



(19) **United States**

(12) **Patent Application Publication**
FLINT et al.

(10) **Pub. No.: US 2010/0081154 A1**

(43) **Pub. Date: Apr. 1, 2010**

(54) **IDENTIFICATION AND USE OF BACTERIAL [2FE-2S] DIHYDROXY-ACID DEHYDRATASES**

(75) Inventors: **DENNIS FLINT**, Newark, DE (US); **Steven Cary Rothman**, Wilmington, DE (US); **Wonchul Suh**, Hockessin, DE (US); **Jean-Francois Tomb**, Wilmington, DE (US); **Rick W. Ye**, Hockessin, DE (US)

Correspondence Address:
E I DU PONT DE NEMOURS AND COMPANY
LEGAL PATENT RECORDS CENTER
BARLEY MILL PLAZA 25/1122B, 4417 LANCASTER PIKE
WILMINGTON, DE 19805 (US)

(73) Assignee: **BUTAMAX(TM) ADVANCED BIOFUELS LLC**, Wilmington, DE (US)

(21) Appl. No.: **12/569,636**

(22) Filed: **Sep. 29, 2009**

Related U.S. Application Data

(60) Provisional application No. 61/100,792, filed on Sep. 29, 2008.

Publication Classification

(51) **Int. Cl.**
C12Q 1/34 (2006.01)
C12N 1/20 (2006.01)
C12N 1/16 (2006.01)
C12P 7/16 (2006.01)
C12P 7/18 (2006.01)
(52) **U.S. Cl.** **435/18**; 435/252.1; 435/255.1; 435/160; 435/158

(57) **ABSTRACT**

A group of bacterial dihydroxy-acid dehydratases having a [2Fe-2S] cluster was discovered. Bacterial [2Fe-2S] DHADs were expressed as heterologous proteins in bacteria and yeast cells, providing DHAD activity for conversion of 2,3-dihydroxyisovalerate to α -ketoisovalerate or 2,3-dihydroxymethylvalerate to α -ketomethylvalerate. Isobutanol and other compounds may be synthesized in pathways that include bacterial [2Fe-2S] DHAD activity.

2Fe-2S		56	129	201
	<i>S. mutans</i> DHAD	TPCNIHL	IGGCDKNM	CPGPGGCGGGM
	<i>S. solfataricus</i> DHAD	GPCNFHT	IGGCDKTT	HPTLGTCSGL
	<i>A. thaliana</i> (plant) DHAD	NTCNMHL	IPGCDKNM	CPGAGACGGM
	<i>N. europaea</i> DHAD	TPCNMGL	IGGCDKNM	CPGAGSCGGM
	<i>Synechocystis</i> DHAD	TPCNMGI	IGGCDKNM	CPGAGSCGGM
	<i>S. thermophilus</i> DHAD	TPCNIHL	IGGCDKNM	CPGPGGCGGGM
	<i>R. metallidurans</i> DHAD	TPCNAGL	IGGCDKNM	CPSTGSCGGM
	<i>R. eutropha</i> DHAD	TPCNAGL	IGGCDKNM	CPSTGSCGGM
	<i>L. lactis</i> DHAD	NPCNMHL	VPGCDKNM	IPGQQGACGGM
4Fe-4S				
	<i>E. coli</i> DHAD	VPGHVHL	ISNCDKIT	CPTCGSCSGM

FIG. 1

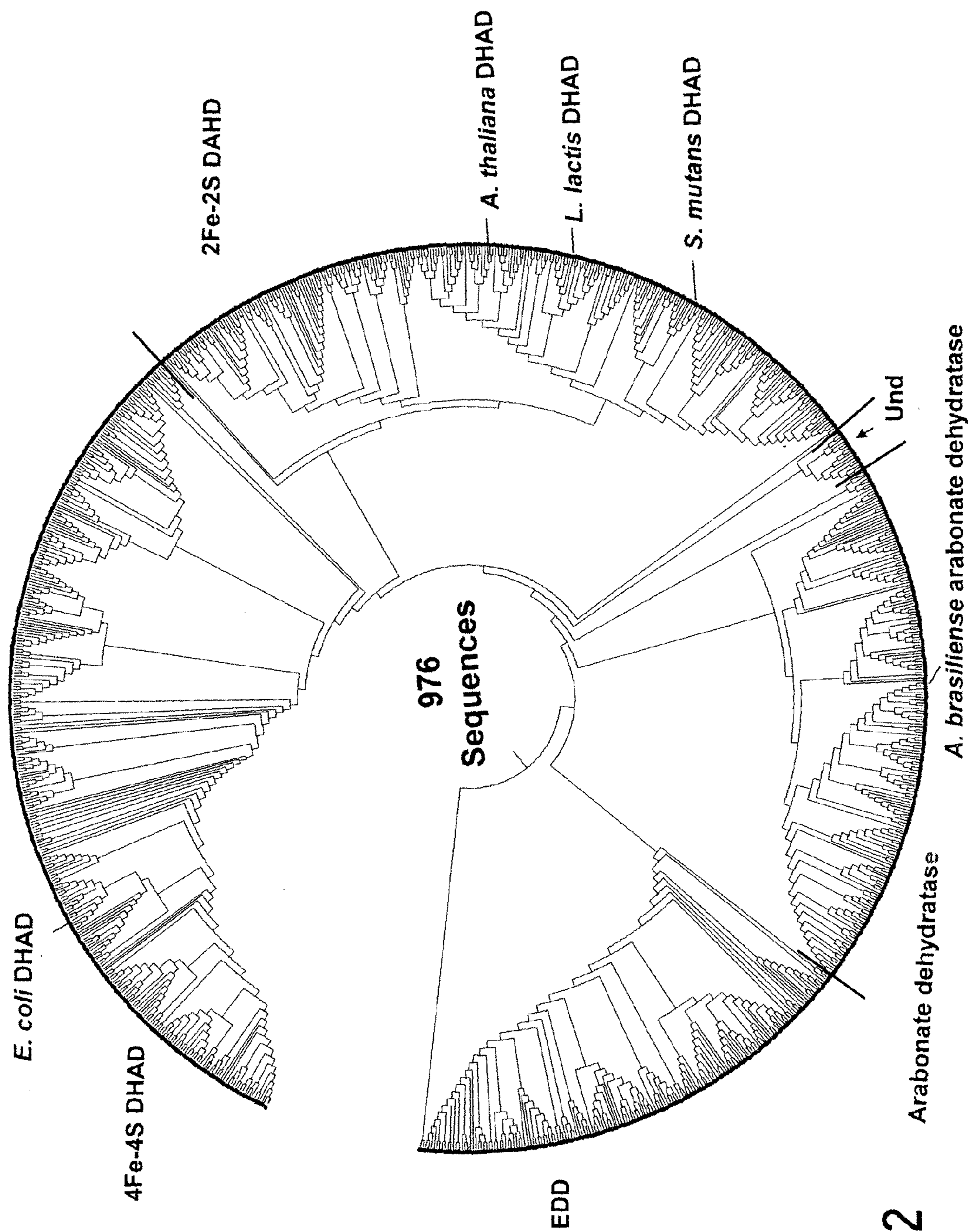


FIG. 2

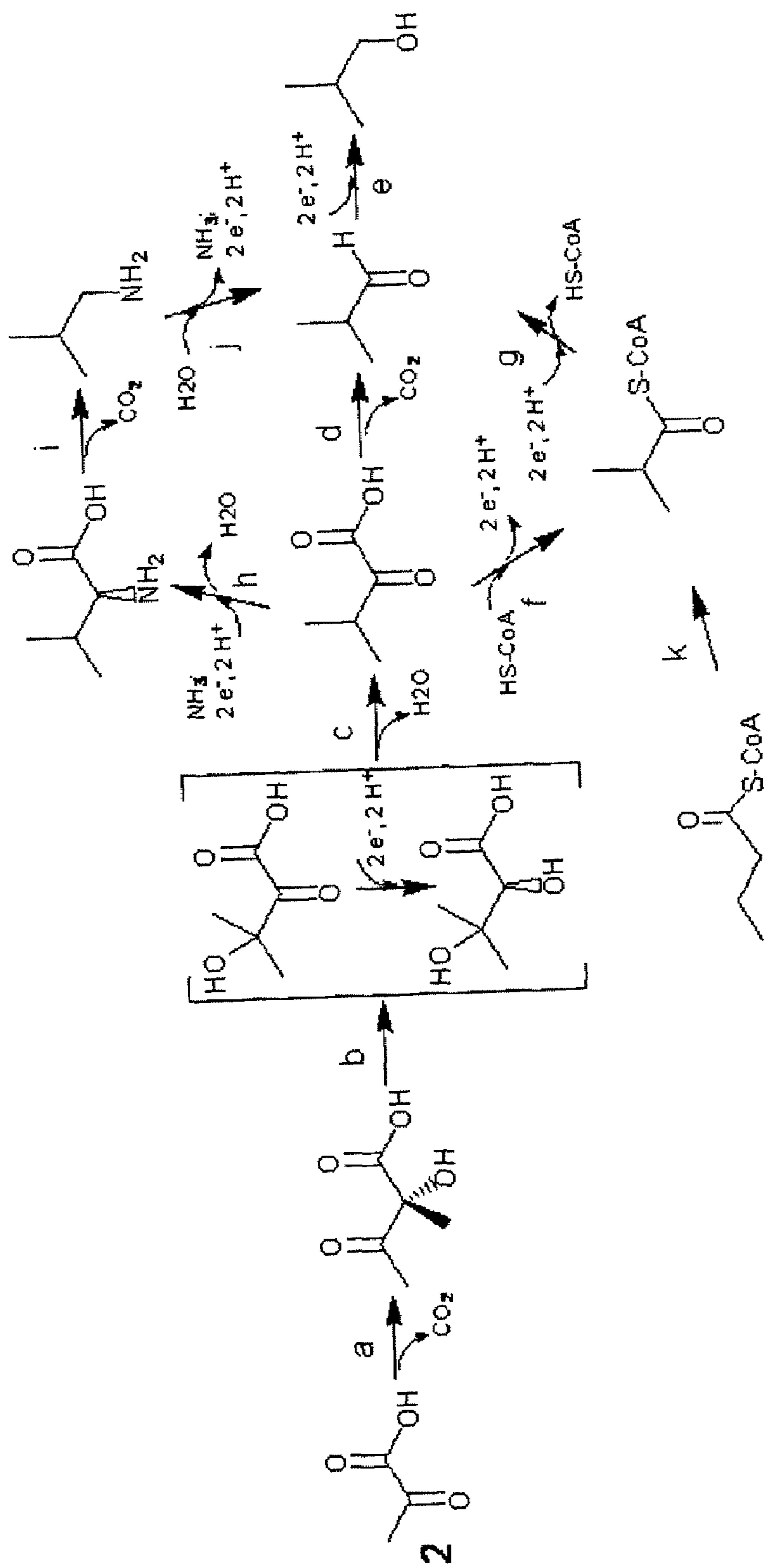


FIG. 3

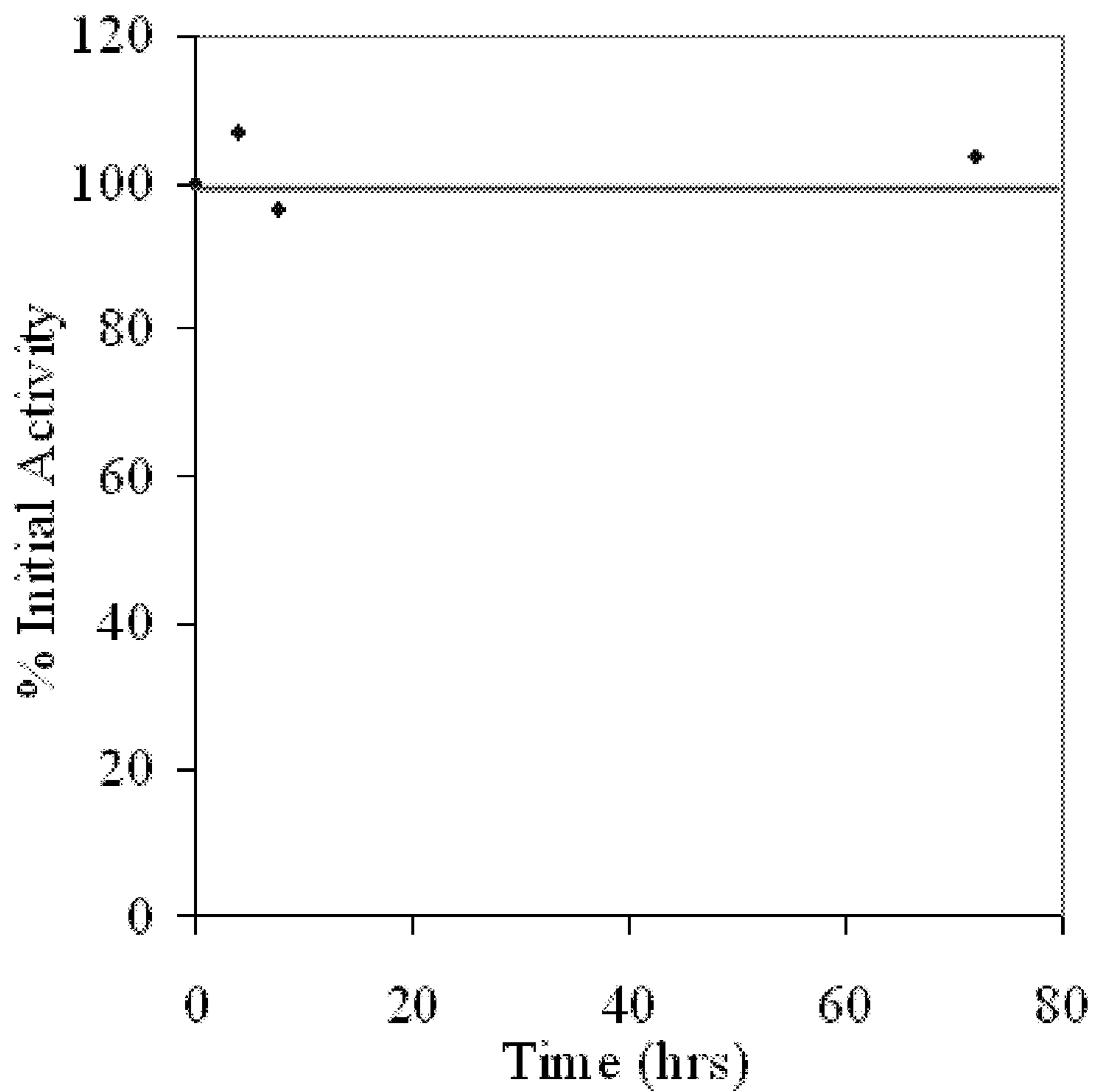


FIG. 4A

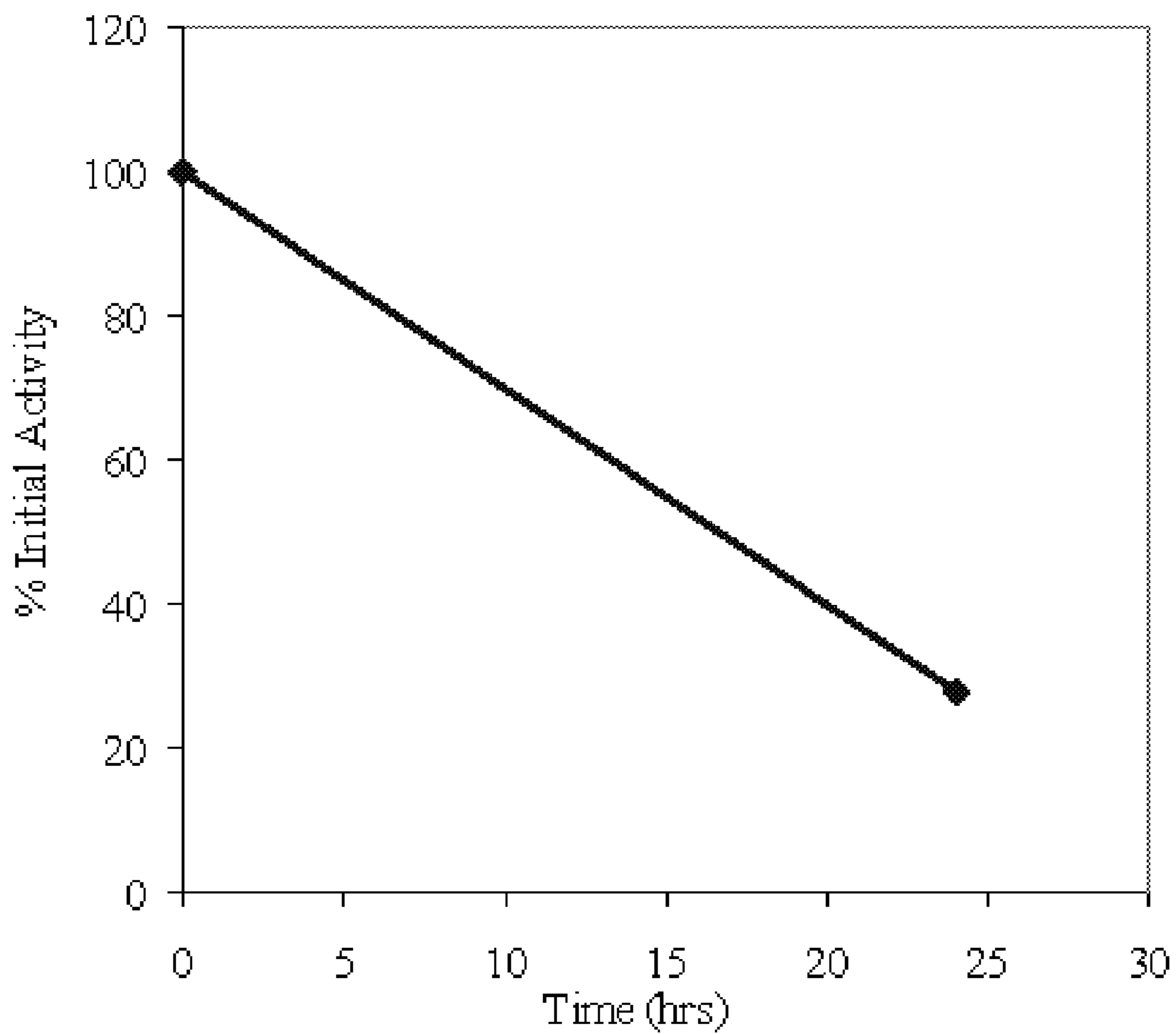


FIG. 4B

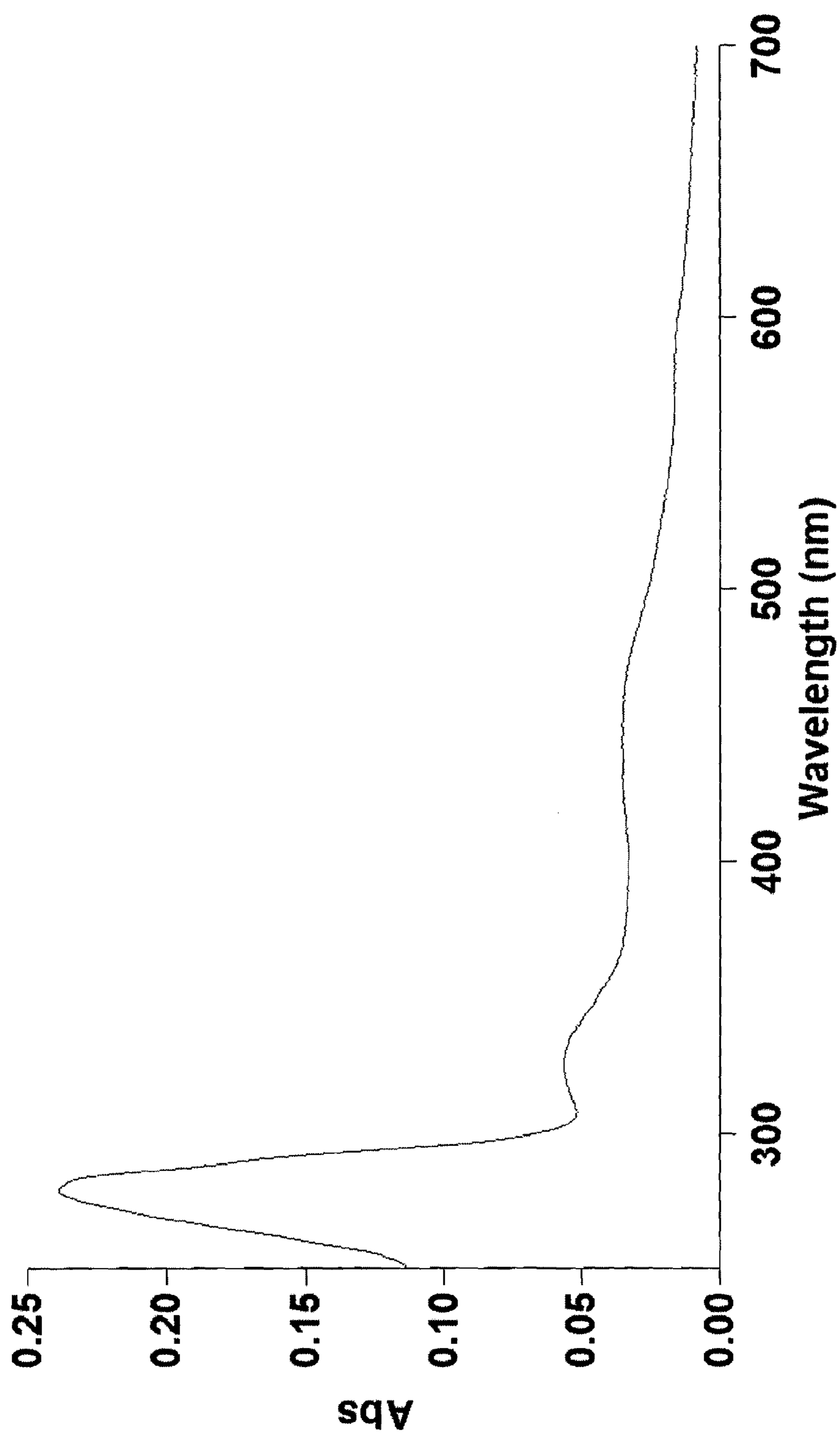


FIG. 5

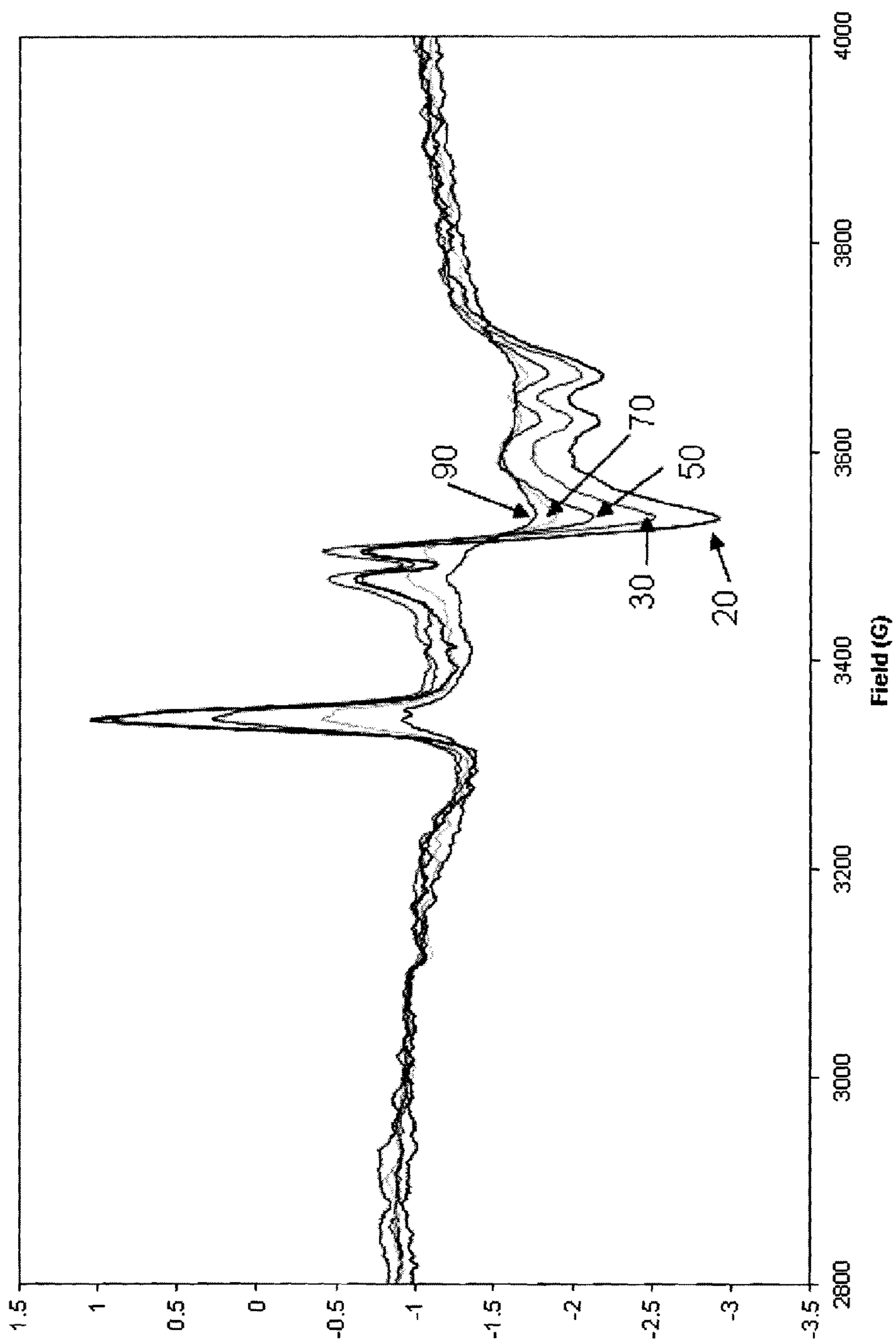


FIG. 6

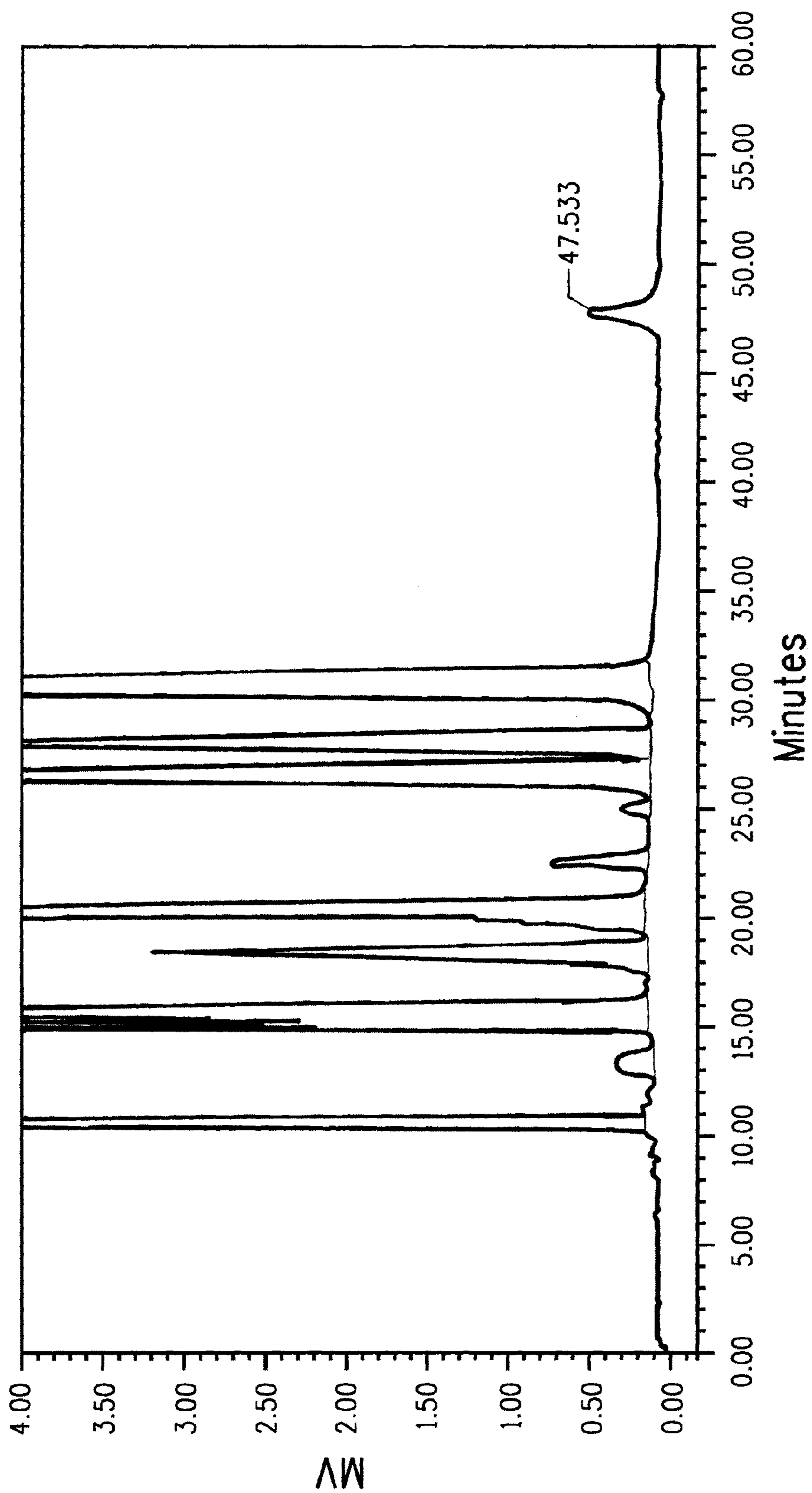


FIG. 7

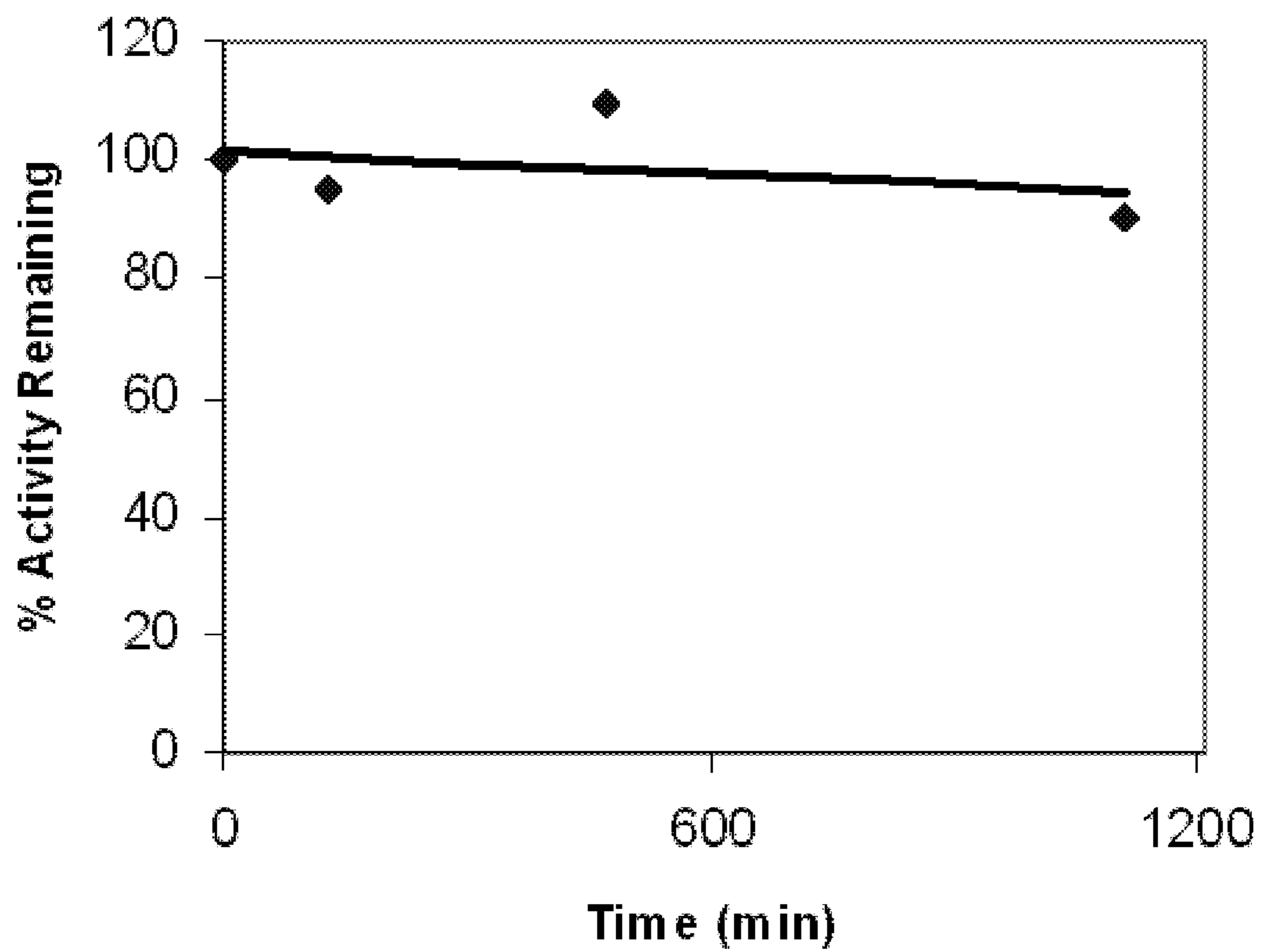


FIG. 8

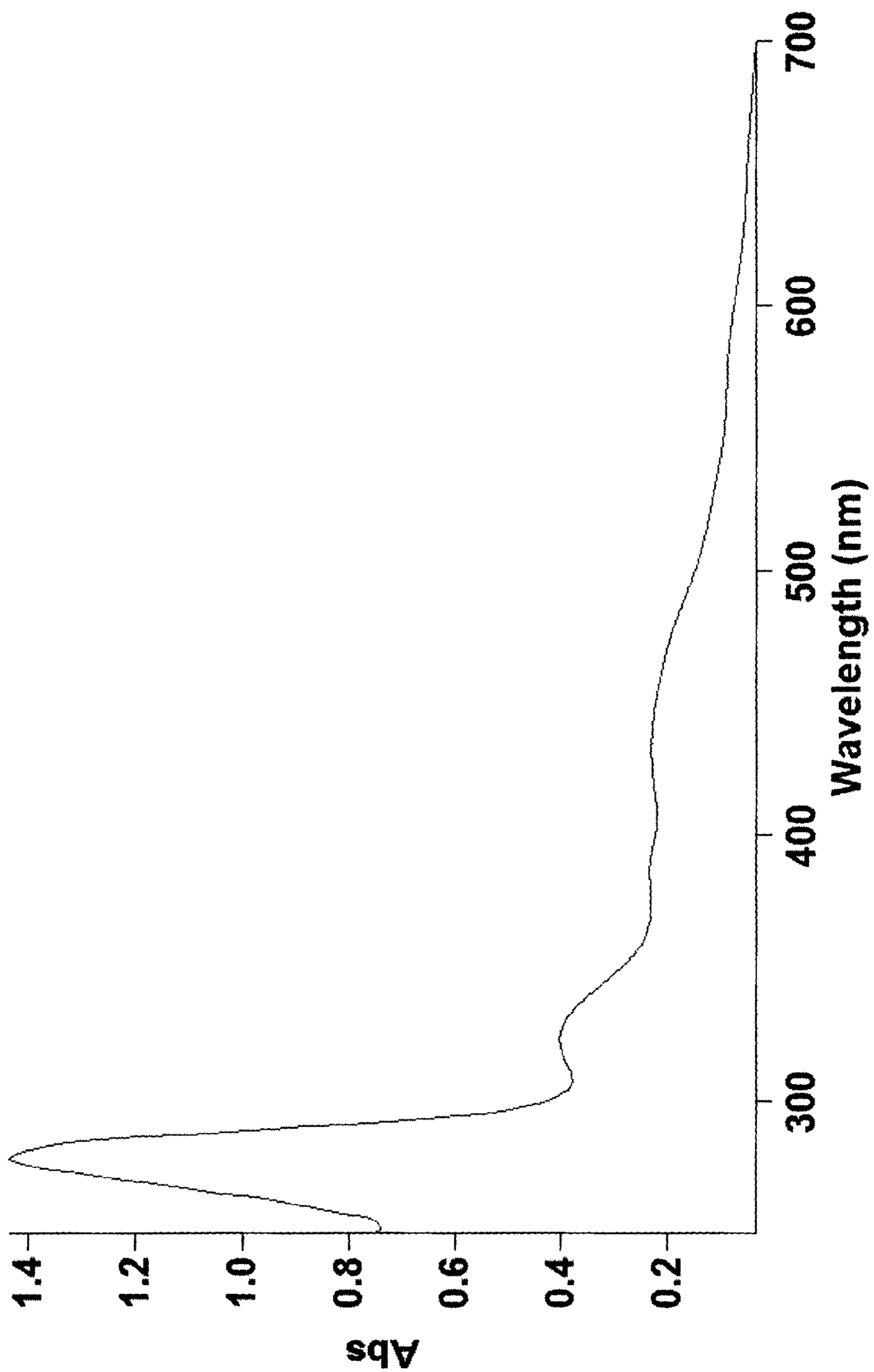


FIG. 9

IDENTIFICATION AND USE OF BACTERIAL [2FE-2S] DIHYDROXY-ACID DEHYDRATASES

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is related to and claims the benefit of priority of U.S. Provisional Application No. 61/100,792, filed Sep. 29, 2008, the entirety of which is herein incorporated by reference.

FIELD OF THE INVENTION

[0002] The invention relates to the field of industrial microbiology and the expression of dihydroxy-acid dehydratase activity. More specifically, bacterial dihydroxy-acid dehydratases with a [2Fe-2S] cluster are identified and expressed as heterologous proteins in bacterial and yeast hosts.

BACKGROUND OF THE INVENTION

[0003] Dihydroxy-acid dehydratase (DHAD), also called acetohydroxy acid dehydratase, catalyzes the conversion of 2,3-dihydroxyisovalerate to α -ketoisovalerate and of 2,3-dihydroxymethylvalerate to α -ketomethylvalerate. The DHAD enzyme, classified as E.C. 4.2.1.9, is part of naturally occurring biosynthetic pathways producing valine, isoleucine, leucine and pantothenic acid (vitamin B5). Increased expression of DHAD activity is desired for enhanced microbial production of branched chain amino acids or of pantothenic acid.

[0004] DHAD catalyzed conversion of 2,3-dihydroxyisovalerate to α -ketoisovalerate is also a common step in the multiple isobutanol biosynthetic pathways that are disclosed in commonly owned and co-pending US Patent Application Publication US 20070092957 A1. Disclosed therein is engineering of recombinant microorganisms for production of isobutanol. Isobutanol is useful as a fuel additive, whose availability may reduce the demand for petrochemical fuels.

[0005] For improved production of compounds synthesized in pathways including dihydroxy-acid dehydratase, it is desirable to express a heterologous enzyme that provides this enzymatic activity in the production host of interest. Obtaining high functional expression of dihydroxy-acid dehydratases in a heterologous host is complicated by the enzyme's requirement for an Fe—S cluster, which involves availability and proper loading of the cluster into the DHAD apo-protein.

[0006] Fe—S cluster requiring DHAD enzymes are known in the art and are found either in the [4Fe-4S] or [2Fe-2S] form. Some bacterial enzymes are known, the best characterized of which is from *E. coli* (Flint, D H, et al. (1993) *J. Biol. Chem.* 268:14732-14742). However these bacterial enzymes are all in the [4Fe-4S] form. The only [2Fe-2S] form reported to date is a spinach enzyme (Flint and Emptage (1988) *J. Biol. Chem.* 263:3558-3564).

[0007] It is desirable to use the [2Fe-2S] form of the enzyme in host cells to enhance the production of introduced biosynthetic pathways as the [2Fe-2S] form creates a lesser burden on Fe—S cluster synthesis and/or assembly. Unfortunately, only one [2Fe-2S] form of this enzyme is known.

[0008] There exists a need therefore to identify new [2Fe-2S] forms of DHAD for use in recombinant host cells where the conversion of 2,3-dihydroxyisovalerate to α -ketoisovalerate is a metabolic pathway step in a desired biosynthetic pathway.

SUMMARY OF THE INVENTION

[0009] Provided herein is a method for identifying [2Fe-2S] DHAD enzymes, said method comprising:

[0010] a) querying one or more amino acid sequences with a Profile Hidden Markov Model prepared using the proteins of SEQ ID NOs:164, 168, 230, 232, 298, 310, 344, and 346, wherein a match with an E-value of less than 10^{-5} provides a first subset of sequences whereby said first subset of sequences correspond to one or more DHAD related proteins;

[0011] b) analyzing the first subset of sequences that correspond to one or more DHAD related proteins of step (a) for the presence of three conserved cysteines that correspond to positions 56, 129, and 201 in the *Streptococcus mutans* dihydroxy-acid dehydratase amino acid sequence (SEQ ID NO: 168) whereby a second subset of sequences encoding [2Fe-2S] DHAD enzymes are identified; and

[0012] c) analyzing said second subset of sequences of step (b) for the presence of signature conserved amino acids at positions corresponding to positions in the *Streptococcus mutans* DHAD amino acid sequence (SEQ ID NO: 168) that are aspartic acid at position 88, arginine or asparagine at position 142, asparagine at position 208, and leucine at position 454 whereby a third subset of sequences encoding [2Fe-2S] DHAD enzymes are further identified.

[0013] In another aspect of the invention, the above method further comprises:

[0014] d) expressing a polypeptide having a sequence identifiable by any one or all of steps a), b), and c) in a cell; and

[0015] e) confirming that said polypeptide has DHAD activity in the cell.

[0016] In another aspect of the invention, the method above further comprises:

[0017] d) purifying a protein encoded by a sequence identifiable by any one or all of steps a), b), and c); and

[0018] e) confirming that said protein is a [2Fe-2S] DHAD enzyme by UV-vis and EPR spectroscopy.

[0019] In another aspect of the invention, the method above further comprises selecting one or more sequences corresponding to bacterial [2Fe-2S] DHAD enzyme sequences identified in any one or all of steps a), b), and c). Said selected sequences may be expressed in a cell; and the DHAD activity in the cell may be confirmed. Said selected sequences may be further purified such that a purified protein is obtained and the [2Fe-2S] DHAD enzyme activity of said purified protein may be confirmed by UV-vis and EPR spectroscopy.

[0020] Another aspect of the invention is directed to a microbial host cell comprising at least one heterologous [2Fe-2S] DHAD enzyme identifiable by the methods described herein. Said cell maybe a bacterial cell or a yeast cell. The cell may also be a recombinant cell that produces isobutanol.

[0021] Another aspect of the invention is a method for the production of isobutanol comprising:

[0022] a) providing the microbial host cell comprising at least one heterologous [2Fe-2S] DHAD enzyme identifiable by the methods described herein wherein said host cell further comprises an isobutanol biosynthetic pathway; and

[0023] b) growing the host cell of step (a) under conditions wherein isobutanol is produced.

[0024] Another aspect of the invention is a method for the conversion of 2,3-dihydroxyisovalerate to α -ketoisovalerate comprising:

[0025] a) providing a microbial host cell comprising at least one heterologous [2Fe-2S] DHAD enzyme identifiable by the methods described herein and a source of 2,3-dihydroxyisovalerate; and

[0026] b) growing the microbial host cell of (a) under conditions where the 2,3-dihydroxyisovalerate is converted to α -ketoisovalerate.

BRIEF DESCRIPTION OF THE FIGURES AND SEQUENCE DESCRIPTIONS

[0027] The invention can be more fully understood from the following detailed description, figure, and the accompanying sequence descriptions, which form a part of this application.

[0028] FIG. 1 shows the conserved cysteine regions, with C for cysteine in bold, in representative bacterial [2Fe-2S] DHADs and a [4Fe-4S] DHAD. Single letter amino acid abbreviations are used.

[0029] FIG. 2 shows a phylogenetic tree of DHAD related proteins. Branches for [4Fe-4S] and [2Fe-2S] DHADs as well as EDDs, aldonic acid dehydratases and an undefined group (Und) are marked. Select individual DHADs are labeled.

[0030] FIG. 3 shows biosynthetic pathways for isobutanol production.

[0031] FIG. 4 shows graphs of stability of activity in air of A) *S. mutans* DHAD, and B) *E. coli* DHAD.

[0032] FIG. 5 shows a plot of the UV-visible spectrum of *S. mutans* DHAD.

[0033] FIG. 6 shows a plot of the electron paramagnetic resonance (EPR) spectrum of *S. mutans* DHAD at different temperatures between 20° K. and 90° K.

[0034] FIG. 7 shows HPLC analysis of an extract of yeast cells that express acetolactate synthase, KARI, and *S. mutans* DHAD genes, with an isobutanol peak at 47.533 minutes.

[0035] FIG. 8 shows a graph of stability of *L. lactis* DHAD in air.

[0036] FIG. 9 shows a plot of the UV-visible spectrum of the purified *L. lactis* DHAD.

[0037] Table 1 is a table of the Profile HMM for dihydroxy-acid dehydratases based on enzymes with assayed function prepared as described in Example 1. Table 1 is submitted herewith electronically and is incorporated herein by reference.

[0038] The following sequences conform with 37 C.F.R. 1.821-1.825 (“Requirements for Patent Applications Containing Nucleotide Sequences and/or Amino Acid Sequence Disclosures—the Sequence Rules”) and are consistent with World Intellectual Property Organization (WIPO) Standard ST.25 (1998) and the sequence listing requirements of the EPO and PCT (Rules 5.2 and 49.5 (a-bis), and Section 208 and Annex C of the Administrative Instructions). The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

TABLE 2a

Representative bacterial [2Fe—2S] DHAD proteins and encoding sequences		
Organism of derivation	SEQ ID NO: Nucleic acid	SEQ ID NO: Peptide
<i>Mycobacterium</i> sp. MCS	1	2
<i>Mycobacterium gilvum</i> PYR-GCK	3	4

TABLE 2a-continued

Representative bacterial [2Fe—2S] DHAD proteins and encoding sequences		
Organism of derivation	SEQ ID NO: Nucleic acid	SEQ ID NO: Peptide
<i>Mycobacterium smegmatis</i> str. MC2 155	5	6
<i>Mycobacterium vanbaalenii</i> PYR-1	7	8
<i>Nocardia farcinica</i> IFM 10152	9	10
<i>Rhodococcus</i> sp. RHA1	11	12
<i>Mycobacterium ulcerans</i> Agy99	13	14
<i>Mycobacterium avium</i> subsp. <i>paratuberculosis</i> K-10	15	16
<i>Mycobacterium tuberculosis</i> H37Ra	17	18
<i>Mycobacterium leprae</i> TN *	19	20
<i>Kineococcus radiotolerans</i> SRS30216	21	22
<i>Janibacter</i> sp. HTCC2649	23	24
<i>Nocardioides</i> sp. JS614	25	26
<i>Renibacterium salmoninarum</i> ATCC 33209	27	28
<i>Arthrobacter aurescens</i> TC1	29	30
<i>Leifsonia xyli</i> subsp. <i>xyli</i> str. CTCB07	31	32
Marine actinobacterium PHSC20C1	33	34
<i>Clavibacter michiganensis</i> subsp. <i>michiganensis</i> NCPPB 382	35	36
<i>Saccharopolyspora erythraea</i> NRRL 2338	37	38
<i>Acidothermus cellulolyticus</i> 11B	39	40
<i>Corynebacterium efficiens</i> YS-314	41	42
<i>Brevibacterium linens</i> BL2	43	44
<i>Tropheryma whipplei</i> TW08/27	45	46
<i>Methylobacterium extorquens</i> PA1	47	48
<i>Methylobacterium chloromethanicum</i>	428	429
<i>Methylobacterium nodulans</i> ORS 2060	49	50
<i>Rhodopseudomonas palustris</i> BisB5	51	52
<i>Rhodopseudomonas palustris</i> BisB18	53	54
<i>Bradyrhizobium</i> sp. ORS278	55	56
<i>Bradyrhizobium japonicum</i> USDA 110	57	58
<i>Fulvimarina pelagi</i> HTCC2506	59	60
<i>Aurantimonas</i> sp. SI85-9A1	61	62
<i>Hoeflea phototrophica</i> DFL-43	63	64
<i>Mesorhizobium loti</i> MAFF303099	65	66
<i>Mesorhizobium</i> sp. BNC1	67	68
<i>Parvibaculum lavamentivorans</i> DS-1	69	70
<i>Loktanella vestfoldensis</i> SKA53	71	72
<i>Roseobacter</i> sp. CCS2	73	74
<i>Dinoroseobacter shibae</i> DFL 12	75	76
<i>Roseovarius nubinhibens</i> ISM	77	78
<i>Sagittula stellata</i> E-37	79	80
<i>Roseobacter</i> sp. AzwK-3b	81	82
<i>Roseovarius</i> sp. TM1035	83	84
<i>Oceanicola batsensis</i> HTCC2597	85	86
<i>Oceanicola granulosus</i> HTCC2516	87	88
<i>Rhodobacteriales bacterium</i> HTCC2150	89	90
<i>Paracoccus denitrificans</i> PD1222	91	92
<i>Oceanibulbus indolifex</i> HEL-45	93	94
<i>Sulfitobacter</i> sp. EE-36	95	96
<i>Roseobacter denitrificans</i> OCh 114	97	98
<i>Jannaschia</i> sp. CCS1	99	100
<i>Caulobacter</i> sp. K31	101	102
<i>Candidatus Pelagibacter ubique</i> HTCC1062	103	104
<i>Erythrobacter litoralis</i> HTCC2594	105	106
<i>Erythrobacter</i> sp. NAP1	107	108
<i>Comamonas testosteroni</i> KF-1	109	110
<i>Sphingomonas wittichii</i> RW1	111	112
<i>Burkholderia xenovorans</i> LB400	113	114
<i>Burkholderia phytofirmans</i> PsJN	115	116
<i>Bordetella petrii</i> DSM 12804	117	118
<i>Bordetella bronchiseptica</i> RB50	119	120
<i>Bradyrhizobium</i> sp. ORS278	121	122
<i>Bradyrhizobium</i> sp. BTA1	123	124
<i>Bradyrhizobium japonicum</i>	125	126
<i>Sphingomonas wittichii</i> RW1	127	128
<i>Rhodobacteriales bacterium</i> HTCC2654	129	130

TABLE 2a-continued

Representative bacterial [2Fe—2S] DHAD proteins and encoding sequences		
Organism of derivation	SEQ ID NO: Nucleic acid	SEQ ID NO: Peptide
<i>Solibacter usitatus</i> Ellin6076	131	132
<i>Roseiflexus</i> sp. RS-1	133	134
<i>Rubrobacter xylanophilus</i> DSM 9941	135	136
<i>Salinispora tropica</i> CNB-440	137	138
<i>Acidobacteria bacterium</i> Ellin345	139	140
<i>Thermus thermophilus</i> HB27	141	142
<i>Maricaulis maris</i> MCS10	143	144
<i>Parvularcula bermudensis</i> HTCC2503	145	146
<i>Oceanicaulis alexandrii</i> HTCC2633	147	148
<i>Plesiocystis pacifica</i> SIR-1	149	150
<i>Bacillus</i> sp. NRRL B-14911	151	152
<i>Oceanobacillus iheyensis</i> HTE831	153	154
<i>Staphylococcus saprophyticus</i> subsp. <i>saprophyticus</i> ATCC 15305	155	156
<i>Bacillus selenitireducens</i> MLS10	157	158
<i>Streptococcus pneumoniae</i> SP6-BS73	159	160
<i>Streptococcus sanguinis</i> SK36	161	162
<i>Streptococcus thermophilus</i> LMG 18311	163	164
<i>Streptococcus suis</i> 89/1591	165	166
<i>Streptococcus mutans</i> UA159	167	168
<i>Leptospira borgpetersenii</i> serovar Hardjo-bovis L550	169	170
Candidatus <i>Vesicomycosocius okutanii</i> HA	171	172
Candidatus <i>Ruthia magnifica</i> str. Cm (<i>Calyptogena magnifica</i>)	173	174
<i>Methylococcus capsulatus</i> str. Bath	175	176
uncultured marine bacterium EB80_02D08	177	178
uncultured marine gamma proteobacterium EBAC31A08	179	180
uncultured marine gamma proteobacterium EBAC20E09	181	182
uncultured gamma proteobacterium eBACHOT4E07	183	184
<i>Alcanivorax borkumensis</i> SK2	185	186
<i>Chromohalobacter salexigens</i> DSM 3043	187	188
<i>Marinobacter algicola</i> DG893	189	190
<i>Marinobacter aquaeolei</i> VT8	191	192
<i>Marinobacter</i> sp. ELB17	193	194
<i>Pseudoalteromonas haloplanktis</i> TAC125	195	196
<i>Acinetobacter</i> sp. ADP1	197	198
<i>Opitutaceae bacterium</i> TAV2	199	200
<i>Flavobacterium</i> sp. MED217	201	202
<i>Cellulophaga</i> sp. MED134	203	204
<i>Kordia algicida</i> OT-1	205	206
<i>Flavobacteriales bacterium</i> ALC-1	207	208
<i>Psychroflexus torquis</i> ATCC 700755	209	210
<i>Flavobacteriales bacterium</i> HTCC2170	211	212
unidentified eubacterium SCB49	213	214
<i>Gramella forsetii</i> KT0803	215	216
<i>Robiginitalea biformata</i> HTCC2501	217	218
<i>Tenacibaculum</i> sp. MED152	219	220
<i>Polaribacter irgensii</i> 23-P	221	222
<i>Pedobacter</i> sp. BAL39	223	224
<i>Flavobacteria bacterium</i> BAL38	225	226
<i>Flavobacterium psychrophilum</i> JIP02/86	227	228
<i>Flavobacterium johnsoniae</i> UW101	229	230
<i>Lactococcus lactis</i> subsp. <i>cremoris</i> SK11	231	232
<i>Psychromonas ingrahamii</i> 37	233	234
<i>Microscilla marina</i> ATCC 23134	235	236
<i>Cytophaga hutchinsonii</i> ATCC 33406	237	238
<i>Rhodopirellula baltica</i> SH 1	239	240
<i>Blastopirellula marina</i> DSM 3645	241	242
<i>Planctomyces maris</i> DSM 8797	243	244
<i>Algoriphagus</i> sp. PR1	245	246
Candidatus <i>Sulcia muelleri</i> str. Hc (<i>Homalodisca coagulata</i>)	247	248
Candidatus <i>Carsonella ruddii</i> PV	249	250

TABLE 2a-continued

Representative bacterial [2Fe—2S] DHAD proteins and encoding sequences		
Organism of derivation	SEQ ID NO: Nucleic acid	SEQ ID NO: Peptide
<i>Synechococcus</i> sp. RS9916	251	252
<i>Synechococcus</i> sp. WH 7803	253	254
<i>Synechococcus</i> sp. CC9311	255	256
<i>Synechococcus</i> sp. CC9605	257	258
<i>Synechococcus</i> sp. WH 8102	259	260
<i>Synechococcus</i> sp. BL107	261	262
<i>Synechococcus</i> sp. RCC307	263	264
<i>Synechococcus</i> sp. RS9917	265	266
<i>Synechococcus</i> sp. WH 5701	267	268
<i>Prochlorococcus marinus</i> str. MIT 9313	269	270
<i>Prochlorococcus marinus</i> str. NATL2A	271	272
<i>Prochlorococcus marinus</i> str. MIT 9215	273	274
<i>Prochlorococcus marinus</i> str. AS9601	275	276
<i>Prochlorococcus marinus</i> str. MIT 9515	277	278
<i>Prochlorococcus marinus</i> subsp. <i>pastoris</i> str. CCMP1986	279	280
<i>Prochlorococcus marinus</i> str. MIT 9211	281	282
<i>Prochlorococcus marinus</i> subsp. <i>marinus</i> str. CCMP1375	283	284
<i>Nodularia spumigena</i> CCY9414	285	286
<i>Nostoc punctiforme</i> PCC 73102	287	288
<i>Nostoc</i> sp. PCC 7120	289	290
<i>Trichodesmium erythraeum</i> IMS101	291	292
<i>Acaryochloris marina</i> MBIC11017	293	294
<i>Lyngbya</i> sp. PCC 8106	295	296
<i>Synechocystis</i> sp. PCC 6803	297	298
<i>Microcystis aeruginosa</i> PCC 7806	426	427
<i>Cyanothece</i> sp. CCY0110	299	300
<i>Thermosynechococcus elongatus</i> BP-1	301	302
<i>Synechococcus</i> sp. JA-2-3B'a(2-13)	303	304
<i>Gloeobacter violaceus</i> PCC 7421	305	306
<i>Nitrosomonas eutropha</i> C91	307	308
<i>Nitrosomonas europaea</i> ATCC 19718	309	310
<i>Nitrospira multififormis</i> ATCC 25196	311	312
<i>Chloroflexus aggregans</i> DSM 9485	313	314
<i>Leptospirillum</i> sp. Group II UBA	315	316
<i>Leptospirillum</i> sp. Group II UBA	317	318
<i>Halorhodospira halophila</i> SL1	319	320
<i>Nitrococcus mobilis</i> Nb-231	321	322
<i>Alkalilimnicola ehrlichii</i> MLHE-1	323	324
<i>Deinococcus geothermalis</i> DSM 11300	325	326
<i>Polynucleobacter</i> sp. QLW-P1DMWA-1	327	328
<i>Polynucleobacter necessarius</i> STIR1	329	330
<i>Azoarcus</i> sp. EbN1	331	332
<i>Burkholderia phymatum</i> STM815	333	334
<i>Burkholderia xenovorans</i> LB400	335	336
<i>Burkholderia multivorans</i> ATCC 17616	337	338
<i>Burkholderia cenocepacia</i> PC184	339	340
<i>Burkholderia mallei</i> GB8 horse 4	341	342
<i>Ralstonia eutropha</i> JMP134	343	344
<i>Ralstonia metallidurans</i> CH34	345	346
<i>Ralstonia solanacearum</i> UW551	347	348
<i>Ralstonia pickettii</i> 12J	349	350
<i>Limnobacter</i> sp. MED105	351	352
<i>Herminiimonas arsenicoxydans</i>	353	354
<i>Bordetella parapertussis</i>	355	356
<i>Bordetella pertussis</i> DSM 12804	357	358
<i>Polaromonas</i> sp. JS666	359	360
<i>Polaromonas naphthalenivorans</i> CJ2	361	362
<i>Rhodoflexus ferrireducens</i> T118	363	364
<i>Verminephrobacter eiseniae</i> EF01-2	365	366
<i>Acidovorax</i> sp. JS42	367	368
<i>Delftia acidovorans</i> SPH-1	369	370
<i>Methylobium petroleiphilum</i> PM1	371	372
gamma proteobacterium KT 71	373	374
<i>Tremblaya princeps</i>	375	376
<i>Blastopirellula marina</i> DSM 3645	377	378
<i>Planctomyces maris</i> DSM 8797	379	380
<i>Salinibacter ruber</i> DSM 13855	387	388

TABLE 2b

Additional representative bacterial [2Fe—2S] DHAD proteins and encoding sequences		
Organism of derivation	Nucleic acid SEQ ID NO:	Amino acid SEQ ID NO:
<i>Burkholderia ambifaria</i> AMMD	443	444
<i>Bradyrhizobium</i> sp. BTAi1	445	446
<i>Delftia acidovorans</i> SPH-1	447	448
<i>Microcystis aeruginosa</i> NIES-843	449	450
uncultured marine microorganism HF4000_APKG8C21	451	452
<i>Burkholderia ubonensis</i> Bu	453	454
<i>Gemmata obscuriglobus</i> UQM 2246	455	456
<i>Mycobacterium abscessus</i>	457	458
<i>Synechococcus</i> sp. PCC 7002	459	460
<i>Burkholderia graminis</i> C4D1M	461	462
<i>Methylobacterium radiotolerans</i> JCM 2831	463	464
<i>Leptothrix cholodnii</i> SP-6	465	466
<i>Verrucomicrobium spinosum</i> DSM 4136	467	468
<i>Cyanothece</i> sp. ATCC 51142	469	470
<i>Opitutus terrae</i> PB90-1	471	472
<i>Leptospira biflexa</i> serovar Patoc strain 'Patoc 1 (Paris)'	473	474
<i>Methylacidiphilum inferorum</i> V4	475	476
<i>Cupriavidus taiwanensis</i>	477	478
<i>Chthoniobacter flavus</i> Ellin428	479	480
<i>Cyanothece</i> sp. PCC 7822	481	482
<i>Phenylobacterium zucineum</i> HLK1	483	484
<i>Leptospirillum</i> sp. Group II '5-way CG'	485	486
<i>Arthrospira maxima</i> CS-328	487	488
<i>Oligotropha carboxidovorans</i> OM5	489	490
<i>Rhodospirillum centenum</i> SW	491	492
<i>Cyanothece</i> sp. PCC 8801	493	494
<i>Thermus aquaticus</i> Y51MC23	495	496
<i>Cyanothece</i> sp. PCC 7424	497	498
<i>Acidithiobacillus ferrooxidans</i> ATCC 23270	499	500
<i>Cyanothece</i> sp. PCC 7425	501	502
<i>Arthrobacter chlorophenolicus</i> A6	503	504
<i>Burkholderia multivorans</i> CGD2M	505	506
<i>Thermomicrobium roseum</i> DSM 5159	507	508
bacterium Ellin514	509	510
<i>Desulfobacterium autotrophicum</i> HRM2	511	512
<i>Thioalkalivibrio</i> sp. K90mix	513	514
<i>Flavobacteria bacterium</i> MS024-3C	515	516
<i>Flavobacteria bacterium</i> MS024-2A	517	518
' <i>Nostoc azollae</i> ' 0708	519	520
<i>Acidobacterium capsulatum</i> ATCC 51196	521	522
<i>Gemmatimonas aurantiaca</i> T-27	523	524
<i>Gemmatimonas aurantiaca</i> T-27	525	526
<i>Rhodococcus erythropolis</i> PR4	527	528
<i>Deinococcus deserti</i> VCD115	529	530
<i>Rhodococcus opacus</i> B4	531	532
<i>Chryseobacterium gleum</i> ATCC 35910	533	534
<i>Thermobaculum terrenum</i> ATCC BAA-798	535	536
<i>Kribbella flavida</i> DSM 17836	537	538
<i>Gordonia bronchialis</i> DSM 43247	539	540
<i>Geodermatophilus obscurus</i> DSM 43160	541	542
<i>Xylanimonas cellulolytica</i> DSM 15894	543	544
<i>Sphingobacterium spiritivorum</i> ATCC 33300	545	546
<i>Meiothermus silvanus</i> DSM 9946	547	548
<i>Meiothermus ruber</i> DSM 1279	549	550
<i>Nakamurella multipartita</i> DSM 44233	551	552
<i>Cellulomonas flavigena</i> DSM 20109	553	554
<i>Rhodothermus marinus</i> DSM 4252	555	556
<i>Planctomyces limnophilus</i> DSM 3776	557	558
<i>Beutenbergia cavernae</i> DSM 12333	559	560
<i>Spirosoma linguale</i> DSM 74	561	562
<i>Sphaerobacter thermophilus</i> DSM 20745	563	564
<i>Lactococcus lactis</i>	565	566
<i>Thermus thermophilus</i> HB8	567	568
<i>Anabaena variabilis</i> ATCC 29413	569	570
<i>Roseovarius</i> sp. 217	571	572
uncultured <i>Prochlorococcus marinus</i> clone HF10-88D1	573	574
<i>Burkholderia xenovorans</i> LB400	575	576

TABLE 2b-continued

Additional representative bacterial [2Fe—2S] DHAD proteins and encoding sequences		
Organism of derivation	Nucleic acid SEQ ID NO:	Amino acid SEQ ID NO:
<i>Saccharomonospora viridis</i> DSM 43017	577	578
<i>Pedobacter heparinus</i> DSM 2366	579	580
<i>Microcoleus chthonoplastes</i> PCC 7420	581	582
<i>Acidimicrobium ferrooxidans</i> DSM 10331	583	584
<i>Rhodobacterales bacterium</i> HTCC2083	585	586
Candidatus <i>Pelagibacter</i> sp. HTCC7211	587	588
<i>Chitinophaga pinensis</i> DSM 2588	589	590
<i>Alcanivorax</i> sp. DG881	591	592
<i>Micrococcus luteus</i> NCTC 2665	593	594
<i>Verrucomicrobiae bacterium</i> DG1235	595	596
<i>Synechococcus</i> sp. PCC 7335	597	598
<i>Brevundimonas</i> sp. BAL3	599	600
<i>Dyadobacter fermentans</i> DSM 18053	601	602
gamma proteobacterium NOR5-3	603	604
gamma proteobacterium NOR51-B	605	606
<i>Cyanobium</i> sp. PCC 7001	607	608
<i>Jonesia denitrificans</i> DSM 20603	609	610
<i>Brachybacterium faecium</i> DSM 4810	611	612
<i>Paenibacillus</i> sp. JDR-2	613	614
<i>Octadecabacter antarcticus</i> 307	615	616
<i>Variovorax paradoxus</i> S110	617	618

TABLE 3

SEQ ID Numbers of Additional Proteins and Encoding sequences		
Description	SEQ ID NO: Encoding seq	SEQ ID NO: protein
<i>Vibrio cholerae</i> KARI	389	390
<i>Pseudomonas aeruginosa</i> PAO1 KARI	422	423
<i>Pseudomonas fluorescens</i> PF5 KARI	391	392
<i>Achromobacter xylosoxidans</i> butanol dehydrogenase sadB	393	394
<i>Escherichia coli</i> str. K-12 substr. MG1655 Phosphogluconate dehydratase	383	384
<i>Azospirillum brasilense</i> arabonate dehydratase	385	386
<i>Escherichia coli</i> str. K-12 substr. MG1655 DHAD	381	382

SEQ ID NOs: 395-409, 412-421, 424, and 431-436 are primers for PCR, cloning or sequencing analysis used as described in the Examples herein. SEQ ID NO: 410 is the nucleotide sequence of the pDM1 vector. SEQ ID NO: 411 is the nucleotide sequence of the pLH532 vector. SEQ ID NO: 425 is the *S. cerevisiae* FBA promoter. SEQ ID NO: 430 is the nucleotide sequence of the pRS423 FBA ilvD(Strep) vector. SEQ ID NO: 437 is the nucleotide sequence of the pNY13 vector. SEQ ID NO: 438 is the alsS coding region from *B. subtilis*. SEQ ID NO: 439 is the *S. cerevisiae* GPD promoter. SEQ ID NO: 440 is the *S. cerevisiae* CYC1 terminator. SEQ ID NO: 442 is the *S. cerevisiae* ILV5 gene.

DETAILED DESCRIPTION OF THE INVENTION

[0039] As disclosed herein, applicants have solved the stated problem through the discovery of methods of identifying [2Fe-2S] DHADs. Through the discovery these enzymes and their use in recombinant hosts, a heretofore unappreciated activity advantage in pathway engineering with DHADs has been identified.

[0040] The present invention relates to recombinant yeast or bacterial cells engineered to provide heterologous expression of dihydroxy-acid dehydratase (DHAD) having a [2Fe-2S] cluster. The expressed DHAD functions as a component of a biosynthetic pathway for production of a compound such as valine, isoleucine, leucine, pantothenic acid, or isobutanol. These amino acids and pantothenic acid may be used as nutritional supplements, and isobutanol may be used as a fuel additive to reduce demand for petrochemicals.

[0041] The following abbreviations and definitions will be used for the interpretation of the specification and the claims.

[0042] As used herein, the terms “comprises,” “comprising,” “includes,” “including,” “has,” “having,” “contains” or “containing,” or any other variation thereof, are intended to cover a non-exclusive inclusion. For example, a composition, a mixture, process, method, article, or apparatus that comprises a list of elements is not necessarily limited to only those elements but may include other elements not expressly listed or inherent to such composition, mixture, process, method, article, or apparatus. Further, unless expressly stated to the contrary, “or” refers to an inclusive or and not to an exclusive or. For example, a condition A or B is satisfied by any one of the following: A is true (or present) and B is false (or not present), A is false (or not present) and B is true (or present), and both A and B are true (or present).

[0043] Also, the indefinite articles “a” and “an” preceding an element or component of the invention are intended to be nonrestrictive regarding the number of instances (i.e. occurrences) of the element or component. Therefore “a” or “an” should be read to include one or at least one, and the singular word form of the element or component also includes the plural unless the number is obviously meant to be singular.

[0044] The term “invention” or “present invention” as used herein is a non-limiting term and is not intended to refer to any single embodiment of the particular invention but encompasses all possible embodiments as described in the specification and the claims.

[0045] As used herein, the term “about” modifying the quantity of an ingredient or reactant of the invention employed refers to variation in the numerical quantity that can occur, for example, through typical measuring and liquid handling procedures used for making concentrates or use solutions in the real world; through inadvertent error in these procedures; through differences in the manufacture, source, or purity of the ingredients employed to make the compositions or carry out the methods; and the like. The term “about” also encompasses amounts that differ due to different equilibrium conditions for a composition resulting from a particular initial mixture. Whether or not modified by the term “about”, the claims include equivalents to the quantities. In one embodiment, the term “about” means within 10% of the reported numerical value, preferably within 5% of the reported numerical value.

[0046] The term “[2Fe-2S] DHAD” refers to DHAD enzymes having a bound [2Fe-2S] cluster.

[0047] The term “[4Fe-4S] DHAD” refers to DHAD enzymes having a bound [4Fe-4S] cluster.

[0048] The term “Dihydroxy-acid dehydratase” will be abbreviated DHAD and will refer to an enzyme that converts 2,3-dihydroxyisovalerate to α -ketoisovalerate.

[0049] The term “isobutanol biosynthetic pathway” refers to an enzyme pathway to produce isobutanol from pyruvate.

[0050] The term “a facultative anaerobe” refers to a micro-organism that can grow in both aerobic and anaerobic environments.

[0051] The term “carbon substrate” or “fermentable carbon substrate” refers to a carbon source capable of being metabolized by host organisms of the present invention and particularly carbon sources selected from the group consisting of monosaccharides, oligosaccharides, polysaccharides, and one-carbon substrates or mixtures thereof. Carbon substrates may include C6 and C5 sugars and mixtures thereof.

[0052] The term “gene” refers to a nucleic acid fragment that is capable of being expressed as a specific protein, optionally including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. “Native gene” refers to a gene as found in nature with its own regulatory sequences. “Chimeric gene” refers to any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. “Endogenous gene” refers to a native gene in its natural location in the genome of an organism. A “foreign gene” or “heterologous gene” refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A “transgene” is a gene that has been introduced into the genome by a transformation procedure.

[0053] As used herein the term “coding region” refers to a DNA sequence that codes for a specific amino acid sequence. “Suitable regulatory sequences” refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, polyadenylation recognition sequences, RNA processing site, effector binding site and stem-loop structure.

[0054] The term “promoter” refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental or physiological conditions. Promoters which cause a gene to be expressed in most cell types at most times are commonly referred to as “constitutive promoters”. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

[0055] The term “operably linked” refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of effecting the expression of that coding sequence

(i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

[0056] The term “expression”, as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide.

[0057] As used herein the term “transformation” refers to the transfer of a nucleic acid fragment into a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as “transgenic” or “recombinant” or “transformed” organisms.

[0058] The terms “plasmid” and “vector” as used herein, refer to an extra chromosomal element often carrying genes which are not part of the central metabolism of the cell, and usually in the form of circular double-stranded DNA molecules. Such elements may be autonomously replicating sequences, genome integrating sequences, phage or nucleotide sequences, linear or circular, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of introducing a promoter fragment and DNA sequence for a selected gene product along with appropriate 3' untranslated sequence into a cell.

[0059] As used herein the term “codon degeneracy” refers to the nature in the genetic code permitting variation of the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. The skilled artisan is well aware of the “codon-bias” exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a gene for improved expression in a host cell, it is desirable to design the gene such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

[0060] The term “codon-optimized” as it refers to genes or coding regions of nucleic acid molecules for transformation of various hosts, refers to the alteration of codons in the gene or coding regions of the nucleic acid molecules to reflect the typical codon usage of the host organism without altering the polypeptide encoded by the DNA.

[0061] As used herein, an “isolated nucleic acid fragment” or “isolated nucleic acid molecule” will be used interchangeably and will mean a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid fragment in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA.

[0062] A nucleic acid fragment is “hybridizable” to another nucleic acid fragment, such as a cDNA, genomic DNA, or RNA molecule, when a single-stranded form of the nucleic acid fragment can anneal to the other nucleic acid fragment under the appropriate conditions of temperature and solution ionic strength. Hybridization and washing conditions are well known and exemplified in Sambrook, J., Fritsch, E. F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory: Cold Spring Harbor, N.Y. (1989), particularly Chapter 11 and Table 11.1 therein (entirely incorporated herein by reference). The conditions of temperature and ionic strength determine the “stringency” of

the hybridization. Stringency conditions can be adjusted to screen for moderately similar fragments (such as homologous sequences from distantly related organisms), to highly similar fragments (such as genes that duplicate functional enzymes from closely related organisms). Post-hybridization washes determine stringency conditions. One set of preferred conditions uses a series of washes starting with 6×SSC, 0.5% SDS at room temperature for 15 min, then repeated with 2×SSC, 0.5% SDS at 45° C. for 30 min, and then repeated twice with 0.2×SSC, 0.5% SDS at 50° C. for 30 min. A more preferred set of stringent conditions uses higher temperatures in which the washes are identical to those above except for the temperature of the final two 30 min washes in 0.2×SSC, 0.5% SDS was increased to 60° C. Another preferred set of highly stringent conditions uses two final washes in 0.1×SSC, 0.1% SDS at 65° C. An additional set of stringent conditions include hybridization at 0.1×SSC, 0.1% SDS, 65° C. and washes with 2×SSC, 0.1% SDS followed by 0.1×SSC, 0.1% SDS, for example.

[0063] Hybridization requires that the two nucleic acids contain complementary sequences, although depending on the stringency of the hybridization, mismatches between bases are possible. The appropriate stringency for hybridizing nucleic acids depends on the length of the nucleic acids and the degree of complementation, variables well known in the art. The greater the degree of similarity or homology between two nucleotide sequences, the greater the value of T_m for hybrids of nucleic acids having those sequences. The relative stability (corresponding to higher T_m) of nucleic acid hybridizations decreases in the following order: RNA:RNA, DNA:RNA, DNA:DNA. For hybrids of greater than 100 nucleotides in length, equations for calculating T_m have been derived (see Sambrook et al., supra, 9.50-9.51). For hybridizations with shorter nucleic acids, i.e., oligonucleotides, the position of mismatches becomes more important, and the length of the oligonucleotide determines its specificity (see Sambrook et al., supra, 11.7-11.8). In one embodiment the length for a hybridizable nucleic acid is at least about 10 nucleotides. Preferably a minimum length for a hybridizable nucleic acid is at least about 15 nucleotides; more preferably at least about 20 nucleotides; and most preferably the length is at least about 30 nucleotides. Furthermore, the skilled artisan will recognize that the temperature and wash solution salt concentration may be adjusted as necessary according to factors such as length of the probe.

[0064] A “substantial portion” of an amino acid or nucleotide sequence is that portion comprising enough of the amino acid sequence of a polypeptide or the nucleotide sequence of a gene to putatively identify that polypeptide or gene, either by manual evaluation of the sequence by one skilled in the art, or by computer-automated sequence comparison and identification using algorithms such as BLAST (Altschul, S. F., et al., *J. Mol. Biol.*, 215:403-410 (1993)). In general, a sequence of ten or more contiguous amino acids or thirty or more nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene specific oligonucleotide probes comprising 20-30 contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., in situ hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12-15 bases may be used as amplification primers in PCR in order to obtain a particular

nucleic acid fragment comprising the primers. Accordingly, a “substantial portion” of a nucleotide sequence comprises enough of the sequence to specifically identify and/or isolate a nucleic acid fragment comprising the sequence. The instant specification teaches the complete amino acid and nucleotide sequence encoding particular proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

[0065] The term “complementary” is used to describe the relationship between nucleotide bases that are capable of hybridizing to one another. For example, with respect to DNA, adenosine is complementary to thymine and cytosine is complementary to guanine.

[0066] The term “percent identity”, as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, “identity” or “sequence identity” also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. “Identity” and “similarity” can be readily calculated by known methods, including but not limited to those described in:

[0067] 1.) *Computational Molecular Biology* (Lesk, A. M., Ed.) Oxford University: NY (1988); 2.) *Biocomputing: Informatics and Genome Projects* (Smith, D. W., Ed.) Academic: NY (1993); 3.) *Computer Analysis of Sequence Data, Part I* (Griffin, A. M., and Griffin, H. G., Eds.) Humana: NJ (1994); 4.) *Sequence Analysis in Molecular Biology* (von Heinje, G., Ed.) Academic (1987); and 5.) *Sequence Analysis Primer* (Gribskov, M. and Devereux, J., Eds.) Stockton: NY (1991).

[0068] Preferred methods to determine identity are designed to give the best match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Sequence alignments and percent identity calculations may be performed using the MegAlign™ program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, Wis.). Multiple alignment of the sequences is performed using the “Clustal method of alignment” which encompasses several varieties of the algorithm including the “Clustal V method of alignment” corresponding to the alignment method labeled Clustal V (described by Higgins and Sharp, *CABIOS*. 5:151-153 (1989); Higgins, D. G. et al., *Comput. Appl. Biosci.*, 8:189-191 (1992)) and found in the MegAlign™ program of the LASERGENE bioinformatics computing suite (DNASTAR Inc.). For multiple alignments, the default values correspond to GAP PENALTY=10 and GAP LENGTH PENALTY=10. Default parameters for pairwise alignments and calculation of percent identity of protein sequences using the Clustal method are KTUPLE=1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. For nucleic acids these parameters are KTUPLE=2, GAP PENALTY=5, WINDOW=4 and DIAGONALS SAVED=4. After alignment of the sequences using the Clustal V program, it is possible to obtain a “percent identity” by viewing the “sequence distances” table in the same program. Additionally the “Clustal W method of alignment” is available and corresponds to the alignment method labeled Clustal W (described by Higgins and Sharp,

CABIOS. 5:151-153 (1989); Higgins, D. G. et al., *Comput. Appl. Biosci.* 8:189-191 (1992) Thompson, J. D., Higgins, D. G., and Gibson T. J. (1994) *Nuc. Acid Res.* 22: 4673-4680) and found in the MegAlign™ v6.1 program of the LASERGENE bioinformatics computing suite (DNASTAR Inc.). Default parameters for multiple alignment (GAP PENALTY=10, GAP LENGTH PENALTY=0.2, Delay Divergen Seqs (%)=30, DNA Transition Weight=0.5, Protein Weight Matrix=Gonnet Series, DNA Weight Matrix=IUB). After alignment of the sequences using the Clustal W program, it is possible to obtain a “percent identity” by viewing the “sequence distances” table in the same program.

[0069] It is well understood by one skilled in the art that many levels of sequence identity are useful in identifying polypeptides, from other species, wherein such polypeptides have the same or similar function or activity. Useful examples of percent identities include, but are not limited to: 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%, or any integer percentage from 55% to 100% may be useful in describing the present invention, such as 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99%. Suitable nucleic acid fragments not only have the above homologies but typically encode a polypeptide having at least 50 amino acids, preferably at least 100 amino acids, more preferably at least 150 amino acids, still more preferably at least 200 amino acids, and most preferably at least 250 amino acids.

[0070] The term “sequence analysis software” refers to any computer algorithm or software program that is useful for the analysis of nucleotide or amino acid sequences. “Sequence analysis software” may be commercially available or independently developed. Typical sequence analysis software will include, but is not limited to: 1.) the GCG suite of programs (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, Wis.); 2.) BLASTP, BLASTN, BLASTX (Altschul et al., *J. Mol. Biol.*, 215:403-410 (1990)); 3.) DNASTAR (DNASTAR, Inc. Madison, Wis.); 4.) Sequencher (Gene Codes Corporation, Ann Arbor, Mich.); and 5.) the FASTA program incorporating the Smith-Waterman algorithm (W. R. Pearson, *Comput. Methods Genome Res.*, [Proc. Int. Symp.] (1994), Meeting Date 1992, 111-20. Editor(s): Suhai, Sandor. Plenum: New York, N.Y.). Within the context of this application it will be understood that where sequence analysis software is used for analysis, that the results of the analysis will be based on the “default values” of the program referenced, unless otherwise specified. As used herein “default values” will mean any set of values or parameters that originally load with the software when first initialized.

[0071] Standard recombinant DNA and molecular cloning techniques used here are well known in the art and are described by Sambrook, J., Fritsch, E. F. and Maniatis, T., *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989) (hereinafter “Maniatis”); and by Silhavy, T. J., Bannan, M. L. and Enquist, L. W., *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1984); and by Ausubel, F. M. et al., *Current Protocols in Molecular Biology*, published by Greene Publishing Assoc. and Wiley-Interscience (1987). Additional methods used here are in *Methods in Enzymology*, Volume

194, *Guide to Yeast Genetics and Molecular and Cell Biology* (Part A, 2004, Christine Guthrie and Gerald R. Fink (Eds.), Elsevier Academic Press, San Diego, Calif.).

Discovery of [2Fe-2S] DHADs

[0072] DHAD proteins are known to contain a bound iron-sulfur (Fe—S) cluster that is required for enzyme activity. The only DHAD with a [2Fe-2S] cluster reported to date is a spinach enzyme (Flint and Emptage (1988) *J. Biol. Chem.* 263:3558-3564). Some bacterial enzymes are also known, the best characterized of which is from *E. coli* (Flint, D H, et al. (1993) *J. Biol. Chem.* 268:14732-14742), which has a [4Fe-4S] cluster.

[0073] Applicants have now determined that there is a class of bacterial DHADs that contain a [2Fe-2S] cluster ([2Fe-2S] DHADs). Applicants have found that the group of [2Fe-2S] DHADs may be distinguished from a group of [4Fe-4S] DHADs by the presence of three conserved cysteine residues in the protein. The three conserved cysteines are similar to three essential cysteines reported in the *Azospirillum brasiliense* arabonate (an aldonic acid) dehydratase (Watanabe, S et al. *J. Biol. Chem.* (2006) 281:33521-33536) that was reported to contain a [4Fe-4S] cluster. In the *Azospirillum brasiliense* arabonate dehydratase protein, cysteines located at amino acid positions 56, 124, and 197 were determined to be essential for enzyme activity and likely involved in coordination with the Fe—S cluster. Surprisingly, applicants have found that three conserved cysteines, which are in corresponding positions to the three essential cysteines of the *Azospirillum brasiliense* arabonate dehydratase, are characteristic of DHADs containing a [2Fe-2S] cluster. Applicants have found that the *E. coli* DHAD, that contains a [4Fe-4S] cluster, has two of the conserved cysteines, but not the third conserved cysteine. Shown in FIG. 1 is a comparison of the sequence regions of the conserved cysteines for the *E. coli* [4Fe-4S] cluster-containing DHADs and for representatives of a phylogenetic group of [2Fe-2S] cluster DHADs that was identified herein in Example 1 and is described below.

[0074] Applicants have developed a method for identifying [2Fe-2S] DHADs. In the present invention, bacterial [2Fe-2S] DHADs, which may be identified by this method, may be used for heterologous expression in bacteria.

[0075] To structurally characterize DHAD enzymes a Profile Hidden Markov Model (HMM) was prepared as described in Example 1 using amino acid sequences of DHAD proteins with experimentally verified function, as determined in Example 2 herein, and is given in Table 1. These DHADs are from *Nitrosomonas europaea* (DNA SEQ ID NO:309; Protein SEQ ID NO:310), *Synechocystis* sp. PCC6803 (DNA SEQ ID:297; Protein SEQ ID NO:298), *Streptococcus mutans* (DNA SEQ ID NO:167; Protein SEQ ID NO:168), *Streptococcus thermophilus* (DNA SEQ ID NO:163; SEQ ID No:164), *Ralstonia metallidurans* (DNA SEQ ID NO:345; Protein SEQ ID NO:346), *Ralstonia eutropha* (DNA SEQ ID NO:343; Protein SEQ ID NO:344), and *Lactococcus lactis* (DNA SEQ ID NO:231; Protein SEQ ID NO:232). In addition the DHAD from *Flavobacterium johnsoniae* (DNA SEQ ID NO:229; Protein SEQ ID NO:230) was found to have dihydroxy-acid dehydratase activity when expressed in *E. coli* and was used in making the Profile. This Profile HMM for DHADs may be used to identify DHAD related proteins. Any protein that matches the Profile HMM with an E value of $<10^{-5}$ is a DHAD related protein, which includes [4Fe-4S] DHADs, [2Fe-2S] DHADs, aldonic acid

dehydratases, and phosphogluconate dehydratases. A phylogenetic tree of sequences matching this Profile HMM is shown in FIG. 2.

[0076] Sequences matching the Profile HMM given herein are then analyzed for the presence of the three conserved cysteines described above. The exact positions of the three conserved cysteines may vary, and these may be identified in the context of the surrounding sequence using multiple sequence alignments performed with the Clustal W algorithm (Thompson, J. D., Higgins, D. G., and Gibson T. J. (1994) *Nuc. Acid Res.* 22: 4673 4680) employing the following parameters: 1) for pairwise alignment parameters, a Gap opening=10; Gap extend=0.1; matrix is Gonnet 250; and mode—Slow-accurate, 2) for multiple alignment parameters, Gap opening=10; Gap extension=0.2; and matrix is Gonnet series. For example, the three conserved cysteines are located at amino acid positions 56, 129, and 201 in the *Streptococcus mutans* DHAD (SEQ ID NO: 168), and at amino acid positions 61, 135, and 207 in the *Lactococcus lactis* DHAD (SEQ ID NO: 232). The exact positions of the three conserved cysteines in other protein sequences correspond to these positions in the *S. mutans* or the *L. lactis* amino acid sequence. One skilled in the art will readily be able to identify the presence or absence of each of the three conserved cysteines in the amino acid sequence of a DHAD protein using pairwise or multiple sequence alignments. In addition, other methods may be used to determine the presence of the three conserved cysteines, such as analysis by eye.

[0077] The DHAD Profile HMM matching proteins that have two but not the third (position 56) conserved cysteine include [4Fe-4S] DHADs and phosphogluconate dehydratases (EDDs). Proteins having the three conserved cysteines include arabonate dehydratases and [2Fe-2S] DHADs, and are members of a [2Fe-2S] DHAD/aldonic acid dehydratase group. The [2Fe-2S] DHADs may be distinguished from the aldonic acid dehydratases by analyzing for signature conserved amino acids found to be present in the [2Fe-2S] DHADs or in the aldonic acid dehydratases at positions corresponding to the following positions in the *Streptococcus mutans* DHAD amino acid sequence. These signature amino acids are in [2Fe-2S] DHADs or in aldonic acid dehydratases, respectively, at the following positions (with greater than 90% occurrence): 88 asparagine vs glutamic acid; 113 not conserved vs glutamic acid; 142 arginine or asparagine vs not conserved; 165: not conserved vs glycine; 208 asparagine vs not conserved; 454 leucine vs not conserved; 477 phenylalanine or tyrosine vs not conserved; and 487 glycine vs not conserved.

[0078] The disclosed methods for identification of [2Fe-2S] DHAD enzymes can be carried out on a single sequence or on a group of sequences. In a preferred embodiment, one or more sequence databases are queried with a Profile HMM as described herein. Suitable sequence databases are known to those skilled in the art and include but are not limited to the Genbank non-redundant protein database, the SwissProt database, or UniProt database or other available databases such as GQPat (GenomeQuest, Westboro, Mass.) and BRENDA (Biobase, Beverly, Mass.).

[0079] Among the [2Fe-2S] DHADs identifiable by this method, bacterial [2Fe-2S] DHADs may readily be identifiable by the natural source organism being a type of bacteria. Any bacterial [2Fe-2S] DHAD identifiable by this method may be suitable for heterologous expression in a microbial host cell. It will also be appreciated that any bacterial [2Fe-

2S] DHAD expressly disclosed herein by sequence may be suitable for heterologous expression in bacterial cells. Preferred bacterial [2Fe-2S] DHAD enzymes can be expressed in a host cell and provide DHAD activity.

[0080] Initially, 193 different bacterial [2Fe-2S] DHADs with sequence identities of less than 95% (greater than 95% identity proteins removed to simplify the analysis) were identified as described in Example 1 and the amino acid and coding sequences of these proteins are provided in the sequence listing, with SEQ ID NOs listed in Table 2a.

[0081] A subsequent analysis described in Example 11 returned 268 different bacterial [2Fe-2S] DHADs. Amino acid and coding sequences that were not identical to any of the 193 bacterial [2Fe-2S] DHADs provided by the initial identification are included in the sequence listing, with SEQ ID NOs listed in Table 2b.

[0082] Any [2Fe-2S] DHAD protein matching a sequence identifiable through the methods disclosed herein or a sequence expressly disclosed herein with an identity of at least about 95%, 96%, 97%, 98%, or 99% is a [2Fe-2S] DHAD that may be used for heterologous expression in bacterial cells as disclosed herein. Among the bacterial [2Fe-2S] DHADs expressly disclosed herein, there is 100% conservation of the signature amino acids at positions: 88 aspartic acid, 142 arginine or asparagine, 208 asparagine, and 454 leucine.

[0083] In addition, bacterial [2Fe-2S] DHADs that may be used in the present invention are identifiable by their position in the [2Fe-2S] DHAD branch of a phylogenetic tree of DHAD related proteins such as that shown in FIG. 2 and described in Example 1. In addition, bacterial [2Fe-2S] DHADs that may be used are identifiable using sequence comparisons with any of the 281 bacterial [2Fe-2S] DHADs whose sequences are provided herein, where sequence identity may be at least about 80%-85%, 85%-90%, 90%-95% or 95%-99%.

[0084] Additionally, the sequences of [2Fe-2S] DHADs provided herein may be used to identify other homologs in nature. For example each of the DHAD encoding nucleic acid fragments described herein may be used to isolate genes encoding homologous proteins. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to: 1.) methods of nucleic acid hybridization; 2.) methods of DNA and RNA amplification, as exemplified by various uses of nucleic acid amplification technologies [e.g., polymerase chain reaction (PCR), Mullis et al., U.S. Pat. No. 4,683,202; ligase chain reaction (LCR), Tabor, S. et al., *Proc. Acad. Sci. USA* 82:1074 (1985); or strand displacement amplification (SDA), Walker, et al., *Proc. Natl. Acad. Sci. U.S.A.*, 89:392 (1992)]; and 3.) methods of library construction and screening by complementation.

[0085] For example, genes encoding similar proteins or polypeptides to the [2Fe-2S] DHAD encoding genes provided herein could be isolated directly by using all or a portion of the instant nucleic acid fragments as DNA hybridization probes to screen libraries from any desired organism using methodology well known to those skilled in the art. Specific oligonucleotide probes based upon the disclosed nucleic acid sequences can be designed and synthesized by methods known in the art (Maniatis, supra). Moreover, the entire sequences can be used directly to synthesize DNA probes by methods known to the skilled artisan (e.g., random primers DNA labeling, nick translation or end-labeling techniques), or RNA probes using available in vitro transcription

systems. In addition, specific primers can be designed and used to amplify a part of (or full-length of) the instant sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full-length DNA fragments by hybridization under conditions of appropriate stringency. Typically, in PCR-type amplification techniques, the primers have different sequences and are not complementary to each other. Depending on the desired test conditions, the sequences of the primers should be designed to provide for both efficient and faithful replication of the target nucleic acid. Methods of PCR primer design are common and well known in the art (Thein and Wallace, "The use of oligonucleotides as specific hybridization probes in the Diagnosis of Genetic Disorders", in *Human Genetic Diseases: A Practical Approach*, K. E. Davis Ed., (1986) pp 33-50, IRL: Herndon, Va.; and Rychlik, W., In *Methods in Molecular Biology*, White, B. A. Ed., (1993) Vol. 15, pp 31-39, PCR Protocols: Current Methods and Applications. Humana: Totowa, N.J.).

[0086] Generally two short segments of the described sequences may be used in polymerase chain reaction protocols to amplify longer nucleic acid fragments encoding homologous genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from the described nucleic acid fragments, and the sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding microbial genes.

[0087] Alternatively, the second primer sequence may be based upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman et al., *PNAS USA* 85:8998 (1988)) to generate cDNAs by using PCR to amplify copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the instant sequences. Using commercially available 3' RACE or 5' RACE systems (e.g., BRL, Gaithersburg, Md.), specific 3' or 5' cDNA fragments can be isolated (Ohara et al., *PNAS USA* 86:5673 (1989); Loh et al., *Science* 243:217 (1989)).

[0088] Alternatively, the provided [2Fe-2S] DHAD encoding sequences may be employed as hybridization reagents for the identification of homologs. The basic components of a nucleic acid hybridization test include a probe, a sample suspected of containing the gene or gene fragment of interest, and a specific hybridization method. Probes are typically single-stranded nucleic acid sequences that are complementary to the nucleic acid sequences to be detected. Probes are "hybridizable" to the nucleic acid sequence to be detected. The probe length can vary from 5 bases to tens of thousands of bases, and will depend upon the specific test to be done. Typically a probe length of about 15 bases to about 30 bases is suitable. Only part of the probe molecule need be complementary to the nucleic acid sequence to be detected. In addition, the complementarity between the probe and the target sequence need not be perfect. Hybridization does occur between imperfectly complementary molecules with the result that a certain fraction of the bases in the hybridized region are not paired with the proper complementary base.

[0089] Hybridization methods are well defined. Typically the probe and sample must be mixed under conditions that will permit nucleic acid hybridization. This involves contacting the probe and sample in the presence of an inorganic or

organic salt under the proper concentration and temperature conditions. The probe and sample nucleic acids must be in contact for a long enough time that any possible hybridization between the probe and sample nucleic acid may occur. The concentration of probe or target in the mixture will determine the time necessary for hybridization to occur. The higher the probe or target concentration, the shorter the hybridization incubation time needed. Optionally, a chaotropic agent may be added. The chaotropic agent stabilizes nucleic acids by inhibiting nuclease activity. Furthermore, the chaotropic agent allows sensitive and stringent hybridization of short oligonucleotide probes at room temperature (Van Ness and Chen, *Nucl. Acids Res.* 19:5143-5151 (1991)). Suitable chaotropic agents include guanidinium chloride, guanidinium thiocyanate, sodium thiocyanate, lithium tetrachloroacetate, sodium perchlorate, rubidium tetrachloroacetate, potassium iodide and cesium trifluoroacetate, among others. Typically, the chaotropic agent will be present at a final concentration of about 3 M. If desired, one can add formamide to the hybridization mixture, typically 30-50% (v/v).

[0090] Various hybridization solutions can be employed. Typically, these comprise from about 20 to 60% volume, preferably 30%, of a polar organic solvent. A common hybridization solution employs about 30-50% v/v formamide, about 0.15 to 1 M sodium chloride, about 0.05 to 0.1 M buffers (e.g., sodium citrate, Tris-HCl, PIPES or HEPES (pH range about 6-9)), about 0.05 to 0.2% detergent (e.g., sodium dodecylsulfate), or between 0.5-20 mM EDTA, FICOLL (Pharmacia Inc.) (about 300-500 kdal), polyvinylpyrrolidone (about 250-500 kdal) and serum albumin. Also included in the typical hybridization solution will be unlabeled carrier nucleic acids from about 0.1 to 5 mg/mL, fragmented nucleic DNA (e.g., calf thymus or salmon sperm DNA, or yeast RNA), and optionally from about 0.5 to 2% wt/vol glycine. Other additives may also be included, such as volume exclusion agents that include a variety of polar water-soluble or swellable agents (e.g., polyethylene glycol), anionic polymers (e.g., polyacrylate or polymethylacrylate) and anionic saccharidic polymers (e.g., dextran sulfate).

[0091] Nucleic acid hybridization is adaptable to a variety of assay formats. One of the most suitable is the sandwich assay format. The sandwich assay is particularly adaptable to hybridization under non-denaturing conditions. A primary component of a sandwich-type assay is a solid support. The solid support has adsorbed to it or covalently coupled to it immobilized nucleic acid probe that is unlabeled and complementary to one portion of the sequence.

Expression of Heterologous Bacterial [2Fe-2S] DHADs in Bacterial and Yeast Hosts

[0092] Applicants have found that a heterologous [2Fe-2S] DHAD provides DHAD activity when expressed in a microbial cell. Any [2Fe-2S] DHAD which may be identified as described herein, may be expressed in a heterologous microbial cell. Expression of any of these proteins provides DHAD activity for a biosynthetic pathway that includes conversion of 2,3-dihydroxyisovalerate to α -ketoisovalerate or 2,3-dihydroxymethylvalerate to α -ketomethylvalerate. Expression of a [2Fe-2S] DHAD, as opposed to a 4Fe-4D DHAD, lowers the requirement for Fe and S in clusters to obtain enzyme activity. In addition, the *S. mutans* [2Fe-2S] DHAD was shown herein to have higher stability in air as compared to the

sensitivity in air of the *E. coli* [4Fe-4S] DHAD, which is desirable for obtaining better activity in a heterologous host cell.

[0093] Bacterial cells that may be hosts for expression of a heterologous bacterial [2Fe-2S] DHAD include, but are not limited to, *Clostridium*, *Zymomonas*, *Escherichia*, *Salmonella*, *Rhodococcus*, *Pseudomonas*, *Bacillus*, *Lactobacillus*, *Enterococcus*, *Pediococcus*, *Alcaligenes*, *Klebsiella*, *Paenibacillus*, *Arthrobacter*, *Corynebacterium*, and *Brevibacterium*, *Lactococcus*, *Leuconostoc*, *Oenococcus*, *Pediococcus*, and *Streptococcus*. Engineering expression of a heterologous bacterial [2Fe-2S] DHAD may increase DHAD activity in a host bacterial cell that naturally expresses a [2Fe-2S] DHAD or a [4Fe-4S] DHAD. Such host cells may include, for example, *E. coli* and *Bacillus subtilis*. Furthermore, engineering expression of a heterologous bacterial [2Fe-2S] DHAD provides DHAD activity in a host bacterial cell that has no endogenous DHAD activity. Such host cells may include, for example, *Lactobacillus*, *Enterococcus*, *Pediococcus* and *Leuconostoc*.

[0094] Specific hosts include: *Escherichia coli*, *Alcaligenes eutrophus*, *Bacillus licheniformis*, *Paenibacillus macerans*, *Rhodococcus erythropolis*, *Pseudomonas putida*, *Lactobacillus plantarum*, *Enterococcus faecium*, *Enterococcus gallinarum*, *Enterococcus faecalis*, and *Bacillus subtilis*.

[0095] A host bacterial cell may be engineered to express a heterologous bacterial [2Fe-2S] DHAD by methods well known to one skilled in the art. The coding region for the DHAD to be expressed may be codon optimized for the target host cell, as well known to one skilled in the art. Vectors useful for the transformation of a variety of host cells are common and commercially available from companies such as EPI-CENTRE® (Madison, Wis.), Invitrogen Corp. (Carlsbad, Calif.), Stratagene (La Jolla, Calif.), and New England Biolabs, Inc. (Beverly, Mass.). Typically the vector contains a selectable marker and sequences allowing autonomous replication or chromosomal integration in the desired host. In addition, suitable vectors comprise a promoter region which harbors transcriptional initiation controls and a transcriptional termination control region, between which a coding region DNA fragment may be inserted, to provide expression of the inserted coding region. Both control regions may be derived from genes homologous to the transformed host cell, although it is to be understood that such control regions may also be derived from genes that are not native to the specific species chosen as a production host.

[0096] Initiation control regions or promoters, which are useful to drive expression of bacterial [2Fe-2S] DHAD coding regions in the desired bacterial host cell are numerous and familiar to those skilled in the art. Virtually any promoter capable of driving these genetic elements is suitable for the present invention including, but not limited to, lac, ara, tet, trp, IP_L , IP_R , T7, tac, and trc promoters (useful for expression in *Escherichia coli*, *Alcaligenes*, and *Pseudomonas*); the amy, apr, and npr promoters, and various phage promoters useful for expression in *Bacillus subtilis*, *Bacillus licheniformis*, and *Paenibacillus macerans*; nisA (useful for expression Gram-positive bacteria, Eichenbaum et al. *Appl. Environ. Microbiol.* 64(8):2763-2769 (1998)); and the synthetic P11 promoter (useful for expression in *Lactobacillus plantarum*, Rud et al., *Microbiology* 152:1011-1019 (2006)).

[0097] Termination control regions may also be derived from various genes native to the preferred hosts. Optionally, a termination site may be unnecessary; however, it is most preferred if included.

[0098] Certain vectors are capable of replicating in a broad range of host bacteria and can be transferred by conjugation. The complete and annotated sequence of pRK404 and three related vectors: pRK437, pRK442, and pRK442(H), are available. These derivatives have proven to be valuable tools for genetic manipulation in Gram-negative bacteria (Scott et al., *Plasmid* 50(1):74-79 (2003)). Several plasmid derivatives of broad-host-range Inc P4 plasmid RSF1010 are also available with promoters that can function in a range of Gram-negative bacteria. Plasmid pAYC36 and pAYC37, have active promoters along with multiple cloning sites to allow for heterologous gene expression in Gram-negative bacteria. Some vectors that are useful for transformation of *Bacillus subtilis* and *Lactobacillus* include pAM β 1 and derivatives thereof (Renault et al., *Gene* 183:175-182 (1996); and O'Sullivan et al., *Gene* 137:227-231 (1993)); pMBB1 and pHW800, a derivative of pMBB1 (Wyckoff et al. *Appl. Environ. Microbiol.* 62:1481-1486 (1996)); pMG1, a conjugative plasmid (Tanimoto et al., *J. Bacteriol.* 184:5800-5804 (2002)); pNZ9520 (Kleerebezem et al., *Appl. Environ. Microbiol.* 63:4581-4584 (1997)); pAM401 (Fujimoto et al., *Appl. Environ. Microbiol.* 67:1262-1267 (2001)); and pAT392 (Arthur et al., *Antimicrob. Agents Chemother.* 38:1899-1903 (1994)). Several plasmids from *Lactobacillus plantarum* have also been reported (van Kranenburg et al., *Appl. Environ. Microbiol.* 71(3):1223-1230 (2005)).

[0099] Chromosomal gene replacement tools are also widely available. For example, a thermosensitive variant of the broad-host-range replicon pWV101 has been modified to construct a plasmid pVE6002 which can be used to effect gene replacement in a range of Gram-positive bacteria (Maguin et al., *J. Bacteriol.* 174(17):5633-5638 (1992)). Additionally, in vitro transposomes are available from commercial sources such as EPICENTRE® to create random mutations in a variety of genomes.

[0100] Yeast cells that may be hosts for expression of a heterologous bacterial [2Fe-2S] DHAD are any yeast cells that are amenable to genetic manipulation and include, but are not limited to, *Saccharomyces*, *Schizosaccharomyces*, *Hansenula*, *Candida*, *Kluyveromyces*, *Yarrowia* and *Pichia*. Suitable strains include, but are not limited to, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Kluyveromyces lactis*, *Kluyveromyces thermotolerans*, *Candida glabrata*, *Candida albicans*, *Pichia stipitis* and *Yarrowia lipolytica*. Most suitable is *Saccharomyces cerevisiae*.

[0101] Expression is achieved by transforming with a gene comprising a sequence encoding any of these [2Fe-2S] DHADs. The coding region for the DHAD to be expressed may be codon optimized for the target host cell, as well known to one skilled in the art. Methods for gene expression in yeast are known in the art (see for example *Methods in Enzymology*, Volume 194, *Guide to Yeast Genetics and Molecular and Cell Biology* (Part A, 2004, Christine Guthrie and Gerald R. Fink (Eds.), Elsevier Academic Press, San Diego, Calif.). Expression of genes in yeast typically requires a promoter, operably linked to a coding region of interest, and a transcriptional terminator. A number of yeast promoters can be used in constructing expression cassettes for genes in yeast, including, but not limited to promoters derived from the following genes: CYC1, HIS3, GAL1, GAL10, ADH1,

PGK, PHO5, GAPDH, ADC1, TRP1, URA3, LEU2, ENO, TPI, CUP1, FBA, GPD, GPM, and AOX1. Suitable transcriptional terminators include, but are not limited to FBAt, GPDt, GPMt, ERG10t, GAL1t, CYC1, and ADH1.

[0102] Suitable promoters, transcriptional terminators, and [2Fe-2S] DHAD coding regions may be cloned into *E. coli*-yeast shuttle vectors, and transformed into yeast cells. These vectors allow strain propagation in both *E. coli* and yeast strains. Typically the vector used contains a selectable marker and sequences allowing autonomous replication or chromosomal integration in the desired host. Typically used plasmids in yeast are shuttle vectors pRS423, pRS424, pRS425, and pRS426 (American Type Culture Collection, Rockville, Md.), which contain an *E. coli* replication origin (e.g., pMB1), a yeast 2 μ origin of replication, and a marker for nutritional selection. The selection markers for these four vectors are His3 (vector pRS423), Trp1 (vector pRS424), Leu2 (vector pRS425) and Ura3 (vector pRS426). Construction of expression vectors with a chimeric gene encoding the described DHADs may be performed by either standard molecular cloning techniques in *E. coli* or by the gap repair recombination method in yeast.

[0103] The gap repair cloning approach takes advantage of the highly efficient homologous recombination in yeast. Typically, a yeast vector DNA is digested (e.g., in its multiple cloning site) to create a "gap" in its sequence. A number of insert DNAs of interest are generated that contain a 21 by sequence at both the 5' and the 3' ends that sequentially overlap with each other, and with the 5' and 3' terminus of the vector DNA. For example, to construct a yeast expression vector for "Gene X", a yeast promoter and a yeast terminator are selected for the expression cassette. The promoter and terminator are amplified from the yeast genomic DNA, and Gene X is either PCR amplified from its source organism or obtained from a cloning vector comprising Gene X sequence. There is at least a 21 by overlapping sequence between the 5' end of the linearized vector and the promoter sequence, between the promoter and Gene X, between Gene X and the terminator sequence, and between the terminator and the 3' end of the linearized vector. The "gapped" vector and the insert DNAs are then co-transformed into a yeast strain and plated on the medium containing the appropriate compound mixtures that allow complementation of the nutritional selection markers on the plasmids. The presence of correct insert combinations can be confirmed by PCR mapping using plasmid DNA prepared from the selected cells. The plasmid DNA isolated from yeast (usually low in concentration) can then be transformed into an *E. coli* strain, e.g. TOP10, followed by mini preps and restriction mapping to further verify the plasmid construct. Finally the construct can be verified by sequence analysis.

[0104] Like the gap repair technique, integration into the yeast genome also takes advantage of the homologous recombination system in yeast. Typically, a cassette containing a coding region plus control elements (promoter and terminator) and auxotrophic marker is PCR-amplified with a high-fidelity DNA polymerase using primers that hybridize to the cassette and contain 40-70 base pairs of sequence homology to the regions 5' and 3' of the genomic area where insertion is desired. The PCR product is then transformed into yeast and plated on medium containing the appropriate compound mixtures that allow selection for the integrated auxotrophic marker. For example, to integrate "Gene X" into chromosomal location "Y", the promoter-coding regionX-terminator

construct is PCR amplified from a plasmid DNA construct and joined to an autotrophic marker (such as URA3) by either SOE PCR or by common restriction digests and cloning. The full cassette, containing the promoter-coding region-terminator-URA3 region, is PCR amplified with primer sequences that contain 40-70 bp of homology to the regions 5' and 3' of location "Y" on the yeast chromosome. The PCR product is transformed into yeast and selected on growth media lacking uracil. Transformants can be verified either by colony PCR or by direct sequencing of chromosomal DNA.

Confirming DHAD Activity

[0105] The presence of DHAD activity in a cell engineered to express a bacterial [2Fe-2S] DHAD can be confirmed using methods known in the art. As one example, and as demonstrated in the Examples herein, crude extracts from cells engineered to express a bacterial [2Fe-2S] DHAD may be used in a DHAD assay as described by Flint and Emptage (J. Biol. Chem. (1988) 263(8): 3558-64) using dinitrophenylhydrazine. In another example, and as demonstrated in the Examples herein, DHAD activity may be assayed by expressing a bacterial DHAD identifiable by the methods disclosed herein in a yeast strain that lacks endogenous DHAD activity. If DHAD activity is present, the yeast strain will grow in the absence of branched-chain amino acids. DHAD activity may also be confirmed by more indirect methods, such as by assaying for a downstream product in a pathway requiring DHAD activity. Any product that has α -ketoisovalerate or α -ketomethylvalerate as a pathway intermediate may be measured as an assay for DHAD activity. A list of such products includes, but is not limited to, valine, isoleucine, leucine, pantothenic acid, 2-methyl-1-butanol, 3-methyl-1-butanol and isobutanol.

Isobutanol Production

[0106] Expression of a bacterial [2Fe-2S] DHAD in bacteria or yeast, as described herein, provides the transformed, recombinant host cell with dihydroxy-acid dehydratase activity for conversion of 2,3-dihydroxyisovalerate to α -ketoisovalerate or 2,3-dihydroxymethylvalerate to α -ketomethylvalerate. Any product that has α -ketoisovalerate or α -ketomethylvalerate as a pathway intermediate may be produced with greater effectiveness in a bacterial or yeast strain disclosed herein having the described heterologous [2Fe-2S] DHAD. A list of such products includes, but is not limited to, valine, isoleucine, leucine, pantothenic acid, 2-methyl-1-butanol, 3-methyl-1-butanol and isobutanol.

[0107] For example, in yeast biosynthesis of valine includes steps of acetolactate conversion to 2,3-dihydroxyisovalerate by acetohydroxyacid reductoisomerase (ILV5), conversion of 2,3-dihydroxyisovalerate to α -ketoisovalerate (also called 2-ketoisovalerate) by dihydroxy-acid dehydratase, and conversion of α -ketoisovalerate to valine by branched-chain amino acid transaminase (BAT2) and branched-chain amino acid aminotransferase (BAT1). Biosynthesis of leucine includes the same steps to α -ketoisovalerate, followed by conversion of α -ketoisovalerate to alpha-isopropylmalate by alpha-isopropylmalate synthase (LEU9, LEU4), conversion of alpha-isopropylmalate to beta-isopropylmalate by isopropylmalate isomerase (LEU1), conversion of beta-isopropylmalate to alpha-ketoisocaproate by beta-IPM dehydrogenase (LEU2), and finally conversion of alpha-ketoisocaproate to leucine by branched-chain amino acid

transaminase (BAT2) and branched-chain amino acid aminotransferase (BAT1). The bacterial pathway is similar, involving differently named proteins and genes. Increased conversion of 2,3-dihydroxyisovalerate to α -ketoisovalerate will increase flow in these pathways, particularly if one or more additional enzymes of a pathway is overexpressed. Thus it is desired for production of valine or leucine to use a strain disclosed herein.

[0108] Biosynthesis of pantothenic acid includes a step performed by DHAD, as well as steps performed by ketopantoate hydroxymethyltransferase and pantothenate synthase. Engineering of expression of these enzymes for enhanced production of pantothenic acid biosynthesis in microorganisms is described in U.S. Pat. No. 6,177,264.

[0109] The α -ketoisovalerate product of DHAD is an intermediate in isobutanol biosynthetic pathways disclosed in commonly owned and co-pending US Patent Publication 20070092957 A1, which is herein incorporated by reference. A diagram of the disclosed isobutanol biosynthetic pathways is provided in FIG. 3. Production of isobutanol in a strain disclosed herein benefits from increased DHAD activity. As disclosed herein, DHAD activity is provided by expression of a bacterial [2Fe-2S] DHAD in a bacterial or yeast cell. As described in US 20070092957 A1, steps in an example isobutanol biosynthetic pathway include conversion of:

[0110] pyruvate to acetolactate (see FIG. 1, pathway step a therein), as catalyzed for example by acetolactate synthase,

[0111] acetolactate to 2,3-dihydroxyisovalerate (see FIG. 1, pathway step b therein) as catalyzed for example by acetohydroxy acid isomerase;

[0112] 2,3-dihydroxyisovalerate to α -ketoisovalerate (see FIG. 1, pathway step c therein) as catalyzed for example by acetohydroxy acid dehydratase, also called dihydroxy-acid dehydratase (DHAD);

[0113] α -ketoisovalerate to isobutyraldehyde (see FIG. 1, pathway step d therein) as catalyzed for example by branched-chain α -keto acid decarboxylase; and

[0114] isobutyraldehyde to isobutanol (see FIG. 1, pathway step e therein) as catalyzed for example by branched-chain alcohol dehydrogenase.

[0115] The substrate to product conversions, and enzymes involved in these reactions, for steps f, g, h, i, j, and k of alternative pathways are described in US 20070092957 A1.

[0116] Genes that may be used for expression of the pathway step enzymes named above other than the bacterial [2Fe-2S] DHADs disclosed herein, as well as those for two additional isobutanol pathways, are described in US 20070092957 A1, and additional genes that may be used can be identified by one skilled in the art through bioinformatics or experimentally as described above. The preferred use in all three pathways of ketol-acid reductoisomerase (KARI) enzymes with particularly high activities is disclosed in commonly owned and co-pending US Patent Application Publication No. US20080261230A1. Examples of high activity KARIs disclosed therein are those from *Vibrio cholerae* (DNA: SEQ ID NO:389; protein SEQ ID NO:390), *Pseudomonas aeruginosa* PAO1, (DNA: SEQ ID NO: 422; protein SEQ ID NO:423), and *Pseudomonas fluorescens* PF5 (DNA: SEQ ID NO:391; protein SEQ ID NO:392).

[0117] Additionally described in US 20070092957 A1 are construction of chimeric genes and genetic engineering of bacteria and yeast for isobutanol production using the disclosed biosynthetic pathways.

Growth for Production

[0118] Recombinant bacteria or yeast hosts disclosed herein are grown in fermentation media which contains suitable carbon substrates. Additional carbon substrates may include but are not limited to monosaccharides such as fructose, oligosaccharides such as lactose maltose, galactose, or sucrose, polysaccharides such as starch or cellulose or mixtures thereof and unpurified mixtures from renewable feedstocks such as cheese whey permeate, cornsteep liquor, sugar beet molasses, and barley malt. Other carbon substrates may include ethanol, lactate, succinate, or glycerol.

[0119] Additionally the carbon substrate may also be one-carbon substrates such as carbon dioxide, or methanol for which metabolic conversion into key biochemical intermediates has been demonstrated. In addition to one and two carbon substrates, methylotrophic organisms are also known to utilize a number of other carbon containing compounds such as methylamine, glucosamine and a variety of amino acids for metabolic activity. For example, methylotrophic yeasts are known to utilize the carbon from methylamine to form trehalose or glycerol (Bellion et al., *Microb. Growth C1 Compd.*, [Int. Symp.], 7th (1993), 415-32, Editor(s): Murrell, J. Collin; Kelly, Don P. Publisher: Intercept, Andover, UK). Similarly, various species of *Candida* will metabolize alanine or oleic acid (Sulter et al., *Arch. Microbiol.* 153:485-489 (1990)). Hence it is contemplated that the source of carbon utilized in the present invention may encompass a wide variety of carbon containing substrates and will only be limited by the choice of organism.

[0120] Although it is contemplated that all of the above mentioned carbon substrates and mixtures thereof are suitable in the present invention, preferred carbon substrates are glucose, fructose, and sucrose, or mixtures of these with C5 sugars such as xylose and/or arabinose for yeasts cells modified to use C5 sugars. Sucrose may be derived from renewable sugar sources such as sugar cane, sugar beets, cassava, sweet sorghum, and mixtures thereof. Glucose and dextrose may be derived from renewable grain sources through saccharification of starch based feedstocks including grains such as corn, wheat, rye, barley, oats, and mixtures thereof. In addition, fermentable sugars may be derived from renewable cellulosic or lignocellulosic biomass through processes of pretreatment and saccharification, as described, for example, in co-owned and co-pending U.S. Patent Application Publication No. 2007/0031918A1, which is herein incorporated by reference. Biomass refers to any cellulosic or lignocellulosic material and includes materials comprising cellulose, and optionally further comprising hemicellulose, lignin, starch, oligosaccharides and/or monosaccharides. Biomass may also comprise additional components, such as protein and/or lipid. Biomass may be derived from a single source, or biomass can comprise a mixture derived from more than one source; for example, biomass may comprise a mixture of corn cobs and corn stover, or a mixture of grass and leaves. Biomass includes, but is not limited to, bioenergy crops, agricultural residues, municipal solid waste, industrial solid waste, sludge from paper manufacture, yard waste, wood and forestry waste. Examples of biomass include, but are not limited to, corn grain, corn cobs, crop residues such as corn husks, corn stover, grasses, wheat, wheat straw, barley, barley straw, hay, rice straw, switchgrass, waste paper, sugar cane bagasse, sorghum, soy, components obtained from milling of grains,

trees, branches, roots, leaves, wood chips, sawdust, shrubs and bushes, vegetables, fruits, flowers, animal manure, and mixtures thereof.

[0121] In addition to an appropriate carbon source, fermentation media must contain suitable minerals, salts, cofactors, buffers and other components, known to those skilled in the art, suitable for the growth of the cultures and promotion of an enzymatic pathway comprising a bacterial [2Fe-2S] DHAD.

Culture Conditions

[0122] Typically cells are grown at a temperature in the range of about 20° C. to about 40° C. in an appropriate medium. Suitable growth media in the present invention are common commercially prepared media such as Luria Bertani (LB) broth, Sabouraud Dextrose (SD) broth, Yeast Medium (YM) broth, or broth that includes yeast nitrogen base, ammonium sulfate, and dextrose (as the carbon/energy source) or YPD Medium, a blend of peptone, yeast extract, and dextrose in optimal proportions for growing most *Saccharomyces cerevisiae* strains. Other defined or synthetic growth media may also be used, and the appropriate medium for growth of the particular microorganism will be known by one skilled in the art of microbiology or fermentation science. The use of agents known to modulate catabolite repression directly or indirectly, e.g., cyclic adenosine 2':3'-monophosphate, may also be incorporated into the fermentation medium.

[0123] Suitable pH ranges for the fermentation of yeast are typically between pH 3.0 to pH 9.0. where pH 5.0 to pH 8.0 is preferred as the initial condition. Suitable pH ranges for the fermentation of other microorganisms are between pH 3.0 to pH 7.5, where pH 4.5.0 to pH 6.5 is preferred as the initial condition.

[0124] Fermentations may be performed under aerobic or anaerobic conditions, where anaerobic or microaerobic conditions are preferred.

Industrial Batch and Continuous Fermentations

[0125] Fermentation may be a batch method of fermentation. A classical batch fermentation is a closed system where the composition of the medium is set at the beginning of the fermentation and not subject to artificial alterations during the fermentation. Thus, at the beginning of the fermentation the medium is inoculated with the desired organism or organisms, and fermentation is permitted to occur without adding anything to the system. Typically, however, "batch" fermentation is batch with respect to the addition of carbon source and attempts are often made at controlling factors such as pH and oxygen concentration. In batch systems the metabolite and biomass compositions of the system change constantly up to the time the fermentation is stopped. Within batch cultures cells moderate through a static lag phase to a high growth log phase and finally to a stationary phase where growth rate is diminished or halted. If untreated, cells in the stationary phase will eventually die. Cells in log phase generally are responsible for the bulk of production of end product or intermediate.

[0126] A variation on the standard batch system is the fed-batch system. Fed-batch fermentation processes are also suitable in the present invention and comprise a typical batch system with the exception that the substrate is added in increments as the fermentation progresses. Fed-batch systems are useful when catabolite repression is apt to inhibit the metabo-

lism of the cells and where it is desirable to have limited amounts of substrate in the media. Measurement of the actual substrate concentration in fed-batch systems is difficult and is therefore estimated on the basis of the changes of measurable factors such as pH, dissolved oxygen and the partial pressure of waste gases such as CO₂. Batch and fed-batch fermentations are common and well known in the art and examples may be found in Thomas D. Brock in *Biotechnology: A Textbook of Industrial Microbiology*, Second Edition (1989) Sinauer Associates, Inc., Sunderland, Mass., or Deshpande, Mukund V., *Appl. Biochem. Biotechnol.*, 36:227, (1992), herein incorporated by reference.

[0127] The fermentation culture may be adapted to continuous fermentation methods. Continuous fermentation is an open system where a defined fermentation medium is added continuously to a bioreactor and an equal amount of conditioned media is removed simultaneously for processing. Continuous fermentation generally maintains the cultures at a constant high density.

[0128] Continuous fermentation allows for the modulation of one factor or any number of factors that affect cell growth or end product concentration. For example, one method will maintain a limiting nutrient such as the carbon source or nitrogen level at a fixed rate and allow all other parameters to moderate. In other systems a number of factors affecting growth can be altered continuously while the cell concentration, measured by the turbidity of the culture medium, is kept constant. Continuous systems strive to maintain steady state growth conditions and thus the cell loss due to the medium being drawn off must be balanced against the cell growth rate in the fermentation. Methods of modulating nutrients and growth factors for continuous fermentation processes as well as techniques for maximizing the rate of product formation are well known in the art of industrial microbiology and a variety of methods are detailed by Brock, supra.

[0129] It is contemplated that batch, fed-batch, continuous processes, or any known mode of fermentation is suitable for growth of the described recombinant microbial host cell. Additionally, it is contemplated that cells may be immobilized on a substrate as whole cell catalysts and subjected to fermentation conditions for isobutanol production.

Methods for Product Isolation from the Fermentation Medium

[0130] Bioproduced isobutanol may be isolated from the fermentation medium using methods known in the art such as for ABE fermentations (see for example, Durre, *Appl. Microbiol. Biotechnol.* 49:639-648 (1998), Groot et al., *Process Biochem.* 27:61-75 (1992), and references therein). For example, solids may be removed from the fermentation medium by centrifugation, filtration, decantation, or the like. Then, the isobutanol may be isolated from the fermentation medium using methods such as distillation, azeotropic distillation, liquid-liquid extraction, adsorption, gas stripping, membrane evaporation, or pervaporation.

[0131] Because isobutanol forms a low boiling point, azeotropic mixture with water, distillation can be used to separate the mixture up to its azeotropic composition. Distillation may be used in combination with another separation method to obtain separation around the azeotrope. Methods that may be used in combination with distillation to isolate and purify butanol include, but are not limited to, decantation, liquid-liquid extraction, adsorption, and membrane-based techniques. Additionally, butanol may be isolated using azeotropic distillation using an entrainer (see for example Doherty and Malone, *Conceptual Design of Distillation Systems*, McGraw Hill, New York, 2001).

[0132] The butanol-water mixture forms a heterogeneous azeotrope so that distillation may be used in combination with decantation to isolate and purify the isobutanol. In this method, the isobutanol containing fermentation broth is distilled to near the azeotropic composition. Then, the azeotropic mixture is condensed, and the isobutanol is separated from the fermentation medium by decantation. The decanted aqueous phase may be returned to the first distillation column as reflux. The isobutanol-rich decanted organic phase may be further purified by distillation in a second distillation column.

[0133] The isobutanol may also be isolated from the fermentation medium using liquid-liquid extraction in combination with distillation. In this method, the isobutanol is extracted from the fermentation broth using liquid-liquid extraction with a suitable solvent. The isobutanol-containing organic phase is then distilled to separate the butanol from the solvent.

[0134] Distillation in combination with adsorption may also be used to isolate isobutanol from the fermentation medium. In this method, the fermentation broth containing the isobutanol is distilled to near the azeotropic composition and then the remaining water is removed by use of an adsorbent, such as molecular sieves (Aden et al. *Lignocellulosic Biomass to Ethanol Process Design and Economics Utilizing Co-Current Dilute Acid Prehydrolysis and Enzymatic Hydrolysis for Corn Stover*, Report NREL/TP-510-32438, National Renewable Energy Laboratory, June 2002).

[0135] Additionally, distillation in combination with pervaporation may be used to isolate and purify the isobutanol from the fermentation medium. In this method, the fermentation broth containing the isobutanol is distilled to near the azeotropic composition, and then the remaining water is removed by pervaporation through a hydrophilic membrane (Guo et al., *J. Membr. Sci.* 245, 199-210 (2004)).

EXAMPLES

[0136] The present invention is further defined in the following Examples. It should be understood that these Examples, while indicating a preferred embodiment of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention and, without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various uses and conditions.

General Methods

[0137] Standard recombinant DNA and molecular cloning techniques described in the Examples are well known in the art and are described by Sambrook, J., Fritsch, E. F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y., (1989) (Maniatis) and by T. J. Silhavy, M. L. Bannan, and L. W. Enquist, *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1984) and by Ausubel, F. M. et al., *Current Protocols in Molecular Biology*, pub. by Greene Publishing Assoc. and Wiley-Interscience (1987) and by *Methods in Yeast Genetics*, 2005, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

[0138] Materials and methods suitable for the maintenance and growth of bacterial cultures are well known in the art. Techniques suitable for use in the following Examples may be found as set out in *Manual of Methods for General Bacteri-*

ology (Phillipp Gerhardt, R. G. E. Murray, Ralph N. Costilow, Eugene W. Nester, Willis A. Wood, Noel R. Krieg and G. Briggs Phillips, eds), American Society for Microbiology, Washington, D.C. (1994)) or by Thomas D. Brock in *Biotechnology: A Textbook of Industrial Microbiology*, Second Edition, Sinauer Associates, Inc., Sunderland, Mass. (1989). All reagents, restriction enzymes and materials used for the growth and maintenance of bacterial cells were obtained from Aldrich Chemicals (Milwaukee, Wis.), BD Diagnostic Systems (Sparks, Md.), Life Technologies (Rockville, Md.), or Sigma Chemical Company (St. Louis, Mo.) unless otherwise specified.

[0139] Microbial strains were obtained from The American Type Culture Collection (ATCC), Manassas, Va., unless otherwise noted. The oligonucleotide primers used in the following Examples were synthesized by Sigma-Genosys (Woodlands, Tex.) or Integrated DNA Technologies (Coralsville, Iowa).

[0140] Synthetic complete medium is described in Amberg, Burke and Strathern, 2005, *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

HPLC

[0141] Analysis for fermentation by-product composition is well known to those skilled in the art. For example, one high performance liquid chromatography (HPLC) method utilizes a Shodex SH-1011 column with a Shodex SH-G guard column (both available from Waters Corporation, Milford, Mass.), with refractive index (RI) detection. Chromatographic separation is achieved using 0.01 M H₂SO₄ as the mobile phase with a flow rate of 0.5 mL/min and a column temperature of 50° C. Isobutanol retention time is about 47.6 minutes.

[0142] The meaning of abbreviations is as follows: “s” means second(s), “min” means minute(s), “h” means hour(s), “psi” means pounds per square inch, “nm” means nanometers, “d” means day(s), “μL” means microliter(s), “mL” means milliliter(s), “L” means liter(s), “mm” means millimeter(s), “nm” means nanometers, “mM” means millimolar, “μM” means micromolar, “M” means molar, “mmol” means millimole(s), “μmol” means micromole(s), “g” means gram(s), “μg” means microgram(s) and “ng” means nanogram(s), “PCR” means polymerase chain reaction, “OD” means optical density, “OD600” means the optical density measured at a wavelength of 600 nm, “kDa” means kilodaltons, “g” means the gravitation constant, “bp” means base pair(s), “kbp” means kilobase pair(s), “~” means about, “% w/v” means weight/volume percent, “% v/v” means volume/volume percent, “HPLC” means high performance liquid chromatography, and “GC” means gas chromatography.

Example 1

Identification of Bacterial Dihydroxy-Acid Dehydratases with [2Fe-2S] Cluster

Phylogenetic Analysis

[0143] Phylogenetic relationships were determined for dihydroxy-acid dehydratases (DHADs) and related proteins. Related proteins were identified through BlastP searches of publicly available databases using amino acid sequences of *E. coli* DHAD (SEQ ID NO:382), *E. coli* phosphogluconate dehydratase (EDD; SEQ ID NO:384; coding region SEQ ID NO:383), and *Azospirillum brasiliense* arabinonate dehy-

dratase (SEQ ID NO:386; coding region SEQ ID NO:385), with the following search parameters: E value=10, word size=3, Matrix=Blosum62, and Gap opening=11 and gap extension=1. Blast searches employing the three different protein sequences generated overlapping sets of sequence matches. Sequences were selected from the search results based on E value cutoff of 10⁻⁵ with removal of 95% identity sequences. Sequences that were shorter than 350 amino acids and sequences that were longer than 650 amino acids were also removed. The resultant set of 976 amino acid sequences included dihydroxy-acid dehydratases, phosphogluconate dehydratases, and aldonic acid dehydratases.

[0144] A profile HMM was generated from the experimentally verified DHADs described in Example 2. See details below on building, calibrating, and searching with this profile HMM. An hmmer search, using this profile HMM as a query, against the 976 sequences matched all sequences with an E value of <10⁻⁵. Multiple sequence alignments of the amino acid sequences were performed with the Clustal W algorithm (Thompson, J. D., Higgins, D. G., and Gibson T. J. (1994) *Nuc. Acid Res.* 22: 4673-4680) employing the following parameters: 1) for pairwise alignment parameters, a Gap opening=10; Gap extend=0.1; matrix is Gonnet 250; and mode—Slow-accurate, 2) for multiple alignment parameters, Gap opening=10; Gap extension=0.2; and matrix is Gonnet series. Phylogenetic trees were generated from sequence alignments based on the Neighbor Joining method. A tree representing phylogenetic relationships among the 976 sequences is shown in FIG. 2. Four main main branches emerged from this analysis. They are labeled “4Fe-4S DHAD”, “2Fe-2S DHAD”, “aldonic acid dehydratase”, and “EDD”, based on the criteria detailed below. A fifth small branch of 17 sequences is marked as “Und” for undefined.

[0145] The aligned sequences were initially analyzed for the presence of three cysteines determined to be essential for enzyme activity and likely involved in Fe—S coordination in the *Azospirillum brasiliense* arabinonate dehydratase, which was reported as a [4Fe-4S] cluster protein (Watanabe, S et al. *J. Biol. Chem.* (2006) 281:33521-33536). Each of the 976 sequences has the cysteines corresponding to two of the *Azospirillum brasiliense* arabinonate dehydratase essential cysteines (at positions 124 and 197). Within the phylogenetic tree, there is a branch of 168 sequences that includes the [4Fe-4S] phosphogluconate dehydratase of *Zymomonas mobilis* (Rodriguez, M. et. al. (1996) *Biochem Mol Biol Int.* 38:783-789). Only the four amino acids alanine, valine or serine or glycine, but not cysteine, were found at the position of the third essential cysteine of the *A. brasiliense* arabinonate dehydratase (position 56). This 168 sequence branch is labeled in FIG. 2 as “EDD”.

[0146] A different branch of the tree contains 322 sequences, among which is the known [4Fe-4S] cluster DHAD of *E. coli*. This 322 sequence branch is labeled in FIG. 2 as “4Fe-4S DHAD”. All sequences within this branch, and in the branch of 17 sequences (“Und”), contain glycine at the position corresponding to the third cysteine. The remaining 469 sequences, which are clustered in two branches and comprise both the aldonic acid recognizing *A. brasiliense* arabinonate dehydratase and a set of DHADs, possess all three cysteines. These cysteines are at positions 56, 129, and 201 in the *S. mutans* DHAD (SEQ ID NO: 168) and at positions 61, 135, and 207 in the *L. lactis* DHAD (SEQ ID NO: 232). Shown in FIG. 1 are examples of regions from multiple sequence alignments that include the conserved cysteines.

[0147] Further analysis of multiple sequence alignments and of phylogenetic trees was performed to identify DHAD-specific residues (signatures) to distinguish DHADs from the arabonate dehydratases and other aldonic acid dehydratases. Among sequences containing the three specified conserved cysteines, a clade of 274 sequences was found to contain the DHADs from *S. mutans* and *L. lactis*. The *A. brasiliense* arabonate dehydratase was found in a separate clade of 195 sequences. Multiple sequence alignments containing the sequences from the DHAD group of 274, the aldonic acid dehydratase group, and the “[4Fe-4S] DHAD” branch were analyzed for conserved residues at each position. A set of residues that are conserved in a majority of both DHAD groups, but not in the aldonic acid dehydratases group, was detected and is shown in Table 4. Additionally, residues that are conserved in the aldonic acid dehydratases, but not in either of the two DHAD groups were also found. Such differentially conserved residues may act as substrate specificity determinants in their respective enzymes.

TABLE 4

Position*	Conserved residues* discriminating DHADs from aldonic acid dehydratases		
	[4Fe-4S] DHAD	DHAD- 274 group	Aldonic acid Dehydratase
88	Asp	Asp	Glu
113	NC**	NC	Glu
142	Arg	Arg or Asn	NC
165	NC	NC	Gly
208	Asn	Asn	NC
454	Leu	Leu	NC
477	Phe	Phe or Tyr	NC
487	Gly	Gly	NC

Residue(s) conserved in a >90% majority of representatives.

*Position numbering is based on the position in the *S. mutans* DHAD

**Not Conserved

[0148] The group of DHADs forming the 274-sequence clade that does not include [4Fe-4S] cluster *E. coli* DHAD, [4Fe-4S] cluster *Z. mobilis* phosphogluconate dehydratase, or the reportedly [4Fe-4S] cluster *A. brasiliense* arabonate dehydratase, was differentially identified from the other groups by phylogeny and conserved residues found in multiple sequence alignments as described above. Consistent with the proposal that the group includes [2Fe-2S] cluster DHADs, the *Arabidopsis thaliana* DHAD and the *S. solfataricus* DHAD are part of the group. Because *Arabidopsis thaliana* is a plant as is spinach, and the spinach DHAD has been identified as a [2Fe-2S] cluster DHAD (Flint and Emptage (1988) *J. Biol. Chem.* 263:3558-3564), the *Arabidopsis thaliana* DHAD may be a [2Fe-2S] cluster DHAD. The *S. solfataricus* DHAD is reported to be oxygen resistant like the spinach [2Fe-2S] cluster DHAD, (Kim and Lee (2006) *J. Biochem.* 139, 591-596) which is an indication that the *S. solfataricus* DHAD may be a [2Fe-2S] cluster DHAD.

[0149] The 274 sequence clade is labeled on FIG. 4 as “2Fe-2S DHAD”. 193 of these sequences are from bacterial sources. The three conserved cysteines and the conserved residues specified in Table 4 can be identified in multiple sequence alignments of the 193 bacterial DHADs, employing the alignment procedure described above. The sequences of the 193 bacterial [2Fe-2S] DHADs are provided in the sequence listing and the SEQ ID NOs are listed in Table 2a.

[0150] Within the [2Fe-2S] DHAD group, the identities of several bacterial proteins have been confirmed by functional analysis as DHADs, which is described in other examples herein. Other examples herein show that proteins in this group, such as the *S. mutans* DHAD and the *L. lactis* DHAD, contain a [2Fe-2S] cluster.

Preparation of Profile Hmm

[0151] Seven bacterial DHADs that were identified as members of the [2Fe-2S] phylogenetic group were expressed in *E. coli* and dihydroxy-acid dehydratase activity was found as described in Example 2 below. These DHADs are from *Nitrosomonas europaea* (SEQ ID NO:310), *Synechocystis* sp. PCC6803 (SEQ ID NO: 298), *Streptococcus mutans* (SEQ ID NO:168), *Streptococcus thermophilus* (SEQ ID NO:164), *Ralstonia metallidurans* (SEQ ID NO:346), *Ralstonia eutropha* (SEQ ID NO:344), and *Lactococcus lactis* (SEQ ID NO:232). In addition the DHAD from *Flavobacterium johnsoniae* (SEQ ID NO:230) was found to have dihydroxy-acid dehydratase activity when expressed in *E. coli*. The amino acid sequences of these experimentally determined functional bacterial DHADs were analyzed using the HMMER software package (The theory behind profile HMMs is described in R. Durbin, S. Eddy, A. Krogh, and G. Mitchison, *Biological sequence analysis: probabilistic models of proteins and nucleic acids*, Cambridge University Press, 1998; Krogh et al., 1994; *J. Mol. Biol.* 235:1501-1531), following the user guide which is available from HMMER (Janelia Farm Research Campus, Ashburn, Va.). The output of the HMMER software program is a Profile Hidden Markov Model (HMM) that characterizes the input sequences. As stated in the user guide, Profile HMMs are statistical models of multiple sequence alignments. They capture position-specific information about how conserved each column of the alignment is, and which amino acid residues are most likely to occur at each position. Thus HMMs have a formal probabilistic basis. Profile HMMs for a large number of protein families are publicly available in the PFAM database (Janelia Farm Research Campus, Ashburn, Va.).

[0152] The Profile HMM was built as follows:

Step 1. Build a Sequence Alignment

[0153] The eight sequences for the functionally verified DHADs listed above were aligned using Clustal W with default parameters.

Step 2. Build a Profile HMM

[0154] The hmmbuild program was run on the set of aligned sequences using default parameters. hmmbuild reads the multiple sequence alignment file, builds a new Profile HMM, and saves the Profile HMM to file. Using this program an un-calibrated profile was generated from the multiple alignment for each set of subunit sequences described above.

[0155] The following information based on the HMMER software user guide gives some description of the way that the hmmbuild program prepares a Profile HMM. A Profile HMM is capable of modeling gapped alignments, e.g. including insertions and deletions, which lets the software describe a complete conserved domain (rather than just a small ungapped motif). Insertions and deletions are modeled using insertion (I) states and deletion (D) states. All columns that contain more than a certain fraction x of gap characters will be assigned as an insert column. By default, x is set to 0.5. Each match state has an I and a D state associated with it. HMMER calls a group of three states (M/D/I) at the same consensus

position in the alignment a “node”. These states are interconnected with arrows called state transition probabilities. M and I states are emitters, while D states are silent. The transitions are arranged so that at each node, either the M state is used (and a residue is aligned and scored) or the D state is used (and no residue is aligned, resulting in a deletion-gap character, ‘-’). Insertions occur between nodes, and I states have a self-transition, allowing one or more inserted residues to occur between consensus columns.

[0156] The scores of residues in a match state (i.e. match state emission scores), or in an insert state (i.e. insert state emission scores) are proportional to $\text{Log}_2(p_x)/(\text{null}_x)$. Where p_x is the probability of an amino acid residue, at a particular position in the alignment, according to the Profile HMM and null_x is the probability according to the Null model. The Null model is a simple one state probabilistic model with pre-calculated set of emission probabilities for each of the 20 amino acids derived from the distribution of amino acids in the SWISSPROT release 24.

[0157] State transition scores are also calculated as log odds parameters and are proportional to $\text{Log}_2(t_x)$. Where t_x is the probability of transiting to an emitter or non-emitter state.

Step 3. Calibrate the Profile HMM

[0158] The Profile HMM was read using `hmmcalibrate` which scores a large number of synthesized random sequences with the Profile (the default number of synthetic sequences used is 5,000), fits an extreme value distribution (EVD) to the histogram of those scores, and re-saves the HMM file now including the EVD parameters. These EVD parameters (μ and λ) are used to calculate the E-values of bit scores when the profile is searched against a protein sequence database. `hmmcalibrate` writes two parameters into the HMM file on a line labeled “EVD”: these parameters are the μ (location) and λ (scale) parameters of an extreme value distribution (EVD) that best fits a histogram of scores calculated on randomly generated sequences of about the same length and residue composition as SWISS-PROT. This calibration was done once for the Profile HMM.

[0159] The calibrated Profile HMM for the DHAD set of sequences is provided in Table 1. The Profile HMM is provided in a chart that gives the probability of each amino acid occurring at each position in the amino acid sequence. The highest probability is highlighted for each position. The first line for each position reports the match emission scores: probability for each amino acid to be in that state (highest score is highlighted). The second line reports the insert emission scores, and the third line reports on state transition scores: M→M, M→I, M→D; I→M, I→I; D→M, D→D; B→M; M→E.

[0160] For example, the DHAD Profile HMM shows that methionine has a 1757 probability of being in the first position, the highest probability which is highlighted. In the second position glutamic acid has the highest probability, which is 1356. In the third position lysine has the highest probability, which is 1569.

Step 4. Test the Specificity and Sensitivity of the Built Profile HMMs

[0161] The Profile HMM was evaluated using `hmmsearch`, which reads a Profile HMM from `hmmfile` and searches a sequence file for significantly similar sequence matches. The

sequence file searched contained 976 sequences (see above). During the search, the size of the database (Z parameter) was set to 1 billion. This size setting ensures that significant E-values against the current database will remain significant in the foreseeable future. The E-value cutoff was set at 10.

[0162] A `hmmsearch` with the Profile HMM generated from the alignment of the eight DHADs with experimentally verified function, matched all 976 sequences with an E value $<10^{-5}$. This result indicates that members of the dehydratase superfamily share significant sequence similarity. A `hmmsearch` with a cutoff of E value 10^{-5} was used to separate DHAD related dehydratases from other more remote but related proteins, as described above.

Example 2

Expression and Characterization of Bacterial [2Fe-2S] Dihydroxy-Acid Dehydratases in *E. coli*

[0163] The *ilvD* coding regions from different bacteria, which from the phylogenetic analysis described in Example 1 are in the [2Fe-2S] group, were expressed under the control of the T7 promoter in vector pET28a (Novagen) in *E. coli*. Each *ilvD* coding region was amplified with a specific forward primer with an *NheI* restriction site and a specific reverse primer with a *NotI* restriction site (listed in Table 5).

TABLE 5

SEQ ID NOs of primers used for PCR of DHAD coding regions from the listed organisms.		
Organism	Forward primer SEQ ID NO	Reverse Primer SEQ ID NO
<i>Nitrosomonas europaea</i> ATCC 19718	424	401
<i>Synechocystis</i> sp. PCC 6803	399	400
<i>Streptococcus mutans</i> UA159 (ATCC 700610)	435	436
<i>Streptococcus thermophilus</i> LMG 18311	395	396
<i>Ralstonia metallidurans</i> CH34	404	405
<i>Ralstonia eutropha</i> H16 (ATCC 17699)	406	407
<i>Lactococcus lactis</i>	420	421

[0164] The genomic DNA of each bacterial strain was used as a template. Genomic DNA was prepared from each strain listed in Table 1 using a MasterPure DNA Purification Kit (Epicentre, Madison, Wis.). The plasmid vector was amplified with primers pET28a-F (*NotI*) (SEQ ID NO:397) and pET28a-R (*NheI*) (SEQ ID NO:398) to remove the his tag region. Both gene and plasmid fragments were digested with *NheI* and *NotI* before ligation. The ligation mixture was transformed into *E. coli* (Top 10) competent cells (Invitrogen). Transformants were grown in LB agar plates supplemented with 50 $\mu\text{g}/\text{ml}$ of kanamycin. Positive clones that were confirmed by sequencing were transformed into the *E. coli* Tuner (DE3) strain (Novagen) for expression. Selected colonies were grown in LB liquid medium supplemented with kanamycin at 30° C. Induction was carried out by adding 0.5 mM of IPTG when the *E. coli* culture reached an O.D. of 0.3 to 0.4 at 600 nm. The culture was harvested after 5 hours of induction. Cell pellets were washed with Tris buffer (pH 8.0).

[0165] Enzymatic activity of the crude extract was assayed at 37° C. as follows. Cells to be assayed for DHAD were suspended in 2-5 volumes of 50 mM Tris, 10 mM MgSO_4 , pH 8.0 (TM8) buffer, then broken by sonication at 0° C. The crude extract from the broken cells was centrifuged to pellet

the cell debris. The supernatants were removed and stored on ice until assayed (initial assay was within 2 hrs of breaking the cells). It was found that the DHADs assayed herein were stable in crude extracts kept on ice for a few hours. The activity was also preserved when small samples were frozen in liquid N₂ and stored at -80° C.

[0166] The supernatants were assayed using the reagent 2,4-dinitrophenyl hydrazine as described in Flint and Emptage (J. Biol. Chem. (1988) 263: 3558-64). When the activity was so high that it became necessary to dilute the crude extract to obtain an accurate assay, the dilution was done in 5 mg/ml BSA in TM8.

[0167] Protein assays were performed using the Pierce Better Bradford reagent (cat 23238) using BSA as a standard. Dilutions for protein assays were made in TM8 buffer when necessary.

[0168] All of the DHADs were active when expressed in *E. coli*, and the specific activities are given in Table 6. The DHAD from *Streptococcus mutans* had the highest specific activity.

TABLE 6

Activities of bacterial [2Fe—2S] DHADs in <i>E. coli</i>		
Organism Source of DHAD	SEQ ID NO of DHAD coding sequence	Specific Activity ($\mu\text{mol min}^{-1} \text{mg}^{-1}$)
<i>Nitrosomonas europaea</i> ATCC 19718	309	3.3
<i>Synechocystis</i> sp. PCC 6803	297	3.5
<i>Streptococcus mutans</i> UA159 (ATCC 700610)	167	7.9
<i>Streptococcus thermophilus</i> LMG 18311	163	6.6
<i>Ralstonia metallidurans</i> CH34	404	2.4
<i>Ralstonia eutropha</i> H16 (ATCC 17699)	406	4.1
<i>Lactococcus lactis</i>	231	2.1
Vector control	N/A	0.07

Example 3

Purification and Characterization of DHAD from *Streptococcus mutans* Expressed in *E. coli*

[0169] The DHAD from *S. mutans* was further purified and characterized. For purification of the *S. mutans* DHAD, six liters of culture of the *E. coli* Tuner strain harboring the pET28a plasmid with the *S. mutans* *ilvD* were grown and induced with IPTG. The enzyme was purified by breaking the cells, as described in Example 2, in a 50 mM Tris buffer pH 8.0 containing 10 mM MgCl₂ (TM8 buffer), centrifuging to remove cell debris, then loading the supernatant of the crude extract on a Q Sepharose (GE Healthcare) column and eluting the DHAD with an increasing concentration of NaCl in TM8 buffer. The fractions containing the DHAD based on the color appearance (brownish color is due to the presence of the Fe—S cluster) were pooled and loaded onto a Sephacryl S-100 (GE Healthcare) column and eluted with TM8 buffer. As judged by SDS gels, the purity of the protein eluted from the Sephacryl column was estimated to be 60-80%. The activity of the partially purified enzyme was assayed at 37° C. as described by Flint et al. (J. Biol. Chem. (1988) 263(8): 3558-64). The specific activity of the purified protein was 40 $\mu\text{mol min}^{-1} \text{mg}^{-1}$. The k_{cat} for the purified enzyme was estimated to be 50-70 sec^{-1} .

[0170] Stability of the purified DHAD in air was studied by incubating the purified enzyme at 23° C. for various time intervals in the presence of ambient air, followed by an activity assay as described above. The activity of the DHAD from *Streptococcus mutans*, in contrast to DHAD purified similarly from *E. coli*, was stable even after 72 hours of incubation as shown in FIG. 4, where (A) shows results from the *S. mutans* DHAD and (B) shows results from the *E. coli* DHAD.

[0171] The UV-visible spectrum of the purified *S. mutans* is shown in FIG. 5. The number of peaks above 300 nm is typical of proteins with [2Fe-2S] clusters. The *S. mutans* DHAD was reduced with sodium dithionite and EPR spectra were obtained at varying temperatures. FIG. 6 show spectra measured at temperatures between 20° K. and 70° K. That the EPR spectrum of *S. mutans* is measurable up to 70° K. is indicative that it contains a [2Fe-2S] cluster. It is well known for example that the EPR spectra of proteins containing [4Fe-4S] clusters are not observable at temperatures much above 10° K. (See, for example, Rupp, et al. Biochimica et Biophysica Acta (1978) 537:255-269.)

Example 4

Construction of Dihydroxy-Acid Dehydratase (DHAD) Expression Cassettes for *Lactobacillus plantarum*

[0172] The purpose of this example is to describe how to clone and express a dihydroxy-acid dehydratase gene (*ilvD*) from different bacterial sources in *Lactobacillus plantarum* PN0512 (ATCC PTA-7727). A shuttle vector pDM1 (SEQ ID NO:410) was used for cloning and expression of *ilvD* genes from *Lactococcus lactis* subsp *lactis* NCDO2118 (NCIMB 702118) [Godon et al., J. Bacteriol. (1992) 174:6580-6589] and *Streptococcus mutans* UA159 (ATCC 700610) in *L. plantarum* PN0512. Plasmid pDM1 contains a minimal pLF1 replicon (~0.7 Kbp) and *pemK-pemI* toxin-antitoxin (TA) from *Lactobacillus plantarum* ATCC14917 plasmid pLF1, a P15A replicon from pACYC184, chloramphenicol marker for selection in both *E. coli* and *L. plantarum*, and P30 synthetic promoter [Rud et al, *Microbiology* (2006) 152:1011-1019]. Plasmid pLF1 (C.-F. Lin et al., GenBank accession no. AF508808) is closely related to plasmid p256 [Sørvig et al., *Microbiology* (2005) 151:421-431], whose copy number was estimated to be ~5-10 copies per chromosome for *L. plantarum* NC7. A P30 synthetic promoter was derived from *L. plantarum* rRNA promoters that are known to be among the strongest promoters in lactic acid bacteria (LAB) [Rud et al. *Microbiology* (2005) 152:1011-1019].

[0173] The *Lactococcus lactis* *ilvD* coding region (SEQ ID NO:231) was PCR-amplified from *Lactococcus lactis* subsp *lactis* NCDO2118 genomic DNA with primers 3T-*ilvDLI* (BamHI) (SEQ ID NO:408) and 5B-*ilvDLI* (NotI) (SEQ ID NO:409). *L. lactis* subsp *lactis* NCDO2118 genomic DNA was prepared with a Puregene Gentra Kit (QIAGEN, CA). The 1.7 Kbp *L. lactis* *ilvD* PCR product (*ilvDLI*) was digested with NotI and treated with the Klenow fragment of DNA polymerase to make blunt ends. The resulting *L. lactis* *ilvD* coding region fragment was digested with BamHI and gel-purified using a QIAGEN gel extraction kit (QIAGEN, CA). Plasmid pDM1 was digested with ApaLI, treated with the Klenow fragment of DNA polymerase to make blunt ends, and then digested with BamHI. The gel purified *L. lactis* *ilvD* coding region fragment was ligated into the BamHI and ApaLI (blunt) sites of the plasmid pDM1. The ligation mix-

ture was transformed into *E. coli* Top10 cells (Invitrogen, CA). Transformants were plated for selection on LB chloramphenicol plates. Positive clones were screened by Sall digestion, giving one fragment with an expected size of 5.3 Kbp. The positive clones were further confirmed by DNA sequencing. The correct clone was named pDM1-ilvD (*L. lactis*).

[0174] The *S. mutans* UA159 (ATCC 700610) ilvD coding region from the plasmid pET28a was cloned on the plasmid pDM1. The construction of pET28a containing the *S. mutans* ilvD was described in Example 2. The plasmid pET28a containing the *S. mutans* ilvD was digested with XbaI and NotI, treated with the Klenow fragment of DNA polymerase to make blunt ends, and a 1,759 bp fragment containing the *S. mutans* ilvD coding region was gel-purified. Plasmid pDM1 was digested with BamHI, treated with the Klenow fragment of DNA polymerase to make blunt ends, and then digested with PvuII. The gel purified fragment containing *S. mutans* ilvD coding region was ligated into the BamHI (blunt) and PvuII sites of the plasmid pDM1. The ligation mixture was transformed into *E. coli* Top10 cells (Invitrogen, CA). Transformants were plated for selection on LB chloramphenicol plates. Positive clones were screened by ClaI digestion, giving one fragment with an expected size of 5.5 Kbp. The correct clone was named pDM1-ilvD (*S. mutans*).

Example 5

Measurement of Expressed DHAD Activity in *L. plantarum* PN0512

[0175] *L. plantarum* PN0512 was transformed with plasmid pDM1-ilvD (*L. lactis*) or pDM1-ilvD (*S. mutans*) by electroporation. Electro-competent cells were prepared by the following procedure. 5 ml of Lactobacilli MRS medium containing 1% glycine was inoculated with PN0512 cells and grown overnight at 30° C. 100 ml MRS medium with 1% glycine was inoculated with the overnight culture to an OD₆₀₀=0.1 and grown to an OD₆₀₀=0.7 at 30° C. Cells were harvested at 3700×g for 8 min at 4° C., washed with 100 ml cold 1 mM MgCl₂, centrifuged at 3700×g for 8 min at 4° C., washed with 100 ml cold 30% PEG-1000 (81188, Sigma-Aldrich, St. Louis, Mo.), recentrifuged at 3700×g for 20 min at 4° C., then resuspended in 1 ml cold 30% PEG-1000. 60 μl of electro-competent cells were mixed with ~100 ng plasmid DNA in a cold 1 mm gap electroporation cuvette and electroporated in a BioRad Gene Pulser (Hercules, Calif.) at 1.7 kV, 25 μF, and 400Ω. Cells were resuspended in 1 ml MRS medium containing 500 mM sucrose and 100 mM MgCl₂, incubated at 30° C. for 2 hrs, and then plated on MRS medium plates containing 10 μg/ml of chloramphenicol.

[0176] *L. plantarum* PN0512 transformants carrying pDM1-ilvD (*L. lactis*) or pDM1-ilvD (*S. mutans*), as well as control transformants with the pDM1 vector alone, were grown overnight in Lactobacilli MRS medium at 30° C. 120 ml of MRS medium supplemented with 100 mM MOPS (pH7.5), 40 μM ferric citrate, 0.5 mM L-cysteine, and 10 μg/ml chloramphenicol was inoculated with overnight culture to an OD₆₀₀=0.1 in a 125 ml screw cap flask, for each overnight sample. The cultures were anaerobically incubated at 37° C. until reaching an OD₆₀₀ of 1-2. Cultures were centrifuged at 3700×g for 10 min at 4° C. Pellets were washed with 50 mM potassium phosphate buffer pH 6.2 (6.2 g/L KH₂PO₄ and 1.2 g/L K₂HPO₄) and re-centrifuged. Pellets were frozen and stored at -80° C. until assayed for DHAD activity. Cell extract samples were assayed for DHAD activity using a dinitrophenylhydrazine based method as described in Example 2. The DHAD activity results are given in Table 7. Specific activity of *L. lactis* DHAD and *S. mutans* DHAD in

L. plantarum PN0512 showed 0.02 and 0.06 μmol min⁻¹ mg⁻¹, respectively, while the vector control sample exhibited no detectable activity.

TABLE 7

DHAD activity in <i>L. plantarum</i> PN0512.		
Source of DHAD	Plasmid	Specific Activity (μmol min ⁻¹ mg ⁻¹)
Vector control	pDM1	0.00
<i>Lactococcus lactis</i> subsp <i>lactis</i> NCDO2118	pDM1-ilvD(<i>L. lactis</i>)	0.02
<i>Streptococcus mutans</i> UA159	pDM1-ilvD(<i>S. mutans</i>)	0.06

Example 6

Expression of Dihydroxy-Acid Dehydratase from *S. mutans* in Yeast

[0177] The shuttle vector pRS423 FBA ilvD (Strep) (SEQ ID NO:430) was used for the expression of DHAD from *Streptococcus mutans*. This shuttle vector contained an F1 origin of replication (1423 to 1879) for maintenance in *E. coli* and a 2 micron origin (nt 8082 to 9426) for replication in yeast. The vector has an FBA promoter (nt 2111 to 3108; SEQ ID NO:425) and FBA terminator (nt 4861 to 5860). In addition, it carries the His marker (nt 504 to 1163) for selection in yeast and ampicillin resistance marker (nt 7092 to 7949) for selection in *E. coli*. The ilvD coding region (nt 3116 to 4828) from *Streptococcus mutans* UA159 (ATCC 700610) is between the FBA promoter and FBA terminator forming a chimeric gene for expression.

[0178] To test the expression of the DHAD from *Streptococcus mutans* in yeast strain BY4741 (known in the art and obtainable from ATCC #201388), the expression vector pRS423 FBA IlvD (Strep) was transformed in combination with empty vector pRS426 into BY4741 cells (obtainable from ATCC, # 201388). The transformants were grown on synthetic medium lacking histidine and uracil (Teknova). Growth on liquid medium for assay was carried out by adding 5 ml of an overnight culture into 100 ml medium in a 250 ml flask. The cultures were harvested when they reached 1 to 2 O.D. at 600 nm. The samples were washed with 10 ml of 20 mM Tris (pH 7.5) and then resuspended in 1 ml of the same Tris buffer. The samples were transferred into 2.0 ml tubes containing 0.1 mm silica (Lysing Matrix B, MP biomedical). The cells were then broken in a bead-beater (BIO101). The supernatant was obtained by centrifugation in a microfuge at 13,000 rpm at 4° C. for 30 minutes. Typically, 0.06 to 0.1 mg of crude extract protein was used in DHAD assay at 37° C. as described by Flint and Emptage (J. Biol. Chem. (1988) 263 (8): 3558-64) using dinitrophenylhydrazine. The dehydratase from *Streptococcus mutans* had a specific activity of 0.24 μmol min⁻¹ mg⁻¹ when expressed in yeast. A control strain containing empty vectors pRS423 and pRS426 had a background of activity in the range of 0.03 to 0.06 μmol min⁻¹ mg⁻¹.

Example 7

Expression of the IlvD Gene from *L. lactis* in Yeast

[0179] The ilvD coding region from *L. lactis* was amplified with the forward primer IlvD (LI)-F (SEQ ID NO:420) and reverse primer IlvD (LI)-R (SEQ ID NO:421). The amplified fragment was cloned into the shuttle vector pNY13by gap

repair. pNY13 (SEQ ID NO:437) was derived from pRS423. This shuttle vector contained an F1 origin of replication (1423 to 1879) for maintenance in *E. coli* and a 2 micron origin (nt 7537 to 8881) for replication in yeast. The vector has an FBA promoter (nt 2111 to 3110) and FBA terminator (nt 4316 to 5315). In addition, it carries the His marker (nt 504 to 1163) for selection in yeast and Ampicillin resistance marker (nt 6547 to 7404) for selection in *E. coli*.

[0180] Positive clones were selected based on amplification with the forward and reverse primers for the *ilvD* coding region and further confirmed by sequencing. The new construct was designated as pRS423 FBA *ilvD* (*L. lactis*). This construct was transformed into yeast strain BY4743 (Δ LEU1) (Open Biosystems, Huntsville, Ala.; Catalog #YSC1021-666629) along with the empty vector pRS426 as described in Example 6. The growth and assay of yeast strains containing the expression vector were also carried out according to the procedures as described in Example 6. The dihydroxy-acid dehydratase activity in yeast strains with the *L. lactis* *ilvD* gene was determined to be in the range of 0.05 to 0.17 $\mu\text{mol min}^{-1} \text{mg}^{-1}$. This activity was slightly above the control. Complementation experiments were carried out to investigate the expression of this DHAD and DHADs from other bacteria (example 8).

Example 8

Complementation of Yeast *ILV3* Deletion Strain with Bacterial DHADs

[0181] The endogenous DHAD enzyme of *S. cerevisiae* is encoded by the *ILV3* gene and the protein is targeted to the mitochondrion. Deletion of this gene results in loss of endogenous DHAD activity and provides a test strain where expression of heterologous cytosolic DHAD activity can be readily assessed. Deletion of *ILV3* results in an inability of the strain to grow in the absence of branched-chained amino acids. Expression of different bacterial DHADs was assayed by determining their ability to complement the yeast *ILV3* deletion strain such that it grows in the absence of branched-chained amino acids.

[0182] Expression shuttle vectors containing *ilvD* gene sequences encoding DHADs from the bacteria listed in Table 8 were constructed. The basic elements of these expression constructs were the same as the pRS423 FBA *ilvD* (strep) vector described in Example 6. Each of the *ilvD* coding regions was prepared by PCR as described in Example 2 and cloned to replace the *Streptococcus mutans* *ilvD* coding region in pRS423 FBA *ilvD* (Strep) creating the plasmids listed in Table 8. These expression constructs were transformed into the *ILV3* deletion strain, BY4741 *ilv3::URA3* and was prepared as follows. An *ilv3::URA3* disruption cassette was constructed by PCR amplification of the *URA3* marker from pRS426 (ATCC No. 77107) with primers “*ILV3::URA3 F*” and “*ILV3::URA3 R*”, given as SEQ ID NOs: 431 and 432. These primers produced a 1.4 kb *URA3* PCR product that contained 70 by 5' and 3' extensions identical to sequences upstream and downstream of the *ILV3* chromosomal locus for homologous recombination. The PCR product was transformed into BY4741 cells (ATCC 201388) using standard genetic techniques (*Methods in Yeast Genetics*, 2005, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., pp. 201-202) and resulting transformants were maintained on synthetic complete media lacking uracil and supplemented with 2% glucose at 30° C. Transformants were screened by PCR using primers “*ILV3 F Check*”

and “*URA3 REV Check*”, given as SEQ ID NOs:433 and 434, to verify integration at the correct site and disruption of the endogenous *ILV3* locus.

[0183] The transformants with bacterial DHADs in Table 8 were selected on plates with yeast synthetic medium lacking histidine. Colonies selected were then patched onto plates lacking valine, leucine and isoleucine. Strains containing the expression vectors listed in Table 8 were able to grow on these plates lacking the branched chain amino acids, while the control strain with the control plasmid did not. This result indicated that DHADs from these bacteria were actively expressed in *S. cerevisiae*.

TABLE 8

Bacterial DHADs tested and Expression vectors		
Organism Source of DHAD	SEQ ID NO. OF DHAD nucleic acid sequence	Vector designation
<i>Nitrosomonas europaea</i> ATCC 19718	309	pRS423 FBA <i>ilvD</i> (<i>europ</i>)
<i>Synechocystis</i> sp. PCC 6803	297	pRS423 FBA <i>ilvD</i> (<i>Synech</i>)
<i>Streptococcus thermophilus</i> LMG 18311	163	pRS423 FBA <i>ilvD</i> (<i>thermo</i>)
<i>Ralstonia eutropha</i> H16 (ATCC 17699)	406	pRS423 FBA <i>ilvD</i> (H16)
<i>Lactococcus lactis</i>	231	pRS423 FBA <i>ilvD</i> (<i>L. lactis</i>)

Example 9

Purification and Characterization of DHAD from *Lactococcus lactis* Expressed in *E. coli*

[0184] The DHAD from *L. lactis* was purified and characterized. For purification of *L. lactis* DHAD, 14 liters of culture of the *E. coli* Tuner (DE3) strain (Novagen) harboring the pET28a plasmid containing the *L. lactis* *ilvD* were grown and induced with IPTG. The enzyme was purified by breaking the cells, as described in Example 2, in 120 mls of 50 mM Tris buffer pH 8.0 containing 10 mM MgCl_2 (TM8 buffer), centrifuging to remove cell debris, then loading the supernatant of the crude extract on a 5x15 cm Q Sepharose (GE Healthcare) column and eluting the DHAD with an increasing concentration of NaCl in TM8 buffer. The fractions containing the DHAD were pooled, made 1 M in $(\text{NH}_4)_2\text{SO}_4$ and loaded onto a 2.6x15 cm phenyl-Sepharose column (GE Healthcare) column equilibrated with 1 M $(\text{NH}_4)_2\text{SO}_4$ in TM8 buffer and eluted with a decreasing gradient of $(\text{NH}_4)_2\text{SO}_4$. The fractions containing DHAD off the phenyl-Sepharose column were pooled and concentrated to 10 ml. This was loaded onto a 3.5x60 cm Superdex-200 column (GE Healthcare) and eluted with TM8. The fractions containing DHAD activity were pooled, concentrated, and frozen as beads in $\text{N}_2(g)$. As judged by SDS gels, the purity of the protein eluted from the Superdex-200 column was estimated to be >80%. The activity of the enzyme was assayed at 37° C. as described by Flint et al. (J. Biol. Chem. (1988) 263(8): 3558-64). The specific activity of the purified protein was 64 $\mu\text{mol min}^{-1} \text{mg}^{-1}$ at pH 8 and 37° C. The k_{cat} for the purified enzyme was 71 sec^{-1} .

[0185] Stability of the purified DHAD in air was studied by incubating the purified enzyme at 23° C. for various time intervals in the presence of ambient air, followed by an activ-

ity assay as described above. The DHAD from *L. lactis* was almost fully active even after 20 hours of incubation in air as shown in FIG. 8.

[0186] The UV-visible spectrum of the purified *L. lactis* is shown in FIG. 9. It is characteristic of proteins with [2Fe-2S] clusters.

Example 10

Use of Dihydroxy-Acid Dehydratase to Construct a Pathway for the Production of Isobutanol in Yeast

[0187] The first three steps of an isobutanol biosynthetic pathway are performed by the enzymes acetolactate synthase, ketol-acid reductoisomerase (KARI), and dihydroxy-acid dehydratase. Acetolactate synthase is encoded by *alsS*. KARI genes are known as *ILV5* in yeast or *ilvC* in bacteria. Once α -ketoisovalerate (KIV) is formed from pyruvate by the reaction of these three enzymes, it can be further converted to isobutanol in yeast by alcohol dehydrogenases.

[0188] Vector pLH532 (SEQ ID NO:411) was constructed to express KARI and *alsS* genes. This vector is derived from the 2 MICRON based vector pHR81. In pLH532 the *alsS* coding region from *B. subtilis* (nt 14216 to 15931) was under the control of the CUP1 promoter (nt. 15939 to 16386). There were two KARI genes in pLH532: the *ilvC* coding region from *P. fluorescens* Pf5 (nt. 10192 to 11208) was under the control of the yeast *ILV5* promoter (nt. 11200 to 12390), and the yeast *ILV5* coding region (nt. 8118-9167) was placed under the control of the FBA promoter (nt. 7454-8110). The selection marker was *URA3* (nt. 3390 to 4190).

[0189] The yeast host for isobutanol production was BY4741 *pdcl::FBAp-alsS-LEU2*. This strain was constructed as follows. First, the expression plasmid pRS426-FBAP-alsS was constructed. The 1.7 kb *alsS* coding region fragment of pRS426::*GPD::alsS::CYC* was isolated by gel purification following *BbvCI* and *PacI* digestion. This plasmid has a chimeric gene containing the *GPD* promoter (SEQ ID NO:439), the *alsS* coding region from *Bacillus subtilis* (SEQ ID NO:438), and the *CYC1* terminator (SEQ ID NO:440) and was described in Example 17 of US Patent Publication # US20070092957A1 which is herein incorporated by reference. The *ILV5* fragment from plasmid pRS426::*FBA::ILV5::CYC*, also described in US20070092957 Example 17, was removed by restriction digestion with *BbvCI* and *PacI* and the remaining 6.6 kb vector fragment was gel purified. This vector has a chimeric gene containing the *FBA* promoter (SEQ ID NO:425) and *CYC1* terminator bounding the coding region of the *ILV5* gene of *S. cerevisiae* (SEQ ID NO:442). These two purified fragments were ligated overnight at 16° C. and transformed into *E. coli* TOP10 chemically competent cells (Invitrogen). Transformants were obtained by plating cells on LB Amp100 medium. Insertion of *alsS* into the vector was confirmed by restriction digestion pattern and PCR (primers N98SeqF1 and N99SeqR2, SEQ ID NOs:412 and 413).

[0190] A *pdcl::FBAp-alsS-LEU2* disruption cassette was created by joining the *FBAP-alsS* segment from pRS426-FBAP-alsS to the *LEU2* gene from pRS425 (ATCC No.

77106) by SOE PCR (as described by Horton et al. (1989) Gene 77:61-68) using as template pRS426-FBAP-alsS and pRS425 plasmid DNAs, with Phusion DNA polymerase (New England Biolabs Inc., Beverly, Mass.; catalog no. F-540S) and primers 112590-48A and 112590-30B through D, given as SEQ ID NOs:414, SEQ ID NOs:415, 416, and 417. The outer primers for the SOE PCR (112590-48A and 112590-30D) contained 5' and 3' 50 by regions homologous to regions upstream and downstream of the *PDC1* promoter and terminator. The completed cassette PCR fragment was transformed into BY4741 (ATCC No. 201388) and transformants were maintained on synthetic complete media lacking leucine and supplemented with 2% glucose at 30° C. using standard genetic techniques (*Methods in Yeast Genetics*, 2005, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., pp. 201-202). Transformants were screened by PCR using primers 112590-30E and 112590-30F, given as SEQ ID NOs:419 and 418, to verify integration at the *PDC1* locus with deletion of the *PDC1* coding region. The correct transformants have the genotype: BY4741 *pdcl::FBAP-alsS-LEU2*.

[0191] To test whether the *ilvD* encoded DHAD from *Streptococcus mutans* could be used for the biosynthesis of isobutanol, the expression vector containing this *ilvD*, pRS423 FBA *ilvD* (strep) prepared in Example 6, was co-transformed with vector pLH532 into yeast strain BY4741 *pdcl::FBAP-alsS-LEU2*. Competent cell preparation, transformation, and growth medium for selection of the transformants were the same as described in Example 6. Selected colonies were grown under oxygen-limiting conditions in 15 ml of medium in 20 ml serum bottles with stoppers. The bottles were incubated at 30° C. in a shaker with a constant speed of 225 rotations per minute. After 48 hours of incubation, the samples were analyzed with HPLC for the presence of isobutanol. Result from the HPLC analysis is shown in FIG. 7. The presence of isobutanol was indicated by a peak with a retention time of 47.533 min. This result showed that expression of the *ilvD* gene from *Streptococcus mutans*, along with expression of *alsS* and KARI genes, led to the production of isobutanol in yeast.

Example 11

Query of Updated Databases to Identify Additional [2Fe-2S]-Dihydroxy-Acid Dehydratases

[0192] A later second query of the updated public database was performed to discover newly sequenced [2Fe-2S] dihydroxy-acid dehydratases. At the 95% identity cutoff, an initial set of 1425 sequences was generated from a database query as described in Example 1, "Phylogenetic analysis". Multiple sequence alignments were then executed with ClustalW as described in Example 1. Sequences were subsequently analyzed for the following conserved residues at the corresponding positions in the *S. mutans* DHAD: cysteines at positions 56, 129, and 201, aspartic acid at position 88, arginine or asparagine at position 142, asparagine at position 208, and leucine at position 454. In addition to the original set of 193, 88 novel [2Fe-2S] dihydroxy-acid dehydratases from bacteria were identified and are listed in Table 2b.

TABLE 1-continued

560(C)	2804	3772	-3185	-3198	-2739	-1303	-2462	-2065	-2882	-2628	-1924	-1927	-2044	-2547	-2727	-661	-799	-1463	-3099	-2886	566
—	-149	-500	233	43	-381	399	106	-626	210	-466	-720	275	394	45	96	359	117	-369	-294	-249	
—	-16	-7108	-8150	-894	-1115	-701	-1378	*	*	*	*	*	*	*	*	*	*	*	*	*	
561(V)	-1771	-1603	-3750	-3689	-2037	-3050	-3231	403	-3479	-1154	-1076	-3246	-3399	-3383	-3437	-2628	-1917	3536	-3074	-2677	567
—	-149	-500	233	43	-381	399	106	-626	210	-466	-720	275	394	45	96	359	117	-369	-294	-249	
—	-16	-7108	-8150	-894	-1115	-701	-1378	*	*	*	*	*	*	*	*	*	*	*	*	*	
562(T)	-1213	-1674	-2755	-2906	-3163	-1922	-2659	-2698	-2788	-3105	-2612	-2311	-2600	-2708	-2753	-1463	3819	-2197	-3286	-3156	568
—	-149	-500	233	43	-381	399	106	-626	210	-466	-720	275	394	45	96	359	117	-369	-294	-249	
—	-16	-7108	-8150	-894	-1115	-701	-1378	*	*	*	*	*	*	*	*	*	*	*	*	*	
563(D)	-2784	-3432	4016	-1200	-4140	-2466	-2197	-4505	-2621	-4365	-3956	-1551	-3014	-2039	-3232	-2593	-2938	-4046	-3710	-3552	569
—	-149	-500	233	43	-381	399	106	-626	210	-466	-720	275	394	45	96	359	117	-369	-294	-249	
—	-21	-6715	-7757	-894	-1115	-701	-1378	*	*	*	*	*	*	*	*	*	*	*	*	*	
564(F)	-525	-445	-2202	-1627	1946	-2001	-744	1247	-1346	952	561	1079	-2030	-1067	-1362	-1067	-465	338	-714	-230	570
—	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
—	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

HMMER2.0 [2.2g]
 NAME dhad_for_hmm
 LENG 564
 ALPH Amino
 MAP yes
 COM/app/public/hmmer/current/bin/hmmbuild -F dhad-exp_hmm dhad_for_hmm.aln
 COM/app/public/hmmer/current/bin/hmmcalibrate dhad-exp_hmm
 NSEQ 8
 DATE Tue Jun 3 10:48:24 2008
 XT -8455 -4 -1000 -1000 -8455 -4 -8455 -4
 NULL -4 -8455
 NULE 595 -1558 85 338 -294 453 -1158 197 249 902 -1085 -142 -21 -313 45 531 201 384 -1998
 EVD -499.650970 0.086142

Program name and version
 Name of the input sequence alignment file
 Length of the alignment: include indels
 Type of residues
 Map of the match states to the columns of the alignment
 Commands used to generate the file: this one means that hmmbuild (default parameters) was applied to the alignment file
 Commands used to generate the file: this one means that hmmcalibrate (default parameters) was applied to the hmm profile
 Number of sequences in the alignment file
 When was the file generated
 The transition probability distribution for the null model (single G state).
 The symbol emission probability distribution for the null model (G state); consists of K (e.g. 4 or 20) integers.
 The null probability used to convert these back to model probabilities is 1/K.
 The extreme value distribution parameters μ and λ respectively; both floating point values. Lambda is positive and nonzero. These values are set when the model is calibrated with hmmlcalibrate.

 SEQUENCE LISTING

The patent application contains a lengthy "Sequence Listing" section. A copy of the "Sequence Listing" is available in electronic form from the USPTO web site (<http://seqdata.uspto.gov/?pageRequest=docDetail&DocID=US20100081154A1>). An electronic copy of the "Sequence Listing" will also be available from the USPTO upon request and payment of the fee set forth in 37 CFR 1.19(b)(3).

What is claimed is:

1. A method for identifying [2Fe-2S] DHAD enzymes comprising:

a) querying one or more amino acid sequences with a Profile Hidden Markov Model prepared using the proteins of SEQ ID NOs:164, 168, 230, 232, 298, 310, 344, and 346, wherein a match with an E-value of less than 10^{-5} provides a first subset of sequences whereby said first subset of sequences correspond to one or more DHAD related proteins;

b) analyzing the first subset of sequences that correspond to one or more DHAD related proteins of step (a) for the presence of three conserved cysteines that correspond to positions 56, 129, and 201 in the *Streptococcus mutans* dihydroxy-acid dehydratase amino acid sequence (SEQ ID NO: 168) whereby a second subset of sequences encoding [2Fe-2S] DHAD enzymes are identified; and

c) analyzing the second subset of sequences of step (b) for the presence of signature conserved amino acids at positions corresponding to positions in the *Streptococcus mutans* DHAD amino acid sequence (SEQ ID NO: 168) that are aspartic acid at position 88, arginine or asparagine at position 142, asparagine at position 208, and leucine at position 454 whereby a third subset of sequences encoding [2Fe-2S] DHAD enzymes are further identified.

2. The method of claim 1 further comprising

d) expressing a polypeptide having a sequence identifiable by any one or all of steps a), b), and c) in a cell; and

e) confirming that said polypeptide has DHAD activity in the cell.

3. The method of claim 1 further comprising

d) purifying a protein encoded by a sequence identifiable by any one or all of steps a), b), and c); and

e) confirming that said protein is a [2Fe-2S] DHAD enzyme by UV-vis and EPR spectroscopy.

4. The method of claim 1 further comprising selecting one or more sequences corresponding to bacterial [2Fe-2S] DHAD enzyme sequences identified in any one or all of steps a), b), and c).

5. The method of claim 2 wherein the cell lacks endogenous DHAD activity.

6. The method of claim 4 further comprising

d) expressing said selected one or more sequences corresponding to bacterial [2Fe-2S] DHAD enzyme sequences in a cell; and

e) confirming that said enzyme sequence has DHAD activity in the cell.

7. The method of claim 4 further comprising

d) purifying a protein encoded by said selected one or more sequences corresponding to bacterial [2Fe-2S] DHAD enzyme sequences whereby a purified protein is produced; and

e) confirming that the protein is a [2Fe-2S] DHAD enzyme by UV-vis and EPR spectroscopy.

8. A microbial host cell comprising at least one heterologous [2Fe-2S] DHAD enzyme identifiable by the method of any one of claims 1-7.

9. The microbial host cell of claim 8 wherein the cell is bacterial cell or a yeast cell.

10. The microbial host cell of claim 9 wherein the bacterial host cell is a member of a genus of bacteria selected from the group consisting of *Clostridium*, *Zymomonas*, *Escherichia*, *Salmonella*, *Rhodococcus*, *Pseudomonas*, *Bacillus*, *Lactobacillus*, *Enterococcus*, *Pediococcus*, *Alcaligenes*, *Klebsiella*, *Paenibacillus*, *Arthrobacter*, *Corynebacterium*, and *Brevibacterium*, *Lactococcus*, *Leuconostoc*, *Oenococcus*, *Pediococcus*, and *Streptococcus*.

11. The microbial host cell of claim 9 wherein the yeast cells a member of a genus of yeast selected from the group consisting of *Saccharomyces*, *Schizosaccharomyces*, *Hansenula*, *Candida*, *Kluyveromyces*, *Yarrowia* and *Pichia*.

12. The microbial host cell of claim 8 wherein the cell produces isobutanol.

13. A method for the production of isobutanol comprising:

a) providing the microbial host cell of claim 8 wherein said host cell comprises an isobutanol biosynthetic pathway; and

b) growing the host cell of step (a) under conditions wherein isobutanol is produced.

14. A method for the conversion of 2,3-dihydroxyisovalerate to α -ketoisovalerate comprising:

a) providing the microbial host of claim 8 and a source of 2,3-dihydroxyisovalerate; and

b) growing the microbial host cell of (a) under conditions where the 2,3-dihydroxyisovalerate is converted to α -ketoisovalerate.

* * * * *