

(19) AUSTRALIAN PATENT OFFICE

(54) Title
Improved glucose dehydrogenase polypeptides and related polynucleotides

(51)⁶ International Patent Classification(s)
C12N 9/04 (2006.01) 20060101AFI2005100
C12N 9/04 8BMEP
PCT/US2004/026194

(21) Application No: 2004288134 (22) Application Date: 2004 .08 .11

(87) WIPO No: W005/045016

(30) Priority Data

(31) Number	(32) Date	(33) Country
60/545,657	2004 .02 .18	US
60/494,300	2003 .08 .11	US

(43) Publication Date : 2005 .05 .19

(71) Applicant(s)
Codexis, Inc.

(72) Inventor(s)
Jenne, Stephane J., Newman, Lisa Marie, Davis, S. Christopher, Krebber, Anke

(74) Agent/Attorney
Spruson & Ferguson, Level 35 St Martins Tower 31 Market Street, Sydney, NSW, 2000

(56) Related Art
W02004/015132 A2
Manjon et al Biotechnology Letters 2002 Vol 24 pp 1227-32
US7,132,267 B1

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

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
19 May 2005 (19.05.2005)

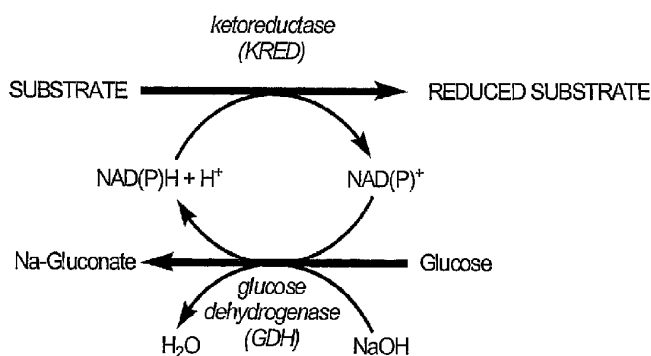
PCT

(10) International Publication Number
WO 2005/045016 A3

- (51) International Patent Classification⁷: **C12N 9/04**, 15/52, 15/63, 5/10 (74) Agent: **POCHOPIEN, Donald J.**; MCANDREWS, HELD & MALLOY, LTD., 500 W. Madison Street, Suite 3400, Chicago, IL 60661 (US).
- (21) International Application Number: PCT/US2004/026194 (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AI., AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CII, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PI., PT, RO, RU, SC, SD, SE, SG, SK, SL., SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.
- (22) International Filing Date: 11 August 2004 (11.08.2004) (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SI., SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CII, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL., PI., PT, RO, SE, SI, SK, TR), OAPI (BI, BJ, CI, CG, CI, CM, GA, GN, GQ, GW, ML., MR, NI, SN, TD, TG).
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/494,300 11 August 2003 (11.08.2003) US
60/545,657 18 February 2004 (18.02.2004) US
- (71) Applicant (for all designated States except US): **CODEXIS, INC.** [US/US]; 200 Penobscot Drive, Redwood City, CA 94063 (US).
- (72) Inventors: **KREBBER, Anke**; 3500 Louis Road, Palo Alto, CA 94303 (US). **NEWMAN, Lisa Marie**; 1002 King Street, Redwood City, CA 94061 (US). **DAVIS, S. Christopher**; 104 Winfield Street, San Francisco, CA 94110 (US). **JENNE, Stephane J.**; 821 El Camino Real #202, Burlingame, CA 94010 (US).
- Published:
— with international search report

[Continued on next page]

(54) Title: IMPROVED GLUCOSE DEHYDROGENASE POLYPEPTIDES AND RELATED POLYNUCLEOTIDES



(57) Abstract: The present invention is directed to glucose dehydrogenase (GDH) polypeptides that have enhanced GDH activity and/or thermostability relative to the backbone wild-type glucose dehydrogenase polypeptide. In addition, the present invention is directed to a polynucleotide that encodes for the GDH polypeptides of the present invention, to nucleic acid sequences comprising the polynucleotides



— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

— with sequence listing part of description published separately in electronic form and available upon request from the International Bureau

(88) Date of publication of the international search report:
10 November 2005

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

IMPROVED GLUCOSE DEHYDROGENASE POLYPEPTIDES AND RELATED POLYNUCLEOTIDES

FIELD OF THE INVENTION

[01] The present invention is related to the field of enzymology, and
5 particularly to the field of glucose dehydrogenase enzymology. More
specifically, the present invention is directed to glucose dehydrogenase
polypeptides having improved enzymatic activity (*i.e.*, high substrate turnover)
and stability, and to polynucleotides sequences encoding for the improved
glucose dehydrogenase polypeptides. The present invention is useful because
10 the glucose dehydrogenase polypeptides can be coupled to oxido- or reductase
enzymes to produce synthetic organic chemicals or precursors in high yields.

BACKGROUND OF THE INVENTION

[02] Glucose dehydrogenase [EC1.1.1.47] or "GDH" catalyzes the
conversion of β -glucose and nicotinamide adenine dinucleotide (NAD) to
15 gluconolactone and reduced nicotinamide adenine dinucleotide (NADH). NAD
serves as a co-factor in this reaction and may be phosphorylated in the form of
NADP. GDH is an important enzyme for use in clinical tests and the food
industry. GDH is also applied as a catalyst for chemical conversions where it
serves a role in the regeneration of NADH and NADPH in enzymatic carbonyl
20 reductions, such as aldehydes and ketones.

[03] *Bacillus* species have been an excellent source of GDH. The enzyme
from *B. megaterium* M1286 was purified to homogeneity and found to be a
homotetramer of 30,000 DA subunits with pH optimum of 8.0-9.0 depending on
buffer conditions and uses either nicotinamide adenine dinucleotide (NAD) or
25 nicotinamide adenine dinucleotide phosphate (NADP) as cofactor (Pauly H.E.
and Pfeleiderer G., Hoppe Seylers Z. Physiol. Chem. 1975 356:1613-23). The
enzyme from *Cryptococcus uniguttulatus* Y 0033 has a pH optimum of 6.0-8.0,
an optimum temperature of 55° C and a molecular weight of 110 kDa (U.S. Pat.
4,877,733). The enzyme from *Pseudomonas* sp. FH1227 has a pH optimum of

8.5-9.0, an optimum temperature of 55° C and a molecular weight of 101 kDa (U.S. Pat. 5,298,411).

- [04] Commercially applied GDHs are primarily derived from microorganisms. Initially, GDH was produced by fermentation of the natural host organisms such as *B. megaterium* ATCC 39118 (U.S. Pat. 4,542,098), *Bacillus cereus* DSM 1644 (US4397952), *Cryptococcus uniguttulatus* Y 0033 (U.S. Pat. 4,877,733) and *Pseudomonas* sp. FH1227 (U.S. Pat. 5,298,411). Since then, GDH encoding genes have been identified, cloned and expressed in heterologous hosts such as *Escherichia coli*.
- 10 [05] The *Bacillus subtilis* 61297 GDH gene was expressed in *E. coli* and exhibited the same physicochemical properties as the enzyme produced in its native host (Vasantha et al. Proc. Natl. Acad. Sci. USA 1983 80:785). The gene sequence of the *B. subtilis* GDH gene was reported by Lampel, K. A., Uratani, B., Chaudhry, R., Ramaley, R. F., and Rudikoff S., "Characterization of the developmentally regulated *Bacillus subtilis* glucose dehydrogenase gene," *J. Bacteriol.* 166, 238-243 (1986) and Yamane, K., Kumano, M. and Kurita, K., "The 25 degrees-36 degrees region of the *Bacillus subtilis* chromosome: determination of the sequence of a 146 kb segment and identification of 113 genes," *Microbiology* 142 (Pt 11), 3047-3056 (1996), and is found in Genbank under Accession Nos. M12276 and D50453.
- 15 [06] Similarly, gene sequences were determined for GDH from *B. cereus* ATCC14579 (Nature 2003 423:87-91; Genbank Acc. No. AE017013) and *B. megaterium* (Eur. J. Biochem. 1988 174:485-490, Genbank Acc. No. X12370; J. Ferment. Bioeng. 1990 70:363-369, Genbank Acc. No. D90044). The GDH enzymes from *B. subtilis* and *B. megaterium* are approximately 85% homologous (J. Theor. Biol. 1986 120:489-497).
- 20 [07] It has been well established that GDH enzymes suffer from limited stability. Ramaley and Vasantha reported that presence of glycerol in extraction and purification buffers is absolutely necessary to retain activity for GDH from *B. subtilis* (J. Biol. Chem. 1983 258:12558-12565). The enzyme instability can be largely attributed to the dissociation of the tetramer into its
- 25
- 30

monomers, which is an equilibrium process that is controlled by environmental factors such as pH and ionic strength (Maurer and Pfeleiderer, Z. Naturforsch. 1987 42:907-915). This has lead to the isolation and studies of GDH from other *Bacillus* sp. such as *B. megaterium*. For instance, U. S. Pats. 5,114,853 and 5,126,256 and Baik *et al.*, *Appl. Microbiol. Biotechnol.* 2003 61:329-335 describe GDH encoding genes from *B. megaterium* and mutants thereof that exhibit increased thermostability and that can be produced in recombinant *E. coli* hosts. However, there remains an industrial need for GDH enzymes that not only have increased thermostability but that also have enhanced enzymatic activity. The above referenced publications and patents, and all other publications and patents referenced herein, are hereby incorporated by reference herein in their entirety.

Brief Summary of the Invention

According to a first embodiment of the invention, there is provided a non-naturally occurring polypeptide capable of converting glucose and nicotinamide adenine dinucleotide (NAD) to gluconolactone and reduced nicotinamide adenine dinucleotide (NADH), or capable of converting glucose and nicotinamide adenine dinucleotide phosphate (NADP) to gluconolactone and reduced nicotinamide adenine dinucleotide phosphate (NADPH) with at least 1.5 times the glucose dehydrogenase (GDH) activity of the wild-type GDH of SEQ ID NO:2, and which comprises an amino acid sequence that is at least 91% identical to SEQ ID NO:54.

According to a second embodiment of the invention, there is provided a fragment of a non-naturally occurring polypeptide in accordance with the first embodiment of the present invention, which is capable of converting glucose and nicotinamide adenine dinucleotide (NAD) to gluconolactone and reduced nicotinamide adenine dinucleotide (NADP) to gluconolactone and reduced nicotinamide adenine dinucleotide phosphate (NADPH) with at least from 1.5 to about 11 times the GDH activity of the wild-type GDH of SEQ ID NO:2.

According to a third embodiment of the invention, there is provided a composition comprising a non-naturally occurring polypeptide in accordance with the first embodiment of the present invention and a suitable carrier.

According to a fourth embodiment of the invention, there is provided a polynucleotide encoding a glucose dehydrogenase polypeptide in accordance with the first embodiment of the present invention.

According to a fifth embodiment of the invention, there is provided an expression vector comprising a polynucleotide in accordance with the fourth embodiment of the present invention operatively linked to a promoter.

According to a sixth embodiment of the invention, there is provided a host cell
5 comprising a polynucleotide in accordance with the fifth embodiment of the present invention, or an expression vector in accordance with the sixth embodiment of the present invention.

According to a seventh embodiment of the invention, there is provided a method of making a non-naturally occurring glucose dehydrogenase polypeptide, comprising:

10 (a) cultivating a host cell comprising a nucleic acid construct, wherein the nucleic acid construct comprises a nucleic acid sequence encoding a polypeptide in accordance with the first embodiment of the present invention, under conditions suitable for production of the polypeptide; and

(b) recovering the polypeptide.

15 The present invention has multiple aspects. In one aspect, the present invention is directed to a polypeptide having at least 1.5 times, typically 1.5 to about 25 times, more typically from 1.5 to about 11 times, the GDH activity of the wild-type GDH of SEQ ID NO:2 (such as determined by the method of Example 4) and being selected from the group consisting of:

20 (a) a polypeptide having an amino acid sequence which has at least 91% homology, preferably at least 95% homology, and more preferably at least 98% homology with the amino acid sequence of SEQ ID NO:54, 74, 84, 160, 164 or 168 (hereinafter "homologous polypeptides");

(b) a polypeptide encoded by a nucleic acid sequence which hybridizes under
25 medium stringency conditions with either

(i) the nucleotide sequence of SEQ ID NO:53, 73, 83, 159, 163 or 167;

(ii) a subsequence of (i) of at least 100 nucleotides, or

(iii) a complementary strand of (i) or (ii) (J. Sambrook, E. F. Fritsch, and T. Maniatis, 1989, Molecular Cloning, A Laboratory Manual, 2d edition, Cold Spring
30 Harbor, N. Y.);

(c) a variant of the polypeptide of SEQ ID NO:54, 74, 84, 160, 164 or 168 comprising a substitution, deletion, and/or insertion of one to six amino acids;

(d) a fragment of (a), (b) or (c) that has from 1.5 to about 11 times the GDH activity of the wild-type GDH of SEQ ID NO: 2, ; and

(e) a polypeptide of (a), (b) or (c) that retains more than 80% of the initial GDH activity after 20 minutes of incubation at 50° C and pH 7. In one embodiment, the present invention is also directed to a variant GDH polypeptide as described herein in isolated and purified form. In another embodiment, the present invention is directed to a variant GDH polypeptide as described herein in lyophilized form. In yet another embodiment, the present invention is directed to a composition comprising a variant GDH polypeptide as described herein and a suitable carrier, typically a buffer solution, more typically a buffer solution having a pH between 6.0 and 8.0.

[09] The novel GDH polypeptides of the present invention have enhanced GDH activity (>1.5 fold) relative to the backbone GDH polypeptide from *B. subtilis* of SEQ ID NO: 2 and typically vary from SEQ ID NO: 2 by 1-7 amino acid residues, more typically by 1-6 amino acid residues, even more typically by 1-5 amino acid residues, and most typically by 1-4 amino acid residues. For purposes of the present invention, the degree of homology between two amino acid sequences was determined using the Needleman Wunsch global alignment algorithm, i.e., using dynamic programming algorithm for Global Alignment Scoring Matrix: PAM 120 matrix with gap penalties for introducing gap = -22.183 and extending gap = -1.396. The percent identity = number of identical residues between the first sequence and the second sequence divided by the length of first sequence in alignment (with gaps)(p) indicates partial match. See Needleman, S.B. & Wunsch, C.D., "A general method applicable to the search for similarities in the amino acid sequence of two proteins," Journal of Molecular Biology, 48:443-453 (1970).

[10] The various residue positions of the *B. subtilis* GDH polypeptide that have been substituted to yield enhanced GDH activity and/or thermostability are summarized in Table 1 herein. The amino acid sequences for a number of the inventive GDH polypeptides that have demonstrated enhanced GDH activity and/or thermostability at 50° C are disclosed herein as SEQ ID NOS: 6, 8, 10,

- 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166 and 168. The polynucleotide sequences encoding for the above described inventive GDH polypeptides have SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 75, 77, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165 and 167, respectively. Another GDH polypeptide (SEQ ID NO: 52) of the present invention, which is encoded by the polynucleotide of SEQ ID NO: 51, is based upon the amino acid sequence of GDH (SEQ ID NO: 4) from *B megaterium* and differs therefrom by six amino acid residues.
- [11] In a preferred embodiment, the present invention is directed to the novel glucose dehydrogenase polypeptides of SEQ ID NOS: 54, 74, 84, 160, 164, and 168 that have enhanced glucose dehydrogenase activity and/or enhanced thermostability relative to the wild-type glucose dehydrogenase of SEQ ID NO: 2.
- [12] In yet another embodiment, the present invention is directed to polypeptide having glucose dehydrogenase enzyme activity and being selected from the group consisting of a GDH polypeptide having at least 84% sequence identity with SEQ ID NO: 52, a GDH polypeptide having at least 98% sequence identity with SEQ ID NO: 72, and a GDH polypeptide having at least 98% sequence identity with SEQ ID NO: 58.
- [13] In its second aspect, the present invention is directed to a polynucleotide sequence that encodes for the correspondingly referenced GDH polypeptide. Given the degeneracy of the genetic code, the present invention is also directed to any polynucleotide that encodes for the above referenced GDH polypeptides of the present invention. In another preferred embodiment, the present invention is directed to certain specific polynucleotides of SEQ ID NOS: 53, 73,

83, 159, 163, and 167 that encode for the novel glucose dehydrogenase polypeptides of SEQ ID NOS: 54, 74, 84, 160, 164 and 168, respectively.

[14] In a third aspect, the present invention is directed to a nucleic acid construct, a vector, or a host cell comprising a polynucleotide sequence
5 encoding a GDH polypeptide of the present invention operatively linked to a promoter.

[15] In a fourth aspect, the present invention is directed to a method of making a GDH polypeptide of the present invention comprising (a) cultivating a host cell comprising a nucleic acid construct comprising a nucleic acid sequence
10 encoding a GDH polypeptide of the present invention under conditions suitable for production of the polypeptide; and (b) recovering the polypeptide.

BRIEF DESCRIPTION OF SEVERAL VIEWS OF THE DRAWINGS

[16] FIG. 1 exemplifies an oxidation-reduction cycle wherein glucose is oxidized by GDH to gluconic acid in the presence of NAD^+ (or NADP^+) to
15 produce the corresponding reduced form NADH (or NADPH), respectively, which in turn drives the reduction of a substrate to a reduced substrate while being oxidized back to NAD (or NADP) by a reductase. The gluconic acid formed in this reaction is neutralized by sodium hydroxide to sodium-gluconate.

[17] FIGS. 2A-2B in combination provide a table comparing the % amino
20 acid identity of the GDH polypeptides of the present invention versus the GDH polypeptides of the indicated prior art references. In col. 4 of FIG. 2, the GDH polypeptide of *B. subtilis* (S06-3) has the same amino acid sequence as disclosed in EP 955375 (col. 8). To generate FIGS 2A-2B, alignments were done using dynamic programming algorithm for Global Alignment Scoring Matrix: PAM
25 120 matrix with gap penalties for introducing gap = -22.183 and extending gap = -1.396. The percent identity = number of identical residues between the first sequence and the second sequence divided by the length of first sequence in alignment (with gaps)(p) indicates partial match. See Needleman, S.B. & Wunsch, C.D., "A general method applicable to the search for similarities in the

amino acid sequence of two proteins," *Journal of Molecular Biology*, 48:443-453 (1970).

[18] FIG. 3 is a 4036 bp expression vector (pCK110900) of the present invention comprising a P15A origin of replication (P15A ori), a lacI repressor, a CAP binding site, a lac promoter (lac), a T7 ribosomal binding site (T7g10 RBS), and a chloramphenicol resistance gene (camR).

[19] The foregoing summary, as well as the following detailed description of certain embodiments of the present invention, will be better understood when read in conjunction with the appended drawings. For the purpose of illustrating the invention, there is shown in the drawings, certain embodiments. It should be understood, however, that the present invention is not limited to the arrangements and instrumentality shown in the attached drawings.

DETAILED DESCRIPTION OF THE INVENTION

[20] The present invention has multiple aspects. In one aspect, the present invention is directed to a polypeptide having at least 1.5 times, typically 1.5 to about 25 times, more typically from 1.5 to about 11 times the GDH activity of the wild-type GDH of SEQ ID NO: 2 (such as determined by the method of Example 4) and being selected from the group consisting of:

(a) a polypeptide having an amino acid sequence which has at least 91% homology, preferably at least 95% homology, and more preferably at least 98% homology with the amino acid sequence of SEQ ID NO: 54, 74, 84, 160, 164 or 168 (hereinafter "homologous polypeptides");

(b) a polypeptide encoded by a nucleic acid sequence which hybridizes under medium stringency conditions with either (i) the nucleotide sequence of SEQ ID NO: 53, 73 or 83, (ii) a subsequence of (i) of at least 100 nucleotides, or (iii) a complementary strand of (i) or (ii) (J. Sambrook, E. F. Fritsch, and T. Maniatis, 1989, *Molecular Cloning, A Laboratory Manual*, 2d edition, Cold Spring Harbor, N.Y.);

(c) a variant of the polypeptide of SEQ ID NO: 54, 74, 84, 160, 164 or 168 comprising a substitution, deletion, and/or insertion of one to six amino acids;

(d) a fragment of (a), (b) or (c) that has from 1.5 to about 11 times the GDH activity of the wild-type GDH of SEQ ID NO: 2; and
(e) a polypeptide of (a), (b) or (c) that retains more than 80% of the initial GDH activity after 20 minutes of incubation at 50° C and pH 7.

5 [21] Unless otherwise noted, as used throughout this specification, the terms "percent identity," "% identity," "percent identical," and "% identical" are used interchangeably herein to refer to the percent amino acid sequence identity that is determined using the Needleman Wunsch global alignment algorithm, i.e., using dynamic programming algorithm for Global Alignment Scoring Matrix:
10 PAM 120 matrix with gap penalties for introducing gap = -22.183 and extending gap = -1.396. The percent identity = number of identical residues between the first sequence and the second sequence divided by the length of first sequence in alignment (with gaps)(p) indicates partial match. See Needleman, S.B. & Wunsch, C.D., "A general method applicable to the search
15 for similarities in the amino acid sequence of two proteins," Journal of Molecular Biology, 48:443-453 (1970).

[22] As used herein, the terms "glucose dehydrogenase" and "GDH" are used interchangeably herein to refer to a polypeptide that has the ability to catalyze the conversion of glucose and nicotinamide adenine dinucleotide (NAD) to
20 gluconolactone and reduced nicotinamide adenine dinucleotide (NADH). Alternatively, the phosphorylated cofactors NADP and NADPH can replace NAD and NADH in the above reaction. In nature, GDH is made up of four subunits that are loosely held together in a homo-tetramer. Based upon the crystal structure of wild-type *B. megaterium* GDH polypeptide (SEQ ID NO: 4)
25 (Yamamoto et al. J. Biochem. 2001 129:303-312), residue positions 188-217 of the polypeptide define a protein loop region that is involved in NAD⁺ and glucose binding.

[23] In use, the enhanced GDH polypeptides of the present invention are preferably coupled to a synthetic reaction as a cofactor regeneration system (See
30 Figure 1) to provide a continuing source of reduced cofactor. As used herein, the term "cofactor" refers to a non-protein compound that operates in

combination with an enzyme that catalyzes a reaction of interest. Suitable cofactors employed with the GDH polypeptides of the present invention include NADP (nicotinamide-adenine dinucleotide phosphate) and NAD (nicotinamide adenine dinucleotide).

5 [24] The term "cofactor regeneration system" refers herein to a set of reactants that participate in a reaction that regenerates a utilized cofactor back to its pre-reaction state. An example is the regeneration of oxidized cofactor regeneration back to reduced cofactor, e.g., NADP to NADPH. The reduced (regenerated) cofactor is then capable of participating in a reaction with a
10 substrate and an enzyme, such as a reducing enzyme, to produce the reduced substrate and the oxidized (utilized) cofactor, which can again be regenerated by the cofactor regeneration system. The above-described operation of the glucose/glucose dehydrogenase cofactor regeneration system is exemplified in Figure 1.

15 [25] In FIG. 1, the reaction catalyzed by the reducing enzyme is shown as being coupled to the glucose dehydrogenase cofactor regeneration system. The term "coupled" is used herein to refer to the use of the reduced form of cofactor in the reduction of a substrate, and the concomitant use of the oxidized form of the same cofactor, generated in the aforementioned reaction, in the oxidation of
20 a component (e.g., glucose) of the cofactor regeneration system, which generates the reduced form of the same cofactor. One possible limiting factor in the overall reaction speed in a coupled system is the speed (activity) of the GDH polypeptide in regenerating cofactor.

[26] The GDH polypeptides of the present invention have enhanced GDH
25 activity (such as measured by the method of Example 4) that is 1.5 fold to about 11 fold greater than the GDH activity of the backbone GDH polypeptide from *B. subtilis* of SEQ ID NO: 2, and typically vary from SEQ ID NO: 2 by 1-7 amino acid residues, more typically by 1-6 amino acid residues, even more typically by 1-5 amino acid residues, and most typically by 1-4 amino acid
30 residues. Preferably, the GDH polypeptides of the present invention have enhanced GDH activity that is 2.5 fold to about 11 fold greater than the GDH

activity of the backbone GDH polypeptide from *B. subtilis* of SEQ ID NO: 2. More preferably, the GDH polypeptides of the present invention have enhanced thermostability after heat treatment at 50° C for 20 minutes that is 1.5 fold to about 15 fold greater than the GDH activity of the backbone GDH polypeptide from *B. subtilis* of SEQ ID NO: 2. Thermostability is determined by measuring the residual GDH activity (such as by the method of Example 4) remaining after heat treatment of the GDH polypeptide at 50° C for 20 minutes.

[27] The amino acid sequences for a number of the inventive GDH polypeptides that have demonstrated enhanced GDH activity and/or thermostability at 50° C are disclosed herein as SEQ ID NOS: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166 and 168. The polynucleotide sequences encoding for the above described inventive GDH polypeptides have SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 75, 77, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165 and 167, respectively.

[28] In yet another aspect, the present invention is directed to GDH polypeptides that have enhanced activity in coupled reactions.

[29] In another embodiment, the present invention is directed to a GDH polypeptide encoded by a nucleic acid sequence which hybridizes under medium stringency conditions with either (i) the nucleotide sequence of SEQ ID NO: 53, 73, 83, 159, 163 or 167 (ii) a subsequence of (i) of at least 100 nucleotides, or (iii) a complementary strand of (i) or (ii) (J. Sambrook, E. F. Fritsch, and T. Maniatis, 1989, Molecular Cloning, A Laboratory Manual, 2d edition, Cold Spring Harbor, N.Y.). For polynucleotides of at least 100 nucleotides in length, low to very high stringency conditions are defined as

prehybridization and hybridization at 42° C in 5x SSPE, 0.3% SDS, 200 µg/ml sheared and denatured salmon sperm DNA, and either 25% formamide for low stringencies, 35% formamide for medium and medium-high stringencies, or 50% formamide for high and very high stringencies, following standard Southern blotting procedures.

[30] For polynucleotides of at least 100 nucleotides in length, the carrier material is finally washed three times each for 15 minutes using 2x SSC, 0.2% SDS at least at 50° C (low stringency), at least at 55° C (medium stringency), at least at 60° C. (medium-high stringency), at least at 65° C (high stringency), and at least at 70° C. (very high stringency).

[31] In another embodiment, the present invention is directed to a variant of the polypeptide of SEQ ID NO: 54, 74, 84, 160, 164 or 168 having a substitution, deletion, and/or insertion of one to six amino acids therefrom, and having from 1.5 to about 11 times the GDH activity of the wild-type GDH of SEQ ID NO: 2, such as determined by the method of Example 4. Preferably, amino acid changes are of a minor nature, that is conservative amino acid substitutions that do not significantly affect the folding and/or activity of the protein; small deletions, typically of one to six amino acids; small amino- or carboxyl-terminal extensions; a small linker peptide; or a small extension that facilitates purification by changing net charge or another function, such as a poly-histidine tract, an antigenic epitope or a binding domain.

[32] Examples of conservative substitutions are within the group of basic amino acids (arginine, lysine and histidine), acidic amino acids (glutamic acid and aspartic acid), polar amino acids (glutamine and asparagine), hydrophobic amino acids (leucine, isoleucine and valine), aromatic amino acids (phenylalanine, tryptophan and tyrosine), and small amino acids (glycine, alanine, serine, threonine, proline, cysteine and methionine). Amino acid substitutions, which do not generally alter the specific activity are known in the art and are described, for example, by H. Neurath and R. L. Hill, 1979, In, The Proteins, Academic Press, New York. The most commonly occurring exchanges are Ala/Ser, Val/Ile, Asp/Glu, Thr/Ser, Ala/Gly, Ala/Thr, Ser/Asn, Ala/Val,

Ser/Gly, Tyr/Phe, Ala/Pro, Lys/Arg, Asp/Asn, Leu/Ile, Leu/Val, Ala/Glu, and Asp/Gly as well as these in reverse.

[33] In another embodiment, the present invention is directed to a fragment of (a), (b) or (c), as described above in the first paragraph of the Detailed Description, that has from 1.5 to about 11 times the GDH activity of the wild-type GDH of SEQ ID NO: 2, such as determined by the method of Example 4. By the term "fragment" is meant that the polypeptide has a deletion of 1 to 10 amino acid residues from the carboxy terminus, the amino terminus, or both. Preferably, the deletion is 1 to 10 residues from the carboxy terminus, more preferably, the deletion is 1 to 5 residues from the carboxy terminus.

[34] In yet another embodiment, the present invention is directed to a GDH polypeptide of (a), (b) or (c), as described above in the first paragraph of the Detailed Description, that retains more than 80% of the initial (pre-incubation) GDH activity after 20 minutes of incubation at 50° C and pH 7. Preferably, the polypeptides of the invention retain at least 85% of the initial residual activity, more preferably at least 90% residual activity after 20 minutes incubation at 50° C and pH 7. The initial GDH activity is readily determined by an assay for GDH activity, such as described in Example 4 herein.

[35] Another GDH polypeptide (SEQ ID NO: 52) of the present invention is based upon the amino acid sequence of the GDH polypeptide (SEQ ID NO: 4) from *B. megaterium* and differs therefrom by six amino acid residues. This GDH polypeptide is encoded by the polynucleotide of SEQ ID NO: 51.

[36] In yet another embodiment, the present invention is directed to polypeptide having glucose dehydrogenase enzyme activity and being selected from the group consisting of a GDH polypeptide having at least 84% sequence identity with SEQ ID NO: 52, a GDH polypeptide having at least 98% sequence identity with SEQ ID NO: 72, and a GDH polypeptide having at least 98% sequence identity with SEQ ID NO: 58. These percent identities were 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 length of the

reference sequence, and using the following default ClustalW parameters to achieve slow/accurate pairwise optimal 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.

Polynucleotides

[37] In its second aspect, the present invention is directed to a polynucleotide sequence that encodes for a GDH polypeptide of the present invention. Given the degeneracy of the genetic code, the present invention is also directed to any polynucleotide that encodes for the above referenced GDH polypeptides of the present invention. In a preferred embodiment, the present invention is directed to certain specific polynucleotides of SEQ ID NOS: 53, 73, 83, 159, 163 and 167 that encode for the novel glucose dehydrogenase polypeptides of SEQ ID NOS: 54, 74, 84, 160, 164 and 168, respectively.

[38] To make the improved GDH polypeptides of the present invention, one starts with one or more wild-type polynucleotides that encode a GDH polypeptide. The term “wild-type” polynucleotide means that the nucleic acid fragment does not comprise any mutations from the form isolated from nature. The term “wild-type” protein means that the protein will be active at a level of activity found in nature and typically will comprise the amino acid sequence as found in nature. Thus, the term “wild type” or “parental sequence” indicates a starting or reference sequence prior to a manipulation of the invention.

[39] Suitable sources of wild-type GDH as a starting material to be improved are readily identified by screening genomic libraries for the GDH activities described herein. *See e.g.*, Example 4. A particularly suitable source of GDH is the *Bacillus sp.* bacteria as found in nature. *See* Example 1. Using the published glucose dehydrogenase gene sequences for *B. subtilis* (*e.g.*, EP 955375) and *B. megaterium* (*e.g.*, U.S. Pat. 5,114,853), primers for amplification of the genes from their respective gene libraries were created using conventional techniques and have the following sequences:

B. subtilis forward primer (SEQ ID NO: 80):

5'-GAATTCGCCCATATGTATCCGGATTAAAAGG-3'

B. subtilis reverse primer (SEQ ID NO: 81):

5'-TGGCCGGATCCTCATTAACCGCGCCTGCCTGGA-3'

5 *B. megaterium* forward primer (SEQ ID NO: 82):

5'-GAATTCGCCCATATGTATAAAGATTTAGAAGG-3'

B. megaterium reverse primer (SEQ ID NO 83):

5'-GGCCGGATCCTCATTATCCGCGTCCTGCTTGGA-3'

[40] Using a forward and reverse primer pair in a conventional polymerase chain reaction (PCR) to amplify the appropriate portion of the gene from *B. subtilis* and *B. megaterium*, several PCR products were obtained. Each PCR product was cloned into an expression vector and operatively linked behind a *lac* promoter. Upon screening the clones for GDH activity (e.g., by the assay of Example 4), several clones were found to express active GDH and these genes were sequenced. A first DNA sequence (SEQ ID NO: 1) designated as S06-3 and encoding a GDH polypeptide identical to the published *Bacillus subtilis* GDH (SEQ ID NO: 2) was obtained from *Bacillus subtilis*. A second DNA sequence (SEQ ID NO: 3), designated as M02-6 and encoding a GDH polypeptide that was 98.5% identical to the published *Bacillus megaterium* GDH (SEQ ID NO: 4), was obtained from *Bacillus megaterium*. These DNA sequences were utilized as the starting material for developing the improved polypeptides and polynucleotides of the present invention.

[41] In addition to the PCR primers described above, other PCR primers of different lengths could be used as well. See, e.g., Innis et al., 1990, PCR: A Guide to Methods and Application, Academic Press, New York. Other conventional nucleic acid amplification procedures such as ligase chain reaction (LCR), ligated activated transcription (LAT) and nucleic acid sequence-based amplification (NASBA) could also be used.

[42] Once a suitable starting material has been identified, a non-naturally occurring and mutated and/or evolved enzyme, having unknown glucose dehydrogenase activity is readily generated using any one of the well-known

mutagenesis or directed evolution methods. See, e.g., Ling, et al., "Approaches to DNA mutagenesis: an overview," Anal. Biochem., 254(2):157-78 (1997); Dale, et al., "Oligonucleotide-directed random mutagenesis using the phosphorothioate method," Methods Mol. Biol., 57:369-74 (1996); Smith, "In vitro mutagenesis," Ann. Rev. Genet., 19:423-462 (1985); Botstein, et al., "Strategies and applications of in vitro mutagenesis," Science, 229:1193-1201 (1985); Carter, "Site-directed mutagenesis," Biochem. J., 237:1-7 (1986); Kramer, et al., "Point Mismatch Repair," Cell, 38:879-887 (1984); Wells, et al., "Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites," Gene, 34:315-323 (1985); Minshull, et al., "Protein evolution by molecular breeding," Current Opinion in Chemical Biology, 3:284-290 (1999); Christians, et al., "Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling," Nature Biotechnology, 17:259-264 (1999); Cramer, et al., "DNA shuffling of a family of genes from diverse species accelerates directed evolution," Nature, 391:288-291; Cramer, et al., "Molecular evolution of an arsenate detoxification pathway by DNA shuffling," Nature Biotechnology, 15:436-438 (1997); Zhang, et al., "Directed evolution of an effective fucosidase from a galactosidase by DNA shuffling and screening," Proceedings of the National Academy of Sciences, U.S.A., 94:45-4-4509; Cramer, et al., "Improved green fluorescent protein by molecular evolution using DNA shuffling," Nature Biotechnology, 14:315-319 (1996); Stemmer, "Rapid evolution of a protein in vitro by DNA shuffling," Nature, 370:389-391 (1994); Stemmer, "DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution," Proceedings of the National Academy of Sciences, U.S.A., 91:10747-10751 (1994); WO 95/22625; WO 97/0078; WO 97/35966; WO 98/27230; WO 00/42651; WO 01/75767 and U.S. Pat. 6,537,746 which issued to Arnold, et al. on March 25, 2003 and is entitled "Method for creating polynucleotide and polypeptide sequences."

[43] Any of these methods can be applied to generate GDH polynucleotides. To maximize any diversity, several of the above-described techniques can be used sequentially. Typically, a library of variant polynucleotides is created by one mutagenic or evolutionary technique and their expression products are

screened to find the polypeptides having the highest GDH activity. Then, a second mutagenic or evolutionary technique is applied to polynucleotides encoding the most active polypeptides to create a second library, which in turn is screened for GDH activity by the same technique. The process of mutating
5 and screening can be repeated as many times as needed, including the insertion of point mutations, to arrive at a polynucleotide that encodes a polypeptide with the desired activity, thermostability, and cofactor preference.

[44] Alternatively, polynucleotides and oligonucleotides of the invention can be prepared by standard solid-phase methods, according to known synthetic
10 methods. Typically, fragments of up to about 100 bases are individually synthesized, then joined (e.g., by enzymatic or chemical ligation methods, or polymerase mediated methods) to form essentially any desired continuous sequence. For example, polynucleotides and oligonucleotides of the invention can be prepared by chemical synthesis using, e.g., the classical phosphoramidite
15 method described by *Beaucage et al.* (1981) *Tetrahedron Letters* 22:1859-69, or the method described by *Matthes et al.* (1984) *EMBO J.* 3:801-05, e.g., as it is typically practiced in automated synthetic methods. According to the phosphoramidite method, oligonucleotides are synthesized, e.g., in an automatic DNA synthesizer, purified, annealed, ligated and cloned in appropriate vectors.

20 [45] In addition, essentially any nucleic acid can be custom ordered from any of a variety of commercial sources, such as The Midland Certified Reagent Company, Midland, TX, The Great American Gene Company (Ramona, CA), ExpressGen Inc., Chicago, IL, Operon Technologies Inc. (Alameda, CA), and many others. Similarly, peptides and antibodies can be custom ordered from
25 any of a variety of sources, such as PeptidoGenic (pkim@ccnet.com), HTI Bio-products, Inc. (<http://www.htibio.com>), BMA Biomedicals Ltd. (U.K.), Bio.Synthesis, Inc., and many others.

[46] Polynucleotides may also be synthesized by well-known techniques as described in the technical literature. See, e.g., *Carruthers et al.*, *Cold Spring Harbor Symp. Quant. Biol.* 47:411-418 (1982), and *Adams et al.*, *J. Am. Chem. Soc.* 105:661 (1983). Double stranded DNA fragments may then be obtained
30

either by synthesizing the complementary strand and annealing the strands together under appropriate conditions, or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

- [47] General texts which describe molecular biological techniques useful herein, including mutagenesis, include Berger and Kimmel, Guide to Molecular Cloning Techniques. Methods in Enzymology, volume 152 Academic Press, Inc., San Diego, CA ("Berger"); *Sambrook et al.*, Molecular Cloning - A Laboratory Manual (2nd Ed.), volumes 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 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 2000) ("Ausubel"). Examples of techniques 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) are found in Berger, Sambrook, and Ausubel, as well as *Mullis et al.*, (1987) U.S. Patent No. 4,683,202; PCR Protocols A Guided to Methods and Applications (*Innis et al.*, eds.) Academic Press Inc. San Diego, CA (1990); Arnheim & Levinson (October 1, 1990) Chemical and Engineering News 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. Natl. 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. Improved methods of cloning *in vitro* amplified nucleic acids are described in *Wallace et al.*, U.S. Pat. No. 5,426,039. Improved methods of amplifying large nucleic acids by PCR are summarized in *Cheng et al.* (1994) Nature 369:684-685 and the references therein, in which PCR amplicons of up to 40kb 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*, Ausubel, Sambrook and Berger, *all supra*.

[48] It will be appreciated by those skilled in the art due to the degeneracy of the genetic code, a multitude of nucleotide sequences encoding GDH polypeptides of the invention may be produced, some of which bear substantial identity to the nucleic acid sequences explicitly disclosed herein.

[49] In the present case, several round No. 1 libraries were created by applying a variety of mutagenic techniques to the coding region of the *B. subtilis gdh* gene (SEQ ID NO: 1) or to the coding region of the *B. megaterium gdh* gene (SEQ ID NO: 3), as obtained by PCR.

[50] To obtain expression of the variant gene encoding a GDH, the variant gene was first operatively linked to one or more heterologous regulatory sequences that control gene expression to create a nucleic acid construct, such as an expression vector or expression cassette. Thereafter, the resulting nucleic acid construct, such as an expression vector or expression cassette, was inserted into an appropriate host cell for ultimate expression of the GDH polypeptide encoded by the shuffled gene. A "nucleic acid construct" is defined herein as a nucleic acid molecule, either single-or double-stranded, which is isolated from a naturally occurring gene or which has been modified to contain segments of nucleic acid combined and juxtaposed in a manner that would not otherwise exist in nature. Thus, in one aspect, the present invention is directed to a nucleic acid construct comprising a polynucleotide encoding a GDH polypeptide of the present invention.

[51] The term "nucleic acid construct" is synonymous with the term "expression cassette" when the nucleic acid construct contains all the control sequences required for expression of a coding sequence of the present invention. The term "coding sequence" is defined herein as a nucleic acid sequence, which directly specifies the amino acid sequence of its protein product. The boundaries of a genomic coding sequence are generally determined by a ribosome binding site (prokaryotes) or by the ATG start codon (eukaryotes) located just upstream of the open reading frame at the 5' end of the mRNA and a transcription

terminator sequence located just downstream of the open reading frame at the 3' end of the mRNA. A coding sequence can include, but is not limited to, DNA, cDNA, and recombinant nucleic acid sequences.

5 [52] An isolated polynucleotide encoding a GDH polypeptide of the present invention may be manipulated in a variety of ways to provide for expression of the polypeptide. Manipulation of the isolated polynucleotide prior to its insertion into a vector may be desirable or necessary depending on the expression vector. The techniques for modifying polynucleotides and nucleic acid sequences utilizing recombinant DNA methods are well known in the art.

10 [53] The term "control sequence" is defined herein to include all components, which are necessary or advantageous for the expression of a polypeptide of the present invention. Each control sequence may be native or foreign to the nucleic acid sequence encoding the polypeptide. Such control sequences include, but are not limited to, a leader, polyadenylation sequence, 15 propeptide sequence, promoter, signal peptide sequence, and transcription terminator. At a minimum, the control sequences include a promoter, and transcriptional and translational stop signals. The control sequences may be provided with linkers for the purpose of introducing specific restriction sites facilitating ligation of the control sequences with the coding region of the 20 nucleic acid sequence encoding a polypeptide.

[54] The term "operably linked" is defined herein as a configuration in which a control sequence is appropriately placed at a position relative to the coding sequence of the DNA sequence such that the control sequence directs the expression of a polypeptide.

25 [55] The control sequence may be an appropriate promoter sequence. The "promoter sequence" is a relatively short nucleic acid sequence that is recognized by a host cell for expression of the longer coding region that follows. The promoter sequence contains transcriptional control sequences, which mediate the expression of the polypeptide. The promoter may be any 30 nucleic acid sequence which shows transcriptional activity in the host cell of choice including mutant, truncated, and hybrid promoters, and may be obtained

from genes encoding extracellular or intracellular polypeptides either homologous or heterologous to the host cell.

[56] For bacterial host cells, suitable promoters for directing the transcription of the nucleic acid constructs of the present invention, include the promoters
5 obtained from the *E. coli* lac operon, *Streptomyces coelicolor* agarase gene (dagA), *Bacillus subtilis* levansucrase gene (sacB), *Bacillus licheniformis* alpha-amylase gene (amyL), *Bacillus stearothermophilus* maltogenic amylase gene (amyM), *Bacillus amyloliquefaciens* alpha-amylase gene (amyQ), *Bacillus licheniformis* penicillinase gene (penP), *Bacillus subtilis* xylA and xylB genes,
10 and prokaryotic beta-lactamase gene (Villa-Kamaroff et al., 1978, Proceedings of the National Academy of Sciences USA 75: 3727-3731), as well as the tac promoter (DeBoer et al., 1983, Proceedings of the National Academy of Sciences USA 80: 21-25). Further promoters are described in "Useful proteins from recombinant bacteria" in Scientific American, 1980, 242: 74-94; and in
15 Sambrook et al., 1989, supra.

[57] For filamentous fungal host cells, suitable promoters for directing the transcription of the nucleic acid constructs of the present invention include promoters obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *Aspergillus niger* neutral alpha-amylase,
20 *Aspergillus niger* acid stable alpha-amylase, *Aspergillus niger* or *Aspergillus awamori* glucoamylase (glaA), *Rhizomucor miehei* lipase, *Aspergillus oryzae* alkaline protease, *Aspergillus oryzae* triose phosphate isomerase, *Aspergillus nidulans* acetamidase, and *Fusarium oxysporum* trypsin-like protease (WO 96/00787), as well as the NA2-tpi promoter (a hybrid of the promoters from the
25 genes for *Aspergillus niger* neutral alpha-amylase and *Aspergillus oryzae* triose phosphate isomerase), and mutant, truncated, and hybrid promoters thereof.

[58] In a yeast host, useful promoters are obtained from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* galactokinase (GAL1), *Saccharomyces cerevisiae* alcohol
30 dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP), and

Saccharomyces cerevisiae 3-phosphoglycerate kinase. Other useful promoters for yeast host cells are described by Romanos et al., 1992, Yeast 8:423-488.

[59] The control sequence may also be a suitable transcription terminator sequence, a sequence recognized by a host cell to terminate transcription. The terminator sequence is operably linked to the 3' terminus of the nucleic acid sequence encoding the polypeptide. Any terminator, which is functional in the host cell of choice, may be used in the present invention.

[60] Preferred terminators for filamentous fungal host cells are obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Aspergillus niger* alpha-glucosidase, and *Fusarium oxysporum* trypsin-like protease.

[61] Preferred terminators for yeast host cells are obtained from the genes for *Saccharomyces cerevisiae* enolase, *Saccharomyces cerevisiae* cytochrome C (CYC1), and *Saccharomyces cerevisiae* glyceraldehyde-3-phosphate dehydrogenase. Other useful terminators for yeast host cells are described by Romanos et al., 1992, supra.

[62] The control sequence may also be a suitable leader sequence, a nontranslated region of an mRNA which is important for translation by the host cell. The leader sequence is operably linked to the 5' terminus of the nucleic acid sequence encoding the polypeptide. Any leader sequence that is functional in the host cell of choice may be used in the present invention. Preferred leaders for filamentous fungal host cells are obtained from the genes for *Aspergillus oryzae* TAKA amylase and *Aspergillus nidulans* triose phosphate isomerase. Suitable leaders for yeast host cells are obtained from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* 3-phosphoglycerate kinase, *Saccharomyces cerevisiae* alpha-factor, and *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP).

[63] The control sequence may also be a polyadenylation sequence, a sequence operably linked to the 3' terminus of the nucleic acid sequence and

- which, when transcribed, is recognized by the host cell as a signal to add polyadenosine residues to transcribed mRNA. Any polyadenylation sequence that is functional in the host cell of choice may be used in the present invention. Preferred polyadenylation sequences for filamentous fungal host cells are
- 5 obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Fusarium oxysporum* trypsin-like protease, and *Aspergillus niger* alpha-glucosidase. Useful polyadenylation sequences for yeast host cells are described by Guo and Sherman, 1995, Molecular Cellular Biology 15: 5983-5990.
- 10 [64] The control sequence may also be a signal peptide coding region that codes for an amino acid sequence linked to the amino terminus of a polypeptide and directs the encoded polypeptide into the cell's secretory pathway. The 5' end of the coding sequence of the nucleic acid sequence may inherently contain a
- 15 signal peptide coding region naturally linked in translation reading frame with the segment of the coding region that encodes the secreted polypeptide. Alternatively, the 5' end of the coding sequence may contain a signal peptide coding region that is foreign to the coding sequence. The foreign signal peptide coding region may be required where the coding sequence does not naturally contain a signal peptide coding region.
- 20 [65] Alternatively, the foreign signal peptide coding region may simply replace the natural signal peptide coding region in order to enhance secretion of the polypeptide. However, any signal peptide coding region that directs the expressed polypeptide into the secretory pathway of a host cell of choice may be used in the present invention.
- 25 [66] Effective signal peptide coding regions for bacterial host cells are the signal peptide coding regions obtained from the genes for Bacillus NCIB 11837 maltogenic amylase, *Bacillus stearothermophilus* alpha-amylase, *Bacillus licheniformis* subtilisin, *Bacillus licheniformis* beta-lactamase, *Bacillus stearothermophilus* neutral proteases (nprT, nprS, nprM), and *Bacillus subtilis*
- 30 prsA. Further signal peptides are described by Simonen and Palva, 1993, Microbiological Reviews 57: 109-137.

[67] Effective signal peptide coding regions for filamentous fungal host cells are the signal peptide coding regions obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* neutral amylase, *Aspergillus niger* glucoamylase, *Rhizomucor miehei* aspartic proteinase, *Humicola insolens* cellulase, and *Humicola lanuginosa* lipase.

[68] Useful signal peptides for yeast host cells are obtained from the genes for *Saccharomyces cerevisiae* alpha-factor and *Saccharomyces cerevisiae* invertase. Other useful signal peptide coding regions are described by Romanos et al., 1992, supra.

[69] The control sequence may also be a propeptide coding region that codes for an amino acid sequence positioned at the amino terminus of a polypeptide. The resultant polypeptide is known as a proenzyme or propolypeptide (or a zymogen in some cases). A propolypeptide is generally inactive and can be converted to a mature active polypeptide by catalytic or autocatalytic cleavage of the propeptide from the propolypeptide. The propeptide coding region may be obtained from the genes for *Bacillus subtilis* alkaline protease (aprE), *Bacillus subtilis* neutral protease (nprT), *Saccharomyces cerevisiae* alpha-factor, *Rhizomucor miehei* aspartic proteinase, and *Myceliophthora thermophila* lactase (WO 95/33836).

[70] Where both signal peptide and propeptide regions are present at the amino terminus of a polypeptide, the propeptide region is positioned next to the amino terminus of a polypeptide and the signal peptide region is positioned next to the amino terminus of the propeptide region.

[71] It may also be desirable to add regulatory sequences, which allow the regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory systems are those which cause the expression of the gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. In prokaryotic host cells, suitable regulatory sequences include the lac, tac, and trp operator systems. In yeast host cells, suitable regulatory systems include the ADH2 system or GAL1 system. In filamentous fungi, suitable regulatory sequences include the TAKA

alpha-amylase promoter, *Aspergillus niger* glucoamylase promoter, and *Aspergillus oryzae* glucoamylase promoter.

- [72] Other examples of regulatory sequences are those which allow for gene amplification. In eukaryotic systems, these include the dihydrofolate reductase gene, which is amplified in the presence of methotrexate, and the metallothionein genes, which are amplified with heavy metals. In these cases, the nucleic acid sequence encoding the GDH polypeptide of the present invention would be operably linked with the regulatory sequence.

Expression Vectors

- [73] In another aspect, the present invention is also directed to a recombinant expression vector comprising a polynucleotide of the present invention (which encodes a GDH polypeptide of the present invention), and one or more expression regulating regions such as a promoter and a terminator, a replication origin, etc., depending on the type of hosts into which they are to be introduced.
- The various nucleic acid and control sequences described above may be joined together to produce a recombinant expression vector which may include one or more convenient restriction sites to allow for insertion or substitution of the nucleic acid sequence encoding the polypeptide at such sites. Alternatively, the nucleic acid sequence of the present invention may be expressed by inserting the nucleic acid sequence or a nucleic acid construct comprising the sequence into an appropriate vector for expression. In creating the expression vector, the coding sequence is located in the vector so that the coding sequence is operably linked with the appropriate control sequences for expression.
- [74] The recombinant expression vector may be any vector (e.g., a plasmid or virus), which can be conveniently subjected to recombinant DNA procedures and can bring about the expression of the polynucleotide sequence. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which the vector is to be introduced. The vectors may be linear or closed circular plasmids.

X₂₁₆ Y I X₂₁₉ X₂₂₀ P X₂₂₂ E I A A V A A W L A S X₂₃₄ E A X₂₃₇ Y V T G X₂₄₂
 T L F A D G G M T X₂₅₂ X₂₅₃ P S F Q A G R G

- The diversity of changes at various residue positions for the GDH polypeptides
 5 of the present invention are shown to the right of the arrow in Table 2 below
 and relative amino acid residues of wild-type GDH of SEQ ID NO: 2 (EP
 055375) which are shown to the left of the arrow:

Table 2

X ₁₁ :	A → D,
X ₁₆ :	A → T
X ₄₆ :	N → D
X ₄₇ :	E → D, K,
X ₅₄ :	K → R
X ₆₃ :	Q → R
X ₆₈ :	K → N,
X ₇₅ :	I → , V
X ₉₂ :	N → S
X ₉₅ :	L → F
X ₉₆ :	E → A
X ₁₁₄ :	G → S
X ₁₆₅ :	I → M, L, V, T
X ₁₇₀ :	E → K
X ₁₇₈ :	P → Q
X ₁₉₄ :	P → T
X ₁₉₇ :	A → K
X ₁₉₈ :	E → A, D, G, K, L
X ₂₀₀ :	F → M, Y
X ₂₀₁ :	A → G, L, S, T
X ₂₀₂ :	D → N
X ₂₀₃ :	P → D, N, R, S
X ₂₀₄ :	K → D, E, L, N, T, Q
X ₂₀₅ :	Q → N
X ₂₀₆ :	K → R
X ₂₀₉ :	V → A
X ₂₁₀ :	E → A, K
X ₂₁₁ :	S → G, T
X ₂₁₂ :	M → K
X ₂₁₆ :	G → L
X ₂₁₉ :	G → A
X ₂₂₀ :	E → Q
X ₂₂₂ :	E → D
X ₂₃₄ :	K → E, S
X ₂₃₇ :	S → C
X ₂₄₂ :	I → V

dropwise on demand by the automatic titrator (a pH of 6.85 was set as a lower limit) to constantly adjust the pH to 7.0. The reaction was complete when no more caustic was needed. The reaction rates were determined by measuring the amount of base added per unit time or by taking samples of the reaction mixture, extracting the sample 3 times with an equal volume of ethyl acetate, and analyzing the combined organic layers by gas chromatography to determine the amount of ethyl-S- 4-chloro-3-hydroxybutyrate produced per unit time.

[104] While the invention has been described with reference to certain embodiments, it will be understood by those skilled in the art that various changes may be made and equivalents may be substituted without departing from the scope of the invention. In addition, many modifications may be made to adapt a particular situation or material to the teachings of the invention without departing from its scope. Therefore, it is intended that the invention not be limited to the particular embodiment disclosed, but that the invention will include all embodiments falling within the scope of the appended claims.

2004288134 13 Nov 2009.

40

31. A method of making a non-naturally occurring polypeptide of claim 1,
substantially as hereinbefore described with reference to any one of the examples.

Dated 13 November, 2009
Codexis, Inc.

5

Patent Attorneys for the Applicant/Nominated Person
SPRUSON & FERGUSON

2389098-1:rec

1/4

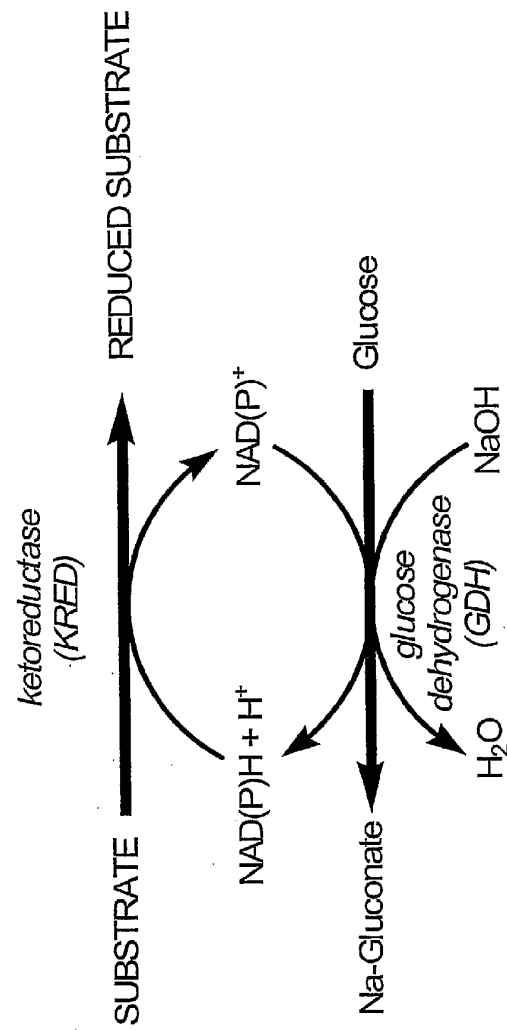


FIG. 1

4/4

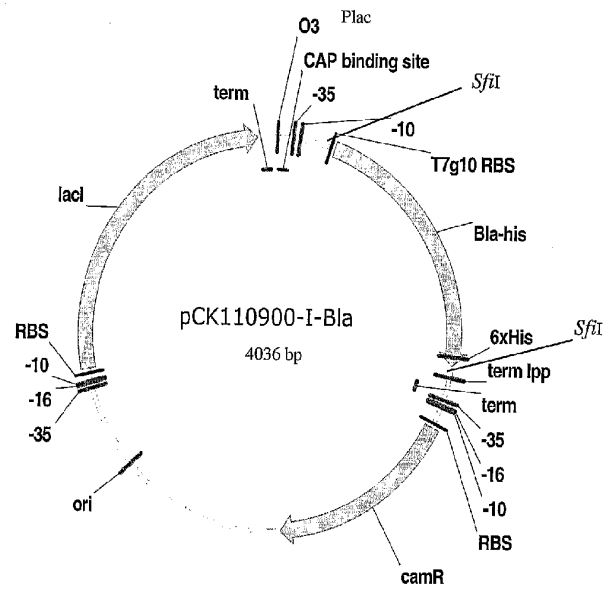


FIG. 3

SEQUENCE LISTING

<110> Krebber, Anke
Newman, Lisa M.
Davis, S. C.
Jenne, Stephane J.

<120> IMPROVED GLUCOSE DEHYDROGENASE POLYPEPTIDES AND RELATED
POLYNUCLEOTIDES

<130> 15076WO03 0352.310WO

<150> 60/494,300

<151> 2003-08-11

<150> 60/545,657

<151> 2004-02-18

<160> 168

<170> PatentIn version 3.2

<210> 1

<211> 786

<212> DNA

<213> Bacillus subtilis

<220>

<221> misc_feature

<223> based upon B. subtilis GDH backbone (SO6-3orf)

<220>

<221> CDS

<222> (1)..(786)

<400> 1

atg tat ccg gat tta aaa gga aaa gtc gtc gct att aca gga gct gct 48
Met Tyr Pro Asp Leu Lys Gly Lys Val Val Ala Ile Thr Gly Ala Ala
1 5 10 15

tca ggg ctc gga aag gcg atg gcc att cgc ttc ggc aag gag cag gca 96
Ser Gly Leu Gly Lys Ala Met Ala Ile Arg Phe Gly Lys Glu Gln Ala
20 25 30

aaa gtg gtt atc aac tat tat agt aat aaa caa gat ccg aac gag gta 144
Lys Val Val Ile Asn Tyr Tyr Ser Asn Lys Gln Asp Pro Asn Glu Val
35 40 45

