydrogenases and methods of producing biochemicals from methanol.

METHANOL DEHYDROGENASES

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Application Serial No. 62/277,849, filed January 12, 2016, which disclosures are incorporated herein by reference.

STATEMENTS REGARDING FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with Government support under DE-AR0000430, awarded by the U.S. Department of Energy. The Government has certain rights in the invention.

TECHNICAL FIELD

[0003] This invention relates to modified enzymes and microorganisms engineered to express such enzymes.

SEQUENCE LISTING

[0004] A Sequence Listing associated with this application is filed in electronic format via EFS-Web and hereby incorporated by reference into the specification in its entirety. The name of the text file containing the Sequence Listing is Sequence_Listing_086W01.txt. The size of the text file is 38,397 bytes, and the text file was created on January 12, 2017.

BACKGROUND

[0005] Methanol utilization by methylotrophic or non-methylotrophic organisms is the first step toward methanol bioconversion to higher carbon-chain chemicals. Methanol oxidation using NAD-dependent methanol dehydrogenase (Mdh) is of particularly interest because it uses NAD⁺ as the electron carrier. Only a limited number of NAD-dependent Mdhs have been reported. The most studied is the *Bacillus methanolicus* Mdh, which exhibits low enzyme specificity to methanol and is dependent on an endogenous activator protein (ACT).

SUMMARY

[0006] The disclosure characterizes and engineers a group III NAD-dependent alcohol dehydrogenase (Mdh2) from *Cupriavidus necator* N-1. This enzyme is the first NAD-dependent Mdh characterized from a Gram-negative, mesophilic, non-methylotrophic organism with a significant activity towards methanol. Interestingly, unlike previously reported Mdh's, Mdh2 is insensitive to *B. methanolicus*

ACT and *Escherichia coli* Nudix hydrolase NudF under mesophilic conditions and exhibited higher or comparable activity and affinity toward methanol relative to the *B. methanolicus* Mdh with or without ACT in a wide range of temperatures. Using directed molecular evolution, variants of Mdh2 were developed that showed a higher K_{cat}/K_m for methanol and 10-fold lower K_{cat}/K_m for *n*-butanol. Thus, these variants represent an NAD-dependent Mdh with much improved catalytic efficiency and specificity toward methanol compared with the existing NAD-dependent Mdh's with or without ACT activation.

[0007] The disclosure provides a recombinant polypeptide comprising a sequence that is at least 70% to 99% identical to SEQ ID NO:2 and has improved methanol dehydrogenase activity compared to a polypeptide consisting of SEQ ID NO:2. In one embodiment, the polypeptide is engineered from *C. necator* N-1. In another embodiment, the polypeptide is at least 70% identical to SEQ ID NO:2 and has one or more mutations at a residue selected from the group consisting of A26, A31, A169 and any combination thereof. In a further embodiment, the one or more mutations independently comprise a substitution with V, I or C. In yet another embodiment, the one or more mutations comprise a substitution with V. In another embodiment, the polypeptide is at least 70% identical to SEQ ID NO:2 and has the mutations A26V, A31V and/or A169V. In yet another embodiment, the polypeptide comprises a catalytic efficiency (K_{cat}/K_m) of greater than 1.6 for methanol and/or less than 903 for *n*-butanol. In yet another embodiment of any of the foregoing, the polypeptide comprises from 1-10 conservative substitutions.

[0008] The disclosure also provides an isolated nucleic acid encoding a polypeptide as described in the preceding paragraph.

[0009] The disclosure also provides a vector comprising the nucleic acid molecule of the disclosure encoding a polypeptide as described above. In one embodiment, the vector is an expression vector. In one embodiment, the vector comprises a nucleic acid sequence encoding a polypeptide of SEQ ID NO:2.

[0010] In another embodiment, the disclosure provides a host cell transfected with an isolated nucleic acid of the disclosure

that encodes a polypeptide comprising at least 70% identity to SEQ ID NO:2.

[0011] In yet another embodiment, the disclosure provides a host cell transfected with a vector of the disclosure.

[0012] In either of the foregoing embodiments, the cell is prokaryotic or eukaryotic.

[0013] The disclosure also provides a recombinant host cell that has been genetically engineered to express a polypeptide comprising at least 70% to 100% identity to SEQ ID NO:2 and which has methanol dehydrogenase activity.

[0014] The disclosure also provides a recombinant host cell comprising a metabolic pathway for the production of chemical (e.g., isobutanol, n-butanol and the like) not normally produced by or in amount greater than in a wild-type microorganism, wherein the microorganism has been engineered to express a polypeptide that is at least 70% identical to SEQ ID NO:2 and has methanol dehydrogenase activity.

[0015] The details of one or more embodiments of the disclosure are set forth in the accompanying drawings and the description below. Other features, objects, and advantages will be apparent from the description and drawings, and from the claims.

BRIEF DESCRIPTION OF DRAWINGS

[0016] **Figure 1A-B** show the effect of (A) pH, (B) thermal stability, on enzymatic activity. Assays (A) were performed using 800 mM methanol and 3 mM NAD⁺ as substrates at 30 °C and pH 9.5; for assays (B), the reaction mixture containing everything except the initiating substrate (methanol) was pre-incubated at temperatures ranging from 25 to 60 °C. Methanol was added to initiate reaction at 30 °C and pH 9.5 to measure remaining activity. The data shown were from triplicate experiments.

[0017] **Figure 2** shows the effect of ions and chelator to Mdh2. Experiments were performed by incubating enzyme with 1mM ions (0.1mM for Cu²⁺ and Zn²⁺) or EDTA for 3 mins, then using 800 mM methanol and 3 mM NAD⁺ as substrates at 30 °C and pH 9.5. The highlighted Ni²⁺ indicates significant activity increase and Cu²⁺ and Zn²⁺ indicates significant activity decrease. The data shown were from triplicate experiments.

[0018] **Figure 3A-B** shows the Effect of activator at different temperatures with **(A)** Mdh2 of *C. necator* N-1 **(B)** Mdh3 of *B. methanolicus* MGA3. Assays were performed using 800 mM methanol and 3 mM NAD⁺ as substrates at 30 °C and pH 9.5. ACT (BM) indicates to ACT of *B. methanolicus* (thermophilic ACT) and NudF (EC) indicates ACT homolog NudF of *E.coli* (mesophilic ACT). The data shown were from triplicate experiments.

[0019] **Figure 4A-C** shows the development of HTS for Mdh. **(A)** Utilizing the colorimetric Nash reaction to measure Mdh activity. The color indicates reaction product diacetyldihydrolutidine and can be quantified by OD₄₀₅ **(B)** Optimization of HTS process by showing improved Z'-factor. Wild-type Mdh2 was used as the positive control and *E.coli* transketolase was used as the negative control. **(C)** Schematic diagram of Mdh HTS process.

[0020] **Figure 5A-B** shows **(A)** C1 to C4 alcohol specificity of Mdh2 and its engineered variants. **(B)** Activity ratio of methanol over longer chain alcohols (C2 to C4). WT: Mdh2, CT1-1: A31V, CT1-2: A169V, CT2-1: A26V, A169V, CT4-1: A26V, A31V, A169V. NAD⁺ 3mM and alcohol concentrations saturate activity of wild-type Mdh2 were chosen for assay conditions and the assays were performed at 30 °C and pH 9.5. The data shown were from triplicate experiments.

[0021] **Figure 6A-B** shows sequence information of *C. necator* N-1 Mdh2. **(A)** Sequence similarity predicted by SWISS-MODEL protein structure homology modeling. Mdh2 was shown as circle in the middle, each template enzyme was shown as a circle which clusters with a group of similar enzymes. The distance between two template enzymes is proportional to the sequence identity. **(B)** Sequence alignment of group III alcohol dehydrogenases/methanol dehydrogenase and recently identified methanol-oxidizing Adhs. Cn, *Cupriavidus necator* N-1 (Cn_Mdh2 (SEQ ID NO:2), Cn_Mdh1 (SEQ ID NO:3)); Zm, *Zymomonas mobilis* ZM4 (SEQ ID NO:4); Kp, *Klebsiella pneumoniae* (SEQ ID NO:5); Bm, *Bacillus methanolicus* MGA3 (SEQ ID NO:6); Dh, *Desulfitobacterium hafniense* Y51 (SEQ ID NO:7); Ls, *Lysinibacillus sphaericus* C3-41 (SEQ ID NO:8); and Lf, *Lysinibacillus fusiformis* ZC1 (SEQ ID NO:9). Amino acid residues that are highly conserved are enclosed by blue boxed and highlighted in yellow. Identical residues are highlighted in red

background. The NAD⁺ binding motif and metal coordination domain are annotated by black stars and triangles, respectively. Predicted residues of substrate binding based on Zm_Adh2 are indicated by blue circles.

[0022] **Figure 7** shows SDS-PAGE analysis to show expression of Mdh1 and Mdh2. L: PageRuler ladder (Thermo Scientific), S: Soluble fraction protein, H: His-tagged purified protein.

[0023] **Figure 8** shows relative activity of Al69 variants measured by Nash assay.

[0024] **Figure 9A-B** shows Mdh2 insensitivity to activation effect. **(A)** Shows the effect of different activator concentration to Mdh2 activity. **(B)** Shows the effect of putative activator proteins of *C. necator* N-1. ACT(BM) indicates that ACT of *B. methanolicus* (thermophilic ACT) and NudF(EC) indicates the ACT homolog NudF of *E. coli* (Mesophilic ACT). Mdh2 activity was measured in the presence of crude extract (50(+) or 150 (++) µg/ml) or 5 µg/ml, purified activator using standard Mdh assay at 30 °C and pH 9.5. pMS4 (CNE_BB1p03180 of *C. necator* N-1), pMS5 (CNE_1c08460 *C. necator* N-1), pMS12 (CNE+1c14320 *C. necator* N-1), pMS13 (CNE_1c04760 *C. necator* N-1), pMS14 (CNE_1c10080 *C. necator* N-1). "+" sign indicates addition of protein in the assay. The data shown were from triplicate experiments.

DETAILED DESCRIPTION

[0025] Before describing the invention in detail, it is to be understood that this invention is not limited to particular compositions or biological systems, which can, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting.

[0026] As used herein and in the appended claims, the singular forms "a," "and," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a polynucleotide" includes a plurality of such polynucleotides and reference to "the microorganism" includes reference to one or more microorganisms, and so forth.

[0027] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to

one of ordinary skill in the art to which this disclosure belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice of the disclosed methods and compositions, the exemplary methods, devices and materials are described herein.

[0028] Also, the use of "or" means "and/or" unless stated otherwise. Similarly, "comprise," "comprises," "comprising" "include," "includes," and "including" are interchangeable and not intended to be limiting.

[0029] It is to be further understood that where descriptions of various embodiments use the term "comprising," those skilled in the art would understand that in some specific instances, an embodiment can be alternatively described using language "consisting essentially of" or "consisting of."

[0030] Any publications discussed above and throughout the text are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior disclosure.

[0031] As used herein an "enzyme" means any substance, composed wholly or largely of protein, that catalyzes or promotes, more or less specifically, one or more chemical or biochemical reactions.

[0032] The term "host cell", as used herein, includes any cell type which is susceptible to transformation with a nucleic acid construct.

[0033] A "mutation" means any process or mechanism resulting in a mutant protein, enzyme, polynucleotide, gene, or cell. This includes any mutation in which a protein, enzyme, polynucleotide, or gene sequence is altered, and any detectable change in a cell arising from such a mutation. Typically, a mutation occurs in a polynucleotide or gene sequence, by point mutations, deletions, or insertions of single or multiple nucleotide residues. A mutation includes polynucleotide alterations arising within a protein-encoding region of a gene as well as alterations in regions outside of a protein-encoding sequence, such as, but not limited to, regulatory or promoter sequences. A mutation in a gene can be "silent", *i.e.*, not reflected in an amino acid alteration upon

expression, leading to a "sequence-conservative" variant of the gene. This generally arises when one amino acid corresponds to more than one codon. A mutation in a sequence gives rise to a variant of such sequence. For example, a mutated Mdh2 provides an Mdh2 variant.

[0034] A "native" or "wild-type" protein, enzyme, polynucleotide, gene, or cell, means a protein, enzyme, polynucleotide, gene, or cell that occurs in nature.

[0035] The terms "polynucleotide," "nucleotide sequence," and "nucleic acid molecule" are used to refer to a polymer of nucleotides (A, C, T, U, G, *etc.* or naturally occurring or artificial nucleotide analogues), *e.g.*, DNA or RNA, or a representation thereof, *e.g.*, a character string, *etc.*, depending on the relevant context. A given polynucleotide or complementary polynucleotide can be determined from any specified nucleotide sequence. It will be readily recognized that if a sequence presents DNA (*i.e.*, containing "T") the RNA sequence is readily derivable and encompassed herein by substituting "T" with "U".

[0036] A "protein" or "polypeptide", which terms are used interchangeably herein, comprises one or more chains of chemical building blocks called amino acids that are linked together by chemical bonds called peptide bonds.

[0037] The term "substrate" or "suitable substrate" means any substance or compound that is converted or meant to be converted into another compound by the action of an enzyme catalyst. In one embodiment of the disclosure the substrate is methanol.

[0038] The term "transformation" means the introduction of a foreign (*i.e.*, extrinsic or extracellular) gene, DNA or RNA sequence to a host cell, so that the host cell will express the introduced gene or sequence to produce a desired substance, typically a protein or enzyme coded by the introduced gene or sequence. The introduced gene or sequence may include regulatory or control sequences, such as start, stop, promoter, signal, secretion, or other sequences used by the genetic machinery of the cell. A host cell that receives and expresses introduced DNA or RNA has been "transformed" and is a "transformant" or a "clone." The DNA or RNA introduced to a host cell can come from any source,

including cells of the same genus or species as the host cell, or cells of a different genus or species.

[0039] Methanol has become an attractive substrate for bioconversion to chemical commodities due to the abundance of methane. Methanol bioconversions to amino acids using methylotrophic bacteria such as *Methylobacterium* sp. for L-serine (Hagishita et al., 1996), and *Methylobacillus glycogenes* for L-threonine (Motoyama et al., 1994), L-glutamate (Motoyama et al., 1993), and L-lysine (Motoyama et al., 2001) have been demonstrated. Despite previous successes, many hurdles remain before industrial applications. In particular, genetic tool development and physiological studies of methylotrophic bacteria are needed for further strain engineering (Schrader et al., 2009). An alternative is to enable methanol assimilation, or even bestow methylotrophic growth on strains suitable for industrial processing. In principle, synthetic methylotrophy can be achieved by overexpressing heterologous enzymes for methanol oxidation and engineering a formaldehyde assimilation pathway to produce central metabolites for growth.

[0040] Methanol oxidation, which is categorized into three groups of enzymes based on their terminal electron acceptors include: (1) Pyrroloquinoline quinone (PQQ) dependent methanol dehydrogenases, (2) methanol oxidases, and (3) NAD-dependent Mdh's. NAD-dependent Mdhs are within metal-containing group III alcohol dehydrogenases (Adh's), named Mdh when the enzymes presents significant methanol activity, such as *B. methanolicus* Mdh's (De Vries et al., 1992). Group III Adh's are structurally unrelated to group I or II Adh's and are highly diverse (Elleuche and Antranikian, 2013). Among the three types of methanol oxidizing enzymes, NAD-dependent Mdh's are the favorable option for synthetic methylotrophy due to their applicability in both aerobic and anaerobic conditions (Whitaker et al., 2015). Furthermore, electrons derived from methanol oxidation are stored in NADH, which can be used to drive production of target metabolites without sacrificing additional carbons. As such, this type of enzyme was used in a redox balanced, methanol condensation cycle (MCC) to achieve conversion of methanol to higher alcohols (Bogorad et al.,

2014 and WO2014/153207, incorporated herein by reference). In addition, an NAD-dependent Mdh of *B. methanolicus* was introduced in *E.coli* and *Corynebacterium glutamicum* to demonstrate methanol assimilation via the ribulose monophosphate pathway (Müller et al., 2015; Witthoff et al., 2015).

[0041] NAD-dependent methanol oxidation presents a principal step in utilizing methanol as a substrate for microbial production of chemicals. Prior reported Mdhs from *B. methanolicus* (Krog et al., 2013) and a few additional homologs (Ochsner et al., 2014) require ACT and thermophilic conditions at 50°C to activate methanol oxidation activity. A recent report (Müller et al., 2015) also presents challenges in activation of recombinant *B. methanolicus* Mdh in *E. coli* under mesophilic conditions. Despite previously reported NAD-dependent, type I Adhs from human liver, horse liver, yeast, and *C. glutamicum* and *Bacillus stearothermophilus*, which exhibited moderate enzymatic activity toward methanol without the requirement for activation (Kotrbova-Kozak et al., 2007; Mani et al., 1970; Sheehan et al., 1988), successful attempts of methanol assimilation were only reported using Mdhs from *B. methanolicus* (Müller et al., 2015; Witthoff et al., 2015). Unfortunately, unlike PQQ-dependent Mdhs, NAD-dependent Mdhs exhibit broad substrate specificity and only show moderate activity towards methanol.

[0042] NAD-dependent Mdh's with relatively high activity have only been reported in the Gram-positive, thermophilic methylotroph, *B. methanolicus* (Arfman et al., 1989; Hektor et al., 2002; Krog et al., 2013), with a few homologs reported from other Gram-positives, both mesophilic and thermophilic bacteria (Ochsner et al., 2014; Sheehan et al., 1988). The existence of NAD-dependent Mdh's in thermophiles is in agreement with the thermodynamic argument that NAD⁺ dependent methanol oxidation is favorable at high temperatures (Whitaker et al., 2015). Their sequences have 45-53% similarity to the NAD-dependent 1,3-propanediol dehydrogenase of *Klebsiella pneumoniae* (Krog et al., 2013). In contrast to the PQQ-dependent Mdh's which exhibit high methanol specificity (Keltjens et al., 2014), NAD-dependent Mdh's have broad substrate specificities, with

optimum activity to 1-propanol or *n*-butanol and marginal activity to methanol (Krog *et al.*, 2013; Sheehan *et al.*, 1988).

[0043] The methanol activity of Mdh's of *B. methanolicus* can be greatly enhanced by an endogenous ACT, which contains a conserved motif for hydrolyzing nucleoside diphosphates linked to a moiety X ("Nudix") (Arfman *et al.*, 1997, 1991, 1989; Hektor *et al.*, 2002; Kloosterman *et al.*, 2002). This activation effect has been found to be widespread among group III Adh's (Ochsner *et al.*, 2014) and results in both increased K_{cat} and decreased K_m . Notably, this activation is general to all substrates, instead of a specific activation for methanol (Krog *et al.*, 2013; Ochsner *et al.*, 2014). ACT activates Mdh by hydrolytically removing the nicotinamide mononucleotide (NMN) moiety of the Mdh-bound NAD, causing a change in its reaction mechanism from the ping-pong type mechanism to the ternary complex mechanism (Arfman *et al.*, 1997). The ACT-Mdh activation model has been proposed to be a reversible process in which the interaction between ACT and Mdh results in conformational change to position NAD⁺ and methanol binding sites closer together, thus enabling direct electron transfer (Kloosterman *et al.*, 2002). However, the detailed mechanism of Mdh activation is still unclear. For the purpose of metabolic engineering, it would be useful to identify an Mdh with high activity under mesophilic or thermophilic conditions without the need for ACT.

[0044] This disclosure describes the characterization of a putative Mdh encoded by *mdh2* in the genome of a non-methylotrophic bacteria *C. necator* N-1. The disclosure demonstrates that Mdh2 is an active NAD-dependent Mdh without the need for ACT. Mdh2 is the first group III Adh identified in Gram-negative, mesophilic bacteria that possesses significant methanol activity. Using directed evolution, the Mdh activity and specificity for methanol was further improved in this polypeptide.

[0045] This disclosure provides a characterized and engineered a NAD-dependent methanol dehydrogenase, Mdh2, from a non-methylotrophic bacteria *C. necator* N-1 that can be expressed in recombinant host cells (e.g., *E. coli*). Mdh2 represents a group III Adh in Gram-negative, mesophilic organism to exhibit significant activity towards methanol. Wild-type Mdh2 exhibits methanol

oxidation activity 0.32 U/mg and K_m value 132mM at 30 °C, and is insensitive to activation under mesophilic temperatures. After protein evolution using HTS, the variant CT4-1 retained methanol oxidation activity with K_m values of 21.6 mM and 120 mM for methanol and *n*-butanol, respectively.

[0046] CT4-1 is an improved NAD-dependent Mdh with respect to methanol specificity, activity, and independence of activation as such it is suitable for metabolic engineering of organisms for synthetic methylotrophy or *in vitro* methanol condensation. A previous study suggested that group III Adh activation by ACT homolog Nudix hydrolases presents a common mechanism, however, Mdh2 was insensitive to the activation facilitated by both. Structural analysis and sequence alignment confirmed that Mdh2 belongs to group III Adh, by the high structural similarities to the 1,3-PDH of *K. pneumoniae* and Adh2 of *Z. mobilis* ZM4, in addition to a putative NAD⁺ binding motif and metal binding residues. Notably, the two most similar enzymes, *Z. mobilis* Adh2 and *K. pneumoniae* 1,3-PDH, do not have methanol oxidation activity. Similarly, *C. necator* N-1 cannot grow on methanol as a carbon source, suggesting that the methanol oxidation may be a gratuitous activity in Mdh2.

[0047] Discovery of *C. necator* N-1 Mdh2 opens up the possibility of searching for useful NAD-dependent Mdhs for synthetic methylotrophy from Gram-negative, mesophilic organisms. General perception of Mdhs in Gram-negative, mesophilic methylotrophs are mostly PQQ-dependent enzymes localized in periplasm. In contrast, NAD-dependent Mdhs are localized in bacterial cytoplasm (Keltjens et al., 2014). Despite the fact that *C. necator* N-1 possesses an active Mdh, this organism cannot utilize methanol as a carbon source. It remains unclear the physiological role of *mdh2* in *C. necator* N-1. A possible explanation can be found in recent study on a non-methylotrophic, Gram-positive bacteria *C. glutamicum* which possesses a AdhA for methanol oxidation to CO₂, where methanol served as an auxiliary carbon source for energy generation, of which four essential enzymes alcohol dehydrogenase, acetaldehyde dehydrogenase, mycothiol-dependent formaldehyde dehydrogenase, and formate dehydrogenase are involved (Witthoff et al., 2013). More detailed

characterizations on physiological growth conditions and genome analysis for *C. necator* N-1 will unveil the role of *mdh2*.

[0048] As used herein an Mdh2 polypeptide of the disclosure can be characterized as having a sequence that is at least 80% identical to the sequence as set forth in SEQ ID NO:2 (see also **Fig. 6B**):

MTHLNIANRV	DSFFIPCVTL	FGPGCARETG	ARARSLGARK	ALIVTDAGLH	KMGLSEVVAG	60
HIREAGLQAV	IFPGAEPNPT	DVNVHDGVKL	FEREECDFIV	SLGGGSSHDC	AKGIGLVTAG	120
GGHIRDYEGI	DKSTVPMTP	ISINTTAGTA	AEMTRFCIIT	NSSNHVKMAI	VDWRCTPLIA	180
IDDPPLMVAM	PPALTAATGM	DALTHAIEAY	VSTAATPITD	ACAIEAIVLI	AEWLKAVAN	240
GDSMEARAAM	CYAQYLAGMA	FNNASLGYVH	AMAHQLGGFY	NLPHGVCNAI	LLPHVSEFNL	300
IAAPERYARI	AELLGENIGG	LSAHDAKAA	VSAIRTLSTS	IGIPAGLAGL	GVKADDHEVM	360
ASNAQKDACM	LTNPRKATLA	QVMAIFAAAM				390

wherein the Mdh2 polypeptide has methanol dehydrogenase activity. As discussed more fully below, the Mdh2 polypeptide can be 80%, 82%, 85%, 87%, 90%, 92%, 95%, 98%, 99% or 100% identical to SEQ ID NO:2 so long as the polypeptide has methanol dehydrogenase activity. Moreover, the disclosure demonstrates a method of screening variant of Mdh2 that are simple and routine.

[0049] The disclosure also provides Mdh2 variants that have improved enzymatic activity compared to the wild-type of SEQ ID NO:2 and have methanol dehydrogenase activity. Such variants include selective mutations at particular residues, but may have conservative substitutions at other residues. In one embodiment, the disclosure provides an Mdh2 variant that is at least 80-99% identical to SEQ ID NO:2 and has a mutation at a residue selected from the group consisting of A26, A31, A169 and any combination thereof. In one embodiment, the residues are substituted with a non-polar amino acid (e.g., P, V, M, I or L) or a polar amino acid (e.g., S, C, N, T or Q). In a specific embodiment, the residue is substituted with a V (e.g., A26V, A31V and/or A169V). In another embodiment, the Mdh2 variant comprises a catalytic efficiency (K_{cat}/K_m) of greater than 1.6 for methanol and/or less than 903 for n-butanol. In another embodiment, the Mdh2 variant comprises SEQ ID NO:2 with a mutation at positions selected from the group consisting of A26, A31, A169 and any combination thereof, and wherein the variant comprises from 1-10 conservative amino acid substitutions at positions other than A26, A31 and A169. Positions that can tolerate conservative substitutions can be identified in

Fig. 6B, wherein non-conserved amino acids are susceptible to conservative substitutions.

[0050] In general, the disclosure includes any polypeptide encoded by a modified Mdh2 polynucleotide derived by mutation, recursive sequence recombination, and/or diversification of the polynucleotide sequences described herein. In some embodiments, a Mdh2 polypeptide or variant thereof, is modified by single or multiple amino acid substitutions, a deletion, an insertion, or a combination of one or more of these types of modifications. Substitutions can be conservative or non-conservative, can alter function or not, and can add new function. Insertions and deletions can be substantial, such as the case of a truncation of a substantial fragment of the sequence, or in the fusion of additional sequence, either internally or at N or C terminal.

[0051] The "activity" of an enzyme is a measure of its ability to catalyze a reaction, *i.e.*, to "function", and may be expressed as the rate at which the product of the reaction is produced. For example, enzyme activity can be represented as the amount of product produced per unit of time or per unit of enzyme (*e.g.*, concentration or weight), or in terms of affinity or dissociation constants. As used interchangeably herein a "methanol dehydrogenase activity", "biological activity of Mdh2" or "functional activity of Mdh2", refers to an activity exerted by a Mdh2 protein, polypeptide or variant on a Mdh2 polypeptide substrate, as determined *in vivo*, or *in vitro*, according to standard techniques. The biological activity of an Mdh2 or variant thereof is described herein.

[0052] "Conservative amino acid substitution" or, simply, "conservative variations" of a particular sequence refers to the replacement of one amino acid, or series of amino acids, with essentially identical amino acid sequences. One of skill will recognize that individual substitutions, deletions or additions which alter, add or delete a single amino acid or a percentage of amino acids in an encoded sequence result in "conservative variations" where the alterations result in the deletion of an amino acid, addition of an amino acid, or substitution of an amino acid with a chemically similar amino acid.

[0053] Conservative substitution tables providing functionally similar amino acids are well known in the art. For example, one conservative substitution group includes Alanine (A), Serine (S), and Threonine (T). Another conservative substitution group includes Aspartic acid (D) and Glutamic acid (E). Another conservative substitution group includes Asparagine (N) and Glutamine (Q). Yet another conservative substitution group includes Arginine (R) and Lysine (K). Another conservative substitution group includes Isoleucine, (I) Leucine (L), Methionine (M), and Valine (V). Another conservative substitution group includes Phenylalanine (F), Tyrosine (Y), and Tryptophan (W).

[0054] Thus, "conservative amino acid substitutions" of a listed polypeptide sequence (e.g., SEQ ID NO:2) include substitutions of a percentage, typically less than 10%, of the amino acids of the polypeptide sequence, with a conservatively selected amino acid of the same conservative substitution group. Accordingly, a conservatively substituted variation of a polypeptide of the disclosure can contain 100, 75, 50, 25, or 10 or less substitutions with a conservatively substituted variation of the same conservative substitution group. Thus, "conservative amino acid substitutions," in one or a few amino acids in an amino acid sequence are substituted with different amino acids with highly similar properties, are readily identified as being highly similar to a disclosed construct. Such conservative variations of each disclosed sequence are a feature of the polypeptides provided herein.

[0055] Non-conservative modifications of a particular polypeptide are those which substitute any amino acid not characterized as a conservative substitution. For example, any substitution which crosses the bounds of the six groups set forth above. These include substitutions of basic or acidic amino acids for neutral amino acids, (e.g., Asp, Glu, Asn, or Gln for Val, Ile, Leu or Met), aromatic amino acid for basic or acidic amino acids (e.g., Phe, Tyr or Trp for Asp, Asn, Glu or Gln) or any other substitution not replacing an amino acid with a like amino acid. Basic side chains include lysine (K), arginine (R), histidine (H); acidic side chains include aspartic acid (D), glutamic acid (E);

uncharged polar side chains include glycine (G), asparagine (N), glutamine (Q), serine (S), threonine (T), tyrosine (Y), cysteine (C); nonpolar side chains include alanine (A), valine (V), leucine (L), isoleucine (I), proline (P), phenylalanine (F), methionine (M), tryptophan (W); beta-branched side chains include threonine (T), valine (V), isoleucine (I); aromatic side chains include tyrosine (Y), phenylalanine (F), tryptophan (W), histidine (H).

[0056] "Conservative variants" are proteins or enzymes in which a given amino acid residue has been changed without altering overall conformation and function of the protein or enzyme, including, but not limited to, replacement of an amino acid with one having similar properties, including polar or non-polar character, size, shape and charge. Amino acids other than those indicated as conserved may differ in a protein or enzyme so that the percent protein or amino acid sequence similarity between any two proteins of similar function may vary and can be, for example, at least 30%, at least 50%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98% or at least 99%, as determined according to an alignment scheme.

[0057] It is understood that the addition of sequences which do not alter the encoded activity of a nucleic acid molecule, such as the addition of a non-functional or non-coding sequence, is a conservative variation of the basic nucleic acid.

[0058] In some embodiments, a polypeptide of the disclosure can include modified amino acids or be mutated to incorporate modified amino acids. Non-limiting examples of a modified amino acid include a glycosylated amino acid, a sulfated amino acid, a prenylated (e.g., farnesylated, geranylgeranylated) amino acid, an acetylated amino acid, an acylated amino acid, a pegylated amino acid, a biotinylated amino acid, a carboxylated amino acid, a phosphorylated amino acid, and the like. References adequate to guide one of skill in the modification of amino acids are replete throughout the literature. Example protocols are found in Walker (1998) Protein Protocols on CD-ROM (Humana Press, Towata, N.J.).

[0059] Recombinant methods for producing and isolating modified Mdh2 polypeptides of the disclosure are described herein. In addition to recombinant production, the polypeptides may be

produced by direct peptide synthesis using solid-phase techniques (e.g., Stewart et al. (1969) Solid-Phase Peptide Synthesis (WH Freeman Co, San Francisco); and Merrifield (1963) J. Am. Chem. Soc. 85: 2149-2154; each of which is incorporated by reference). Peptide synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer, Foster City, Calif.) in accordance with the instructions provided by the manufacturer.

[0060] A "parent" protein, enzyme, polynucleotide, gene, or cell, is any protein, enzyme, polynucleotide, gene, or cell, from which any other protein, enzyme, polynucleotide, gene, or cell, is derived or made, using any methods, tools or techniques, and whether or not the parent is itself native or mutant. A parent polynucleotide or gene encodes for a parent protein or enzyme. In one embodiment, a parent Mdh2 comprises the sequence of SEQ ID NO:2. In another embodiment, a parent cell or microorganism comprises a polynucleotide that encodes or comprises a polypeptide of SEQ ID NO:2 or an Mdh2 variant.

[0061] In other embodiments, isolated nucleic acid molecules are provided. In one embodiment, the disclosure provides a novel family of isolated or recombinant polynucleotides referred to herein as "Mdh2 polynucleotides" or "Mdh2 nucleic acid molecules." Mdh2 polynucleotide sequences are characterized by the ability to encode a Mdh2 polypeptide. In general, the disclosure includes any nucleotide sequence that encodes any of the novel Mdh2 or Mdh2 variant polypeptides described herein. In some embodiments, a Mdh2 polynucleotide that encodes a Mdh2 variant polypeptide is provided.

[0062] In one embodiment, the Mdh2 polynucleotides comprise recombinant or isolated forms of naturally occurring nucleic acids isolated from an organism, e.g., a bacterial strain. Exemplary Mdh2 polynucleotides include those encoding the polypeptide of SEQ ID NO:2 and variants thereof. In another embodiment, Mdh2 polynucleotides are produced by diversifying, e.g., recombining and/or mutating one or more naturally occurring, isolated, or recombinant Mdh2 polynucleotides. As described in more detail elsewhere herein, it is often possible to generate diversified Mdh2

polynucleotides encoding Mdh2 polypeptides or variants thereof with superior functional attributes, *e.g.*, increased catalytic function, increased stability, or higher expression level, than a parent Mdh2 polynucleotide used as a substrate in the diversification process.

[0063] The polynucleotides of the disclosure have a variety of uses in, for example recombinant production (*i.e.*, expression) of the Mdh2 polypeptides or variants of the disclosure and as substrates for further diversity generation, *e.g.*, recombination reactions or mutation reactions to produce new and/or improved Mdh2 variants, and the like.

[0064] It is important to note that certain specific, substantial and credible utilities of Mdh2 polynucleotides do not require that the polynucleotide encode a polypeptide with substantial Mdh2 activity or even variant Mdh2 activity. For example, Mdh2 polynucleotides that do not encode active enzymes can be valuable sources of parental polynucleotides for use in diversification procedures to arrive at Mdh2 polynucleotide variants, or non-Mdh2 polynucleotides, with desirable functional properties (*e.g.*, high k_{cat} or k_{cat}/K_m , low K_m , high stability towards heat or other environmental factors, high transcription or translation rates, resistance to proteolytic cleavage, *etc.*).

[0065] Mdh2 polynucleotides, including nucleotide sequences that encode Mdh2 polypeptides and variants thereof, fragments of Mdh2 polypeptides, related fusion proteins, or functional equivalents thereof, are used in recombinant DNA molecules that direct the expression of the Mdh2 polypeptides or variants in appropriate host cells, such as bacterial cells. Due to the inherent degeneracy of the genetic code, other nucleic acid sequences which encode substantially the same or a functionally equivalent amino acid sequence can also be used to clone and express the Mdh2 polynucleotides.

[0066] Thus, the disclosure also provides polynucleotides that encode a polypeptide of the disclosure. The polynucleotide can be DNA or RNA. The polynucleotide can encode a polypeptide comprising SEQ ID NO:2 or variant thereof. In one embodiment, the polynucleotide is selected from the group consisting of: (i) SEQ ID NO:1; (ii) a polynucleotide that hybridizes to a sequence

consisting of SEQ ID NO:1 and encodes a polypeptide comprising SEQ ID NO:2; (iii) a polynucleotide that hybridizes to a sequence consisting of SEQ ID NO:1 and encodes a polypeptide having having Mdh2 activity; (iv) a polynucleotide that encodes a polypeptide of SEQ ID NO:2; (v) a polynucleotide that hybridizes to a sequence consisting of SEQ ID NO:1 and encodes a polypeptide of SEQ ID NO:2 having mutations at A26, A31 and/or A169 and having Mdh2 activity; (vi) any of (i) to (v) wherein the polynucleotide is RNA or DNA; and (vii) a sequence of SEQ ID NO:1 wherein T is U. In one embodiment, the polynucleotide encodes a polypeptide that is at least 80%, 85%, 90%, 95%, 98%, 99% or 100% identical to SEQ ID NO:2. In another embodiment, the polynucleotide encodes a polypeptide of SEQ ID NO:2 having mutations at A26V, A31V, and/or A169V.

[0067] As will be understood by those of skill in the art, it can be advantageous to modify a coding sequence to enhance its expression in a particular host. The genetic code is redundant with 64 possible codons, but most organisms preferentially use a subset of these codons. The codons that are utilized most often in a species are called optimal codons, and those not utilized very often are classified as rare or low-usage codons (see, e.g., Zhang et al. (1991) Gene 105:61-72; incorporated by reference herein). Codons can be substituted to reflect the preferred codon usage of the host, a process sometimes called "codon optimization" or "controlling for species codon bias."

[0068] Optimized coding sequences containing codons preferred by a particular prokaryotic or eukaryotic host (see also, Murray et al. (1989) Nucl. Acids Res. 17:477-508; incorporated by reference herein) can be prepared, for example, to increase the rate of translation or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, as compared with transcripts produced from a non-optimized sequence. Translation stop codons can also be modified to reflect host preference. For example, preferred stop codons for *S. cerevisiae* and mammals are UAA and UGA, respectively. The preferred stop codon for monocotyledonous plants is UGA, whereas insects and *E. coli* prefer to use UAA as the stop codon (Dalphin et al. (1996) Nucl. Acids

Res. 24: 216-218; incorporated by reference herein). Methodology for optimizing a nucleotide sequence for expression in a plant is provided, for example, in U.S. Pat. No. 6,015,891, and the references cited therein (incorporated herein by reference).

[0069] A nucleic acid of the disclosure can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to nucleotide sequences can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

[0070] In another embodiment, an isolated nucleic acid molecule of the disclosure hybridizes under stringent conditions to a nucleic acid molecule that encodes a polypeptide of SEQ ID NO:2 and wherein the nucleic acid molecule that hybridizes to a polynucleotide encoding a polypeptide consisting of SEQ ID NO:2 has methanol dehydrogenase activity. Nucleic acid molecules are "hybridizable" to each other when at least one strand of one polynucleotide can anneal to another polynucleotide under defined stringency conditions. Stringency of hybridization is determined, e.g., by (a) the temperature at which hybridization and/or washing is performed, and (b) the ionic strength and polarity (e.g., formamide) of the hybridization and washing solutions, as well as other parameters. Hybridization requires that the two polynucleotides contain substantially complementary sequences; depending on the stringency of hybridization, however, mismatches may be tolerated. Typically, hybridization of two sequences at high stringency (such as, for example, in an aqueous solution of 0.5 X SSC at 65°C) requires that the sequences exhibit some high degree of complementarity over their entire sequence. Conditions of intermediate stringency (such as, for example, an aqueous solution of 2 X SSC at 65°C) and low stringency (such as, for example, an aqueous solution of 2 X SSC at 55°C), require correspondingly less overall complementarity between the hybridizing sequences (1 X SSC is 0.15 M NaCl, 0.015 M Na citrate). Nucleic acid molecules that hybridize include those which anneal under suitable stringency

conditions and which encode polypeptides or enzymes having the same function. Further, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 30%, 40%, 50%, or 60% homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 70%, more preferably at least about 80%, even more preferably at least about 85% or 90% homologous to each other typically remain hybridized to each other. In some cases, an isolated nucleic acid molecule of the disclosure that hybridizes under stringent conditions to a nucleic acid sequence encoding a polypeptide set forth in SEQ ID NO:2, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

[0071] One of skill in the art will appreciate that many conservative variations of the polynucleotides of the disclosure can provide a functionally identical polynucleotide or encode a polypeptide having Mdh2 activity. For example, owing to the degeneracy of the genetic code, "silent substitutions" (*i.e.*, substitutions in a polynucleotide sequence which do not result in an alteration in an encoded polypeptide) are an implied feature of every nucleic acid sequence which encodes an amino acid.

[0072] "Silent variations" are one species of "conservatively modified variations." As mentioned elsewhere herein one of skill in the art will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine) can be modified by standard techniques to encode a functionally identical polypeptide. Accordingly, each silent variation of a polynucleotide which encodes a polypeptide is implicit in any described sequence. The disclosure provides each and every possible variation of a polynucleotide sequence encoding a polypeptide of the disclosure that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide encoding a Mdh2 polypeptide or variant of the disclosure. All such

variations of every nucleic acid herein are specifically provided and described by consideration of the sequence in combination with the genetic code. Any variant can be produced as noted herein.

[0073] It will be appreciated by those skilled in the art that due to the degeneracy of the genetic code, a multitude of nucleotide sequences encoding modified Mdh2 polypeptides of the disclosure may be produced, some of which bear substantial identity to the polynucleotide sequences explicitly disclosed herein. For instance, codons AGA, AGG, CGA, CGC, CGG, and CGU all encode the amino acid arginine. Thus, at every position in the nucleic acids of the invention where an arginine is specified by a codon, the codon can be altered to any of the corresponding codons described above without altering the encoded polypeptide. It is understood that U in an RNA sequence corresponds to T in a DNA sequence.

[0074] The skilled artisan will appreciate that changes can be introduced by mutation into the nucleotide sequences of any nucleic acid sequence encoding a polypeptide of SEQ NO:2, thereby leading to changes in the amino acid sequence of the encoded polypeptide. In some cases the alteration will lead to altered function of the polypeptide. In other cases the change will not alter the functional ability of the encoded polypeptide. In general, substitutions that do not alter the function of a polypeptide include nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues. Generally these substitutions can be made in without altering the methanol dehydrogenase activity of the polypeptide. A "non-essential" amino acid residue is a residue that can be altered from the parent sequence without altering the biological activity of the resulting polypeptide, e.g., catalyzing the conversion of methane to methanol.

[0075] Mutational methods of generating diversity include, for example, site-directed mutagenesis (Ling *et al.* (1997) "Approaches to DNA mutagenesis: an overview" *Anal Biochem.* 254(2): 157-178; Dale *et al.* (1996) "Oligonucleotide-directed random mutagenesis using the phosphorothioate method" *Methods Mol. Biol.* 57:369-374; Smith (1985) "*In vitro* mutagenesis" *Ann. Rev. Genet.* 19:423-462; Botstein & Shortle (1985) "Strategies and applications of *in vitro*

mutagenesis" *Science* 229:1193-1201; Carter (1986) "Site-directed mutagenesis" *Biochem. J.* 237:1-7; and Kunkel (1987) "The efficiency of oligonucleotide directed mutagenesis" in *Nucleic Acids & Molecular Biology* (Eckstein, F. and Lilley, D. M. J. eds., Springer Verlag, Berlin)); mutagenesis using uracil containing templates (Kunkel (1985) "Rapid and efficient site-specific mutagenesis without phenotypic selection" *Proc. Natl. Acad. Sci. USA* 82:488-492; Kunkel *et al.* (1987) "Rapid and efficient site-specific mutagenesis without phenotypic selection" *Methods in Enzymol.* 154, 367-382; and Bass *et al.* (1988) "Mutant Trp repressors with new DNA-binding specificities" *Science* 242:240-245); oligonucleotide-directed mutagenesis (*Methods in Enzymol.* 100: 468-500 (1983); *Methods in Enzymol.* 154: 329-350 (1987); Zoller & Smith (1982) "Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment" *Nucleic Acids Res.* 10:6487-6500; Zoller & Smith (1983) "Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors" *Methods in Enzymol.* 100:468-500; and Zoller & Smith (1987) "Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template" *Methods in Enzymol.* 154:329-350); phosphorothioate-modified DNA mutagenesis (Taylor *et al.* (1985) "The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA" *Nucl. Acids Res.* 13: 8749-8764; Taylor *et al.* (1985) "The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA" *Nucl. Acids Res.* 13: 8765-8787; Nakamaye & Eckstein (1986) "Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis" *Nucl. Acids Res.* 14: 9679-9698; Sayers *et al.* (1988) "Y-T Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis" *Nucl. Acids Res.* 16:791-802; and Sayers *et al.* (1988) "Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide" *Nucl. Acids Res.* 16: 803-814); mutagenesis using gapped duplex DNA (Kramer *et al.* (1984) "The gapped duplex DNA approach to oligonucleotide-directed mutation construction" *Nucl.*

Acids Res. 12: 9441-9456; Kramer & Fritz (1987) Methods in Enzymol. "Oligonucleotide-directed construction of mutations via gapped duplex DNA" 154:350-367; Kramer et al. (1988) "Improved enzymatic *in vitro* reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations" Nucl. Acids Res. 16: 7207; and Fritz et al. (1988) "Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions *in vitro*" Nucl. Acids Res. 16: 6987-6999) (each of which is incorporated by reference).

[0076] Additional suitable methods include point mismatch repair (Kramer et al. (1984) "Point Mismatch Repair" Cell 38:879-887), mutagenesis using repair-deficient host strains (Carter et al. (1985) "Improved oligonucleotide site-directed mutagenesis using M13 vectors" Nucl. Acids Res. 13: 4431-4443; and Carter (1987) "Improved oligonucleotide-directed mutagenesis using M13 vectors" Methods in Enzymol. 154: 382-403), deletion mutagenesis (Eghtedarzadeh & Henikoff (1986) "Use of oligonucleotides to generate large deletions" Nucl. Acids Res. 14: 5115), restriction-selection and restriction-purification (Wells et al. (1986) "Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin" Phil. Trans. R. Soc. Lond. A 317: 415-423), mutagenesis by total gene synthesis (Nambiar et al. (1984) "Total synthesis and cloning of a gene coding for the ribonuclease S protein" Science 223: 1299-1301; Sakamar and Khorana (1988) "Total synthesis and expression of a gene for the α -subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin)" Nucl. Acids Res. 14: 6361-6372; Wells et al. (1985) "Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites" Gene 34:315-323; and Grundstrom et al. (1985) "Oligonucleotide-directed mutagenesis by microscale `shot-gun` gene synthesis" Nucl. Acids Res. 13: 3305-3316); double-strand break repair (Mandecki (1986); Arnold (1993) "Protein engineering for unusual environments" Current Opinion in Biotechnology 4:450-455; and "Oligonucleotide-directed double-strand break repair in plasmids of *Escherichia coli*: a method for site-specific mutagenesis" Proc. Natl. Acad. Sci. USA, 83:7177-7181) (each of which is incorporated by reference). Additional

details on many of the above methods can be found in Methods in Enzymology Volume 154, which also describes useful controls for trouble-shooting problems with various mutagenesis methods.

[0077] Additional details regarding various diversity generating methods can be found in the following U.S. patents, PCT publications, and EPO publications: U.S. Pat. No. 5,605,793 to Stemmer (Feb. 25, 1997), "Methods for *In vitro* Recombination;" U.S. Pat. No. 5,811,238 to Stemmer *et al.* (Sep. 22, 1998) "Methods for Generating Polynucleotides having Desired Characteristics by Iterative Selection and Recombination;" U.S. Pat. No. 5,830,721 to Stemmer *et al.* (Nov. 3, 1998), "DNA Mutagenesis by Random Fragmentation and Reassembly;" U.S. Pat. No. 5,834,252 to Stemmer, *et al.* (Nov. 10, 1998) "End-Complementary Polymerase Reaction;" U.S. Pat. No. 5,837,458 to Minshull, *et al.* (Nov. 17, 1998), "Methods and Compositions for Cellular and Metabolic Engineering;" WO 95/22625, Stemmer and Cramer, "Mutagenesis by Random Fragmentation and Reassembly;" WO 96/33207 by Stemmer and Lipschutz "End Complementary Polymerase Chain Reaction;" WO 97/20078 by Stemmer and Cramer "Methods for Generating Polynucleotides having Desired Characteristics by Iterative Selection and Recombination;" WO 97/35966 by Minshull and Stemmer, "Methods and Compositions for Cellular and Metabolic Engineering;" WO 99/41402 by Punnonen *et al.* "Targeting of Genetic Vaccine Vectors;" WO 99/41383 by Punnonen *et al.* "Antigen Library Immunization;" WO 99/41369 by Punnonen *et al.* "Genetic Vaccine Vector Engineering;" WO 99/41368 by Punnonen *et al.* "Optimization of Immunomodulatory Properties of Genetic Vaccines;" EP 752008 by Stemmer and Cramer, "DNA Mutagenesis by Random Fragmentation and Reassembly;" EP 0932670 by Stemmer "Evolving Cellular DNA Uptake by Recursive Sequence Recombination;" WO 99/23107 by Stemmer *et al.*, "Modification of Virus Tropism and Host Range by Viral Genome Shuffling;" WO 99/21979 by Apt *et al.*, "Human Papillomavirus Vectors;" WO 98/31837 by del Cardayre *et al.* "Evolution of Whole Cells and Organisms by Recursive Sequence Recombination;" WO 98/27230 by Patten and Stemmer, "Methods and Compositions for Polypeptide Engineering;" WO 98/13487 by Stemmer *et al.*, "Methods for Optimization of Gene Therapy by Recursive Sequence Shuffling and Selection;" WO 00/00632, "Methods for

Generating Highly Diverse Libraries;" WO 00/09679, "Methods for Obtaining *in vitro* Recombined Polynucleotide Sequence Banks and Resulting Sequences;" WO 98/42832 by Arnold et al., "Recombination of Polynucleotide Sequences Using Random or Defined Primers;" WO 99/29902 by Arnold et al., "Method for Creating Polynucleotide and Polypeptide Sequences;" WO 98/41653 by Vind, "An *in vitro* Method for Construction of a DNA Library;" WO 98/41622 by Borchert et al., "Method for Constructing a Library Using DNA Shuffling;" WO 98/42727 by Pati and Zarling, "Sequence Alterations using Homologous Recombination;" WO 00/18906 by Patten et al., "Shuffling of Codon-Altered Genes;" WO 00/04190 by del Cardayre et al. "Evolution of Whole Cells and Organisms by Recursive Recombination;" WO 00/42561 by Cramer et al., "Oligonucleotide Mediated Nucleic Acid Recombination;" WO 00/42559 by Selifonov and Stemmer "Methods of Populating Data Structures for Use in Evolutionary Simulations;" WO 00/42560 by Selifonov et al., "Methods for Making Character Strings, Polynucleotides & Polypeptides Having Desired Characteristics;" WO 01/23401 by Welch et al., "Use of Codon-Varied Oligonucleotide Synthesis for Synthetic Shuffling;" and WO 01/64864 "Single-Stranded Nucleic Acid Template-Mediated Recombination and Nucleic Acid Fragment Isolation" by Affholter (each of which is incorporated by reference).

[0078] Also provided are recombinant constructs comprising one or more of the nucleic acid sequences as described above. The constructs comprise a vector, such as, a plasmid, a cosmid, a phage, a virus, a bacterial artificial chromosome (BAC), a yeast artificial chromosome (YAC), or the like, into which a polynucleotide sequence of the disclosure has been inserted, in a forward or reverse orientation. In a one embodiment, the construct further comprises regulatory sequences including, for example, a promoter operably linked to the polynucleotide sequence to be expressed. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available.

[0079] Accordingly, in other embodiments, vectors that include a polynucleotide of the disclosure are provided. In other embodiments, host cells transfected with a polynucleotide (e.g.,

SEQ ID NO:1) of the disclosure, or a vector that includes a polynucleotide of the disclosure, are provided. Host cells include eucaryotic cells such as yeast cells, insect cells, or animal cells. Host cells also include procaryotic cells such as bacterial cells.

[0080] The terms "vector", "vector construct" and "expression vector" mean the vehicle by which a DNA or RNA sequence (e.g. a foreign gene) can be introduced into a host cell, so as to transform the host and promote expression (e.g. transcription and translation) of the introduced sequence. Vectors typically comprise the DNA of a transmissible agent, into which foreign DNA encoding a protein is inserted by restriction enzyme technology. A common type of vector is a "plasmid", which generally is a self-contained molecule of double-stranded DNA that can readily accept additional (foreign) DNA and which can readily introduced into a suitable host cell. A large number of vectors, including plasmid and fungal vectors, have been described for replication and/or expression in a variety of eukaryotic and prokaryotic hosts. Non-limiting examples include pKK plasmids (Clontech), pUC plasmids, pET plasmids (Novagen, Inc., Madison, Wis.), pRSET or pREP plasmids (Invitrogen, San Diego, Calif.), or pMAL plasmids (New England Biolabs, Beverly, Mass.), and many appropriate host cells, using methods disclosed or cited herein or otherwise known to those skilled in the relevant art. Recombinant cloning vectors will often include one or more replication systems for cloning or expression, one or more markers for selection in the host, e.g., antibiotic resistance, and one or more expression cassettes.

[0081] The terms "express" and "expression" mean allowing or causing the information in a gene or DNA sequence to become manifest, for example producing a protein by activating the cellular functions involved in transcription and translation of a corresponding gene or DNA sequence. A DNA sequence is expressed in or by a cell to form an "expression product" such as a protein. The expression product itself, e.g. the resulting protein, may also be said to be "expressed" by the cell. A polynucleotide or polypeptide is expressed recombinantly, for example, when it is expressed or produced in a foreign host cell under the control of a foreign or

native promoter, or in a native host cell under the control of a foreign promoter.

[0082] Polynucleotides provided herein can be incorporated into any one of a variety of expression vectors suitable for expressing a polypeptide. Suitable vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, pseudorabies, adenovirus, adeno-associated viruses, retroviruses and many others. Any vector that transduces genetic material into a cell, and, if replication is desired, which is replicable and viable in the relevant host can be used.

[0083] Vectors can be employed to transform an appropriate host to permit the host to express an inventive protein or polypeptide. Examples of appropriate expression hosts include: bacterial cells, such as *E. coli*, *B. subtilis*, *Streptomyces*, and *Salmonella typhimurium*; fungal cells, such as *Saccharomyces cerevisiae*, *Pichia pastoris*, and *Neurospora crassa*; insect cells such as *Drosophila* and *Spodoptera frugiperda*; mammalian cells such as CHO, COS, BHK, HEK 293 br Bowes melanoma; or plant cells or explants, etc.

[0084] As referred to herein, "sequence similarity" means the extent to which nucleotide or polypeptide sequences are related. The extent of similarity between two sequences can be based on percent sequence identity and/or conservation. "Sequence identity" herein means the extent to which two nucleotide or amino acid sequences are invariant. "Sequence alignment" means the process of lining up two or more sequences to achieve maximal levels of identity for the purpose of assessing the degree of similarity (see, e.g., **Fig. 6B**). Numerous methods for aligning sequences and assessing similarity/identity are known in the art such as, for example, the Cluster Method, wherein similarity is based on the MEGALIGN algorithm, as well as BLASTN, BLASTP, and FASTA (Lipman and Pearson, 1985; Pearson and Lipman, 1988). When using all of these programs, the preferred settings are those that results in the highest sequence similarity. For example, the "identity" or "percent identity" with respect to a particular pair of aligned

amino acid sequences can refer to the percent amino acid sequence identity that is obtained by ClustalW analysis (version W 1.8 available from European Bioinformatics Institute, Cambridge, UK), counting the number of identical matches in the alignment and dividing such number of identical matches by the greater of (i) the length of the aligned sequences, and (ii) 96, and using the following default ClustalW parameters to achieve slow/accurate pairwise alignments--Gap Open Penalty: 10; Gap Extension Penalty: 0.10; Protein weight matrix: Gonnet series; DNA weight matrix: IUB; Toggle Slow/Fast pairwise alignments=SLOW or FULL Alignment.

[0085] Two sequences are "optimally aligned" when they are aligned for similarity scoring using, e.g., a defined amino acid substitution matrix (e.g., BLOSUM62), gap existence penalty and gap extension penalty so as to arrive at the highest score possible for that pair of sequences. Amino acid substitution matrices and their use in quantifying the similarity between two sequences are well-known in the art and described, e.g., in Dayhoff *et al.* (1978) "A model of evolutionary change in proteins" in "Atlas of Protein Sequence and Structure," Vol. 5, Suppl. 3 (ed. M. O. Dayhoff), pp. 345-352. Natl. Biomed. Res. Found., Washington, D.C. and Henikoff *et al.* (1992) Proc. Nat'l. Acad. Sci. USA 89: 10915-10919 (each of which is incorporated by reference). The BLOSUM62 matrix is often used as a default scoring substitution matrix in sequence alignment protocols such as Gapped BLAST 2.0. The gap existence penalty is imposed for the introduction of a single amino acid gap in one of the aligned sequences, and the gap extension penalty is imposed for each additional empty amino acid position inserted into an already opened gap. The alignment is defined by the amino acids positions of each sequence at which the alignment begins and ends, and optionally by the insertion of a gap or multiple gaps in one or both sequences so as to arrive at the highest possible score. While optimal alignment and scoring can be accomplished manually, the process is facilitated by the use of a computer-implemented alignment algorithm, e.g., gapped BLAST 2.0, described in Altschul *et al.* (1997) Nucl. Acids Res. 25: 3389-3402 (incorporated by reference herein), and made available to the public at the National Center for Biotechnology Information (NCBI) Website

(www.ncbi.nlm.nih.gov). Optimal alignments, including multiple alignments, can be prepared using, e.g., PSI-BLAST, available through the NCBI website and described by Altschul *et al.* (1997) Nucl. Acids Res. 25:3389-3402 (incorporated by reference herein).

[0086] With respect to an amino acid sequence that is optimally aligned with a reference sequence, an amino acid residue "corresponds to" the position in the reference sequence with which the residue is paired in the alignment. The "position" is denoted by a number that sequentially identifies each amino acid in the reference sequence based on its position relative to the N-terminus. Owing to deletions, insertion, truncations, fusions, etc., that must be taken into account when determining an optimal alignment, in general the amino acid residue number in a test sequence as determined by simply counting from the N-terminal will not necessarily be the same as the number of its corresponding position in the reference sequence. For example, in a case where there is a deletion in an aligned test sequence, there will be no amino acid that corresponds to a position in the reference sequence at the site of deletion. Where there is an insertion in an aligned reference sequence, that insertion will not correspond to any amino acid position in the reference sequence. In the case of truncations or fusions there can be stretches of amino acids in either the reference or aligned sequence that do not correspond to any amino acid in the corresponding sequence.

[0087] A polynucleotide, polypeptide, or other component is "isolated" or "purified" when it is partially or completely separated from components with which it is normally associated (other proteins, nucleic acids, cells, synthetic reagents, etc.). A nucleic acid or polypeptide is "recombinant" when it is artificial or engineered, or derived from an artificial or engineered protein or nucleic acid. For example, a polynucleotide that is inserted into a vector or any other heterologous location, e.g., in a genome of a recombinant organism, such that it is not associated with nucleotide sequences that normally flank the polynucleotide as it is found in nature is a recombinant polynucleotide. A protein expressed *in vitro* or *in vivo* from a recombinant polynucleotide is an example of a recombinant polypeptide. Likewise, a polynucleotide

sequence that does not appear in nature, for example a variant of a naturally occurring gene, is recombinant. For example, an "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term "isolated" includes nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. Typically, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

[0088] As described elsewhere herein the disclosure further provides engineered host cells that are transduced (transformed or transfected) with a vector provided herein (*e.g.*, a cloning vector or an expression vector), as well as the production of polypeptides of the disclosure by recombinant techniques. The vector may be, for example, a plasmid, a viral particle, a phage, *etc.* The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants, and the like. Culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to those skilled in the art and in the references cited herein, including, *e.g.*, Sambrook, Ausubel and Berger, as well as *e.g.*, Freshney (1994) *Culture of Animal Cells: A Manual of Basic Technique*, 3rd ed. (Wiley-Liss, New York) and the references cited therein.

[0089] For example, an Mdh2 polypeptide or variant thereof can be used in a recombinant microorganism as part of a biochemical

pathway for the production of a desired chemical or intermediate. For example, in one embodiment, the microorganism can be an *E. coli* microorganism. In a further embodiment, the *E. coli* expresses, is engineered to express or engineered to overexpress a phosphoketolase in combination with an Mdh2 polypeptide or variant of the disclosure. In another embodiment, the phosphoketolase is Fpk, Xpk or a bifunctional F/Xpk enzyme or homolog thereof. In yet another embodiment of any of the foregoing, the microorganism is engineered to heterologously expresses one or more of the following enzymes: (a) a phosphoketolase; (b) a transaldolase; (c) a transketolase; (d) a ribose-5-phosphate isomerase; (e) a ribulose-5-phosphate epimerase; (f) a hexulose-6-phosphate synthase; (g) a hexulose-6-phosphate isomerase; (h) a dihydroxyacetone synthase; and/or (i) a fructose-6-phosphate aldolase. In yet another embodiment of any of the foregoing, the microorganism expresses, is engineered to express or engineered to overexpress a phosphoketolase derived from *Bifidobacterium adolescentis*. In yet another embodiment of any of the foregoing, the phosphoketolase is a bifunctional F/Xpk. In yet another embodiment of any of the foregoing, the microorganism expresses, is engineered to express or engineered to overexpress a hexulose-6-phosphate synthase. In a further embodiment, the hexulose-6-phosphate synthase is Hps or a homolog thereof. In yet another embodiment of any of the foregoing, the microorganism is engineered expresses or engineered to overexpress a hexulose-6-phosphate isomerase. In a further embodiment, the hexulose-6-phosphate isomerase is Phi or a homolog thereof. In yet another embodiment of any of the foregoing, the microorganism expresses, is engineered to express or engineered to overexpress a dihydroxyacetone synthase. In a further embodiment, the dihydroxyacetone synthase is Das or a homolog thereof. In yet another embodiment of any of the foregoing, the microorganism expresses, is engineered to express or engineered to overexpress a fructose-6-phosphate aldolase. In a further embodiment, the fructose-6-phosphate aldolase is Fsa or a homolog thereof. In yet another embodiment of any of the foregoing, the microorganism expresses, is engineered to express or engineered to overexpress a ribulose-5-phosphate epimerase. In a further embodiment, the

ribulose-5-phosphate epimerase is Rpe or a homolog thereof. In yet another embodiment of any of the foregoing, the microorganism expresses, is engineered to express or engineered to overexpress a ribose-5-phosphate isomerase. In a further embodiment, the ribose-5-phosphate isomerase is Rpi or a homolog thereof. In yet another embodiment of any of the foregoing, the microorganism expresses, is engineered to express or engineered to overexpress a transaldolase. In a further embodiment, the transaldolase is Tal or a homolog thereof. In yet another embodiment of any of the foregoing, the microorganism expresses, is engineered to express or engineered to overexpress transketolase. In a further embodiment, the transketolase is Tkt or a homolog thereof. In yet another embodiment of any of the foregoing, the microorganism expresses, is engineered to express or engineered to overexpress a methanol dehydrogenase comprising an Mdh2 or a variant thereof. In still a further embodiment, the methanol dehydrogenase has at least 70% identity to SEQ ID NO:2 and has methanol dehydrogenase activity. In one embodiment, the Mdh2 comprises a sequence as set forth in SEQ ID NO:2 and has a A26, A31 and/or A169 mutation, wherein the mutation replaces the amino acids at those locations, each independently with V, I or C. In yet another embodiment of any of the foregoing, the microorganism expresses, is engineered to express or engineered to overexpress an alcohol oxidase. In a further embodiment, the alcohol oxidase is Aox or a homolog thereof. In yet another embodiment, the microorganism converts a C1 alcohol to an aldehyde. In yet a further embodiment, the microorganism converts methanol to formaldehyde. In yet another embodiment of any of the foregoing, the microorganism is further engineered to have a reduction or knockout of expression of one or more of ldhA, frdBC, adhE, ackA, pflB, frmA, frmB/yeiG and gapA. In yet another embodiment, the microorganism is further engineered to produce isobutanol or *n*-butanol. In a further embodiment, the microorganism expresses or over expresses a phosphate acetyltransferase that converts acetyl phosphate to acetyl-CoA. In yet a further embodiment, the microorganism produced isobutanol and comprises expression or over expression of one or more enzymes selected from the group consisting of: acetyl-CoA

acetyltransferase, an acetoacetyl-CoA transferase, an acetoacetate decarboxylase) and an adh (secondary alcohol dehydrogenase). In another embodiment, the microorganism comprises one or more deletions or knockouts in a gene encoding an enzyme that catalyzes the conversion of acetyl-coA to ethanol, catalyzes the conversion of pyruvate to lactate, catalyzes the conversion of acetyl-coA and phosphate to coA and acetyl phosphate, catalyzes the conversion of acetyl-coA and formate to coA and pyruvate, or condensation of the acetyl group of acetyl-CoA with 3-methyl-2-oxobutanoate (2-oxoisovalerate). In another embodiment, the microorganism produces n-butanol and comprises expression or over expression of one or more enzymes selected from the group consisting of: a keto thiolase or an acetyl-CoA acetyltransferase activity, a hydroxybutyryl-CoA dehydrogenase activity, a crotonase activity, a crotonyl-CoA reductase or a butyryl-CoA dehydrogenase, and an alcohol dehydrogenase. In another embodiment, the microorganism produces n-butanol and comprises expression or over expression of one or more enzymes that convert acetyl-CoA to malonyl-CoA, malonyl-CoA to Acetoacetyl-CoA, and at least one enzyme that converts (a) acetoacetyl-CoA to (R)- or (S)-3-hydroxybutyryl-CoA and (R)- or (S)-3-hydroxybutyryl-CoA to crotonyl-CoA, crotonyl-CoA to butyryl-CoA, butyryl-CoA to butyraldehyde and butyraldehyde to 1-butanol. In a further embodiment, the microorganism expresses an acetyl-CoA carboxylase and an acetoacetyl-CoA synthase and one or more enzymes selected from the group consisting of (a) hydroxybutyryl CoA dehydrogenase, (b) crotonase, (c) trans-2-enoyl-CoA reductase, and (d) an alcohol/aldehyde dehydrogenase. Pathways that utilize methanol dehydrogenase are further described in International Publication No. WO 2014/153207, which is incorporated herein by reference. The methanol dehydrogenase of the disclosure (e.g., SEQ ID NO:2 and variants thereof) can be used in any of the pathways described in WO 2014/153207.

[0090] As previously discussed, general texts which describe molecular biological techniques useful herein, including the use of vectors, promoters and many other relevant topics, include Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology Volume 152, (Academic Press, Inc., San Diego, Calif.)

("Berger"); Sambrook et al., Molecular Cloning--A Laboratory Manual, 2d ed., Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1989 ("Sambrook") and Current Protocols in Molecular Biology, F. M. Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 1999) ("Ausubel") (each of which is incorporated by reference). Examples of protocols sufficient to direct persons of skill through *in vitro* amplification methods, including the polymerase chain reaction (PCR), the ligase chain reaction (LCR), Q β -replicase amplification and other RNA polymerase mediated techniques (e.g., NASBA), e.g., for the production of the homologous nucleic acids of the invention are found in Berger, Sambrook, and Ausubel, as well as in Mullis et al. (1987) U.S. Pat. No. 4,683,202; Innis et al., eds. (1990) PCR Protocols: A Guide to Methods and Applications (Academic Press Inc. San Diego, Calif.) ("Innis"); Arnheim & Levinson (Oct. 1, 1990) C&EN 36-47; The Journal Of NIH Research (1991) 3: 81-94; Kwoh et al. (1989) Proc. Natl. Acad. Sci. USA 86: 1173; Guatelli et al. (1990) Proc. Nat'l. Acad. Sci. USA 87: 1874; Lomell et al. (1989) J. Clin. Chem 35: 1826; Landegren et al. (1988) Science 241: 1077-1080; Van Brunt (1990) Biotechnology 8: 291-294; Wu and Wallace (1989) Gene 4:560; Barringer et al. (1990) Gene 89:117; and Sooknanan and Malek (1995) Biotechnology 13: 563-564 (each of which is incorporated by reference). Improved methods for cloning *in vitro* amplified nucleic acids are described in Wallace et al., U.S. Pat. No. 5,426,039. Improved methods for amplifying large nucleic acids by PCR are summarized in Cheng et al. (1994) Nature 369: 684-685 and the references cited therein (incorporated by reference herein), in which PCR amplicons of up to 40 kb are generated. One of skill will appreciate that essentially any RNA can be converted into a double stranded DNA suitable for restriction digestion, PCR expansion and sequencing using reverse transcriptase and a polymerase. See, e.g., Ausubel, Sambrook and Berger, all supra.

[0091] The invention is illustrated in the following examples, which are provided by way of illustration and are not intended to be limiting.

EXAMPLES

[0092] All chemicals were purchased from Sigma-Aldrich or Fisher Scientifics unless otherwise specified. KOD Xtreme DNA polymerases were purchased from EMD biosciences (MA, USA). Phusion Hot Start II High-Fidelity DNA polymerases were purchased from Thermo Scientific (MA, USA). DpnI enzymes were purchased from New England Biolabs (MA, USA).

[0093] The complete strain list used is shown in Table 1. *E. coli* XL-1 blue was used as the cloning strain to propagate all plasmids. *C. necator* N-1 strain (ATCC43291) was purchased from ATCC (VA, USA) and the genomic DNA was extracted by Qiagen (CA, USA) DNeasy Blood & Tissue Kit. The insert *mdh1* (CNE_2c07940) gene of pCT23 was amplified from *C. necator* N-1 genomic DNA using primers CT74 and CT75. The insert *mdh2* (CNE_2c13570) gene of pCT20 was amplified from *C. necator* N-1 genomic DNA using primers CT64 and CT65. The insert *nudF* gene of pTW195 was amplified from *E. coli* MG1655 genomic DNA using primers T1478, T1479. The backbones of pTW195, pCT20 and pCT23 were amplified from a modified pZE12-luc (Lutz and Bujard, 1997), pIB4, of which a lacI repressor was included using primers T989 and T990. For pQE9-Act(Bm), the insert *act* gene was amplified from *B. methanolicus* PBl genomic DNA using primers IWB445 and IWB446, whereas vector backbone was amplified from pQE9 acquired from Qiagen (Valencia, CA) using primers IWB094 and IWB141. Polymerase chain reactions (PCRs) were conducted using Phusion Hot Start II High-Fidelity or KOD Xtreme DNA Polymerases, follow with DpnI digestion. The DNA products were purified by Zymo DNA clean&concentrator kit (Zymo Research, CA, USA). The purified backbone and insert were assembled in a 10 µL reaction using isothermal DNA assembly method (Gibson et al., 2009) at 50 °C for 20mins, 5 µL of the reaction mixture was transformed in 50 µL Zymo Z-competent XL-1 blue competent cell (Zymo research) and plated on LB agar plates containing the appropriate antibiotic. Positive transformants were verified by colony PCR and Sanger sequencing.

[0094] **Table 1** List of plasmids and primers used:

Plasmids	Genotype	Reference
pZE12-luc	AmpR ; ColE1 ori ; P _{LlacO-1} :: <i>luc(PP)</i>	(Lutz and Bujard, 1997)
pIB4	AmpR ; ColE1 ori ; P _{LlacO-1} :: <i>fbp(EC)-fxpk(BA)</i> , with lacI	(Bogorad et al., 2013)

pCT20	Amp ^R ; Derivative of pIB4 with <i>mdh2</i> (<i>C. necator</i> N-1)	This Study
pCT23	Amp ^R ; Derivative of pIB4 with <i>mdh1</i> (<i>C. necator</i> N-1)	This Study
pCT20_10C1 2	Amp ^R ; Derivative of pIB4 with <i>mdh2</i> (A31V)	This Study
pCT20_4D8	Amp ^R ; Derivative of pIB4 with <i>mdh2</i> (A169V)	This Study
pCT20_15E9	Amp ^R ; Derivative of pIB4 with <i>mdh2</i> (A26V, A169V)	This Study
pTW212	Amp ^R ; Derivative of pIB4 with <i>mdh2</i> (A26V)	This Study
pCT20_S1	Amp ^R ; Derivative of pIB4 with <i>mdh2</i> (A26V, A31V, A169V)	This Study
pCT20_A169 I	Amp ^R ; Derivative of pIB4 with <i>mdh2</i> (A169I)	This Study
pCT20_A169 L	Amp ^R ; Derivative of pIB4 with <i>mdh2</i> (A169L)	This Study
pCT20_A169 M	Amp ^R ; Derivative of pIB4 with <i>mdh2</i> (A169M)	This Study
pCT20_A169 P	Amp ^R ; Derivative of pIB4 with <i>mdh2</i> (A169P)	This Study
pCT20_A169 C	Amp ^R ; Derivative of pIB4 with <i>mdh2</i> (A169C)	This Study
pQE9- Act(Bm)	Amp ^R ; Derivative of pIB4 with <i>act</i> (<i>B. methanolicus</i> PB1)	This Study
pTW113	Amp ^R ; Derivative of pIB4 with <i>act</i> (<i>B. methanolicus</i> PB1)	This Study
pTW195	Amp ^R ; Derivative of pIB4 with <i>adhA</i> (<i>C. glutamicum</i> 534)	This Study
pMS4	Amp ^R ; Derivative of pIB4 with CNE_1c08460 (<i>C. necator</i> N-1)	This Study
pMS4	Amp ^R ; Derivative of pIB4 with CNE_1c14320 (<i>C. necator</i> N-1)	This Study
pMS4	Amp ^R ; Derivative of pIB4 with CNE_1c04760 (<i>C. necator</i> N-1)	This Study
pMS4	Amp ^R ; Derivative of pIB4 with CNE_1c10080 (<i>C. necator</i> N-1)	This Study
pMS4	Amp ^R ; Derivative of pIB4 with CNE_BB1p03180 (<i>C. necator</i> N-1)	This Study

Primer name	Sequence 5'→3'
T989	TCTAGAGGCATCAAATAAAACGAAA (SEQ ID NO:10)
T990	TCCCTGAAAATACAGGTTTTTCGGAT (SEQ ID NO:11)
T1151	atccgaaaacctgtattttcagggaATGACCACTGCTGCACCCCA (SEQ ID NO:12)
T1152	tttcgtttatttgatgcctctagaTTAGAAACGAATCGCCACAC (SEQ ID NO:13)
T1478	ATCCGAAAACCTGTATTTTCAGGGAATGCTTAAGCCAGACAACCT (SEQ ID NO:14)
T1479	TTTCGTTTTTATTTGATGCCTCTAGATTATGCCCACTCATTTTTTA (SEQ ID NO:15)

IWB094	TCTAGAGGCATCAAATAAAACGAAAGGC (SEQ ID NO:16)
IWB141	TCCCTGAAAATACAGGTTTTTCGGATCCGTGATGGTGATGGTGATG CGATCC (SEQ ID NO:17)
IWB445	TCCGAAAACCTGTATTTTCAGGGAATGGGAAAATTATTTGAGGAA AAAACAATTAAAAC (SEQ ID NO:18)
IWB446	GCCTTTTCGTTTTATTTGATGCCTCTAGATCATTTATGTTTGAGAGC CTCTTGAAGCTGC (SEQ ID NO:19)
CT64	GGATCCGAAAACCTGTATTTTCAGGGAATGACCCACCTGAACATC GCTA (SEQ ID NO:20)
CT65	GAGCCTTTCGTTTTATTTGATGCCTCTAGATTACATCGCCGCAGCG AAGATTGCC (SEQ ID NO:21)
CT74	CGGATCCGAAAACCTGTATTTTCAGGGAATGATCCATGCCTACCA CAACC (SEQ ID NO:22)
CT75	CCTTTCGTTTTATTTGATGCCTCTAGACTAGGCAGACACGGCGCCG ATAAA (SEQ ID NO:23)
CT291	CGAGCAATCATGTGAAGATGNNKATCGTCGACTGGCGTTGCAC (SEQ ID NO:24)
CT292	GTGCAACGCCAGTCGACGATMNNCATCTTCACATGATTGCTCG (SEQ ID NO:25)
MS14	aaaacctgtatttcagggaGAAGTTTATCAAAAGCACTCACATG (SEQ ID NO:26)
MS15	ttttatttgatgcctctagaTCAACGATCAGGCAAGACTCTTTCA (SEQ ID NO:27)
MS16	aaaacctgtatttcagggaCGTCCTGCTTTCGATCCCGAATCCC (SEQ ID NO:28)
MS17	ttttatttgatgcctctagaTCAGGCCGCCAGCAGGTGGTAAAGA (SEQ ID NO:29)
MS33	aaaacctgtatttcagggaATGAAATTCTGCTCGAACTGTGGTC (SEQ ID NO:30)
MS34	ttttatttgatgcctctagaTCAGGGCGTGACCGTGGCCCGGCTG (SEQ ID NO:31)
MS35	aaaacctgtatttcagggaATGTCTTACAAGATCCCGGAATCCG (SEQ ID NO:32)
MS36	ttttatttgatgcctctagaTCATGGCTGCGCTCCGTACACCGCC (SEQ ID NO:33)
MS37	aaaacctgtatttcagggaATGACCGACAAGATCCAACGCGGCA (SEQ ID NO:34)
MS38	ttttatttgatgcctctagaCTATATGGCGTAATGCGGCAGCGGC (SEQ ID NO:35)

[0095] Protein purification and SDS-PAGE. The Mdh1 and Mdh2 were synthesized from His-tag plasmids pCT20 (Mdh2) and pCT23 (Mdh1) in *E. coli* strain XL-1. The XL-1 strains were cultured 16-20 hrs aerobically at 37 °C in Luria-Bertani (LB) media supplemented with appropriate antibiotics. The next day, 1% of overnight culture was inoculated into LB medium with antibiotics and cultured at 37 °C for 2 to 3 hrs until OD₆₀₀ around 0.4-0.8, followed by the addition of 0.1mM IPTG (isopropyl-β-D-thiogalactopyranoside) induction at room temperature (22 - 25 °C) for 16 to 20 hrs. Cells were harvested by centrifugation at 4 °C and either used directly or stored in -80 °C for later protein purification. The purification was conducted with Ni-NTA column using Glycylglycine based buffers at room temperature. Protein concentration was measured by Coomassie Plus Assay (Thermo Scientific) at OD₅₉₅. The purified proteins were analyzed on 12% Mini-PROTEAN TGX gel (Bio-rad, CA, USA) and the gel was stained with SimplyBlue SafeStain (Life technologies, MA, USA).

[0096] Enzyme assays. Mdh activity assays were carried out in a 200 µl assay mixture containing 100 mM sodium bicarbonate buffer (pH 9.5), 30 µg of Mdh, 5 mM of Mg²⁺, 3 mM NAD⁺, and 800 mM of methanol at 30 °C. For C2-C4 alcohol affinity assays, ethanol 10mM and Mdh 1 µg, 1-propanol 5 mM and Mdh 0.5 µg, and n-butanol 100 mM and Mdh 0.5 µg were used instead. For pH assays, buffers pH 6 (2-(N-morpholino)ethanesulfonic acid), pH 7 (potassium phosphate), pH 8.5 (glycylglycine), pH 9.5 (sodium bicarbonate), and pH 10.5 (sodium bicarbonate) were used with the same recipe of Mdh activity assay stated above. For thermal stability assays, the reaction mixture containing everything except the initiating substrate (methanol) was pre-incubated at temperatures ranging from 25 to 60 °C in a Bio-rad PCR machine for 10 minutes before initiating the assay at 30 °C. For temperature activity profile assays, the reaction mixture containing all the components except the enzyme and methanol was pre-incubated at assay temperature for 5 mins before starting the assay. All assays were initiated by adding methanol. 1µg of purified his-tagged ACT or NudF was used to test Mdh activation. It should be noted that NAD⁺ needs to be added before mixing with Mdh, otherwise significant inhibition will be

observed. The activity was defined by the reduction rate of NAD⁺ at OD₃₄₀ using Bio-Tek Eon microplate spectrophotometer. One unit (U) of Mdh activity was defined as the amount of enzyme that converts 1 μmol of substrate into product per minute. The K_m values and V_{max} of Mdh were calculated by Prism 6 (GraphPad Software).

[0097] Nash reaction based screening. Cells were grown overnight in LB medium supplemented with 20 mM MgCl₂, 0.1 mM IPTG and appropriate antibiotics. Nash reagent was prepared by dissolving 5 M ammonium acetate and 50 mM acetylacetone in M9 buffer. Before the assay, cell density was determined by OD₅₉₅. The assay was started by mixing 100 μL of overnight cell culture, 80 μL Nash reagent, and 20 μL 5 M methanol in 96-well plate (#3370, Corning, NY, USA). After 3 hours of incubation in 37 °C shaker (250 rpm), the reaction mixture was centrifuged at 3500 × rpm (Allegra X14-R centrifuge, rotor SX4750, Beckman Coulter, CA, USA) for 10 minutes. 100 μL of supernatant was transferred to a fresh 96-well plate from which OD₄₀₅ measurement was taken. All the OD measurements were accomplished on Victor 3V plate reader (Perkin Elmer, MA). For the quantification of the Nash reaction and cell density OD₅₉₅ and OD₄₀₅ were substituted for OD₆₀₀ and OD₄₁₂, respectively.

[0098] Library Construction and High-Throughput Screening. The random mutagenesis libraries for HTS were constructed by GeneMorph II EZClone mutagenesis kit (Agilent, CA, USA) following manufacturer's protocol. In short, 10 ng of parent Mdh DNA was used as the template for primers CT64 and CT65 for error-prone PCR. The resulting error rate was 2 nt/kb. The error prone PCR product was gel-purified (Zymoclean gel DNA recovery kit, Zymo Research) and assembled to a backbone based on pCT20. The assembled library was transformed to the *E. coli* strain DH10B (Life Technologies, CA, USA) by electroporation and plated on Bioassay QTrays (Molecular Devices, CA) containing 200 mL LB agar (1.5% w/vol). From the Bioassay QTrays, single colonies were picked by a QBot colony picker (Molecular Devices, CA, USA) and inoculated into 96-well low profile plates (X6023, Molecular Devices, CA) containing 150 μL of LB supplemented with ~15% (v/v) glycerol, 1% (w/vol) glucose and appropriate antibiotics. As positive control, 96 colonies

containing the wild-type Mdh or parent Mdh was picked into a single 96-well plate and processed together with other plates. Similarly, colonies containing *E. coli* transketolase (Tkt) was used as the negative control. The plates were covered with a lid and grown overnight in a 37 °C the incubator. Subsequently, plates were used to re-inoculate a fresh 96-well plate (#3370, Corning) filled with 200 µL LB supplemented with 20 mM MgCl₂, 0.1 mM IPTG and appropriate antibiotics. The new plate was sealed with aluminum sealing film (#6569, Corning) and incubated in 37 °C shaker (250 rpm), while the old plate was kept in -80 °C as stock. After about 16 hours of growth, the culture plates were transferred to the BenchCel 4R system with Vprep Velocity11 liquid handler (Agilent, CA, USA) using a 96 LT head. The cells were gently resuspended and 100 µL of the samples were aliquoted to a fresh 96-well plate (#3370, Corning). Cell density was assessed by OD₅₉₅ at this point. After the measurement, Mdh variants were assessed by the Nash reaction in a 96-well format at 405nm using the Victor 3V plate reader as above.

[0099] Site-saturation mutagenesis. The site-saturation mutagenesis on Mdh2 A169 site was constructed by Quikchange II site-directed mutagenesis kit (Agilent, CA) with primers CT291 and CT292. The degenerate codons on the primers generate all possible amino acid substitutions. The library was transformed to *E. coli* DH10β strain and single colonies containing all 19 amino acid substitution variants were isolated for further analysis.

[00100] Mdh2 Sequence Analysis. The Mdh2 amino acid sequence was uploaded to SWISS-MODEL web server ([<http://swissmodel.expasy.org/>]) (Guex et al., 2009), which performed the structure analysis and generated a 2D plot to present Mdh2 homologs of existing protein structure repository. The Mdh2 and its homologs were aligned using T-coffee ([<http://www.tcoffee.org/Projects/tcoffee/>]) (Notredame et al., 2000) and visualized by ESPript 3.0 web server (ESPrpt - [<http://esprpt.ibcp.fr>]) (Robert and Gouet, 2014).

[00101] Expression, purification and characterization of *C. necator* N-1 Mdh2. To determine whether the two putative Mdhs in *C. necator* N-1, coded by *mdh1* and *mdh2* genes, exhibit catalytic

activity towards methanol, these genes were cloned and expressed from the His-tag plasmid pCT20 (Mdh2) and pCT23 (Mdh1) in *E. coli* XL-1, and the Mdh1 and Mdh2 proteins were purified. SDS-PAGE analysis showed both the purified Mdh1 and Mdh2 were detected with molecular masses of approximately 40 kDa, which is close to the predicted sizes 38.8 kDa and 40.7 kDa for Mdh1 and Mdh2, respectively (**Fig. 7**). To test whether Mdh1 or Mdh2 shows the desired activity, a methanol dehydrogenase activity assay was performed by monitoring NAD(P) reduction. Mdh1 methanol-linked oxidation was not observed when using either NAD⁺ or NADP⁺ the electron acceptor. However, Mdh2 showed significant specific activity 0.32 U/mg (Table 2) when NAD⁺ was used as the electron acceptor, whereas no methanol oxidation activity was detected when NADP⁺ was used. The K_m values of Mdh2 for methanol and NAD⁺ were 132 mM and 2.2 mM, respectively. The specific activity and K_m values of Mdh2 at 30 °C without ACT were comparable to the ACT activated *B. methanolicus* Mdhs at 45 °C (Krog et al., 2013). Examination of the catalytic activity of Mdh2 to C1-C4 alcohols, showed that Mdh2 exhibits broad substrate specificity, with highest specificity towards 1-propanol and low affinity to methanol, similar to previously reported Mdhs (**Table 2**) (Krog et al., 2013).

[00102] Table 2 Substrate specificity to C1-C4 alcohols and kinetic constants of recombinant Mdh2 in vitro:

Substrate	K_m (mM)	K_{cat} (s ⁻¹)	K_{cat}/K_m (M ⁻¹ s ⁻¹)	V_{max} (U/mg)	V_{max} (U/mg) (MGA3 Mdh3, 45 °C)
Methanol	132.1±15.4	0.22±0.01	1.6	0.32	0.07
Ethanol	0.77±0.1	11.1±0.3	14483	16	1.3
Propanol	0.54±0.1	9.6±0.2	17759	14.2	2.8
<i>n</i> -butanol	7.2±1	6.5±0.2	906	9.6	2.6
NAD ⁺	2.1±0.2	0.36±0.02	170	0.54	-
NADP ⁺	N.D.	N.D.	N.D.	N.D.	-

[00103] The values shown indicate mean ± standard deviation. Triplicate experiments were performed. Mdh assays to determine K_m for different alcohols were performed using various alcohol concentrations and 3 mM NAD⁺ as substrates at 30 °C and pH 9.5. To determine kinetic parameters of NAD⁺, 800mM of methanol was used

and the rest of conditions remained unchanged. The $V_{\max}$ values of *B. methanolicus* MGA3 Mdh3 were obtained from Krog et al., 2013.

[00104] Effects of pH, temperature, and ions on Mdh2. Further characterization was conducted using methanol as a substrate to investigate the effect of pH, thermal stability, and different metal ions on Mdh2. As shown in **Fig. 1A**, the Mdh2 was active from pH 6 to 10.5, with its optimum at pH9.5. The thermal stability assay revealed that Mdh2 enzyme was inactivated (**Fig. 1B**) when incubating for 10 min at temperatures higher than its physiological growth condition (30 °C). In particular, 13% of activity remained after pre-incubation at 55 °C (**Fig. 1B**) and the activity was abolished at 60 °C. To detect the enzyme sensitivity to various metal ions and EDTA, the assay was performed in the presence of these additives. **Fig. 2** shows the activity of Mdh2 was activated by the addition of 1 mM Ni^{2+} , and was strongly inhibited by 0.1 mM of Cu^{2+} or Zn^{2+} .

[00105] Insensitivity of Mdh2 to ACT. Activation of type III Adhs by the activator protein ACT or its homolog Nudix hydrolase is common, and may even be general for all enzymes in this class (Ochsner et al., 2014). The activation results in drastic improvement in $V_{\max}$ and K_m . For instance, six Mdhs of *B. methanolicus* PB1 and MGA3 can be strongly activated by ACT (Krog et al., 2013) at their physiological temperature, 45 °C. To test if Mdh2 from *C. necator* N-1 can be activated by ACT, a his-tagged thermophilic ACT from *B. methanolicus* PB1 and its mesophilic homolog Nudix hydrolase NudF from *E. coli* was cloned and purified (Ochsner et al., 2014). The ACT and NudF were used to activate Mdh2 at various temperatures from 25-65 °C. Since *B. methanolicus* PB1 ACT is a thermophilic enzyme, the *E. coli* NudF was used to ensure the activator protein was active under mesophilic temperatures. Mdh3 of *B. methanolicus* MGA3, which was previously shown to be activated by ACT (Krog et al., 2013), served as a positive control. Interestingly, Mdh2 was largely insensitive to ACT and NudF between 25-40 °C (**Fig. 3A**). The activity of Mdh2 was mildly increased by ACT or NudF at 55°C and 60 °C where it reached the optimum specific activity at 55 °C with 70% improvement. In contrast, Mdh3 was significantly activated when assay temperature was above 42°C. At

its optimum temperature 60 °C, the specific activity improved more than 15 fold to 0.35 U/mg (**Fig. 3B**).

[00106] To verify if Mdh2 is insensitive to the Nudix proteins from *B. methanolicus* and *E. coli*, the activator concentration was varied by 10-fold (upto 50µg/mL). No activation effect was observed under the conditions tested (**Fig. 9A**). In addition, K_m and K_{cat} for methanol and NAD⁺ remained unchanged in the presence of *B. methanolicus* ACT or *E. coli* NudF at 30 °C (**Table 3**). Next, Mdh2 activity was assayed using purified Mdh2 incubated with *C. necator* N-1 crude extracts at concentrations 50 and 150 µg/mL. However, no activity improvement was observed (**Fig. 9B**). To investigate this possibility further, five Nudix family proteins annotated in the *C. necator* N-1 genome: two hydrolase family proteins (coded by CNE_BB1p03180, CNE_1c08460) and three pyrophosphatases (code by CNE_1c14320, CNE_1c04760, CNE_1c10080) were identified. A BLAST analysis was also used to identify *B. methanolicus* ACT or *E. coli* NudF homologs in *C. necator* N-1 and obtained no additional possibilities. These five putative individual proteins were his-tag cloned, expressed and purified using *E. coli* for the activation tests. Consistent with crude extract results, none of these putative Nudix proteins can activate Mdh2 (**Fig. 9B**).

[00107] Table 3: Effect of activator proteins on kinetic parameters of recombinant Mdh2 in vitro.

Substr.	Activator protein	K_m (mM)	K_{cat} (s ⁻¹)	K_{cat}/K_m (M ⁻¹ s ⁻¹)	V_{max} (U/mg)
Methanol	--	132.1±15.4	0.22±0.01	1.6	0.32
Methanol	<i>B. methanolicus</i> ACT	100.7±16.1	0.22±0.01	2.2	0.33
Methanol	<i>E. coli</i> NudF	122.1±18.2	0.23 ± 0.01	1.9	0.35
NAD ⁺	--	0.93±0.079	0.24±0.005	258	0.36
NAD ⁺	<i>B. methanolicus</i> ACT	0.93±0.11	0.20±0.006	215	0.30
NAD ⁺	<i>E. coli</i> NudF	1.3±0.33	0.21±0.02	162	0.31

[00108] Development of automatic high throughput screening (HTS) for Mdh evolution. As Mdh2 exhibited significant methanol activity without activator protein in mesophilic conditions, this enzyme represents a promising choice for engineering synthetic methylotrophy. However, the activity and substrate specificity

remain low for methanol (**Table 2**). To solve this problem, Mdh2 was engineered for better performance. An NAD⁺ binding site mutation S97G of *B. methanolicus* C1 Mdh had been shown to increase activity and reduce K_m toward methanol significantly. However, the mutation of the corresponding residue (S106G) on Mdh2 did not show methanol oxidation activity. Another group modified *B. stearrowthermophilus* Adh, which has methanol oxidation activity, to become hydrogel forming enzyme by outfitting it with cross-linking domains (Kim et al., 2013). Although the modification remarkably increased the *in vitro* methanol oxidation activity, the feasibility of applying hydrogel forming enzymes in metabolic engineering needs to be further investigated.

[00109] To engineer Mdh, an automatic high throughput screening (HTS) strategy was developed based on automatic liquid handling, colony picking, incubation, and whole cell assay without lysis. Nash reaction (Nash, 1953) allows Mdh assay without cell lysis, which detects formaldehyde produced from methanol oxidation by reacting with acetylacetone and ammonium acetate. The reaction product, diacetyldihydro-lutidine, exhibits yellow color and can be quantified by absorbance at 405 nm (**Fig. 4A**). Since formaldehyde is able to diffuse through the cell membrane, Nash reaction-based screening does not require cell lysis. This greatly simplified the screening procedure, and bypassed the background interference in cell crude extracts.

[00110] The scale of the screening was enhanced with utilizing automated colony picker and liquid handler. After integrating all equipment into the work flow, the initial design was capable to screen 6,000 colonies in a single round using 384-well plates to carry the samples. The readout of Nash reaction was normalized to cell density (OD₅₉₅). Although the process successfully displayed Mdh activity in colorimetric reading, no improved Mdh was obtained from the first few testing rounds due to high false positive rate. The setback prompted us to inspect the screening accuracy of the initial design. Zhang et al. had developed a standard measure to evaluate and validate the quality of HTS assays (Zhang, 1999). The so-called Z'-factor is a statistical characteristic of any given assay with the value between 0 to 1. The Z'-factor was calculated

from the positive control and negative control of an assay, a value larger than 0.5 indicates a large separation between the populations of the measured signals and the assay will be considered as high quality. To evaluate our HTS system, strains containing wild-type Mdh2 or transketolase (Tkt) was tested as positive and negative controls, respectively. 384 single colonies of each control were picked and assayed by Nash reaction. The resulting Z'-factor was 0.23 (**Fig. 4B**), suggesting the low quality of the initial HTS design.

[00111] While revisiting the details of the process, it was hypothesized that the small well dimension of 384-well plates might be constraining mixing during Nash reaction even with shaking. To test the hypothesis that the inaccuracy of the system originated from mixing issue during Nash reaction, 96-well plates were used to replace 384-well plate in the HTS process. After adjusting the process according to the new plate, the Z'-factor improved to 0.76 (**Fig. 4B**). This Z' indicated that the HTS system was suitable for screening for Mdh mutants of high activity and was properly validated. The optimized HTS process is shown in **Fig. 4C**.

[00112] **Directed evolution of Mdh2.** The Mdh evolution was started with error-prone PCR-generated library using the wild-type *mdh2* from *C. necator* N-1 as the template. The first round of screening generated 8 possible positive variants with at least 50% activity improvement based on Nash reaction out of 2000 variants screened. These variants were sequenced and tested by NADH-based assay for crude extract activity to eliminate the false positives. Variants CT1-1 and CT1-2 displayed the highest improved activity based on the crude extract assay, and were selected for purification and further characterization. Purified variant CT1-2 showed a 5-fold decrease in K_m , while CT1-1 improved marginally in K_m and K_{cat} (**Table 4**). However, K_{cat} of CT1-2 decreased by almost 50% compares to the wild-type.

[00113] In the second round of screening, CT1-2 was used as the parent to generate another error-prone PCR library. Seven possible positive variants with at least 70% activity improvement by Nash reaction were obtained from total of 2,000 screened. After confirmation by sequencing and crude extract activity assay, only

variant CT2-1 variant was selected for characterization. Variant CT2-1 restored wild-type K_{cat} while maintaining the K_m improvement (**Table 4**). In addition to the mutation A169V originated from the previous screen, CT2-1 included another mutation, A26V. To determine the effect of A26V, this mutation was introduced into the wild-type *mdh2*, and its effect determined after purification. Interestingly, the mutation A26V alone demolished Mdh activity (**Table 4**), suggesting a synergistic effect of mutation A26V and A169V in enzyme function.

[00114] After these rounds of HTS, a chimeric variant CT4-1 was created by recombining three mutations found so far (A169V, A31V, and A26V). The K_m value of methanol was further lowered to 21.6 mM and K_{cat} remained unchanged (**Table 4**). Variant CT4-1 represented the best performing variant from the series of engineering with about 6-fold higher K_{cat}/K_m ratio towards methanol compared to the wild-type.

[00115] **Table 4** Kinetic Parameters of engineered Mdh2 variants to methanol and n-butanol, using NADH as cofactor:

Mdh2 variant	Library	Mutation	Nash activity	methanol			n-butanol		
				K_m (mM)	K_{cat} (s^{-1})	K_{cat}/K_m ($M^{-1} s^{-1}$)	K_m (mM)	K_{cat} (s^{-1})	K_{cat}/K_m ($M^{-1} s^{-1}$)
WT	-	-	1.00	132±15.4	0.22±0.006	1.6	7.2±2.1	6.5±0.1	903
CT1-1	Round 1	A31V	1.45	129±11.9	0.25±0.01	1.9	11.3±1.3	7.2±0.2	637
T1-2	Round 1	A169V	1.47	26.9±2.7	0.13±0.003	4.8	66.5±11.8	3.3±0.1	50
CT2-1	Round 2	A26V, A169V	2.96	30.7±3.6	0.21±0.01	6.8	75.9±1.1	5.0±0.2	66
CT2-2	-	A26V	ND	ND	ND	ND	ND	ND	ND
CT4-1	Recomb	A26V, A31V, A169V	3.46	21.6±1.5	0.20±0.01	9.3	120±17.9	5.7±0.4	48

[00116] **Substrate specificity of the evolved Mdh2.** To characterize the kinetic parameters of Mdh variants toward longer chain alcohols, n-butanol was chosen as an example to measure K_m and K_{cat} . Results indicates that K_m values for n-butanol were increased by 10-fold or higher (**Table 4**) for variants CT1-2, CT2-1,

CT4-1. The increased K_m towards *n*-butanol was concomitant with the decrease of K_m towards methanol. The best variant, CT4-1, displayed the most significant 19-fold decrease in K_{cat}/K_m towards *n*-butanol among all variants. To further investigate substrate preferences on other higher alcohols, the specific activities towards ethanol and propanol were measured at the concentrations that saturate wild-type Mdh2 activity (**Fig. 5A**). Consistent with the *n*-butanol data, variants CT1-2, CT2-1 and CT4-1 showed 5 to 10-fold lower specific activity towards ethanol and 6 to 8-fold lowered for propanol. As summarized in **Fig. 5B**, CT4-1 significantly improved its methanol over C2 to C4 alcohol activity ratio compares to wild-type.

[00117] Sequence Analysis. The Mdh2 amino acid sequence was uploaded to SWISS-MODEL server (Guex et al., 2009) to predict a hypothetical model based on structural information in the database. The server returned a plot of sequence similarity as shown in **Fig. 6A**. The most structurally similar enzymes were *K. pneumoniae* 1,3-propanediol dehydrogenase (1,3-PDH) and *Zymomonas mobilis* ZM4 alcohol dehydrogenase 2 (Adh2) with 55% and 54% sequence identities, respectively. Both enzymes were categorized as group III metal-dependent dehydrogenases and contained Fe^{2+} in their catalytic centers (Marçal et al., 2009; Moon et al., 2011). Notably, Mdh2 also has 55% sequence identity to *B. methanolicus* MGA3 Mdh, which also belongs to group III dehydrogenases. The structure of group III dehydrogenases can be divided into N-terminal domain and C-terminal domains, which are responsible for $NAD(P)^+$ and metal ion binding, respectively. The metal ion coordination motif composed mainly of 2 to 3 histidine residues (Carpenter et al., 1998; Marçal et al., 2009; Montella et al., 2005; Moon et al., 2011; Ruzheinikov et al., 2001). In *K. pneumoniae* 1,3-PDH and *Z. mobilis* ZM4 Adh2, there were three histidine and one aspartic acid. These four amino acid residues were conserved in all of the enzymes aligned (**Fig. 6B**), corresponding to Asp201, His205, His270, and His284 in Mdh2. On the other hand, the NAD-binding motif (GGGSX₂DX₂K; (SEQ ID NO:36) was also observed in the alignment (**Fig. 6B**) (Wierenga et al., 1986). Taken together, the similarity of both amino acid sequences and functional domains indicated that Mdh2 belongs to group III metal-

dependent dehydrogenases. Previously, NAD-dependent Adhs from *L. sphaericus* C3-41, *L. fusiformis* ZC1, and *D. hafniense* Y51 were identified with methanol oxidation activity (Muller et al., 2015). Alignment with other Mdhs revealed that these methanol-oxidizing dehydrogenases shared the common NAD-binding domain and metal coordination motif (**Fig. 6B**).

[00118] Among the mutations acquired during Mdh2 evolution, A169V contributed significantly to the K_m decrease in methanol. The same mutation also greatly reduced the activity toward C2-C4 aliphatic alcohols. *Z. mobilis* Adh2 is one of the most structurally-similar enzymes to Mdh2 and its binding pocket had been predicted (Moon et al., 2011). Residue A169 is one of the predicted binding pocket residues, which were conserved between *Z. mobilis* Adh2 and Mdh2 (**Fig. 6B**). Therefore, it was hypothesized that the change of alanine to bulkier valine reduces the binding space and subsequently hinders larger substrate binding. To test this hypothesis and explore the best possible amino acid substitution at A169, a site-saturation mutagenesis library of A169 was constructed. The specific activities of all 19 variants were measured by Nash reaction (**Fig. 8**), six best variants were selected for characterization. Although three of the variants with bulkier side chains (A169V, A169I, A169C) showed lower K_m for methanol, the others displayed the opposite (**Table 4**). Presumably the K_m improvement is determined by both size and functional groups in the amino acid side chain. In agreement with this note, extremely large (Phe, Trp, Tyr) or small (Gly) amino acids at A169 showed no activity (**Fig. 8**). Interestingly, A169P displayed higher K_{cat} as well as K_m (**Table 3**). Although protein structure characterization would be required to define the role of residue A169, the results here showed that A169 is crucial to Mdh2 activity and substrate preference.

[00119] **Table 3** Effect of A169 replacement to Mdh2 methanol specificity:

Mdh2 variant	K_m (mM)	K_{cat} (S^{-1})	K_{cat}/K_m ($M^{-1}S^{-1}$)	K_{cat}/K_m Fold change
WT	132±15.4	0.22±0.01	1.6	-
A169V	26.9±2.65	0.13±0.003	4.8	3

A169I	68.9±9.0	0.064±0.002	0.93	0.58
A169L	201±31.5	0.093±0.004	0.46	0.29
A169M	203±25.5	0.073±0.002	0.36	0.23
A169P	303±34.9	0.41±0.014	1.4	0.88
A169C	46.0±3.34	0.16±0.003	3.5	2.2

[00120] A number of embodiments of the invention have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the invention. Accordingly, other embodiments are within the scope of the following claims.

WHAT IS CLAIMED IS

1. An isolated recombinant polypeptide comprising a sequence that has at least 70%, 80%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO:2 and comprises at least one mutation at a residue selected from the group consisting of A26, A31, and A169 and has improved methanol dehydrogenase activity compared to a polypeptide consisting of SEQ ID NO:2.
2. The isolated recombinant polypeptide of claim 1, wherein the polypeptide is engineered from *C. necator* N-1.
3. The isolated recombinant polypeptide of claim 1, wherein each of the at least one mutations independently comprise a substitution with V, I or C.
4. The isolated recombinant polypeptide of claim 1, wherein each of the at least one mutations comprise a substitution with V.
5. The isolated recombinant polypeptide of claim 1, wherein the polypeptide is at least 70% identical to SEQ ID NO:2 and has a mutation selected from the group consisting of A26V, A31V, A169V and any combination thereof.
6. The isolated recombinant polypeptide of claim 1, wherein the polypeptide comprises a catalytic efficiency (K_{cat}/K_m) of greater than 1.6 for methanol and/or less than 903 for n-butanol.
7. The isolated recombinant polypeptide of any one of claims 1 to 6, wherein the isolated recombinant polypeptide comprises from 1-10 conservative amino acid substitutions.
8. An isolated nucleic acid encoding the isolated recombinant polypeptide of any one of claims 1 to 6.
9. A vector comprising the nucleic acid molecule of claim 8.
10. The vector of claim 10, which is an expression vector.

11. A host cell transfected with an isolated nucleic acid of claim 8.
12. A host cell transfected with a vector of claim 9.
13. The host cell of claim 11, wherein the cell is prokaryotic.
14. The host cell of claim 11, wherein the cell is eukaryotic.
15. A recombinant host cell that has been genetically engineered to express a heterologous polypeptide comprising at least 70% sequence identity to SEQ ID NO:2 and which has methanol dehydrogenase activity.
16. A recombinant vector comprising a polynucleotide encoding a polypeptide of SEQ ID NO:2.
17. A recombinant host cell comprising a metabolic pathway for the production of chemical not normally produced by or not normally produced in an amount in a wild-type microorganism, wherein the microorganism has been engineered to express a polypeptide that is at least 70% identical to SEQ ID NO:2 and has methanol dehydrogenase activity.
18. A recombinant host cell of claim 17, wherein the polypeptide is at least 95% identical to SEQ ID NO:2 and has at least one mutation at a residue selected from the group consisting of A26, A31, and A169 and has improved methanol dehydrogenase activity compared to a polypeptide consisting of SEQ ID NO:2.
19. The recombinant host cell of claim 18, wherein each of the at least one mutations independently comprise a substitution with V, I or C.
20. The recombinant host cell of claim 18, wherein each of the at least one mutations comprise a substitution with V.

21. The recombinant host cell of claim 18, wherein the host cell produces acetyl-phosphate through carbon rearrangement of E4P and/or G3P and metabolizes methanol.

22. The recombinant host cell of claim 18, wherein the host cell expresses (i) an enzyme having phosphoketolase activity and (ii) (a) a hexulose-6-phosphate synthase and hexulose-6-phosphate isomerase, or (b) dihydroxyacetone synthase and a fructose-6-phosphate aldolase and one or more of the following enzymes:

- (a) a transaldolase (Tal);
- (b) a transketolase (Tkt);
- (c) a ribose-5-phosphate isomerase (Rpi);
- (d) a ribulose-5-phosphate epimerase (Rpe); and
- (e) a methanol dehydrogenase (Mdh).

23. The recombinant host cell of claim 18, wherein expresses enzymes that catalyze the conversion described in (i)-(x):

- (i) the production of acetyl-phosphate and erythrose-4-phosphate (E4P) from fructose- 6-phosphate and/or the production of acetyl-phosphate and glyceraldehyde 3-phosphate (G3P) from xylulose 5-phosphate;
- (ii) the reversible conversion of fructose- 6-phosphate and E4P to sedoheptulose 7-phosphate (S7P) and (G3P);
- (iii) the reversible conversion of S7P and G3P to ribose-5-phosphate and xylulose-5-phosphate;
- (iv) the reversible conversion of ribose-5-phosphate to ribulose-5-phosphate;
- (v) the reversible conversion of ribulose-5-phosphate to xylulose-5-phosphate;
- (vi) the reversible conversion of xylulose-5-phosphate and E4P to fructose- 6-phosphate and glyceraldehyde-3-phosphate;
- (vii) the conversion of formaldehyde and ribulose-5-phosphate to D-arabino-3-Hexulose 6-phosphate;
- (viii) the reversible conversion of D-arabino-3-Hexulose 6-phosphate to fructose-6-phosphate;

(ix) the conversion of formaldehyde and xylulose-5-phosphate to glyceraldehyde-3-phosphate and dihydroxyacetone; and

(x) the conversion of glyceraldehyde-3-phosphate and dihydroxyacetone to fructose-6-phosphate;

wherein the microorganism produces acetyl-phosphate, or compounds derived from acetyl-phosphate using methanol.

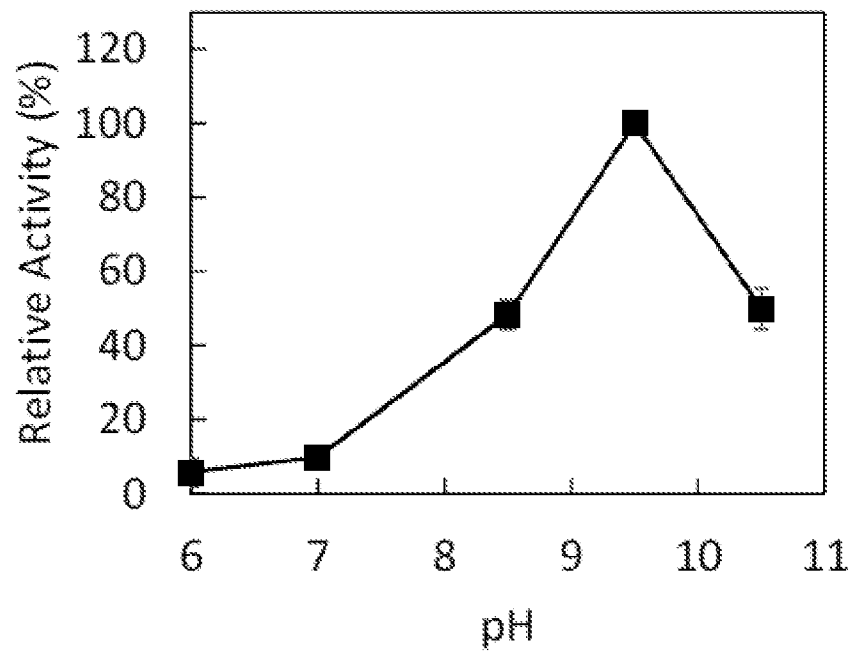


FIG. 1A

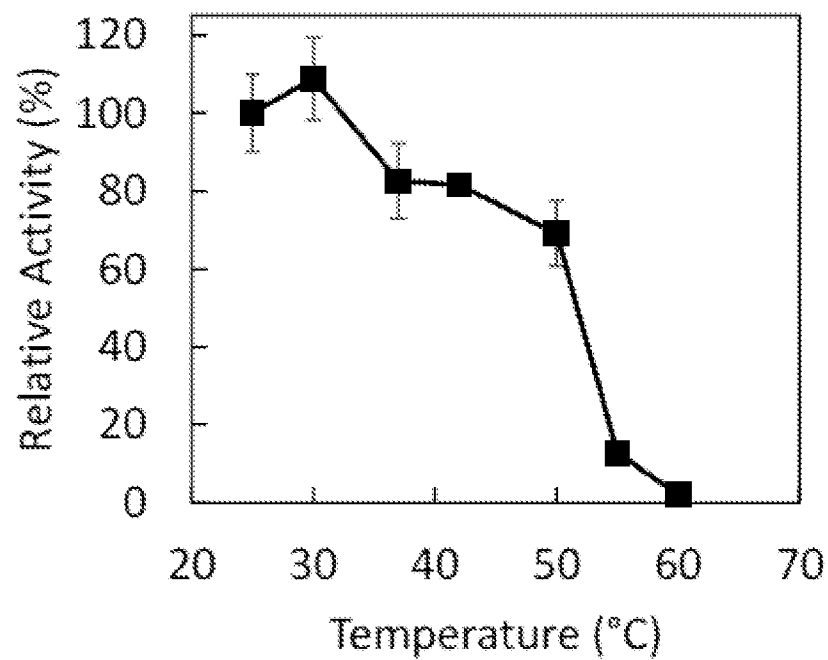


FIG. 1B

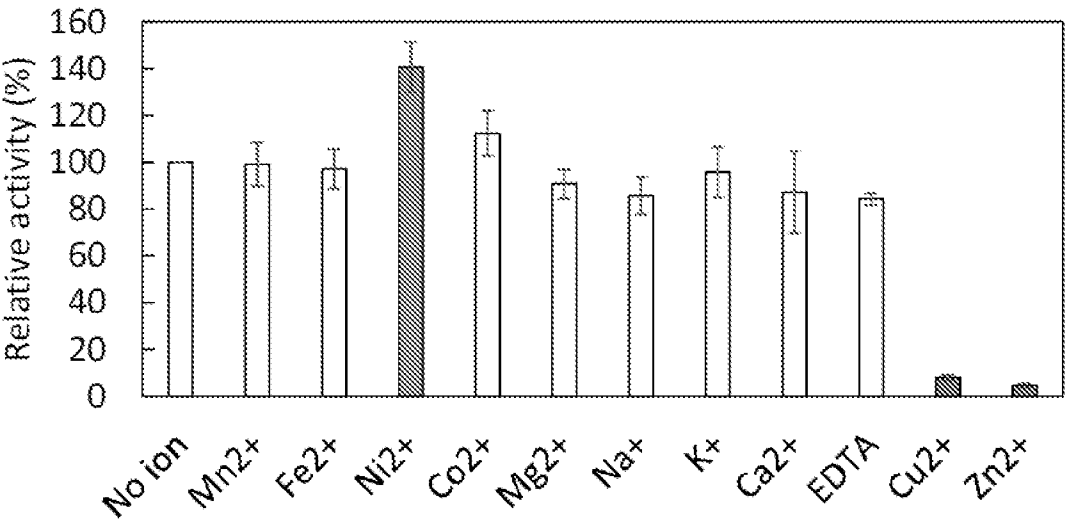


FIG. 2

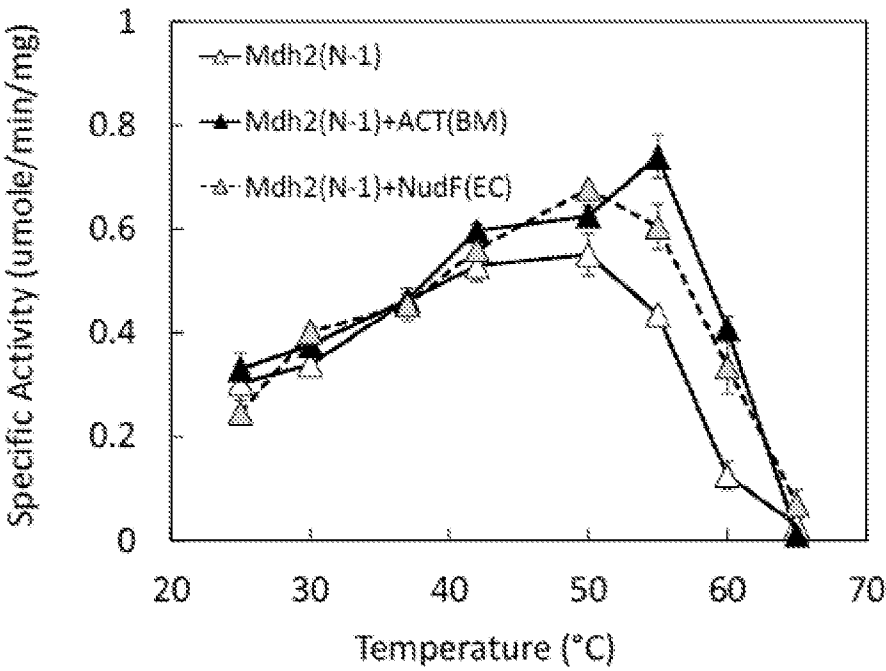


FIG. 3A

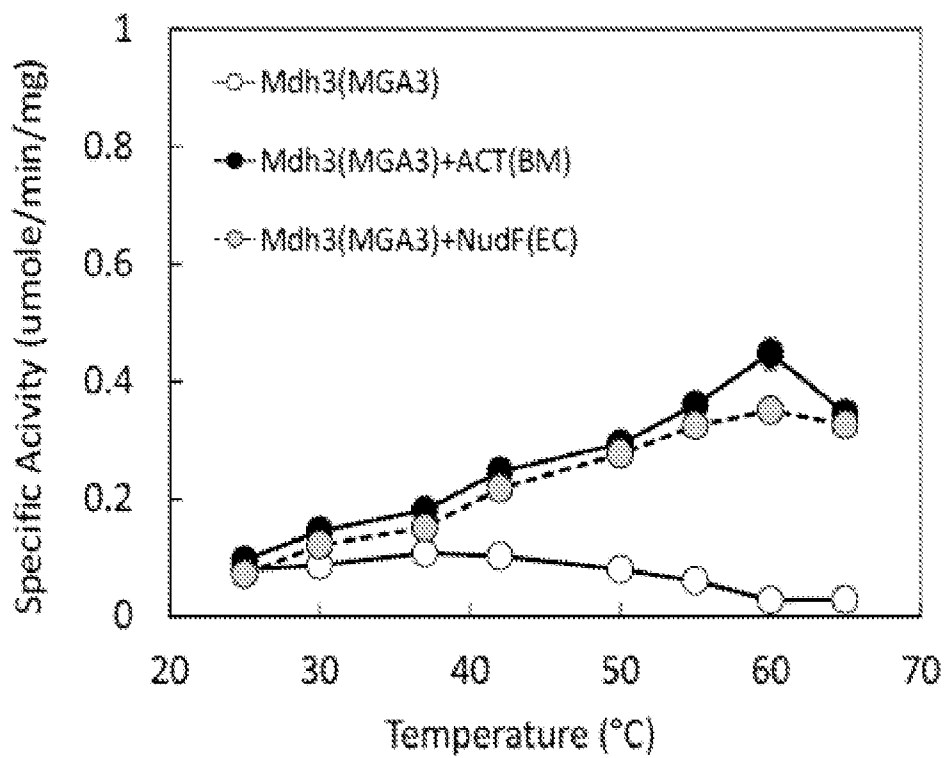


FIG. 3B

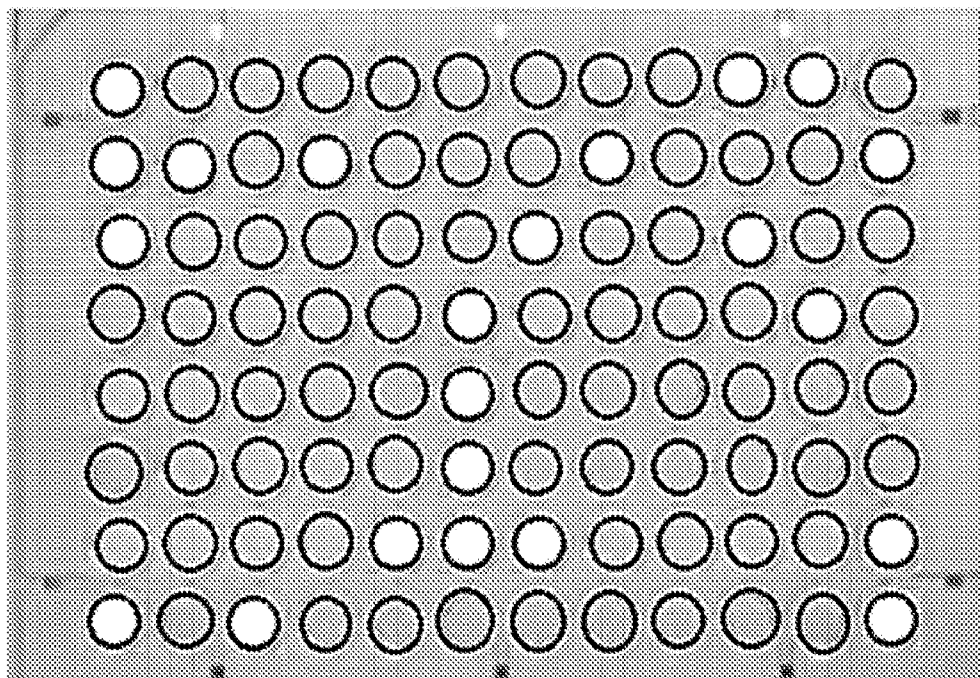


FIG. 4A

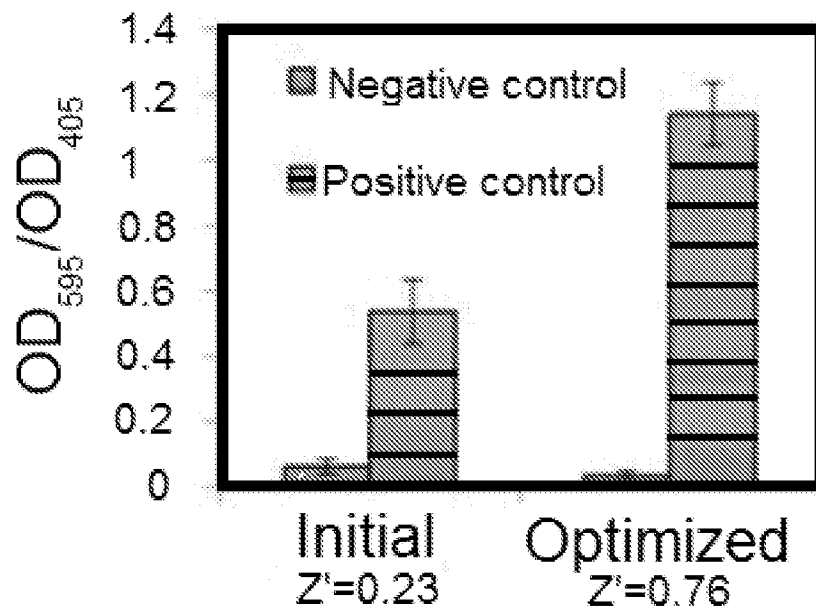


FIG. 4B

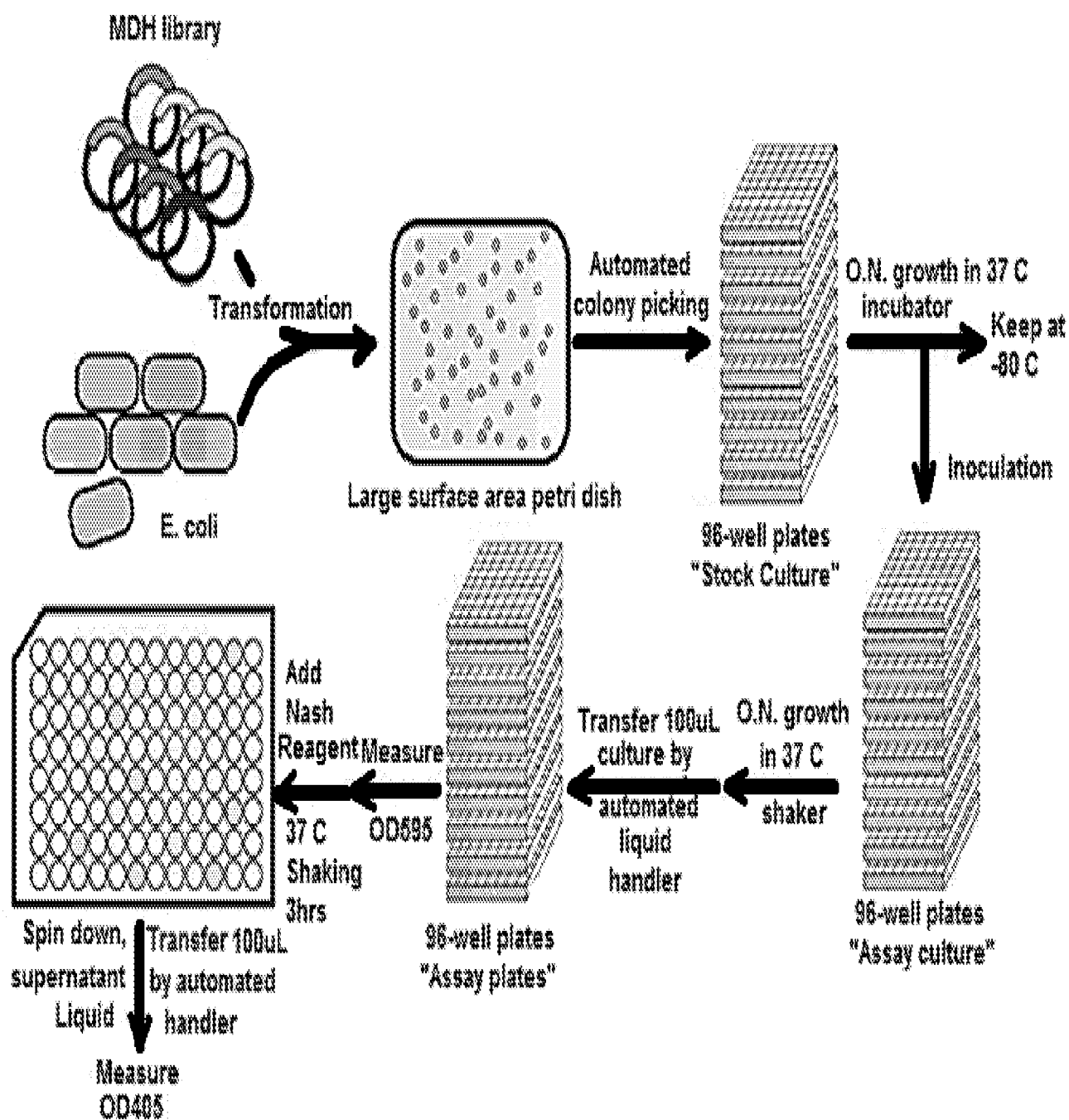


FIG. 4C

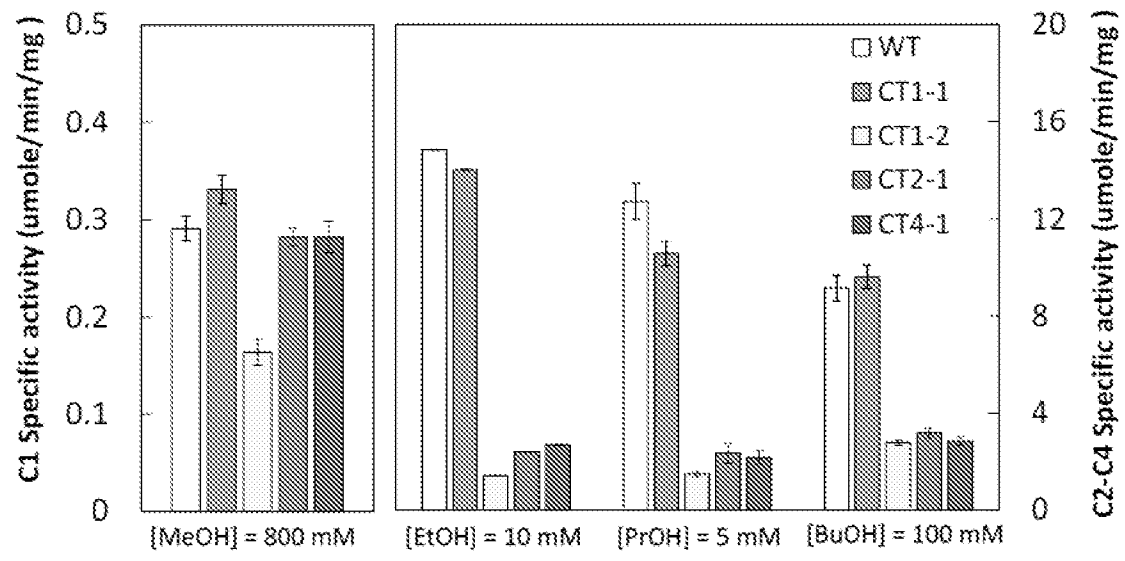


FIG. 5A

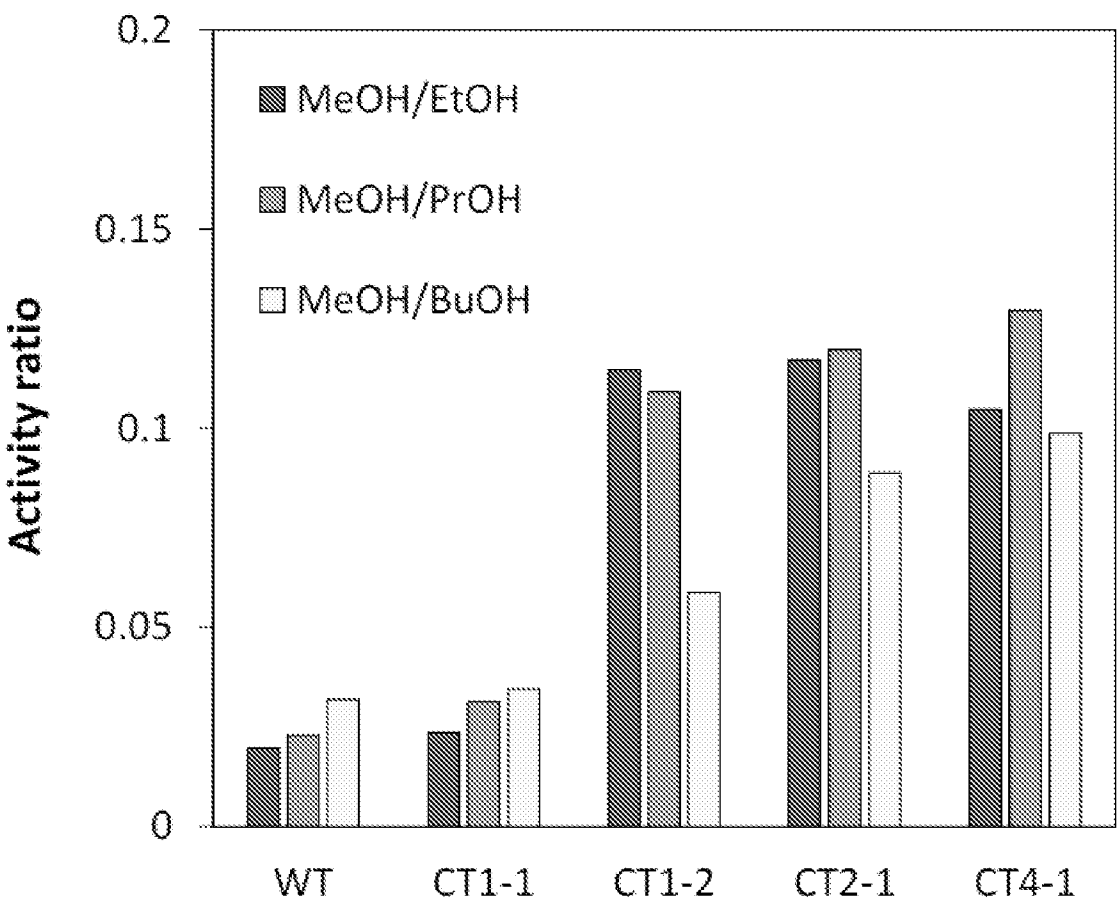


FIG. 5B

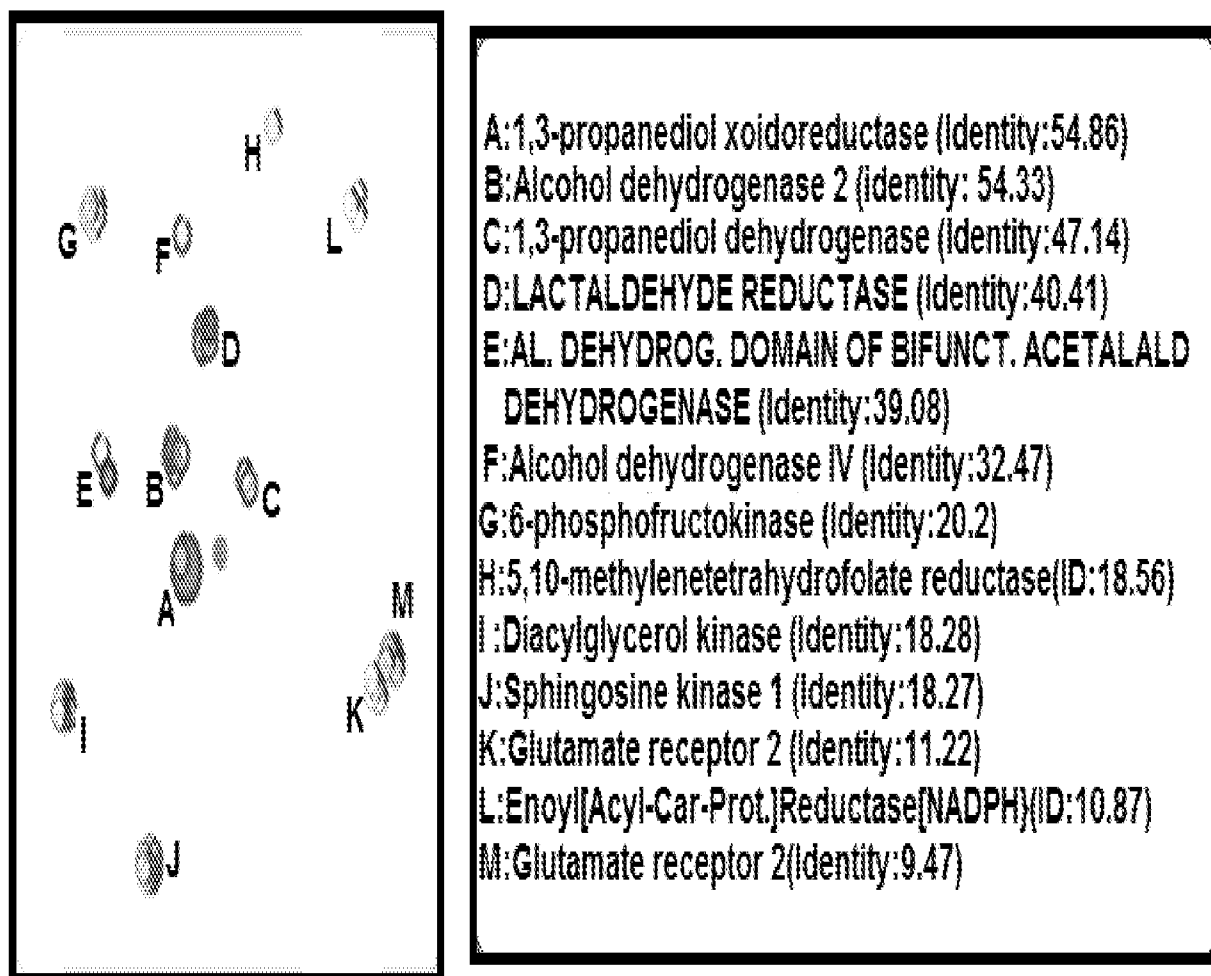


FIG. 6A

FIG. 6B

	250	260	270	280	290	300
Cn_Mdh2	GDSMEARAAMCYAQYLAGMAFNNASLGYVHAMAHQLGGFY	NLPHGVCNAILLPEVSEFN..L				
Cn_Mdh1	LDLIALRRMALAALKAGLAFSNTKTALAHSSISYEMTLRH	GLPHGFIACSFTLPMVLEKA..W				
Zm_Adh2	GKDMFAREAMAYAOFLAGMAFNNASLGYVHAMAHQLGGFY	NLPHGVCNAVLLPEVLAYN..A				
Kp_1,3-Pdh	GSNLOAREYDAYASLLAGMAFNANLGYVHAMAHQLGGFY	DMPHGVCNAVLLPEVARYN..L				
Bm_Mdh	GEDIAREKMAAYAOYMAQVAFNNGGLGLVHSISHQVGGVY	KLQHGICNSVNMPEVCAFN..L				
Dh_Adh	GRDMEAREGMAFANTLSGVVMTVSVTAEHSLEHAMSAYHPELPHGAGLIMISRAFYEFFIEK					
Ls_Adh	GENKOARSDMSYGSYLAGLAFFNAGVAGVHALAYPLGGOF	HLAHGDSNAVLLPYVMGYI..R				
Lf_Adh	GQDKQARTDMSYGSYLAGLAFFNAGVAGVHALAYPLGGOF	HLAHGDSNAVLLPYVMGYI..R				
	310	320	330	340	350	360
Cn_Mdh2	IAAFERYARIAELLGENIGGLSAHDAAKAAVSARTLSTSIGIPA	GLAGLGVKADDHEVMA				
Cn_Mdh1	ARRPDRDRTLQRLFGPVLID.....HGOORLRAFLOGLGVKT	EFADYGVTDDEAGAMI				
Zm_Adh2	SVVAGRLKDVGVAMCLDIANLGDKEGAEATIQAVRDLAASIGIPA	NLTGLAKKEDVPLLA				
Kp_1,3-Pdh	IANPEKFADIAELMGENITGLSTLDAAEKAIATRLSMDIGIPO	HLRDLGVKETDFPYMA				
Bm_Mdh	IAKTERFAHIAELLGENVAGLSTAAAAERAIVALERINKSEGIPS	GYAEMGVKEEDIELLA				
Dh_Adh	HACDQRFVRMAQALCMKDA....ARPEDFITMLVKLQENCGVADLKMSDYGMTPEDEFDKLA					
Ls_Adh	QSCEKRMKDILDAMGLSSAYLSQEEASYKCVDAQQQLVQDVNIPS	TLKGFDPPEALEQLT				
Lf_Adh	QSCEKRMKDILDAMGLSSAYLSQEEASYKCVDTLQQLVQDVKIPS	TLQGFNPASALEQLT				
	370	380	390			
Cn_Mdh2	SNQKDAC.ML.TNPRKATLAQVMAIFAAAM.....					
Cn_Mdh1	TRMLDGAR.GK.N.....FIGAV....SA....					
Zm_Adh2	DHALKDAC.AL.TNPRQGDQKEVEELFLSAF.....					
Kp_1,3-Pdh	EMELKDGN.AF.SNPRKGNEQEIAAIFRQAF.....					
Bm_Mdh	KNAYEDVC.TQ.SNPRVPTVQDIAQLIKNAM.....					
Dh_Adh	ANARETMGGLFAANPYAMTHEDCVEVYRKSYR.....					
Ls_Adh	DDETKQTR.ILARSPMLEREDIYAIYRAAFDGEVREPDAS					
Lf_Adh	EDETKQTR.ILARSPMLEREDIYQIYQAADFQGVVRE...P					

●●●

FIG. 6B (continued)

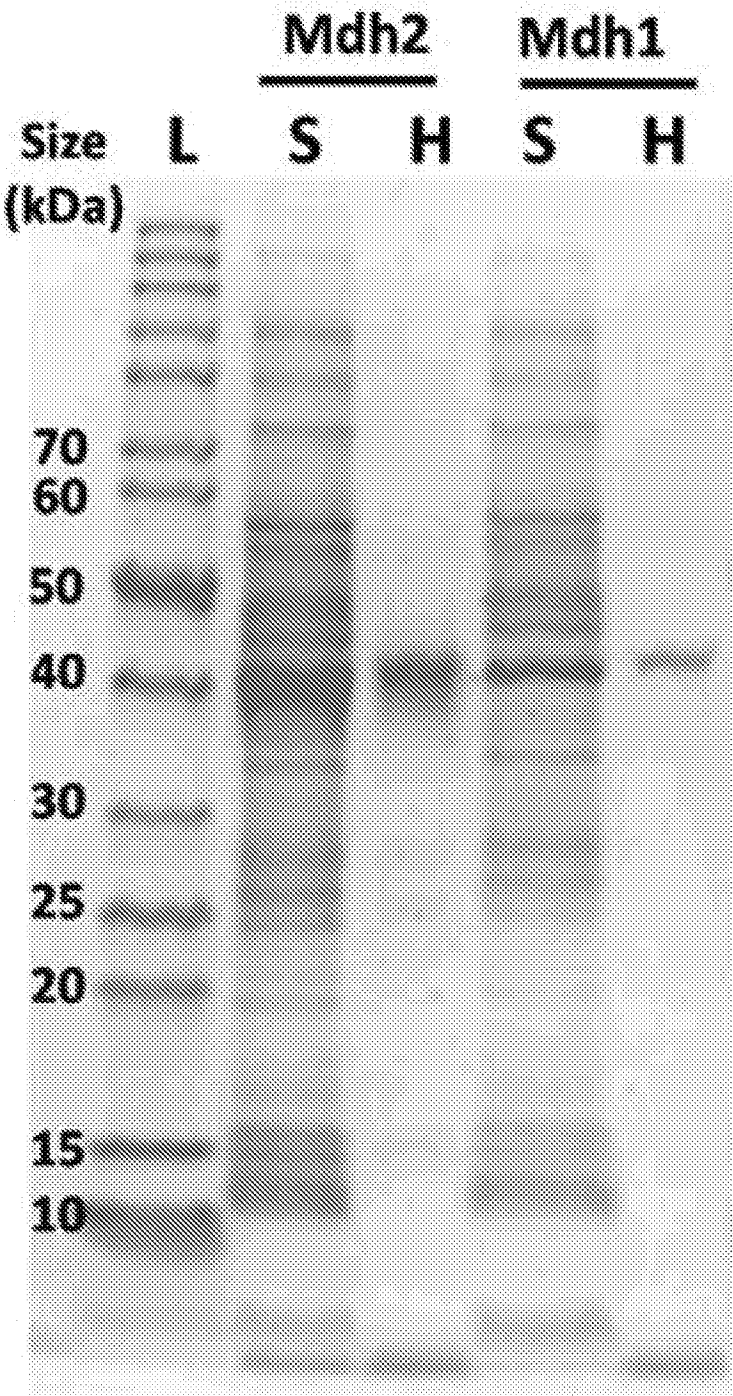


FIG. 7

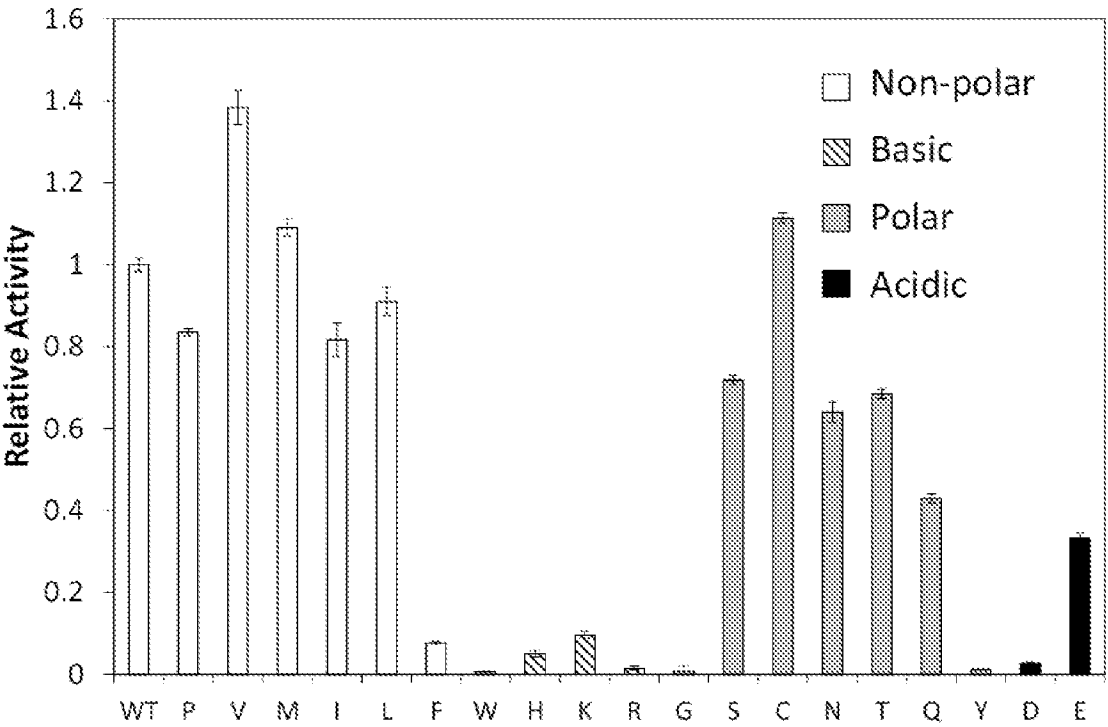


FIG. 8

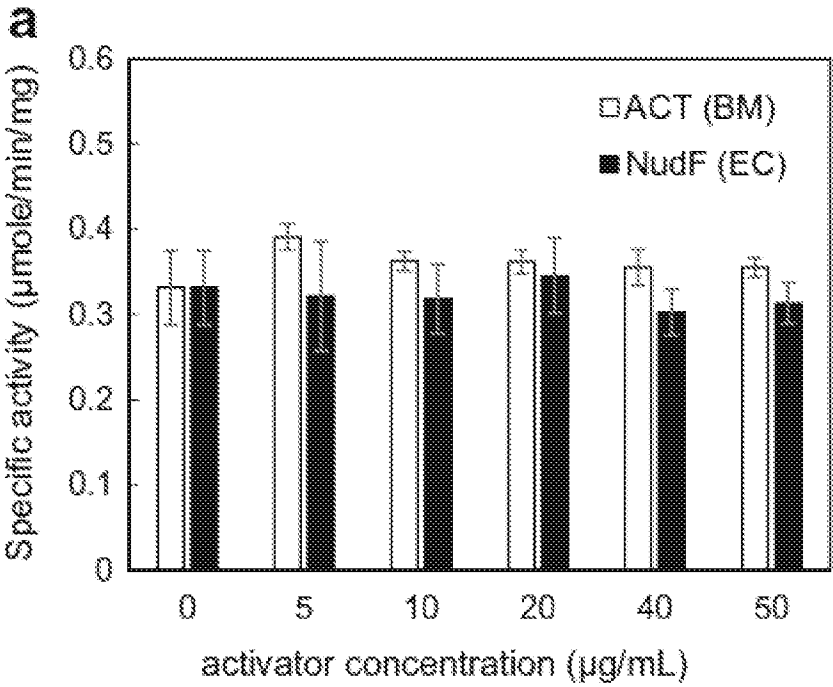


FIG. 9A

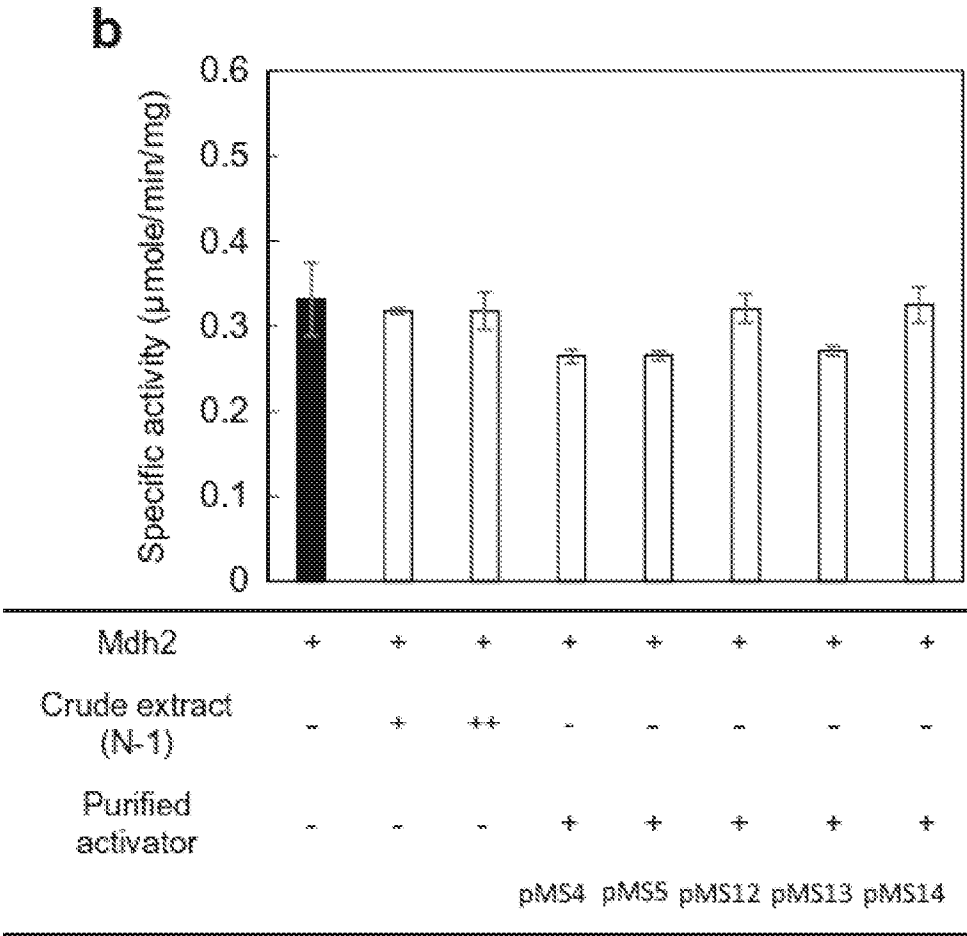


FIG. 9B

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2017/013211

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12N 1/21; C12N 9/04; C12P 1/00; C12P 5/00; C12P 7/16; C12P 7/24 (2017.01)

CPC - C12N 9/0006; C12P 7/18; C12P 7/24; C12Y 101/01001; C12Y 101/01244 (2017.02)

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)

See Search History document

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

USPC - 435/147 (keyword delimited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2015/051298 A2 (GENOMATICA, INC.) 09 April 2015 (09.04.2015) entire document	15-17
A	WO 2011/060920 A2 (BASF PLANT SCIENCE COMPANY GMBH et al) 26 May 2011 (26.05.2011) entire document; SEQ ID NO:130847	1-23
A	KANG et al. "Furfural reduction mechanism of a zinc-dependent alcohol dehydrogenase from <i>Cupriavidus necator</i> JMP134," Molecular Microbiology, 20 November 2011 (20.11.2011), Vol. 83, Pgs. 85-95. entire document	1-23
A	WO 2014/165763 A1 (LYNCH) 09 October 2014 (09.10.2014) entire document	1-23

☐ Further documents are listed in the continuation of Box C.☐ 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

13 March 2017

Date of mailing of the international search report

10 APR 2017

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, VA 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2017/013211

Box No. 1 Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

- a. ☒ forming part of the international application as filed:
 - ☒ in the form of an Annex C/ST.25 text file.
 - ☐ on paper or in the form of an image file.
 - b. ☐ furnished together with the international application under PCT Rule 13ter.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
 - c. ☐ furnished subsequent to the international filing date for the purposes of international search only:
 - ☐ in the form of an Annex C/ST.25 text file (Rule 13ter.1(a)).
 - ☐ on paper or in the form of an image file (Rule 13ter.1(b) and Administrative Instructions, Section 713).
2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

SEQ ID NO:2 was searched.