



(43) International Publication Date
5 October 2017 (05.10.2017)

(10) International Publication Number
WO 2017/168161 A1

(51) International Patent Classification:

C12N 9/10 (2006.01) *C12P 7/02* (2006.01)
C12N 15/52 (2006.01) *C12N 9/02* (2006.01)

(21) International Application Number:

PCT/GB2017/050901

(22) International Filing Date:

30 March 2017 (30.03.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

1605354.8 30 March 2016 (30.03.2016) GB

(71) Applicant: **ZUVASYNTHA LIMITED** [GB/GB];
Biopark Hertfordshire, Broadwater Road, Welwyn Garden
City, Hertfordshire AL7 3AX (GB).

(72) Inventor: **GRADLEY, Michelle**; 22A Nunnery Road,
Canterbury Kent CT1 3LU (GB).

(74) Agents: **GILL JENNINGS & EVERY LLP** et al.; The
Broadgate Tower, 20 Primrose Street, London, EC2A 2ES
(GB).

(81) Designated States (*unless otherwise indicated, for every
kind of national protection available*): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,

BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN,
KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA,
MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG,
NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS,
RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY,
TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN,
ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every
kind of regional protection available*): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ,
TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU,
TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE,
DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU,
LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,
SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments (Rule 48.2(h))*

(54) Title: MODIFIED ENZYME

(57) Abstract: The present invention provides a modified 2-deoxyribose phosphate aldolase (DERA) enzyme variant comprising one or more mutations that improve the aldolase catalytic performance for synthesis of 3-hydroxybutanal or crotonaldehyde, relative to the parent DERA enzyme from which the variant originates and which does not comprise such a modification.



WO 2017/168161 A1

MODIFIED ENZYME

Technical field

5 The present invention provides a modified enzyme that has improved activity for catalysing the formation of 3-hydroxybutanal. The modified enzyme is also capable of converting acetaldehyde to crotonaldehyde, via an aldol condensation. The current invention also relates generally to microorganisms, and related materials and methods, which have been modified to express said enzyme, in order to to
10 enhance their ability to produce commodity chemicals, for example, 1,3-butanediol and derivatives thereof, which can be produced in the microorganisms via the intermediates acetaldehyde and 3-hydroxybutanal.

Background art

15 1,3-butanediol (1,3-BDO) is a four carbon diol which has a number of uses, including in the food, chemical and manufacturing industries.

1,3-BDO has traditionally been produced from petroleum derived acetylene via its
20 hydration. The resulting acetaldehyde is then converted to 3-hydroxybutanal which is subsequently reduced to form 1,3-BDO. In more recent years, acetylene has been replaced by the less expensive ethylene as a source of acetaldehyde. However, as crude oil has become relatively more expensive than natural gas, many ethylene cracking operations are using lighter natural gas feedstocks to earn
25 higher margins, leading to significantly lower quantities of C4 chemicals and rising prices.

Increasing the flexibility of inexpensive and readily available feedstocks while minimizing the environmental impact of chemical production are two goals of a
30 sustainable chemical industry. Feedstock flexibility relies on the introduction of methods that enable access and use of a wide range of materials as primary feedstocks for chemical manufacturing. The reliance on petroleum based feedstocks for either acetylene or ethylene warrants the development of renewable,

or cheaper, or non-petroleum derived feedstock based routes to 1,3-butanediol, butadiene and other valuable chemicals such as methylethylketone.

5 Publications in which microorganisms have been modified in such a way as to affect 1,3-BDO accumulation include the following: US20130109064; US20120329113; EP2495305A1; US8268607; US20110201068; US20100330635 and WO2014036140.

10 Nevertheless it can be seen that developing microorganisms and methods of their use to ferment sustainable and/or cheaper than traditional petroleum based feedstocks to 1,3-butanediol and other important chemicals, would provide a contribution to the art.

Disclosure of the invention

15

The present invention relates to the engineering of organisms to imbue or enhance the ability to convert the central metabolic intermediates acetyl CoA and pyruvate to the common pathway intermediate acetaldehyde, which is then subject to an enzymatically catalysed aldol coupling, ultimately yielding 1,3-butanediol or other
20 products. More specifically, in modified organisms of the invention, acetaldehyde derived from acetyl CoA or pyruvate as the primary pathway product is supplied as the substrate for an aldolase capable of the coupling of two molecules of acetaldehyde to form 3-hydroxybutanal.

25 The 3-hydroxybutanal, which is the product of this aldol coupling, can be directed to other products or other intermediates which can then in turn enter other natural or unnatural metabolic pathways.

30 Example intermediates include 2-hydroxyisobutyryl CoA, 3-hydroxybutyryl CoA, Crotonyl CoA, Crotonaldehyde, Butyryl CoA, Butanal, Acetoacetyl CoA, and acetoacetate.

Desirable downstream products include 2-hydroxisobutyrate, Crotyl alcohol, Crotonic acid, Butanol, Butyrate, 3-hydroxybutyrate, 1,3-butanediol, 3-hydroxybutylamine, Polyhydroxybutyrate, Acetone, and Isopropanol.

- 5 These products can, where desired, be recovered and used to make yet further commodities – for example butadiene, methacrylic acid, 2-methyl-1,4-butanediol, methyltetrahydrofuran, isoprene.

Any of these intermediate products, downstream products, and commodities may be referred to herein as “downstream products” or “products” herein for brevity.

10

A preferred product is 1,3-butanediol (1,3-BDO).

The modified organisms of the invention are typically microorganisms capable of using renewable or inexpensive feedstocks or energy sources such as sunlight, carbohydrates, methanol, synthesis gas (syngas) and/or other gaseous carbon sources such as methane to generate the appropriate metabolic intermediates.

15

As explained in more detail hereinafter, imbuing or enhancing the production of acetaldehyde from the central metabolic intermediate, or increasing its availability to the aldolase, will typically involve one or more of:

20

- (i) introducing a heterologous gene encoding an enzyme having an activity utilised in generation of acetaldehyde from one or more of the central metabolic intermediates;
 - 25 (ii) up-regulating at least one endogenous enzyme having an activity utilised in generation of acetaldehyde from one or more of the central metabolic intermediates; and/or
 - (iii) down-regulating or inactivating an endogenous enzyme which utilises acetaldehyde as a substrate (thereby making the acetaldehyde more available to the aldolase).
- 30

Typically, the aldolase capable of the *in vivo* coupling of two molecules of acetaldehyde to 3-hydroxybutanal will itself also be the product of genetic engineering e.g. via the introduction of a heterologous aldolase as described below.

A genetic modification combining these changes thus serves to increase the flux of central metabolic intermediates to the 3-hydroxybutanal via the acetaldehyde intermediate. In preferred embodiments this 3-hydroxybutanal is subsequently
5 directed to a downstream product as described below.

Thus in one aspect there is provided a non-naturally occurring microbial organism which includes a genetic modification in its genome which enhances production of 3-hydroxybutanal by the microbial organism from at least one endogenous central
10 metabolic intermediate via a 3-hydroxybutanal synthetic pathway in which two molecules of acetaldehyde are coupled to form 3-hydroxybutanal using an aldolase capable of accepting an aldehyde as both the acceptor and donor in an aldol coupling.

As will be well understood by those skilled in the art, "enhanced production" or production of an "increased amount" in the context of an intermediate product of a pathway should not be taken as requiring an increase in the absolute concentration, or steady stage concentration, of a product, in the microbial cell, although that may well result from increased production. Rather it will be understood to include a faster
15 production of the product in question (i.e. a higher pathway flux through it) even where the product does not accumulate, but is subsequently converted to a further product.

Preferred organisms are those in which the modification enhances production of 1,3-butanediol (1,3-BDO) via a 1,3-BDO synthetic pathway in which the 3-hydroxybutanal is reduced to 1,3-BDO. In the disclosure below, therefore, particular
25 emphasis is given to this embodiment. However it will also be understood that the invention applies likewise to modifications and pathways in which 3-hydroxybutanal is converted to other downstream products and, unless context demands otherwise, each of the embodiments relating to 1,3-BDO will be understood to apply mutatis
30 mutandis to these other products.

The genetic modification will be such that said modified organism produces a greater flux of or through 3-hydroxybutanal (and hence also of or through a

downstream intermediate) to a product thereof such as 1,3-BDO) compared to a corresponding reference microbial organism not including said genetic modification, when grown on the same feedstock or energy source under the same conditions. For example the modified organism may produce at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200 times as much 3-hydroxybutanal or, more preferably, downstream product such as 1,3-BDO compared to the reference organism under the same conditions.

“Non-naturally occurring” in the present disclosure denotes the fact that the relevant modification, which may be any modification described herein, for example a modification that increase the flux to 3-hydroxybutanal or downstream product such as 1,3-BDO, is introduced to a reference organism by human intervention.

As explained above, a microbial organism of the invention preferably includes one or more of the following modifications within its genome:

- (i) a modification which increases the flux of a feedstock described herein to acetaldehyde, or makes acetaldehyde more available to the aldolase,
- (ii) a modification which increases the flux of a feedstock described herein to 3-hydroxybutanal or a downstream product thereof.

For example the invention embraces a modification which confers on the microorganism the capability to convert a feedstock described herein to 1,3-BDO, wherein the microorganism lacks the ability to carry out that conversion in the absence of said modification. It also embraces a modification which increases the flux of a feedstock described herein to 1,3-BDO, in a microorganism where that flux is initially very low or negligible.

The modification typically relates to an aldolase enzyme as described herein, as well as a pathway providing acetaldehyde to it.

Preferably said enzyme is deoxyribose phosphate aldolase, EC 4.1.2.4 (“DERA”) or a variant thereof, or other enzyme sharing the ability to accept an aldehyde

(acetaldehyde) as both the acceptor and donor in an aldol coupling. Most preferably, said enzyme is a DERA variant, as described and defined herein.

5 It will be appreciated that “genetic modification” can include more than one modification of the genome of the microbial organism in question.

Microbial organisms of the present invention may include any of the following genetic modifications in respect of the aldolase:

- (i) introduction of at least one heterologous gene encoding the enzyme;
- 10 (ii) up-regulation of at least one endogenous gene encoding the enzyme.

A preferred embodiment is a microbial organism wherein said modification is introduction of a heterologous nucleic acid encoding the enzyme.

15 Optionally the heterologous gene encoding the enzyme may encode a fusion protein encoding also one or more other enzymes present in a 3-hydroxybutanal pathway – for example those involved in the provision of the aldolase substrate acetaldehyde.

20 Optionally the heterologous gene encoding the enzyme may encode a fusion protein encoding also one or more other enzymes present in a downstream product pathway – for example those involved in the conversion of 3-hydroxybutanal to another product or intermediate.

25 The term “3-hydroxybutanal pathway” or “3-hydroxybutanal synthetic pathway” in the present context refers to a series of enzymatically catalysed reactions occurring in a cell which convert one or more principle chemical starting materials or substrates (feedstocks) to central metabolic intermediates comprising one or more of: pyruvate or acetyl CoA which are in turn converted to the common pathway
30 intermediate acetaldehyde which is condensed to form 3-hydroxybutanal. A “3-hydroxybutanal (synthetic) pathway” may also include an activity involved in the conversion of 3-hydroxybutanal directly or indirectly to a downstream product derived from 3-hydroxybutanal, such as 1,3-BDO.

The term “3-hydroxybutanal pathway” enzyme should be construed accordingly i.e. an enzyme providing an activity in a “3-hydroxybutanal pathway”.

5 Thus an example of a “3-hydroxybutanal pathway” is a “1,3-BDO pathway” in which a series of enzymatically catalysed reactions occurring in a cell which convert one or more principle chemical starting materials or substrates (feedstocks) to 1,3-BDO via central metabolic intermediates comprising one or more of: pyruvate or acetyl CoA which are in turn converted to the common pathway intermediate acetaldehyde which is condensed to form the 1,3-BDO precursor, 3-hydroxybutanal, which is in
10 turn converted to 1,3-BDO.

The conversion of central metabolic intermediates to the common pathway intermediate acetaldehyde may require 1 or more steps (e.g. 2, 3, 4 steps).

15 The invention embraces the introduction of all enzymes relevant to the 3-hydroxybutanal (e.g. 1,3-BDO) pathway, including those relating to early substrate utilisation and generation of the central metabolic intermediates themselves, as well as those involved in conversion of the central metabolic intermediates to the common intermediate.

20 Thus in addition to a modification relating to an aldolase, microbial organisms may include one or more other modifications within its genome:

For example said microbial organism may comprise two, three, four, five, six, seven,
25 eight, nine, ten or more exogenous nucleic acids, each encoding a 3-hydroxybutanal (e.g. 1,3-BDO) pathway enzyme.

As explained in more detail below, the invention also embraces the knockout or other impairment of enzyme activities which would otherwise direct flux away from
30 the pathway of choice e.g. direct flux of acetaldehyde away from the aldolase.

In one aspect the invention provides, *inter alia*, a non-naturally occurring microorganism that through genetic engineering gains the ability to produce 1,3-BDO or other downstream product derived from 3-hydroxybutanal from acetyl-coA,

or gains the ability to produce an increased flux of 1,3-BDO or other downstream product derived from 3-hydroxybutanal from acetyl-coA, such that the 1,3-BDO or other downstream product accumulates and can be recovered or further converted enzymatically or chemically without recovery

5

As explained in more detail hereinafter, acetyl CoA may optionally be utilised via acetate. By way of non-limiting example, acetate can then be converted to acetaldehyde via carboxylic acid reductase activity, for example, EC 1.2.7.5 or EC. 1.2.99.6, ATP or ferredoxin driven or EC 1.2.1.30 or EC 1.2.1.3. Or can be converted to acetyl CoA via EC 6.2.1.1 or EC 2.8.3.8 and subsequently converted to acetaldehyde via EC 1.2.1.10.

10

Alternatively acetyl CoA may be utilised via pyruvate (see below, using enzymes such as EC 1.2.7.1 and EC 4.1.1.1) or via direct synthesis of acetaldehyde from acetyl CoA using an aldehyde dehydrogenase (acylating), for example, acetaldehyde dehydrogenase EC 1.2.1.10.

15

In another aspect the invention provides, *inter alia*, a non-naturally occurring microorganism that through genetic engineering gains the ability to produce 1,3-BDO or other downstream product derived from 3-hydroxybutanal from pyruvate, or gains the ability to produce an increased flux of 1,3-BDO or other downstream product derived from 3-hydroxybutanal from pyruvate, such that the 1,3-BDO or other downstream product derived from 3-hydroxybutanal accumulates and can be recovered or further converted enzymatically or chemically without recovery.

20

As explained in more detail hereinafter, pyruvate can be converted to acetaldehyde via acetyl CoA using enzymes such as EC 1.2.7.1 or EC 1.2.1.51 or EC 1.2.4.1 and EC 1.2.1.10. Alternatively, pyruvate can be converted to acetaldehyde, directly via pyruvate decarboxylase (EC 4.1.1.1).

25

It will be appreciated that although "pyruvate" may be referred to herein, depending on the pH and other conditions, it may likewise be present as pyruvic acid, and therefore all these descriptors are used interchangeably, unless context demands

30

otherwise. This applies *mutatis mutandis* to other salts or acids described herein – e.g. acetic acid etc.

Also provided is a process or method for producing a microbial organism according to the invention, which comprises making a genetic modification as described herein.

The invention further provides a method for increasing the flux of 1,3-BDO or other downstream product derived from 3-hydroxybutanal produced by a microbial organism, which method comprises introducing one or more of the genetic modifications described herein into its genome.

Thus in some embodiments the present invention relates, amongst other things, to the generation of microorganisms that are effective at producing 1,3-butanediol from alternative substrates to traditional petroleum-based products.

Methods of producing such a microorganism will typically comprise the step expressing, or causing or allowing the expression of, a heterologous nucleic acid (for example, encoding at least an aldolase as described herein) within the host, following an earlier step of introducing the nucleic acid into the host or an ancestor of either. Suitable heterologous nucleic acids are discussed hereinafter. In other embodiments the methods may include the step of up-regulating native enzymes using genetic engineering and/or repressing enzymes to reduce flux to competing pathways.

Since the central intermediates acetyl CoA and pyruvate are present in all microbial systems, the actual choice of microbe utilised in the present invention will generally be based on the choice of feedstock or energy source which it is desired to use, along with the amenability of the microbe to genetic modification or introduction of a 1,3-BDO (or other downstream product derived from 3-hydroxybutanal) pathway. Preferred processes disclosed herein involve sustainable manufacturing practices that utilise renewable feedstocks, though other feedstocks which may provide cost or environmental benefits compared to traditional petroleum products may also be used, for example natural gas derived methanol.

For example the processes disclosed herein may utilise feedstocks such as syngas, CO₂, CO, and H₂, methane and methanol (shale gas or biomass/ waste derived) to reduce energy intensity and cost and lower greenhouse gas emissions. Other feedstocks are discussed elsewhere herein.

Syngas is a mixture of primarily H₂ and CO that can be obtained via gasification of any organic feedstock, such as coal, coal oil, natural gas, biomass, or waste organic matter.

It will be appreciated that – unless context demands otherwise- where the term “syngas” is used, the embodiments of the invention will apply *mutatis mutandis* to other mixtures of carbon dioxide, carbon monoxide and/or hydrogen, and other substrates such as methane and methanol.

Thus the present invention preferably utilises microorganisms capable of utilizing syngas or other gaseous carbon sources (CO₂, CO) with or without methanol, methane or sugar co-utilisation or by use of methanol, methane or sugars directly as sole feedstocks. Or waste streams containing acetate. Photosynthetic organisms (e.g. algae) capable of using sunlight as an energy source are also expressly included.

In another aspect of the invention there is disclosed a method for producing 1,3-BDO or other downstream product derived from 3-hydroxybutanal that includes culturing the aforementioned non-naturally occurring microbial organisms under conditions and for a sufficient period of time to produce 1,3-BDO or other downstream product derived from 3-hydroxybutanal.

The term “cultured” or “culturing” on a feedstock as used herein is being used in a general sense to mean that the microbial organism utilises the feedstock in question for the production of the relevant product, and should not be taken to imply that the biomass of the microbial organism actually increases during the process.

In another aspect of the invention there is provided a process for producing 1,3-BDO or other downstream product derived from 3-hydroxybutanal, which process comprises culturing a microbial organism of the invention on a reaction feedstock as described herein so that it metabolises the feedstock to produce 1,3-BDO or other downstream product derived from 3-hydroxybutanal from central metabolic intermediates.

In some embodiments the microbe may be cultured in the presence of an additional energy source e.g. a carbohydrate such as a hexose, or sunlight.

The processes of the invention may further comprise recovering some or all of the 1,3-BDO or other downstream product derived from 3-hydroxybutanal e.g. by one or more of electrodialysis, solvent extraction, distillation, or evaporation. However more preferably 1,3-BDO or other downstream product derived from 3-hydroxybutanal may be converted chemically or enzymatically *in situ* to a downstream product or products, which may in turn be recovered by similar means.

The processes of the invention may further comprise converting the 1,3-BDO or other downstream product derived from 3-hydroxybutanal into a pharmaceutical, cosmetic, food, feed or chemical product, which may optionally be an unsaturated alcohol, alkene, carboxylic acid, ether, ester, or ketone e.g. methylethyl ketone, 1-butanol, 2-butanol, butadiene, isoprene and so on.

Thus in various aspects the invention provides non-naturally occurring microorganisms comprising one or more heterologous proteins conferring to the microorganism the capability to convert central intermediates to 1,3-BDO or other downstream product derived from 3-hydroxybutanal as described herein. Alternatively the heterologous protein may be directed at increasing the flux of reaction feedstocks such as syngas or other substrates described herein to 1,3-BDO or other downstream product derived from 3-hydroxybutanal, in a microorganism where that flux is initially very low or negligible under relevant industrial culture conditions.

In various aspects the invention provides a non-naturally occurring microorganism which has been modified to up-regulate (increase expression of) a native protein, or to modify the localisation of a native protein, or to modify the activity or specificity of a native protein, thereby conferring to the microorganism the capability to convert
5 syngas or other substrates described herein to 1,3-BDO or other downstream product derived from 3-hydroxybutanal, wherein the microorganism lacks the ability to carry out that conversion in the absence of said modification. Alternatively the heterologous protein may be directed at increasing the flux of metabolic intermediates from the feedstock being utilised in a microorganism where that flux is
10 initially very low or negligible.

Thus the invention provides a non-naturally occurring microbial organism having a genetically modified 3-hydroxybutanal or 1,3-BDO biosynthetic pathway and the competence to metabolise syngas or other feedstocks or energy source described
15 herein to produce 1,3-BDO or other downstream product derived from 3-hydroxybutanal.

Some aspects and embodiments of the invention will now be discussed in more detail:

20 *Metabolic pathways leading to central metabolites, and increased production or availability of acetaldehyde*

Pyruvate and acetyl CoA are products of a considerable range of different central
25 metabolic pathways for assimilation of carbon. They are converted to important cellular building blocks essential for life. In the present invention they are utilised within a 1,3-BDO biosynthetic pathway, which pathway is at least in part the result of genetic engineering of the microbial organism.

30 As explained above, typically an organism is selected according to the feedstock it is desired to utilise. The organism may be selected to have in its genome a particular metabolic pathway leading to acetyl CoA and/or pyruvate. Example metabolic pathways include:

- the Wood-Ljungdahl pathway (Figure 5)
 - the ribulose monophosphate (RuMP) pathway (Figure 4)
 - the reverse TCA cycle (Figure 6)
 - the serine cycle (Figure 7)
- 5
- glycolysis and the pentose phosphate pathway
 - the Calvin cycle via 3-phosphoglycerate (Figure 9)
 - other pathways such as the 3-hydroxypropionate/4-hydroxybutyrate cycle

10 Pathways generating these central metabolic intermediates are well reported and well understood by those skilled in the art. The Wood-Ljungdahl pathway, reverse TCA cycle, the serine cycle, the RuMP pathway and the Calvin cycle, are examples of C1 (gas and liquid) fixation pathways. In some case these pathways can be used alongside glycolysis.

15 For example, the Wood-Ljungdahl pathway is important for redox balancing by using the reducing equivalents generated from glycolysis and pyruvate decarboxylation to acetyl CoA, to fix the released 2 CO₂ into a further molecule of acetyl CoA.

20 The serine or the RuMP pathways are generally used by methanotropic and methylotrophic organisms for assimilation of C1 feedstocks such as methanol, methane and CO₂. These pathways are well described and well known in the art. The product of the RuMP pathway is pyruvate which would normally be converted primarily to biomass.

25 Intercepting pyruvate via, for example, decarboxylation to acetaldehyde would redirect flux towards 1,3-butanediol synthesis using the described invention. The Calvin cycle is used by photosynthetic organisms such as algae for assimilation of CO₂ using light energy.

30 The serine cycle primarily produces acetyl CoA which normally enters the ethylmalony CoA pathway for synthesis of C4 building blocks for biomass synthesis. If acetyl-CoA is required as the biosynthetic precursor of membrane fatty acids or the storage compound poly 3- hydroxybutyrate for example, then the EMC pathway

is not required for oxidation of acetyl-CoA, (Anthony, C. 2011. Science Progress, 94, 109). Hence, acetyl CoA can be tapped off to other more useful compounds such as 1,3-butanediol via conversion of acetyl CoA to acetaldehyde using a pathway described herein.

5

The same principle can be used for all metabolic pathways producing pyruvate or acetyl CoA as the product. Another example would be the reverse TCA cycle which again produces acetyl CoA from fixation of two molecules of CO₂. All these central metabolic pathways producing acetyl CoA or pyruvate are well understood in the art.

10

A diverse range of microorganisms with known pathways capable of utilising syngas, or gases such as CO, CO₂ and H₂.

15

Microorganisms which are CO utilisers are termed "carboxydophilic microorganisms". Such organisms can be aerobes and anaerobes.

20

Anaerobic examples of these microorganisms fall into 3 main groups: those producing mainly acids (e.g. acetic acid, termed "acetogens"), those producing mainly methane and those producing mainly hydrogen. The first group is of particular interest in the present invention:

25

Carboxydophilic acetogens are acetogenic microorganisms capable of utilising the syngas components CO and H₂ via the Wood-Ljungdahl pathway (Figure 2) producing the key intermediate acetyl CoA.

30

The Wood-Ljungdahl pathway is well known in the art (see Figure 5) and can be separated into two branches: the methyl branch (reductive branch) and the carbonyl branch. The methyl branch converts syngas (CO or CO₂) to methyl-tetrahydrofolate (methyl-THF) whereas the carbonyl branch supplies a molecule of CO which along with methyl-THF is converted to acetyl-CoA.

"Acetogens" as used herein refers to anaerobic organisms able to reduce CO₂/CO to acetate via this pathway. Acetogens can grow on a variety of different substrates

such as, for example, hexoses [glucose, fructose and xylose], C2 and C1 compounds [gas and liquid] including methanol (see Figure 2), CO₂/H₂ and CO gases. Acetogens are also known to utilise acetate directly.

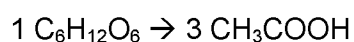
- 5 Over one hundred acetogenic species, representing twenty-two genera, have been isolated so far from various habitats such as sediments, sludge, soils and the intestinal tracts of many animals, including termites and humans. From the twenty-two genera, *Clostridium* and *Acetobacterium* harbour the most known acetogenic species (Drake et al., *Ann. N. Y. Acad. Sci* 1125:100-108 (2008)). However it
10 should be noted that the Wood-Ljungdahl pathway is actually not restricted to acetogens and is present in many anaerobic bacteria as a means of fixing CO₂.

While defined as strict anaerobes some acetogens have been isolated from different aerobic or microaerobic environments (Annals New York Academy of Sci. 2008,
15 1125, 100). It has been demonstrated that these types of bacteria are equipped with an assortment of oxidative stress enzymes and that certain acetogens can even reduce oxygen by different means (Tirado-Acevedo, O. PhD thesis 2010). *Clostridium ljungdahlii* has been shown to tolerate up to 8% oxygen in the gas phase (*ibid*).

20 Acetogens are becoming a significant focus for the biotech industry, as important bulk chemicals can potentially be produced from autotrophic growth at the expense of CO₂ via syngas fermentation or via coupling with methanol, which can serve as a source of carbon and energy in the absence of syngas. in the presence of a more
25 oxidised substrate such as CO₂.

Acetogens can utilise hexoses (e.g. glucose, fructose and xylose) and other sugars as substrates.

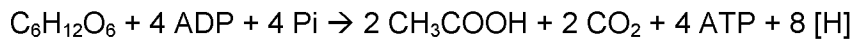
- 30 In many acetogens acetate is the primary product of hexose consumption:



The pathway of hexose consumption starts with their oxidation via the Embden-Meyerhof-Parnas pathway to pyruvate, which is then oxidized by

pyruvate:ferredoxin oxidoreductase to acetyl-CoA, reduced ferredoxin, and CO₂. The acetyl-CoA is then converted to acetate via acetyl phosphate.

5 This oxidative branch of the pathway to pyruvate is coupled to the synthesis of 4 mol of ATP by SLP:



10 Second, the reducing equivalents 8[H], gained during glycolysis and pyruvate:ferredoxin- oxidoreductase are reoxidized by reducing the two mol of CO₂ to another mol of acetate via the Wood-Ljungdahl pathway

15 $2 \text{ CO}_2 + 8 [\text{H}] + n\text{ADP} + n\text{Pi} \rightarrow \text{CH}_3\text{COOH} + n\text{ATP}$. The exact amount of net ATP generated other than from acetyl CoA to acetate via SLP, is not yet known.

During growth on sugars or other organic substrates, CO₂ can be formally regarded as an electron sink and per se, there is no need that the Wood-Ljungdahl pathway is coupled to energy conservation. Energy is gained during glycolysis, and the redox balance is maintained by operation of the Wood-Ljungdahl pathway (Figure 2).
20 However, it is important to note that the Wood-Ljungdahl pathway also enables growth on CO₂ and hydrogen or carbon monoxide or on methanol and CO₂ or formate. Particularly in the case of the gaseous substrates it must be coupled to net ATP synthesis. The overall free energy change of the reaction ($\Delta G_o' = -95 \text{ kJ/mol}$) could allow for the synthesis of 1 to 2 mol of ATP.

25 The energetics of this pathway are finely balanced. One mole of ATP is produced by SLP in the acetate kinase reaction via acetyl phosphate from acetyl CoA, but one mole of ATP is consumed in the formyl-H₄F synthetase reaction. Therefore, the net ATP gain by SLP is zero, and ion gradient-driven phosphorylation is known to occur
30 as well contributing to the generation of more ATP.

The above discussion clearly highlights under most circumstances the important role of SLP level phosphorylation ATP synthesis via conversion of acetyl CoA to acetate, to balance the Wood-Ljungdahl pathway and indeed considerable flux to

acetate is seen in growing acetogens. Driving flux away from acetate synthesis to any more useful chemical in high yield, impacts to varying degrees on an acetogen's energy balance as acetate synthesis is a key ATP synthesis step. Even when growing on methanol and CO₂ where the energetics may be more favourable, in many cases preventing synthesis of acetate may still have some impact. Further, the metabolism of many acetogens has evolved primarily for the efficient synthesis of either biomass or acetate from the central intermediate acetyl CoA. Hence, industrial products derived directly from acetyl CoA or acetate may be preferred for process development using acetogens.

Employing an aldolase capable of accepting the product acetaldehyde as a substrate (both as acceptor and donor) permits the efficient generation of useful chemical products from acetate which acetogens naturally accumulate in high yield and high concentration.

Routes for utilising acetyl-CoA derived acetate for generation of acetaldehyde are explained in more detail below. A preferred route of acetaldehyde generation would be based on utilisation of ferredoxin driven aldehyde ferredoxin oxidoreductase. This enzymatic conversion does not require ATP so may be particularly important for bacteria growing on C1 gases such as CO, CO₂/H₂ or syngas for the energetics reason described above. Since the natural SLP step involving the conversion of acetyl CoA to acetate can be conserved, the energetics of the Wood-Ljungdahl pathway would be unchanged. However other routes for acetate reduction, including those which require ATP, may be utilised where the energetics of the system permit. Overall, the ability to create a C4 compound such as 3-hydroxybutanal from a C2 compound such as acetaldehyde without the normal requirement for energy, is highly desirable when efficiently exploiting microbial metabolism for industrial chemical synthesis.

As noted above, organisms capable of utilizing CO and syngas also generally have the capability of utilizing CO₂ and CO₂/H₂ mixtures through the same basic set of enzymes and transformations encompassed by the Wood-Ljungdahl pathway. H₂-dependent conversion of CO₂ to acetate by microorganisms was recognized long

before it was revealed that CO also could be used by the same organisms and that the same pathways were involved.

5 Thus in one embodiment the invention provides a non-naturally occurring microorganism having the Wood-Ljungdahl pathway and the capability of utilising syngas naturally and that through genetic engineering gains the ability to produce 1,3-BDO or gains the ability to produce an increased flux of 1,3-BDO.

10 Many acetogens have been shown to grow in the presence methanol if another more oxidised co substrate such as CO₂, is present. The provision of reducing equivalents is crucial to drive this process to either natural products such as acetate, or unnatural acetogen products such as 1,3-butanediol.

15 Hydrogen is a major source of reducing equivalents, but equally for example, dissimilation of methanol can also generate reducing equivalents (6[H]) and ATP energy. Further, as shown in Figure 2, methanol utilisation confers an energetic advantage because it provides a preformed methyl group for synthesis of acetyl CoA eliminating the need for an ATP for conversion of formate to formyl-THF catalysed by formyl-THF synthetase (Bainotti, A.E and Nishio, N. 2000. J. Appl. Microbiol, 88, 191). The specific energetic requirements of the transfer of the methyl group from methanol is not clearly understood. However, catalytic amounts of ATP are cited in the literature and Stupperich, E and Konle, R, 1993, Appl. Environ. Microbiol. 59, 3110, describe a requirement of approximately 0.3 ATP per methyl group transfer. Hence, a saving of approximately 0.7ATP may be achieved when a preformed methyl group is provided instead of CO₂. If only a catalytic amount of ATP is required, the energy benefit may be greater. However, as highlighted, a clear understanding of methanol utilising energetics is not yet established in the art.

30 In the case of CO₂, additional sources include, but are not limited to, production of CO₂ as a byproduct in ammonia and hydrogen plants, where methane is converted to CO₂; combustion of wood and fossil fuels; production of CO₂ as a byproduct of fermentation of sugar in the brewing of beer, whisky and other alcoholic beverages, or other fermentative processes; thermal decomposition of limestone, CaCO₃, in the manufacture of lime, CaO; production of CO₂ as byproduct of sodium phosphate

manufacture; and directly from natural carbon dioxide springs, where it is produced by the action of acidified water on limestone or dolomite.

The ability of acetogens to utilise methanol requires specific methyltransferases. Where such acetogen methyltransferases are not naturally present, an acetogen can be engineered with heterologous methyltransferases and other associated proteins to allow it to utilise methanol as well as the other feedstocks discussed above.

Examples of enzymes required to give an acetogen the ability to grow on methanol include:

methanol methyltransferase (MtaB)

Corrinoid protein (MtaC),

Methyltetrahydrofolate:corrinoid protein methyltransferase (MtaA),

Methyltetrahydrofolate:corrinoid protein methyltransferase (AcsE),

Corrinoid iron-sulfur protein (AcsD)

The methylotrophs and methanotrophs also naturally grow on methanol and/or methane, utilising for example, the RuMP (Figure 4) or serine cycle (Figure 7) pathways for C1 metabolism. As for the Wood-Ljungdahl pathway, both the serine cycle of C1 metabolism and the RuMP pathway are well described and well understood in the art.

Thus in one embodiment the invention provides a non-naturally occurring microorganism having the RUMP or serine cycle pathway encoded in its genome and the capability of utilising methanol or methane naturally and that through genetic engineering gains the ability to produce 1,3-BDO or gains the ability to produce an increased flux of 1,3-BDO.

Photosynthetic organisms such as microalgae or cyanobacteria are autotrophs or heterotrophs able to utilise sunlight (light energy) for CO₂ fixation via the Calvin cycle. A product is glyceraldehyde-3-phosphate which can be converted to sugar or to pyruvate and acetyl CoA.

Many diverse microorganisms are heterotrophic and can utilise sugars as a source of carbon and energy via glycolytic pathways such as the Entner doudoroff pathway, Embden meyerhof pathway or pentose phosphate pathway. All sugar assimilation pathways are well understood in the art. A product of these pathways is pyruvate which may be converted to acetyl CoA for example, for entry into the TCA cycle for supply of cellular building blocks such as malate, oxaloacetate, succinate or fumarate.

As discussed above in relation to acetogens, many organisms not normally considered heterotrophs (e.g. acetogens or methylotrophs) are also capable of heterotrophic growth if sugars are supplied.

Photoheterotrophs are heterotrophic phototrophs - that is, they are organisms that use light for energy, but cannot use carbon dioxide as their sole carbon source, instead using carbohydrates, fatty acids, and alcohols and so on. Examples of photoheterotrophic organisms include purple non-sulfur bacteria, green non-sulfur bacteria, and heliobacteria. Thus feedstocks when utilised in the present invention may where desired may include or indeed consist of sugars as all or part of the source of carbon and/or energy.

Enzymes suitable for converting the metabolic intermediates to acetaldehyde are discussed in more detail in the Examples below, and in Figure 3.

Route 1 proceeds from acetyl CoA through acetate, (a natural product of acetogenic microorganisms), to acetaldehyde via carboxylic acid reductase activity (Activity A), using for example, EC 1.2.7.5 or EC. 1.2.99.6, ATP or ferredoxin driven or EC 1.2.1.30 or EC 1.2.1.3.

Route 2 involves direct synthesis of acetaldehyde from acetyl CoA (Activity B) using an aldehyde dehydrogenase (acylating), for example, acetaldehyde dehydrogenase EC 1.2.1.10.

Route 3 involves the conversion of pyruvate to acetaldehyde via acetyl CoA (Activity C and B) using enzymes such as EC 1.2.7.1 or EC 1.2.1.51 or EC 1.2.4.1 and EC 1.2.1.10.

- 5 Route 4 involves the conversion of pyruvate to acetaldehyde, (Activity D) directly via for example, pyruvate decarboxylase (EC 4.1.1.1).

Route 5 involves the conversion of acetyl CoA to acetaldehyde via pyruvate (Activity E and D) using enzymes such as EC 1.2.7.1 and EC 4.1.1.1.

10

Route 6 involves the conversion of acetate to acetaldehyde via acetyl CoA (Activity F and B) using enzymes such as EC 6.2.1.1 or EC 2.8.3.8 and EC 1.2.1.10.

It will be appreciated that in any given organism of the invention, more than one of these routes may be utilised.

15

The enzyme aldehyde ferredoxin oxidoreductase (EC 1.2.7.5) can be found in many acetogens and other organisms and has been shown to be capable of reducing unactivated carboxylic acids to the corresponding aldehyde (White, H *et al.* Biol. Chem Hoppe Seler 1991, 372 (11) 999; White, H and Simon, H. Arch. Microbiol, 1992, 158, 81; Fraisse. L and Simon, H. Arch. Microbiol. 1988, 150,381; Basen et. al. 2014. PNAS, 111 (49), 17618). Kopke, M. et al. PNAS,2010, 107, 15305, describes genes capable of reduction of acetate to acetaldehyde in the acetogen *Clostridium ljungdahlii*.

20

- 25 Alternative means of generating acetaldehyde in acetogens for supply to an aldolase which could be used separately or in conjunction with reduction of acetate are as follows:

30

Acetaldehyde dehydrogenase (EC 1.2.1.10) or any aldehyde dehydrogenase capable of converting acetyl CoA to acetaldehyde directly may be used. As explained above, in this case the SLP step from conversion of acetyl CoA to acetate would be lost. Some compensation for this loss could be achieved from ion gradient phosphorylation from the Wood-Ljungdahl pathway when growing on gases such as CO, CO₂/H₂ or syngas. ATP may also be synthesised via NAD(P) reduction coupled

to reduced ferredoxin, but growth on methanol and CO₂ or another more oxidised co substrate may be most suited due to the potential more favourable energetics and potential for supply of reducing equivalents and ATP from methanol dissimilation. Another example of an alternative would involve a carboxylic acid reductase (CAR) enzyme (EC 1.2.99.6). These enzymes catalyse reduction of carboxylic acids to the corresponding aldehyde via activation with ATP. The energetics of this route would be similar to that described for acetaldehyde synthesis from acetyl CoA via acetaldehyde dehydrogenase. The use of a carboxylic acid reductase in a 1,3-butanediol pathway for synthesis of a corresponding aldehyde is described in US8268607, albeit that the pathway for synthesis is different.

The other enzymes utilised in these routes are discussed in more detail in Examples hereinafter.

15 *Down regulation of native enzymes*

The modified organisms of the invention may be engineered to target (down-regulate, knockout or inhibit) the activity of enzymes which may otherwise direct the flux of intermediates in the 3-hydroxybutanal pathway to other products or biomass. Methods of targeting genes in this way are known in the art, and also discussed below.

In acetogens, if the bioenergetics allow loss of ATP synthesis from acetyl CoA conversion to acetate, acetate accumulation can be reduced by targeting of phosphotransacetylase (pta) or acetate kinase (ack) genes. This can enhance the level of acetyl CoA, which can be utilised directly or via pyruvate. Thus where utilising Route 2, it may be desired to target EC 2.3.1.8 (phosphotransacetylase) or EC 2.7.2.1 (acetate kinase) or both.

Where utilising Route 3 or 4 or 5 or 6, it may be desired to target LDH activity (EC 1.1.1.27 or 1.1.1.37; the latter is malate dehydrogenase but is known to accept pyruvate as a substrate) to prevent loss of pyruvate to lactate, and \or (where appropriate) to target pyruvate formate lyase (EC 2.3.1.54). In all cases the purpose is to prevent or minimise loss of pyruvate to other products.

Alternatively or additionally it may be desired to increase the availability of acetaldehyde to the aldolase by down-regulation or inactivation of an endogenous enzyme (e.g. an alcohol dehydrogenase) which utilises acetaldehyde as a substrate for some other purpose e.g. production of ethanol. Increasing the availability to the aldolase of the acetaldehyde increases production of the 3-hydroxybutanal from the aldolase.

For way of further example, it may particularly be desired to target any alcohol dehydrogenase with a preference for reduction of acetaldehyde to ethanol relative to reduction of 3-hydroxybutanal to 1,3-BDO. These acetaldehyde to ethanol enzymes are generally classified in EC 1.1.1.1.

Aldolase enzymes

Acetaldehyde derived from acetyl CoA or pyruvate is used to supply the substrate for a DERA type aldolase (deoxyribose phosphate aldolase, EC 4.1.2.4, DERA or 'DERA like' enzyme) capable of the coupling of two molecules of acetaldehyde to form 3-hydroxybutanal (via "Reaction G" in Figure 3). Example enzymes are given in Table 6.

The natural deoxyribose phosphate aldolase (DERA) reaction is:



The phosphorylated substrate is preferred but most wild type enzymes will catalyse the coupling of two non-phosphorylated aldehyde molecules. Primarily acetaldehyde and another aldehyde. An example of a DERA which accepts phosphorylated and non-phosphorylated substrates with approximately equal preference is described by Zhong-Yu, Y. et al. (J. Ind. Microbiol Biotech. 2013, 40, 29).

Evolution and development of DERA for synthesis of key pharmaceutical intermediates has received considerable focus over the past 25 years (DeSantis, G et al. Bioorg & Medicinal Chem. 2003, 11, 43). However, it has not previously been

suggested to integrate a DERA type enzyme into an unnatural pathway for synthesis of 1,3-butanediol or other downstream products.

Under certain conditions many DERA enzymes are also capable of catalysing a tandem sequential coupling of three acetaldehydes, which will preferably be avoided in the present context. Fortunately the undesired sequential reaction involving two aldol couplings of acetaldehyde (Figure 8) is generally not the dominant reaction for DERA catalysed aldol couplings. The desired monoaldol product (3-hydroxybutanal) accumulates and high levels of wild type DERA enzymes are required to drive the reaction to accept a second acetaldehyde addition (Green Chemistry in the Pharmaceutical industry, 2010, John Wiley and sons). Hence the interception of the monoaldol (3-hydroxybutanal) and direction to other products is feasible, for example, by reduction to 1,3-butanediol. Furthermore, DERA enzymes are known (and can be generated) which are very inefficient at catalysing the second aldol coupling, thereby producing just the monoaldol 3-hydroxybutanal (see e.g. US7402710) which describes a DERA enzyme from the organism *Pyrobaculum aerophilum*). US7402710 describes synthesis of C4 hydroxyaldehydes such as 3-hydroxybutanal but not in the context of a metabolic pathway or where acetaldehyde is supplied by an enzyme in the pathway. In US7402710, acetaldehyde is added exogenously to an isolated DERA enzyme preparation

Many DERAs are known to be inactivated at aldehyde concentrations above 100mM and may be sensitive to concentrations below this concentration, for both acetaldehyde and 3-hydroxybutanal and this has been a limitation for application of DERA for synthesis of statin intermediates via sequential coupling of chloroacetaldehyde and two molecules of acetaldehyde (Green Chemistry in the Pharmaceutical industry, 2010, John Wiley and sons). In the context of the current invention DERA is used as part of an unnatural pathway for synthesis of 1,3-butanediol and other valuable chemicals where the substrate acetaldehyde is provided via de novo synthesis from a preceding pathway enzyme. Thus aldehyde concentrations in the processes of the invention will never approach 100mM, and sensitivity to this concentration of acetaldehyde is therefore immaterial. Both acetaldehyde and 3-hydroxybutanal are intermediates in the pathway and accumulation of these intermediates will be avoided by ensuring adequate activity of

pathway enzymes to maximise carbon flux to 1,3-butanediol or other target chemicals.

5 Wild type DERA aldolase has been overexpressed in *E.coli* and run as a high intensity process for synthesis of chiral lactol intermediates for the statin pharmaceuticals (Oslaj, M. *et al* Plos one, 8 (5), 1). The process involves a fed batch approach involving the coupling of a 2-substituted acetaldehyde and acetaldehyde to the corresponding lactols in a tandem sequential synthesis. Although this process was run as a whole cell system, the reactants were fed to the
10 cells and were not generated *in situ*. Furthermore there were no modifications made which would have enhanced production or availability of endogenous acetaldehyde from central metabolic intermediates.

The processes of the present invention do not utilise batch feeding of the microbial
15 organisms with 2-substituted acetaldehyde and/or acetaldehyde.

Naturally-occurring DERA enzymes are not optimised for industrial applications of coupling or condensation of two molecules of acetaldehyde. Therefore, there is a need for catalytic improvement to increase the activity of these enzymes and to fully
20 realise their industrial potential.

DERA Variants

Over the past 20 years, 2-deoxyribose phosphate aldolase, EC 4.1.2.4 ("DERA")
25 has been a focus for evolution towards improved synthesis of side chains of statin pharmaceuticals in intensified reaction systems. The objective has been primarily two-fold; improving the stability of the enzyme towards high (molar) concentrations of the aldehyde reactants and promoting the coupling of acetaldehyde with two other aldehydes through a double aldol reaction to form a C6 (or greater) aldol
30 product. As far as can be determined, there has been no description in the art regarding evolution of DERA towards the specific improvement for the coupling of two molecules of acetaldehyde to form 3-hydroxybutanal, where the double aldol is undesired, or where the target is to utilise the improved enzyme within an unnatural metabolic pathway.

An aspect of the present invention relates to DERA variants, or 'DERA type' enzyme variants, having improved activity for catalysing the coupling of two molecules of acetaldehyde to form 3-hydroxybutanal which is then released from the active site.

5

In addition, the invention relates to engineering a DERA enzyme, with a goal of converting two molecules of acetaldehyde to crotonaldehyde via an aldol condensation. This is achieved by the dehydration of 3-hydroxybutanal within the enzyme active site via β -elimination of a molecule of water such that crotonaldehyde rather than 3-hydroxybutanal is released. Furthermore, the engineered aldolase may release 3-hydroxybutanal from the active site and then rebind it to carry out the dehydration step to crotonaldehyde. Further, the engineered aldolase may bind 3-hydroxybutanal produced from a separate enzyme and then carry out the dehydration to form crotonaldehyde which is released from the active site.

10

15

As mentioned, DERA enzymes are known to be capable of coupling two molecules of acetaldehyde, but there are no reported aldolase enzymes capable of crotonaldehyde synthesis from acetaldehyde directly, or via binding or rebinding and dehydration of 3-hydroxybutanal.

20

For the purpose of the present invention, the desire is to improve the operation of the enzyme at acetaldehyde concentrations relevant for *in vivo* application.

25

According to one aspect, the present invention provides a modified 2-deoxyribose phosphate aldolase (DERA) enzyme variant, comprising one or more mutations that improve the aldolase catalytic performance for synthesis of 3-hydroxybutanal or crotonaldehyde, relative to the parent DERA enzyme from which the variant originates and which does not comprise such a modification.

30

A 'DERA type' enzyme may be any enzyme capable of the coupling or condensation of acetaldehyde to 3-hydroxybutanal or crotonaldehyde respectively. DERA enzymes are reported as the only type of aldolase capable of utilising aldehydes as both a donor and acceptor in an aldol coupling. For the purpose of this

invention, any enzyme capable of utilising an aldehyde as both donor and acceptor in an aldol coupling is considered a DERA.

5 The parent DERA enzyme may be a wild-type enzyme or it may be a derivative of the wild-type, which has itself been modified, for example including other modifications described herein. The DERA variant that is modified according to this aspect of the invention shows improved aldolase catalytic performance for synthesis of 3-hydroxybutanal or crotonaldehyde compared with the equivalent aldolase catalytic performance of the parent enzyme.

10 In order to achieve a DERA variant according to this aspect of the invention, the present inventors have developed a consensus sequence, which can be used to facilitate appropriate engineering of any available DERA enzyme.

15 It is desirable to be able to locate residues equivalent to those within regions of interest that may influence the catalytic activity within any DERA sequence, or to identify any additional regions of interest that may influence the catalytic activity within any DERA sequence. The present invention provides a unique tool by which this can be achieved.

20 A consensus sequence is the calculated order of most frequent amino acid residues found at each position in a sequence alignment. It can be used to represent in a concise manner the "average" sequence of a population. In this way it serves as a tool against which any other amino acid sequence (e.g. that obtained from the translation of a nucleotide sequence) can be compared in order to identify homology and similarity. This is crucial when identifying residues to target for substitution by mutagenesis of the relevant position in the underlying gene when presented with a polypeptide sequence of unknown origin. The method by which the consensus of DERA amino acid sequences was created for this invention is described in Example 15.

25

30

The new consensus sequence captures regions that are highly conserved in all DERA enzymes known to date. It can be reasonably assumed that this level of conservation is directly related to efficient processing of the natural substrates and

is strongly biased towards non-preferable acceptance of a molecule of acetaldehyde in the acceptor binding site. Acetaldehyde is the natural donor for the aldol coupling with the natural acceptor glyceraldehyde-3-phosphate. Hence, in order to improve the efficiency of DERA enzymes for commercial applications, one important aspect is to improve the ability for acetaldehyde to act as a both an efficient donor (already sufficiently present in wild type enzymes) and an efficient acceptor (required by evolution) in an aldol coupling.

The new consensus sequence is the following sequence, or a variant thereof:

MAKLIDHTLLKPDATDEDIKLCHEAKEYNFASVCVNPRFVPLAAETLEGDVKVCTVI
GFPLGADTTAVKAFETKAAIANGADEIDMVINIGALKAGNEDVVEEDIKAVVEACGV
LLKVIIETCLLTDEEIVKACEISIKAGADFVKSTSTGFSTGGATVEDVRLMRETVGPDV
GVKASGGVRTAEDALAMIEAGATRIGASAGVAIVKSGY (SEQ ID NO: 1)

Variants of the sequences disclosed herein preferably share at least 55%, 56%, 57%, 58%, 59%, 60%, 65%, or 70% identity, most preferably at least about 80%, 90%, 95%, 96%, 97%, 98% or 99% identity. Such variants may be referred to herein as "substantially similar".

According to a preferred embodiment of this aspect of the invention, a DERA variant according to the invention comprises a polypeptide sequence which when aligned to a consensus sequence having at least 80% sequence identity with the sequence of SEQ ID NO: 1 shows an alignment with amino acid residues at 9 or more of 15 defined positions in the consensus sequence.

Preferably, at least one of said one or more mutations are mutations of residues that are at equivalent positions to one or more of the amino acid residues at 9 or more of the 15 defined positions in the consensus sequence. Preferably, residues at equivalent positions to those of the consensus sequence are identified by alignment of the parent DERA polypeptide sequence with the consensus sequence.

Example 16 below provides details of how any query sequence (e.g. the parent DERA sequence) can be aligned with the consensus sequence to identify residues that are at equivalent positions within the two sequences.

5 Preferably, the one or more mutations in the DERA polypeptide sequence are at residues equivalent to positions T8, L10, C35, V57, F60, D86, K147, G151, K176, S178, G179 G180, G199, A200 and/or S201 of the consensus sequence, wherein equivalent positions are determined by alignment to the consensus sequence.

10

Preferably the one or more mutations are not at residues equivalent to positions D86, K147 and K176 of the consensus sequence. These amino acids have been demonstrated to be critical for the catalytic activity. They are essential for forming a Schiff base as a catalytic functionality, which enables binding of the substrate and performing the aldol coupling and/or condensation reactions. Mutations in these catalytic essential amino acids will ultimately lead to inactivation of the enzyme and are therefore not suitable as target for enzyme engineering.

15

20 Preferably the modification is substitution of one amino acid residue for another. The following are examples of preferred modifications to the DERA polypeptide sequence. A DERA variant according to the invention may comprise one or more of these modifications:

25 If the one or more mutations is at a residue equivalent to position 8, the modification introduces a hydrophobic residue.

If the one or more mutations is at a residue equivalent to position 10, the modification introduces a positively or negatively charged residue, or introduces a hydrophobic residue. Preferably, if the DERA variant is used to produce 3-hydroxybutanal then a modification at a residue equivalent to position 10 introduces a hydrophobic residue. Preferably, if the DERA variant is used to produce crotonaldehyde via a dehydration reaction then a modification at a

30

residue equivalent to position 10 introduces a positively or negatively charged residue.

If the one or more mutations is at a residue equivalent to position 35, the modification introduces a hydrophobic residue.

5 If the one or more mutations is at a residue equivalent to position 57, the modification introduces a hydrophobic residue.

If the one or more mutations is at a residue equivalent to position 60, the modification introduces a hydrophobic residue.

10 If the one or more mutations is at a residue equivalent to position 151, the modification introduces a hydrophobic or negatively charged, .

If the one or more mutations is at a residue equivalent to position 178, the modification introduces a positively or negatively charged residue, or introduces a hydrophobic residue. Preferably, if the DERA variant is used to produce 3-hydroxybutanal then a modification at a residue equivalent to position 178
15 introduces a hydrophobic residue. Preferably, if the DERA variant is used to produce crotonaldehyde via a dehydration reaction then a modification at a residue equivalent to position 178 introduces a positively or negatively charged residue.

20 If the one or more mutations is at a residue equivalent to position 179, the modification introduces a negatively charged or a hydrophobic residue.

If the one or more mutations is at a residue equivalent to position 180 the modification introduces a negatively charged, or a hydrophobic residue.

If the one or more mutations is at a residue equivalent to position 199, the modification introduces a positively or negatively charged residue, or
25 hydrophobic residue;

If the one or more mutations is at a residue equivalent to position 200, the modification introduces a negatively charged, or a hydrophobic residue.

If the one or more mutations is at a residue equivalent to position 201, the modification introduces a negatively charged, or a hydrophobic residue..

30

Examples of specific amino acid substitutions that may be made at positions in the DERA enzyme sequence that are equivalent to one or more of the 15 positions in the consensus sequence according to the invention are detailed in Table 8 below.

This is not an exhaustive list; alternative residues may also be selected for modification. 'Hotspots' are regions of interest within DERA sequences for targeted substitutions that improve the synthesis of 3-hydroxybutanal or crotonaldehyde. The meaning of 'hotspots' is explained in more detail below.

5

Table 9 (also below) provides a preferred sub-set of the modifications detailed in Table 8.

10 In an additional or alternative embodiment of this aspect of the invention, the one or more mutations include mutations of amino acid residues that are at positions that are not equivalent to any of said 15 defined positions in the consensus sequence. In one embodiment, none of said mutations are mutations of amino acid residues that are at positions that are equivalent to any of said 15 defined positions in the consensus sequence.

15

In a DERA variant according to the present invention, it is preferred that one or more mutations improve coordination of a substrate carbonyl group. Alternatively or additionally, it is preferred that one or more mutations improve coordination of a substrate methyl group. Alternatively or additionally, it is preferred that one or more mutations reduce the coordination of a substrate phosphate group. Alternatively or additionally, it is preferred that one or more mutations increase the negative charge in the active site. Alternatively or additionally, it is preferred that one or more mutations increase the hydrophobicity in the active site.

20

DERA variants according to this aspect of the invention can be introduced into any microbial organism described herein, and can be used as a substitute for any other aldolase enzymes described herein.

25

30 A separate aspect of this invention provides an isolated polypeptide comprising a polypeptide sequence which when aligned to a consensus sequence having at least 80% sequence identity with the sequence of SEQ ID NO: 1 shows an

alignment with amino acid residues at 9 or more of 15 defined positions in the consensus sequence.

Preferably, said 15 defined positions in the consensus sequence are positions 8,
5 10, 35, 57, 60, 86, 147, 151, 176, 178, 179 180, 199, 200 and 201.

Preferably, the polypeptide sequence comprises one or more of the following residues at positions that are equivalent to one or more of positions 8, 10, 35, 57, 60, 151, 178, 179 180, 199, 200 and/or 201 of the consensus sequence,
10 wherein equivalent positions are determined by alignment to the consensus sequence:

at position 8, there is a hydrophobic residue selected from any of L, I, V, F, or A,

at position 10, there is a H or a negatively charged residue
15 selected from D or E,

at position 35, there is a hydrophobic residue selected from any of L, I, V, F, or A,

at position 57, there is a hydrophobic residue selected from any of L, I, F, or A,

at position 60, there is a hydrophobic residue selected from any of L, I, V, or A,

at position 151, there is a hydrophobic residue selected from any of A, , L, I, V, F, or W, or negatively charged D.,

at position 178, there is a H or there is a negatively charged
25 residue selected from either of D and E,

at position 179, there is a negatively charged residue, D or a hydrophobic residue selected from any of A, L, I, V, F, or W.

at position 180, there is a negatively charged residue, D or a hydrophobic residue selected from any of A, L, I, V, F, or W. at position

30 199, there is a H, or a negatively charged residue selected from D or E,

at position 200, there is a negatively charged residue, D or a hydrophobic residue selected from any of L, I, V, F, or W. and/or

at position 201, there is a negatively charged residue, D or a hydrophobic residue selected from any of A, L, I, V, F, or W.

5 The present invention also relates to the use of an isolated polypeptide as defined above to improve the aldolase catalytic performance for synthesis of 3-hydroxybutanal or crotonaldehyde of a 2-deoxyribose phosphate aldolase (DERA) enzyme.

10 The present invention also provides an isolated polynucleotide sequence encoding said polypeptide sequences.

The present invention also provides an expression system comprising the isolated polynucleotide sequence of the invention, operably linked to suitable control sequences.

15 The present invention also provides a recombinant microorganism transformed with said expression system.

20 **Example regions of interest within DERA sequences for targeted mutations that improve the synthesis of 3-hydroxybutanal or crotonaldehyde**

Based on the *E. coli* DERA protein crystal structure (PDB 1P1X), which was used as a model due to its high resolution (0.99Å), four initial conserved regions of interest in the DERA protein structure have been identified. These regions, defined by this invention, are important for carbonyl binding (Figure 18 - hotspot B) and for methyl group binding (Figure 18 - hotspot A) of the incoming acetaldehyde acceptor prior to reaction with the Schiff base, enzyme bound, donor acetaldehyde molecule. The 3-position hydroxyl group of 3-hydroxybutanal is derived from the carbonyl of the acetaldehyde acceptor molecule. A "region of interest" is defined as any single amino acid or group of amino acids that when substituted impart an efficiency influence with respect to the coupling or condensation of two molecules of acetaldehyde to form either 3-hydroxybutanal or crotonaldehyde.

25

30

As would be expected in the parent DERA enzyme, which does not carry out the efficient coupling of acetaldehyde in natural metabolism, amino acids in these regions of interests may not promote efficient acetaldehyde acceptor binding, and hence may not promote sufficient orientation with respect to reaction with the Schiff
5 base enzyme bound, natural acetaldehyde donor molecule. Although acetaldehyde can be a substrate in this acceptor position, it is at lower efficiency than that required for commercial application of the enzyme, where the commercial application is for the coupling and/ or condensation of two molecules of acetaldehyde.

10 A third example region of interest (Figure 18 - hotspot C) is known to be important for binding of the phosphate group of the natural substrate acceptor molecule glyceraldehyde-3-phosphate or the natural substrate 2-deoxyribose-5-phosphate (depending on the direction of the reaction). Amino acid changes in this area may
15 be additionally useful to suppress natural substrate binding where these natural molecules are prevalent (e.g growth on sugar based feedstocks). Changes may also promote further 3-hydroxybutanal/crotonaldehyde synthesis efficiency For example the insertion of bulky amino acids may suppress the coupling of three acetaldehyde molecules that produces the undesirable product 2,4,6 trideoxy-D-
20 erythrohexapyranoside. Some DERA enzymes are known to carry out this so called double aldol reaction, where 3-hydroxybutanal may return to the active site and may couple with another donor acetaldehyde molecule. Further, 3-hydroxybutanal may remain within the active site prior to a further acetaldehyde coupling. Contrary to this disclosure, promoting the double aldol reaction has been the target for DERA
25 evolution over the past 20 years, for application to the synthesis of statin side chains.

Regions of interest within enzyme active sites may simply be regions of specific properties. For example, with respect to protease enzymes, trypsin has a
30 negatively charged pocket able to bind positively charged residues such as arginine. The hydrophobic pocket of chymotrypsin attracts hydrophobic residues such as tyrosine, tryptophan, and phenylalanine. Therefore, creating a predominantly hydrophobic environment within the DERA active site may improve acetaldehyde binding within the acceptor site. Further, reducing the size of the active site to

favour binding of a small molecule such as acetaldehyde may impart further benefit. As described above, the use of large bulky amino acids such as phenylalanine may have particular utility for this purpose, for example if placed in the region of interest identified as Hotspot C in Figure 18. The concept for addition of hydrophobic amino acids to other example hotspot regions, is further described within this document.

Reducing the size of the active site may also serve to reduce a DERA's potential to carry out the undesired double aldol reaction where a second molecule of acetaldehyde may react with a bound C4 aldol product such as 3-hydroxybutanal.

Another example region of interest (Figure 18 – Hotspot D) consists of the single amino acid Leu20 in the example *E. coli* sequence (position 10 of SEQ ID NO:1), which is a position in correct proximity to catalyse (if required) a deprotonation step to facilitate the β -elimination of a molecule of water from enzyme bound 3-hydroxybutanal to form crotonaldehyde.

Amino acid residues occupying the described example regions of interest are typically within 4Å distance of bound 3-hydroxybutanal as a Schiff-base in the enzyme's catalytic site and any amino acid change introduced into these regions may be expected to influence the enzyme's ability to bind its substrates and thus influence catalytic activity. Therefore, to increase the catalytic activity of the acetaldehyde coupling reaction to 3-hydroxybutanal, whether or not further dehydration takes place, amino acids in these regions may be targeted. Targeting amino acids outside of these regions of interest, including those outside the active site may give further additional improvement in DERA activity.

Example amino acids influencing 3-hydroxybutanal synthesis

The identified example regions of interest were selected according to their functionality in the target acetaldehyde as acceptor substrate binding (Figure. 18). Amino acids residues in the first example region (Figure 18 – Hotspot A, T18, C47, V73 and F76 in the *E. coli* DERA sequence) have the positional capability to interact with the methyl group of the incoming acceptor acetaldehyde. Residues in the second example region (Figure 18 - Hotspot B A203 and G236 in the *E. coli* DERA

sequence) are positioned to interact with the carbonyl group of the acceptor acetaldehyde.

5 By analysis of available wild type *E.coli* DERA enzyme crystal structures, the region of the active site implicated in binding of the acceptor molecule contain amino acid residues which offer very limited directional positioning of acetaldehyde when it acts as an acceptor substrate. This is expected because acetaldehyde is not the natural acceptor for the DERA aldolase, it is only the natural donor.

10 Studies, of the 4-hydroxy-2-oxovalerate aldolase (EC 4.1.3.39, PDB 1nvm) and L - Threonine aldolases (EC 4.1.2.48, EC 4.1.2.5, EC 4.1.2.49), which naturally, efficiently use acetaldehyde as an acceptor, but glycine and pyruvate as donor (instead of acetaldehyde as in the case of DERA), show that nature has developed options to efficiently bind and arrange acetaldehyde as an acceptor aldehyde. One
15 aspect of this invention provides a means whereby a novel aldolase can be created which is able to exploit in one enzyme what nature currently offers in independent proteins i.e. an aldolase capable of binding acetaldehyde efficiently in both the donor and acceptor sites. For example, orientation of the carbonyl group of acetaldehyde is facilitated by a histidine and a tyrosine residue in these natural
20 acetaldehyde acceptor enzymes. These residues act in a bi-functional way by using electrostatic attractions to the carbonyl group of the acceptor acetaldehyde molecule to coordinate the acetaldehyde molecule in the catalytically active orientation and position. Further, these amino acids also act as an acid base catalyst by providing a proton for the carbonyl group during aldol coupling. The
25 carbonyl group is thereby converted to an alcohol function. In contrast, in the reverse retro-aldol reaction, these amino acids act as a base and accept the proton from the alcohol to facilitate cleavage of acetaldehyde from the 3-hydroxybutanal Schiff-base complex.

30 Studies of the 4-hydroxy-2-oxovalerate aldolase (EC 4.1.3.39, PDB 1nvm) and L - Threonine aldolases (EC 4.1.2.48, EC 4.1.2.5, EC 4.1.2.49) have shown that in addition to the histidine and tyrosine motif that is involved in interaction with the carbonyl group of the acceptor aldehyde, leucine residues are used by these enzyme classes to interact with the acetaldehyde methyl group for further substrate

stabilisation and orientation. Hence, in some cases a hydrophobic pocket exists for interaction with the acetaldehyde methyl moiety.

Introducing these catalytic motifs, or alternative chemically similar residues, into these example regions of interest (Figure 18 - Hotspots A and B), may be expected to increase the production of 3-hydroxybutanal of a DERA variant via the more efficient binding of acetaldehyde as an acceptor aldehyde, which, depending on further introduction of amino acids to promote the dehydration of enzyme bound 3-hydroxybutanal, can be released from the active site.

Example amino acids influencing the ability of DERA to synthesise the undesired by product 2,4,6 trideoxy-D-erythrohexapyranoside and deoxyribose-5-phosphate

It may be additionally desirable to avoid competition with the natural acceptor substrate glyceraldehyde-3-phosphate, which ultimately could lead to formation of the natural DERA product deoxyribose-5-phosphate particularly in processes relying on a sugar feedstock. Hence, mutations in the region of interest labelled Hotspot C in Figure 18 may be desirable. Introducing for example, bulky hydrophobic amino acid side chains in this region is expected to alter the hydrogen bond network and significantly increase the K_m value of glyceraldehyde-3-phosphate, thus promoting 3-hydroxybutanal synthesis under conditions of acetaldehyde availability. In a further example it is known that the point mutation S238D in *E.coli* DERA substantially affects activity towards the natural substrate deoxyribose-5-phosphate, while promoting activity towards the non-phosphorylated substrate deoxyribose (DeSantis, G. *et al.* Bioorg. Med. Chem. 11 (2003) 43–52). Synthesis of 3-hydroxybutanal was not examined. Further, with respect to DERA enzymes which may significantly carry out the double aldol coupling where 3-hydroxybutanal may return to the active site or remain within it, to couple with an additional molecule of acetaldehyde, alterations in this region are expected to influence this potential and prevent the enzyme facilitated synthesis of this 2,4,6 trideoxy-D-erythrohexapyranoside by product. As described earlier, mutations in other positions aligning with the consensus sequence may offer further advantage.

Thus, the present invention provides either a DERA variant which when characterized under identical conditions is capable of the coupling of two molecules of acetaldehyde to form 3-hydroxybutanal with an improved activity over the activity of the DERA enzyme from which it is derived, or a DERA variant which when
5 characterized is capable of the condensation of two molecules of acetaldehyde to form crotonaldehyde either directly and/or via binding or rebinding of 3-hydroxybutanal. Characterisation of the enzyme may be achieved by assessment of the enzyme's performance relative to the wild type or to an enzyme from which it is derived, either within a metabolic pathway *in vitro* or *in vivo*, or as an isolated
10 protein.

An improved enzyme variant or an enzyme variant capable of catalysing a reaction with increased activity is defined as an enzyme variant which differs from the wild type enzyme or an enzyme from which it is derived and which catalyses the
15 respective coupling or condensation reaction to form 3-hydroxybutanal or crotonaldehyde so that the specific activity of the enzyme variant is higher than the specific activity of the wild type enzyme for at least one given concentration of the substrate acetaldehyde, (preferably any acetaldehyde concentration higher than 0 M and up to 1M). A specific activity may be defined as the number of moles of
20 substrate converted to moles of product by unit of time by mole of enzyme or by weight of enzyme.

Design of experiments statistical software

25 Engineering of any enzyme where multiple substitutions in multiple positions may need to be explored prior to achieving an optimal solution may create unmanageable numbers of variants requiring analysis. However, those skilled in the art will understand that Design of Experiments computer software (e.g JMP from SAS) offers a practical approach for exploring the multifactor opportunity spaces
30 that exist in almost all real-world situations

The power of using a rational and defined multifactorial design is two-fold. Firstly, by defining the library at the design stage, there is no need to oversample a pool of

unknown mutants to ensure total library coverage - use of resources is minimised as only those mutations that are necessary are generated.

5 For example, if a library of mutants was designed to contain three separate residues to mutate, with seven different mutations at each residue, the total number of different permutations of mutant in this library is seven to the power of three (7^3) which equals 343. However, if the method of mutant generation in this library is based on random mutagenesis, it would be necessary to oversample and generate 1030 samples to ensure a 95% coverage of the experimental space (i.e. 1030 variants to be statistically certain of generating 95% of all variants in the library). For a defined rational library, it is possible to simply generate 343 mutants.

10 Secondly, a multifactorial approach allows results from one run to be directly compared to any other, meaning a significantly smaller number of mutants need to be assayed to cover the same experimental space of a more classical 'one factor at a time' (OFAT) approach where every mutation is assayed individually and cannot be compared statistically to one another.

20 This further reduces the number of mutants it is necessary to generate. Referring to the above scenario, by using a multifactorial approach it is possible to use only 127 of the 343 mutants generated to produce a statistically significant result that can inform further iterations of library design. As the number of factors being investigated increases, as do the advantages of using this approach: a library of ten factors, each at seven levels, contains 2.82×10^8 permutations. It is possible to cover this experimental space to a statistically significant degree with only 1687 runs using Design of Experiment and multifactorial design.

25 As discussed above, there are no aldolase enzymes reported to be capable of crotonaldehyde synthesis from acetaldehyde directly, or via binding or rebinding and dehydration of 3-hydroxybutanal.

30 A separate aspect of the present invention provides a modified 2-deoxyribose phosphate aldolase (DERA) enzyme variant comprising mutations that increase the ability of the enzyme to perform dehydration of 3-hydroxybutanal to form crotonaldehyde, relative to the activity of the parent DERA enzyme from which

the variant originates and which does not comprise such modifications. A DERA variant according to this aspect of the invention comprises a polypeptide sequence which, when aligned to a consensus sequence having at least 80% sequence identity with the sequence of SEQ ID NO: 1, shows an alignment with amino acid residues at 9 or more of 15 defined positions in the consensus sequence, and wherein said mutations are substitutions of residues at a position equivalent to positions 10, 178 and 199 in the consensus sequence, wherein at positions 10, 178 and 199 there is a H or a negatively charged residue selected from D or E.

Example amino acids influencing 3-hydroxybutanal dehydration to crotonaldehyde

Amino acids in the region of interest labelled Hotspot B in Figure 18 may be mutated to facilitate the aldol condensation reaction to crotonaldehyde. For example, these residues may be mutated to a combination of Histidine and Glutamate or Aspartate. Glutamate or Aspartate are able to act as acid catalysts to protonate the hydroxyl group of the bound substrate, thus converting it to a good leaving group in the β -elimination reaction. Histidine as previously described for natural acetaldehyde acceptor aldolases, is able to orientate the acetaldehyde acceptor molecule to the catalytic active position.

Example region of interest contains Leu20 in the *E. coli* sequence (Figure 18 – Hotspot D). This amino acid is in close proximity to the β -hydrogen in the 3-hydroxybutanal Schiff-base complex. By introducing an amino acid side chain in the position of Leu20 that is capable of acting as a base, abstraction of the β -hydrogen would be expected which would assist the dehydration and formation of crotonaldehyde bound to the Schiff-base complex. Crotonaldehyde would then be released from the active site.

As described above, the present invention also provides a novel consensus sequence (SEQ ID NO:1) which captures regions that are highly conserved in all DERA enzymes known to date. The novel consensus sequence serves as a tool against which any other amino acid sequence can be compared in order to identify

homology and similarity. The consensus sequence therefore enables the identification of the example regions of interest as described above in any query sequence which aligns at any position in the consensus sequence (single amino acid or group of amino acids) and which imparts an efficiency influence with respect to the coupling or condensation of two molecules of acetaldehyde to form either 3-hydroxybutanal or crotonaldehyde. As the regions of interest described above are intended only as examples, the consensus sequence may be used to identify other regions of interest outside of the given examples which when mutated enhance the coupling of two molecules of acetaldehyde or the dehydration of the product thereof relative to the parent DERA enzyme that does not contain the modifications.

Accordingly, a further aspect of the invention provides a non-naturally occurring polypeptide comprising an amino acid sequence having at least 80% identity to the amino acid sequence of SEQ ID NO. 1. In a preferred embodiment, said has at least 90%, preferably 95% identity to SEQ ID NO: 1. The non-naturally occurring polypeptide may preferably comprise the amino acid sequence of SEQ ID NO. 1, or alternatively, it may consist of the amino acid sequence of SEQ ID NO. 1.

The present invention also relates to the use of the non-naturally occurring polypeptide of the invention to identifying regions within a DERA enzyme that can be mutated to influence the catalytic activity of the DERA enzyme, wherein said catalytic activity is the coupling or condensation of two molecules of acetaldehyde to form 3-hydroxybutanal and/or crotonaldehyde. Examples of such use are provided herein.

A separate aspect of the invention provides a method for identifying one or more residues within a DERA polypeptide sequence that can influence catalytic activity of the DERA enzyme, comprising:

- a. aligning the DERA polypeptide sequence to a consensus sequence having at least 80% homology to SEQ ID No 1; and
- b. identifying residues within the DERA sequence that are in alignment with residues of the consensus sequence,

wherein residues of the DERA sequence that are in alignment with residues within the consensus sequence are identified as residues that can influence catalytic activity. Preferably, said catalytic activity is: coupling of two aldehyde molecules to produce 3-hydroxybutanal; and/ or dehydration of 3-hydroxybutanal to crotonaldehyde.

In a preferred embodiment of this method, step (b) comprises identifying residues within the DERA sequence that are in alignment with residues at 9 or more of the 15 positions within the consensus sequence which include: 8, 10, 35, 57, 60, 86, 147, 151, 176, 178, 179 180, 199, 200 and 201. If the DERA sequence comprises at least 9 residues at positions equivalent to 8, 10, 35, 57, 60, 86, 147, 151, 176, 178, 179 180, 199, 200 and/or 201 of the consensus sequence, it can be concluded that the DERA sequence can be modified to improve the catalytic activity of the enzyme, relative to the parent DERA enzyme.

A further aspect of the invention provides a method of identifying regions within a DERA enzyme that can be mutated to influence the coupling or condensation of two molecules of acetaldehyde to form 3-hydroxybutanal and/or crotonaldehyde, comprising aligning the polypeptide sequence of the DERA enzyme with a consensus sequence having at least 80% identity to the amino acid sequence of SEQ ID NO: 1, and identifying regions in the DERA enzyme sequence that align with known regions within the consensus sequence.

Preferably, the known regions within the consensus sequence comprise one or more of the following residues: 10, 35, 57, 60, 86, 147, 151, 176, 178, 179 180, 199, 200 and/or 201.

A further aspect of the invention provides a method of increasing the catalytic activity of a DERA enzyme, comprising:

- (i) aligning the polypeptide sequence of the DERA enzyme with a consensus sequence having at least 80% identity to the amino acid sequence of SEQ ID NO. 1,

- (ii) identifying regions in the DERA sequence that align with regions within the consensus sequence, and
- (iii) mutating the DNA codons encoding one more or amino acid residues within said regions in order to increase the catalytic activity of the enzyme.

Preferably, said regions are single amino acids or groups of amino acids that can be substituted to influence the coupling or condensation of two molecules of acetaldehyde to form either 3-hydroxybutanal or crotonaldehyde.

Preferably, said mutation of one more or amino acid residues results in increased synthesis of 3-hydroxybutanal and/or synthesis of crotonaldehyde.

A further related aspect of the invention is directed to the use of a polypeptide having at least 80% sequence identity with the sequence of SEQ ID NO: 1 in the method for identifying regions a DERA enzyme that can be mutated to influence the coupling or condensation of two molecules of acetaldehyde to form 3-hydroxybutanal and/or crotonaldehyde or in the method for increasing the catalytic activity of a DERA enzyme according to the present invention.

A further aspect of the invention provides a method of increasing the catalytic activity of a DERA enzyme, comprising:

- (i) aligning the polypeptide sequence of the DERA enzyme with a consensus sequence having at least 80% identity to the amino acid sequence of SEQ ID NO. 1,
- (ii) identifying regions in the DERA sequence that align with regions within the consensus sequence, and
- (iii) mutating the DNA codons encoding one more or amino acid residues within said regions in order to increase the catalytic activity of the enzyme.

A further aspect of the invention provides a method of increasing the catalytic activity of a DERA enzyme, comprising:

- (i) aligning the polypeptide sequence of the DERA enzyme with a consensus sequence having at least 80% identity to the amino acid sequence of SEQ ID NO. 1,
- (ii) identifying residues in the DERA sequence that align with positions 10, 178 and 199 within the consensus sequence, and
- (iii) mutating the DNA codons encoding these three residues in order to increase the catalytic activity of the enzyme,
- wherein said mutations are substitutions of residues at a position equivalent to positions 10, 178 and 199 in the consensus sequence, wherein at positions 10, 178 and 199 there is a H or a negatively charged residue selected from D or E, and
- wherein said catalytic activity is the dehydration of 3-hydroxybutanal to form crotonaldehyde.
- A further aspect of the invention provides an assay for screening for DERA variants having improved aldolase catalytic performance for the synthesis of 3-hydroxybutanal or downstream products thereof, comprising determining an increase in the rate of H_2O_2 formation from the oxidation of 3-hydroxybutanal or downstream products thereof in the presence of an alcohol oxidase.
- Preferably, the rate of H_2O_2 formation is compared to a control value.
- Preferably, the downstream product is 1,3-butanediol formed by the selective reduction of 3-hydroxybutanal. Preferably, selective reduction is performed by introducing a heterologous enzyme with alcohol dehydrogenase or aldehyde reductase activity which shows a preference for reduction of 3-hydroxybutanal over acetaldehyde.
- Preferably, the heterologous enzyme is selected from Table 7, or is a variant of an enzyme from Table 7, preferably wherein the heterologous enzyme is GOX 1615, GRE2, or BdhB, or a variant thereof.

Preferably, the assay is performed in a microbial host, preferably wherein the host is *E.coli*. Alternatively, the assay may be performed in a cell lysate or may be performed *in vitro* using recombinant proteins.

5 *Fusions*

As explained above, the aldolase such as DERA may be provided as a fusion protein encoding also one or more other enzymes involved in the provision of the aldolase substrate acetaldehyde, or linked to such other enzymes using chemical or other means (e.g. scaffoldins or dockerins). Examples include fusions of DERA with an acetaldehyde dehydrogenase or pyruvate decarboxylase or a carboxylic acid reductase such as AOR which catalyse reactions B, D and A described herein (see Tables 2, 4, 1). Example 11 demonstrates the production of a DERA-EutE fusion.

In other embodiments, the aldolase such as DERA may be provided as a fusion protein encoding also one or more other enzymes involved in a downstream product pathway, or linked to such other enzymes using chemical or other means (e.g. scaffoldins or dockerins). The enzyme may, for example, be one involved in the conversion of 3-hydroxybutanal to another product or intermediate. Examples include enzymes listed in Table 7.

25 *Production of 1-3 BDO and downstream products, and utilities for 1,3-BDO*

Reduction of the aldehyde moiety of the aldolase catalysed acetaldehyde coupling product 3-hydroxybutanal to the corresponding alcohol prevents further coupling of a third acetaldehyde. 3-Hydroxybutanal is reduced to (via "Reaction H" in Figure 3) 1,3-butanediol by an alcohol dehydrogenase or aldehyde reductase, for example, using enzymes categorised in EC 1.1.1.-.

This reaction is preferably catalysed by a medium chain alcohol dehydrogenase or aldehyde reductase, ideally which shows preference for alcohols of C4 or greater, for example see Appl. Environ. Microbiol, 2000, 66, 5231. More specifically the

enzyme preferably shows a preference for reduction of 3-hydroxybutanal to 1,3-BDO relative to reduction of acetaldehyde to ethanol (Example 9). An alcohol dehydrogenase described by Wales, M and Fewson, C. Microbiol 1994, 140, 173 again shows preference for longer chain alcohols. Although measured in the oxidative direction, the dehydrogenase also accepts 1,4-butanediol as a substrate. 2,3- butanediol is not a substrate, clearly demonstrating the desired primary alcohol as opposed to secondary alcohol specificity for application to 3-hydroxybutanal reduction. Other enzymes which it may be desired to utilise for conversion of 3-Hydroxybutanal to 1,3-butanediol are described in Example 3 and Table 7 hereinafter.

The use of a 3-hydroxybutanal reductase (alcohol dehydrogenase) within a 1,3-butanediol pathway is described in US20130109064 but there is no suggestion therein of the presently disclosed pathway for synthesis of 3-hydroxybutanal.

1,3-BDO has numerous utilities in industry. For example, 1,3-BDO is commonly used as an organic solvent for food flavoring agents. It is also used as a co-monomer for polyurethane and polyester resins and is widely employed as a hypoglycaemic agent. Optically active 1,3-BDO is a useful starting material for the synthesis of biologically active compounds and liquid crystals. Another use of 1,3-butanediol is that its dehydration affords 1,3-butadiene and other important chemicals such as methylethyl ketone (Ichikawa et al., *J. Molecular Catalysis A-Chemical*, 231:181-189 (2005); Ichikawa et al., *J. Molecular Catalysis A-Chemical*, 256:106-112 (2006)), 1,3-butadiene is an important chemical used to manufacture synthetic rubbers (e.g. tyres), latex, and resins. 1,3-Butadiene and further examples of products produced by chemical conversion of 1,3-butanediol are shown in Figure 1.

Production of other downstream products from 3-Hydroxybutanal

It will be appreciated that the aldol coupling product 3-hydroxybutanal can also be directed to products other than 1,3-BDO. In that sense, 3-hydroxybutanal can be considered a branch point for a number of possible unnatural DERA-based pathways leading to a variety of immediate or further downstream products.

By way of non-limiting example, 3-Hydroxybutanal can be converted (e.g. oxidised, reduced) to:

- 3-hydroxybutyrate
- 5 • 1,3-butanediol

which can in turn be utilised or recovered *per se*, or converted to other products via other natural or unnatural metabolic pathways. 3-hydroxybutyrate has utility as a biodegradable plastics monomer.

- 10 Further, 3-hydroxybutanal can be also be converted to metabolic intermediates such as 3-hydroxybutyryl CoA using for example butanal dehydrogenase (EC 1.2.1.57) or another aldehyde dehydrogenase such as EC1.2.1.10 which can allow metabolic access to a range of other products. Aldehyde dehydrogenases have been
- 15 mutated to improve their preference for C4 aldehydes relative to C2 aldehydes (e.g. acetaldehyde). For example Baker et al. describe a mutant with a preference for butanal relative to acetaldehyde, Biochemistry. 2012 Jun 5; 51(22):4558-67. Epub 2012 May 21. This enzyme may have utility in the conversion of 3-hydroxybutanal to 3-hydroxybutyryl CoA. Several other enzymes have a natural preference for a C4 aldehyde. Yan, R.T and Chen, J. S. 1990 Appl Environ Microbiol 56, (9) 2591. Any
- 20 of these enzymes could, if desired, be further engineered to optimise their activity in generating 3-hydroxybutyryl CoA in the context of the present invention.

Downstream products from 3-hydroxybutyryl CoA include:

- 2-hydroxyisobutyric acid (an intermediate for methacrylic acid synthesis).
- 25 • polyhydroxybutyrate (which also has utility in biodegradable plastics).
- Crotyl alcohol (which can be converted enzymatically or chemically to 1,3-butadiene).

- 30 Butanol (which can be converted enzymatically or chemically to 1,3-butadiene) Other downstream products include Crotonic acid, butyrate, 3-hydroxybutyrate, 3-hydroxybutylamine, Polyhydroxybutyrate, Acetone, and isopropanol.

The synthesis of products from 3-hydroxybutyryl CoA is established biochemistry. For example as described in the following references: Toshiyuki, U. et al. 2014,

mbio 5, (5), 1 (butyrate); Torben, H. et al. 2010, Appl. Microbiol Biotechnol. 88, 477 (2-hydroxyisobutyric acid); Nadya, Y. et. al. 2012. J. Biol. Chem. 287 (19) 15502 (2-hydroxyisobutyric acid); Rehm, B.H. 2007. Curr. Issues. Mol. Biol. 9 (1), 41 (polyhydroxybutyrate);. Lee, S.Y et. al. 2008. Biotechnol Bioeng. 101 (2), 209 (butanol); US 8580543; WO2013057194 (crotyl alcohol).

A route to 3-hydroxybutyryl CoA via acetate in acetogens allows for generation of this intermediate without sacrificing the ATP energy which would otherwise be lost if 3-hydroxybutyryl CoA was provided via, for example, acetyl CoA to acetoacetyl CoA. This is because preventing acetate formation from acetyl CoA loses the molecule of ATP generated from the acetate kinase reaction. Generation of acetoacetyl CoA is energetically unfavourable under most conditions. The pathway to 3-hydroxybutyryl CoA via acetate through the DERA pathway described herein, retains the energetics of acetogenesis. Hence, the same product is reached through a more energetically favourable route. The key intermediate branch point is the DERA product 3-hydroxybutanal.

Genetic modification and hosts

Based on the guidance provided herein, those skilled in the art will understand that the number of encoding nucleic acids to introduce in an expressible form will reflect any 3-hydroxybutanal or downstream product derived therefrom (e.g.1,3-BDO) pathway deficiencies of the selected microbial host. Therefore, a non-naturally occurring microorganism of the invention can have one, two, three, or more, up to all nucleic acids encoding the enzymes or proteins constituting a 3-hydroxybutanal or downstream product derived therefrom (e.g.1,3-BDO) pathway revealed herein. In some examples, the non-naturally occurring microorganisms can also include other genetic modifications that facilitate or optimise 1,3-BDO (or other downstream product derived from 3-hydroxybutanal) biosynthesis or that confer other useful functions onto the host microorganism.

Preferred microbial organisms are discussed herein and include, *inter alia*, yeasts such as *Saccharomyces cerevisiae*, *Kluveromyces lactis* *Candida boidinii*, *Pichia angusta*, *Ogataea polymorpha*, *Komagataella pastoris*.

Other preferred microbial organisms include bacteria such as *Moorella thermoacetica*, *Moorella thermoautotrophica*, *Thermoacetogenium phaeum*, *Thermoanaerobacter kivu*, *Acetobacterium woodii*, *Clostridium carboxidivorans*,
 5 *Clostridium drakei*, *Clostridium formicoaceticum*, *Clostridium glycolicum*, *Clostridium magnum*, *Clostridium mayombeii*, *Clostridium ljungdahlii*, *Clostridium ragsdalei*, *Clostridium aceticum*, *Clostridium autoethanogenum*, *Clostridium scatologenes*, *Acetitomaculum ruminis*, *Acetogenium kivui*, *Eubacterium limosum*, *Oxobacter pfennigii*, *Acetobacterium tundrae*, *Acetobacterium noterae*, *Acetobacterium carbinolicum*,
 10 *Acetobacterium dehalogenans*, *Acetobacterium fimetarium*, *Acetobacterium malicum*, *Acetobacterium paludosum*, *Acetobacterium wieringae*, *Acetohalobium arabicum*, *Acetonema longum*, *Acetitomaculum ruminis*, *Acetoanaerobium noterae*, *Acetobacterium bakii*, *Butyribacterium metholytrophicum*, *Blautia hydrogenotrophica*, *Blautia coccoides*, *Blautia producta*,
 15 *Blautia schinkii*, *Peptostreptococcus productus*, *Sporomusa acidovorans*, *Sporomusa aerivorans*, *Sporomusa malonica*, *Sporomusa ovate*, *Sporomusa paucivorans*, *Sporomusa rhizae*, *Sporomusa silvacetica*, *Sporomusa spaeroides*, *Sporomusa termitida*. *Bacillus methanolicus*, *Methylobacterium extorquens*, *Methylobacillus flagellates*, *Methylobacillus glycogenes*, *Methylobacillus pratensis*,
 20 *Hydrogenobacter thermophilus*, *Acidomonas methanolica*, *Methylococcus capsulatus*, *Methylophilus methylotrophus*, *Methylophilus flavus*, *Methylophilus luteus*, *Methylacidiphilum infernorum*, *Methylibium petroleiphilum*, *Hydrogenibacillus schlegelii*, *Lactococcus. sp.* *Lactobacillus.sp.*, *Bacillus sp.* *Geobacillus sp.* *Corynebacterium. sp.* *Klebsiella, oxytoca*, *Ralstonia. sp.*,
 25 *Alcaligenes. sp.* *Cupriavidus. sp.*

In one embodiment, the host is not *E. coli*.

Sources of encoding nucleic acids for use in the present invention can include any species where the encoded gene product is capable of catalysing the referenced
 30 reaction. Such species include both prokaryotic and eukaryotic organisms including, but not limited to, bacteria, including archaea and eubacteria, and eukaryotes, including yeast, plant, insect, animal and mammal, including human.

Exemplary sources of nucleic acids are described herein. However, with the large number of complete genome sequences available, the identification of genes encoding the requisite 1,3-BDO biosynthetic activity (e.g. the aldolase-type enzymes described herein) for one or more genes in related or distant species, including for example, homologs, orthologs, paralogs and non-orthologous gene displacements of known genes, and the interchange of genetic alterations between organisms is routine and well known to those skilled in the art, and can be carried out in the present context in the light of the teaching herein.

Consequently, in the light of the present disclosure, the metabolic modifications enabling biosynthesis of 3-hydroxybutanal or downstream product derived therefrom (e.g. 1,3-BDO) described herein with reference to a particular organism such as *Moorella thermoacetica* can be readily applied to other microorganisms. Those skilled in the art will know that a metabolic modification exemplified in one organism can be applied equally to other organisms.

Those skilled in the art will recognise that whenever a particular protein or nucleic acid is referred to herein e.g. with reference to an accession number or other deposit identification, that a functional variant of that sequence may also be used. Since the present invention is primarily concerned with enzyme activities, it will be appreciated that a functional variant will be one which catalyses the same substrate to product reaction as that catalysed by the enzyme referred to, but has a different sequence.

Non-limiting examples of variants include the following:

(i) Novel, naturally occurring, nucleic acids, isolatable using the recited or referred to sequence. These may include alleles (which will include polymorphisms or mutations at one or more bases), paralogues, isogenes, or other homologous genes belonging to the same families as the relevant enzymes. Also included are orthologues or homologues from different microbial or other species.

Thus, included within the scope of the present invention are uses of nucleic acid molecules which encode amino acid sequences which are homologues of the genes referred to herein. Homology may be at the nucleotide sequence and/or amino acid

sequence level, as discussed below. A homologue from a different species or strain encodes a product which causes a phenotype similar to that caused by the recited sequence.

- 5 (ii) Artificial nucleic acids, which can be prepared by the skilled person in the light of the present disclosure. Such derivatives may be prepared, for instance, by site directed or random mutagenesis, or by direct synthesis. Preferably the variant nucleic acid is generated either directly or indirectly (e.g. via one or more amplification or replication steps) from an original nucleic acid having all or part of
10 the sequence referred to herein.

Changes may be desirable for a number of reasons. For instance they may introduce or remove restriction endonuclease sites or alter codon usage.

- 15 Alternatively changes to a sequence may produce a derivative by way of one or more (e.g. several) of addition, insertion, deletion or substitution of one or more nucleotides in the nucleic acid, leading to the addition, insertion, deletion or substitution of one or more (e.g. several) amino acids in the encoded polypeptide.
- 20 Other desirable mutations may be random or site directed mutagenesis in order to alter or evolve the activity (e.g. specificity) or stability of the encoded polypeptide. Changes may be by way of conservative variation, i.e. substitution of one hydrophobic residue such as isoleucine, valine, leucine or methionine for another, or the substitution of one polar residue for another, such as arginine for lysine,
25 glutamic for aspartic acid, or glutamine for asparagine. As is well known to those skilled in the art, altering the primary structure of a polypeptide by a conservative substitution may not significantly alter the activity of that peptide because the side-chain of the amino acid which is inserted into the sequence may be able to form similar bonds and contacts as the side chain of the amino acid which has been
30 substituted out. This is so even when the substitution is in a region which is critical in determining the peptides conformation. Also included are variants having non-conservative substitutions. As is well known to those skilled in the art, substitutions to regions of a peptide which are not critical in determining its conformation may not greatly affect its activity because they do not greatly alter the

peptide's three dimensional structure. In regions which are critical in determining the peptides conformation or activity such changes may confer advantageous properties on the polypeptide. Indeed, changes such as those described above may confer slightly advantageous properties on the peptide e.g. altered stability or specificity.

The term 'variant' nucleic acid as used herein encompasses all of these possibilities. When used in the context of polypeptides or proteins it indicates the encoded expression product of the variant nucleic acid.

Some of the aspects of the present invention relating to variants will now be discussed in more detail.

Sequence identity may be assessed as using BLASTp (proteins) or Megablast (nucleic acids) from NCBI (<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>) using default settings, as used in the Examples.

Variants of the sequences disclosed herein preferably share at least 55%, 56%, 57%, 58%, 59%, 60%, 65%, or 70%, or 80% identity, most preferably at least about 90%, 95%, 96%, 97%, 98% or 99% identity. Such variants may be referred to herein as "substantially homologous".

Nucleic acid fragments may encode particular functional parts of the enzyme (i.e. encoding a biological activity of it). Thus the present invention provides for the production and use of fragments of the full-length polypeptides disclosed herein, especially active portions thereof. An "active portion" of a polypeptide means a peptide which is less than said full length polypeptide, but which retains its essential biological activity.

Generally speaking, those skilled in the art are well able to construct vectors and design protocols for the recombinant genetic manipulations described herein. Suitable vectors can be chosen or constructed, containing appropriate regulatory sequences, including promoter sequences, terminator fragments, polyadenylation sequences, enhancer sequences, marker genes and other sequences as

appropriate. For further details see, for example, *Molecular Cloning: a Laboratory Manual*: 2nd edition, Sambrook *et al*, 1989, Cold Spring Harbor Laboratory Press or *Current Protocols in Molecular Biology*, Second Edition, Ausubel *et al.* eds., John Wiley & Sons, 1992.

5

A "vector" as used herein need not include a promoter or other regulatory sequence, particularly if the vector is to be used to introduce nucleic acid into cells for recombination into the genome.

10

However for expression purposes the nucleic acid in the vector will typically be under the control of, and operably linked to, an appropriate promoter or other regulatory elements for transcription in a microbial host cell. It may include a native promoter. In the case of cDNA this may be under the control of an appropriate promoter or other regulatory elements for expression in the host cell.

15

By "promoter" is meant a sequence of nucleotides from which transcription may be initiated of DNA operably linked downstream (i.e. in the 3' direction on the sense strand of double-stranded DNA). "Operably linked" means joined as part of the same nucleic acid molecule, suitably positioned and oriented for transcription to be initiated from the promoter. DNA operably linked to a promoter is "under transcriptional initiation regulation" of the promoter. In one embodiment, the promoter is an inducible promoter. The term "inducible" as applied to a promoter is well understood by those skilled in the art. In essence, expression under the control of an inducible promoter is "switched on" or increased in response to an applied stimulus. The nature of the stimulus varies between promoters. Some inducible promoters cause little or undetectable levels of expression (or no expression) in the absence of the appropriate stimulus. Other inducible promoters cause detectable constitutive expression in the absence of the stimulus. Whatever the level of expression is in the absence of the stimulus, expression from any inducible promoter is increased in the presence of the correct stimulus.

20
25
30

The present disclosure teaches how pathways may be engineered into an organism by selection of the appropriate enzymes, cloning their corresponding genes into a production host, optimising the stability and expression of these genes, attenuation

or functional deletion of the competitive pathways, optimising fermentation conditions for the genetically engineered strain to produce the desired product, and assaying for product formation following fermentation.

5 The term "heterologous" is used broadly herein to indicate that the gene/sequence of nucleotides in question (e.g. encoding an aldolase) has been introduced into a host cell or an ancestor thereof, using genetic engineering, i.e. by human intervention. Nucleic acid heterologous to a host cell will be non-naturally occurring in cells of that type, variety or species. Thus the heterologous nucleic acid may comprise a coding sequence of or derived from a microorganism, placed within a different microorganism. A further possibility is for a nucleic acid sequence to be placed within a cell in which it or a homologue is found naturally, but wherein the nucleic acid sequence is linked and/or adjacent to nucleic acid which does not occur naturally within the cell, such as operably linked to one or more regulatory sequences, such as a promoter sequence, for control of expression.

20 "Transformed" in this context means that the nucleotide sequences of the heterologous nucleic acid alter one or more of the cell's characteristics and hence phenotype e.g. with respect to 3-hydroxybutanal or downstream product derived therefrom (e.g. 1,3-BDO).

25 "Nucleic acid" when used in the present invention may include cDNA, RNA, genomic DNA and modified nucleic acids or nucleic acid analogs (e.g. peptide nucleic acid). Where a DNA sequence is specified, e.g. with reference to a figure, unless context requires otherwise the RNA equivalent, with U substituted for T where it occurs, is encompassed. Nucleic acid molecules according to the present invention may be provided isolated and/or purified from their natural environment, in substantially pure or homogeneous form, or free or substantially free of other nucleic acids of the species of origin, and double or single stranded. Where used herein, the term "isolated" encompasses all of these possibilities. The nucleic acid molecules may be wholly or partially synthetic. In particular they may be recombinant in that nucleic acid sequences which are not found together in nature (do not run contiguously) have been ligated or otherwise combined artificially.

Nucleic acids may comprise, consist, or consist essentially of, any of the sequences discussed hereinafter.

5 In the methods herein any shuttle vectors available for Gram-positive bacteria that carry at least one nucleotide sequence homologous to one gene encoding the desired enzyme can be employed for transformation of *M. thermoacetica* or other microorganism of interest.

10 An expression plasmid is obtained by inserting at least a gene responsible for replication of the plasmid in Gram-positive and more specifically in Clostridia species or acetogens. The plasmid capable of introducing the desired gene into an acetogen is not particularly limited as long as it contains at least a gene responsible for replication and amplification in acetogenic bacteria. Specific examples thereof include pAK201 (Kim, A. and Blashek, H. P., *Appl. Environ. Microbiol.* 55 (2):360-365 (1988), pHB101 (Blaschek H. P. *et al.*, *J. Bacteriol.* 147(1):262-266 (1981)), any of the series modular plasmids pMTL8000 (Heap, J.T. *et al.*, *J. Microbiol. Methods* 78:79-85 (2009), pMS1, pMS2, pMS3, pMS4, pKV12 (Staetz, M. *et al.*, *Appl. Environ. Microbiol.* 1033-1037 (1994), pUB110 (McKenzie *et al.*, 1984), pIMP1 (Mermelstein, L *et al.* 1992), pITF (Dong, H. *et al.* 2010). A further example of a plasmid known to be suitable for use in acetogens is pMTL80000 (Kopke, M. *et al.*, *Appl. Environ. Microbiol.* 3394-3403, 2014) available from The University of Nottingham, United Kingdom.

25 Novel shuttle vectors, which are chimeras of pUB110 or any of the above mentioned plasmids and a general *E. coli* cloning vectors such as pUC19 (Yanisch-Perron, C. *et al.*, *Gene* 33:103-119 (1985)) or pBluescript II SK (+/-) can be easily generated and tested. These chimera plasmids are propagated in *E. coli* for plasmid isolation and employed for the genetic engineering work of *M. thermoacetica* or another acetogen or Gram-positive bacteria which is naturally sensitive towards the antibiotic gene expressed by the plasmid. If needed, sub-cloning can be employed
30 to replace the antibiotic resistance cassettes on the existing plasmids with suitable ones based on the antibiotic sensitivity of the target organism. Standard techniques for DNA amplification using a high-fidelity DNA polymerase and molecular sub-

cloning, including restriction enzyme digestion, ligation and *E. coli* transformation can be used for engineering of the plasmids (Sambrook, 1989).

5 The antibiotic resistance of *M. thermoacetica* has been tested in liquid cultures and on plates and selection conditions have been identified. In one embodiment kanamycin and chloramphenicol may be utilised as antibiotic markers for selection of the genetic engineered *M. thermoacetica* strains.

10 The operon or one gene of the operon encoding the required activity can be ligated into the multiple cloning site between two convenient restriction sites.

In order to achieve optimum gene(s) expression for the heterologous genes introduce, the heterologous genes can be codon optimised for the target organism with techniques well known to those skilled in the art.

15 To ease the detection and quantification of the gene product(s) expression, an N-or C-terminus tag sequence can be added to the gene sequences cloned as understood by those skilled in the art.

20 Many *Clostridia* species have been successfully transformed with prior methylated DNA vectors. The methylation of the transformable DNA protects it from being degraded by the host. *In vivo* methylation of the transformable DNA is achieved by its propagation in methylation *E. coli* strains such as Top10 (pAN2) (Kuit *et al.*, *Appl. Microbiol. Biotechnol.* 94:729-741 (2012)).

25 Heterologous (or exogenous, the terms are used interchangeably) gene(s) can be introduced into the chosen host cell, exemplified herein by *M. thermoacetica* and *Acetobacterium woodii*, using techniques well known in the art including, but not limited to, conjugation, electroporation, chemical transformation, transduction, transfection and electrofusion. For electroporation and conjugation, published
30 protocols of *Clostridium perfringens*, *Clostridium. acetobutylicum*, *Clostridium. cellulolyticum* and *Acetobacterium woodii* may be used.

In some embodiments it may be desired to target or inactivate genes in the host microbial cell, for example to increase flux of target metabolic intermediates and/or

1,3-BDO, or divert metabolic pathways away from biomass generation. An example is to minimise loss of pyruvate away from a 3-hydroxybutanal pathway.

5 To permanently inactivate genes in *M. thermoacetica* or other acetogens, a plasmid can be constructed for gene deletion by integrational mutagenesis or gene replacement techniques well known in the art. Integrational mutagenesis and gene replacement can selectively inactivate undesired genes from host genomes. Such methods have been developed and successfully used to create metabolically engineered mutants of Clostridial strains (Green et al., 1996). In this technique, a
10 fragment of the target gene is cloned into a non-replicative vector with a selection marker, resulting in the non-replicative integrational plasmid. The partial gene in the non-replicative plasmid can recombine with the internal homologous region of the original target gene in the parental chromosome (double crossover), which results in the insertional inactivation of the target gene, *ldh* locus in this particular example.
15 The use of gene replacement (by double recombination) is preferred to insertional inactivation (single recombination) since it permits the generation of more stable engineered strains, without the need to maintain selection of vectors. An example describing a double crossover in an acetogen is shown in Example 5.
20 Using this technique, in the same manner non-natural microorganisms can be generated having complete or partial deletion of one, two, three, four, five, or more genes in order to remove competitive pathways.

25 Reduction of expression of the target genes can also be used as an alternative to gene disruption. This may be achieved using expression of antisense RNA for the target gene, which will inhibit but not completely abolish gene expression. The antisense RNA system serves as a convenient approach of gene knock-down of a desired gene with the advantage that it can reduce expression of genes for which complete inactivation could be damaging or lethal to the organism.

30 In using anti-sense genes or partial gene sequences to down-regulate gene expression, a nucleotide sequence is placed under the control of a promoter in a "reverse orientation" such that transcription yields RNA which is complementary to

normal mRNA transcribed from the "sense" strand of the target gene. See, for example, Rothstein et al, 1987; Smith *et al*, (1988) *Nature* 334, 724-726.

5 The complete sequence corresponding to the coding sequence (in reverse orientation for anti-sense) need not be used. For example fragments of sufficient length may be used. It is a routine matter for the person skilled in the art to screen fragments of various sizes and from various parts of the coding sequence to optimise the level of anti-sense inhibition. It may be advantageous to include the initiating ATG codon, and perhaps one or more nucleotides upstream of the
10 initiating codon. A further possibility is to target a conserved sequence of a gene, e.g. a sequence that is characteristic of one or more genes, such as a regulatory sequence.

The sequence employed may be about 500 nucleotides or less, possibly about 400
15 nucleotides, about 300 nucleotides, about 200 nucleotides, or about 100 nucleotides. It may be possible to use oligonucleotides of much shorter lengths, 14-23 nucleotides, although longer fragments, and generally even longer than about 500 nucleotides are preferable where possible, such as longer than about 600 nucleotides, than about 700 nucleotides, than about 800 nucleotides, than about
20 1000 nucleotides or more.

It may be preferable that there is complete sequence identity in the sequence used for down-regulation of expression of a target sequence, and the target sequence, although total complementarity or similarity of sequence is not essential. One or
25 more nucleotides may differ in the sequence used from the target gene. Thus, a sequence employed in a down-regulation of gene expression in accordance with the present invention may be a wild-type sequence (e.g. gene) selected from those available, or a variant of such a sequence in the terms described above. The sequence need not include an open reading frame or specify an RNA that would be
30 translatable.

The transformation, expression and application of antisense RNA inhibition tools have been demonstrated for mesophilic Clostridia such as: *Clostridium acetobutylicum* (Desai R. *et al. Appl. Environ & Environ Microbiol.* 65(3):936-945

(1999)) Fierro-Monti IP *et. al.*, *J Bacteriol.* 174(23):7642-7647 (1992)) and *Clostridium cellulolyticum* (Perret S, *et al.*, *Mol. Microbiol.* 51(2):599-607 (2004)) as well as for thermophiles such as *Thermus thermophilus* (Moreno, R. *et al.*, *J. Bacteriol.*, 7804–7806(2004) and may be applied herein.

5

An attractive approach for down-regulation expression of a target gene is to replace the native promoter with a less active promoter for example one from another gene. This can be achieved by double-recombination/gene replacement techniques well known in the art. Alternatively, expression can be reduced by altering the ribosome binding site or the spacing between the RBS and the translation initiation start codon, or using a less efficient start codon.

10

The results of these studies permit for phenotypic characterisation of the mutants generated as well as allow genetic engineering of *M. thermoacetica* or other acetogens. Further optimisation can be performed to develop genetic systems by varying methods, plasmids and conditions to achieve optimum results. Specifically, the metabolic modifications enabling biosynthesis of 1,3-BDO described herein with reference to a particular organism such as *M. thermoacetica* can be readily applied to other microorganisms, including prokaryotic and eukaryotic organisms alike.

15

20

Any sub-titles herein are included for convenience only, and are not to be construed as limiting the disclosure in any way.

The invention will now be further described with reference to the following non-limiting Figures and Examples. Other embodiments of the invention will occur to those skilled in the art in the light of these.

25

The disclosure of all references cited herein, and the abstract appended hereto, inasmuch as it may be used by those skilled in the art to carry out the invention, is hereby specifically incorporated herein by cross-reference.

30

Figures

Figure 1. Example of chemical transformation of 1,3-butanediol into industrially important chemicals including butadiene and methylethyl ketone. Ichikawa et al., *J. Molecular Catalysis A-Chemical*, 256:106-112 (2006)

- 5 Figure 2. Shows the Wood Ljungdahl pathway for synthesis of 3 acetyl CoA (3 acetate), from gaseous carbon sources with or without methanol, showing the entry point for methanol. Associated equations are: $4\text{CH}_3\text{OH} + 2\text{CO}_2 \rightarrow 3\text{CH}_3\text{COOH}$; $12\text{CO} + 6\text{H}_2\text{O} \rightarrow 3\text{CH}_3\text{COOH} + 6\text{CO}_2$; $12\text{H}_2 + 6\text{CO}_2 \rightarrow 3\text{CH}_3\text{COOH} + 6\text{H}_2\text{O}$. The Wood Ljungdahl pathway can also fix CO_2 derived from the glycolytic pathway (pyruvate decarboxylation) using reducing equivalents derived from glycolysis and pyruvate decarboxylation.
- 10

Figure 3. Shows metabolic pathways (route 1, 2, 3, 4, 5 and 6) for the synthesis of 1,3-butanediol from the central metabolic intermediates acetyl CoA or pyruvate, via the common intermediate acetaldehyde. Enzyme activities required to catalyse these steps are listed as Activity A, B, C, D, E, F, G, and H. Example gene sequences coding for these activities can be found in Tables 1, 2, 3, 4, 5, 6, and 7.

15

Route 1 proceeds from acetyl CoA through acetate (a natural product of acetogenic microorganisms) to acetaldehyde via carboxylic acid reductase activity, for example, EC 1.2.7.5 or EC. 1.2.99.6, ATP or ferredoxin driven or EC 1.2.1.30 or EC 1.2.1.3.

20

Route 2 involves direct synthesis of acetaldehyde from acetyl CoA using an aldehyde dehydrogenase (acylating), for example, acetaldehyde dehydrogenase EC 1.2.1.10.

25

Route 3 involves the conversion of pyruvate to acetaldehyde via acetyl CoA using enzymes such as EC 1.2.7.1 or EC 1.2.1.51 or EC 1.2.4.1 and EC 1.2.1.10.

30 Route 4 involves the conversion of pyruvate to acetaldehyde, directly via pyruvate decarboxylase (EC 4.1.1.1).

Route 5 involves the conversion of acetyl CoA to acetaldehyde via pyruvate using enzymes such as EC 1.2.7.1 and EC 4.1.1.1.

Route 6 involves the conversion of acetate to acetaldehyde via acetyl CoA using enzymes such as EC 6.2.1.1 or EC 2.8.3.8 and EC 1.2.1.10.

- 5 Two molecules of acetaldehyde are coupled to form 3-hydroxybutanal using an aldolase capable of accepting an aldehyde as both the acceptor and donor in an aldol coupling, for example, deoxyribose phosphate aldolase (DERA, EC 4.1.2.4). 3-Hydroxybutanal is reduced to 1,3-butanediol by an alcohol dehydrogenase or aldehyde reductase, for example, using enzymes categorised in EC 1.1.1.1, EC
10 1.1.1.2, EC 1.1.1.72 or EC 1.1.1.265 or EC 1.1.1.283.

- Figure 4. Shows the RuMP pathway and its association with the TCA cycle (modified from Appl. Environ Microbiol. 2003 69, 3986). Pyruvate is the primary product of the RuMP pathway which is converted to acetyl CoA prior to entry to the
15 TCA cycle. Either pyruvate or acetyl CoA can be converted directly to the common intermediate acetaldehyde thereby supplying substrate for a DERA type aldolase capable of accepting acetaldehyde as both the donor and acceptor in an aldol coupling for synthesis of 1,3-butanediol.

- 20 Figure 5. Shows the Wood Ljungdahl pathway. Either pyruvate or acetyl CoA can be converted directly to the common intermediate acetaldehyde thereby supplying substrate for a DERA type aldolase capable of accepting acetaldehyde as both the donor and acceptor in an aldol coupling for synthesis of 1,3-butanediol. Modified from Fung Min Liew, Michael Köpke and Séan Dennis Simpson (2013). Gas
25 Fermentation for Commercial Biofuels Production, Liquid, Gaseous and Solid Biofuels - Conversion Techniques, Prof. Zhen Fang (Ed.), ISBN: 978-953-51-1050-7, InTech, DOI: 10.5772/52164. Acetate derived from acetyl CoA can also be directly reduced to acetaldehyde for supply to the aldolase.

- 30 Figure 6. Shows the reverse TCA cycle. Either pyruvate or acetyl CoA can be converted directly to the common intermediate acetaldehyde thereby supplying substrate for a DERA type aldolase capable of accepting acetaldehyde as both the donor and acceptor in an aldol coupling for synthesis of 1,3-butanediol. Modified from Mar. Drugs. 2011, 9, 719.

Figure 7. Shows the serine cycle. Acetyl CoA can be converted directly to the common intermediate acetaldehyde supplying substrate for a DERA type aldolase capable of accepting acetaldehyde as both the donor and acceptor in an aldol coupling for synthesis of 1,3-butanediol. Central metabolism also converts PEP (phosphoenol pyruvate) into pyruvate which can be decarboxylated to acetaldehyde as described previously.

Figure 8. Shows the coupling of acetaldehyde catalysed by deoxyribose phosphate aldolase (DERA). The coupling product 3-hydroxybutanal accumulates without further coupling, or is subject to a second acetaldehyde addition depending on the enzyme and the reaction conditions.

Figure 9. Shows the Cavin cycle linked to sugar synthesis (or utilisation) and or conversion to pyruvate or acetyl CoA directly. Either pyruvate or acetyl CoA can be converted directly to the common intermediate acetaldehyde thereby supplying substrate for a DERA type aldolase capable of accepting acetaldehyde as both the donor and acceptor in an aldol coupling for synthesis of 1,3-butanediol.

Figure 10. Shows *Acetobacterium woodii* grown on an agar plate containing 0.1 g/L MUG (4-Methylumbelliferyl- β -D-glucopyranosiduronic acid) demonstrating successful expression of a heterologous gene in an acetogen. This system can also act as a reporter to confirm expression of other heterologous genes.

Key.

A1: Colony 1 of *A. woodii* carrying plasmid pEP55

A2: Colony 2 of *A. woodii* carrying plasmid pEP55

B1: Colony 1 of *A. woodii* carrying plasmid pEP56

B2: Colony 2 of *A. woodii* carrying plasmid pEP56

C: Negative Control, *A. woodii* carrying the pEP plasmid expressing an unrelated gene

Figure 11. Cloning strategy to construct an *A. woodii* LDH knockout mutant by replacing the LDH gene with an Erythromycin resistance marker.

Figure 12. Cloning strategy to construct an *A. woodii* LDH knockout mutant by disrupting the LDH gene via single cross-over recombination event and integration of the complete plasmid.

5

Figure 13. Growth of *A. woodii* wildtype and *A. woodii* mutants in the presence of 20 mM Fructose and 40 mM DL-Lactate. Aw = *A. woodii* wildtype, Plasmid = *A. woodii* transformant harboring plasmid pUC19-Ery-pAM β 1, dLDH = double cross-over LDH knockout. SR = Single cross-over LDH knockout.

10

Figure 14. Utilization of Fructose and Acetate production by *A. woodii* wildtype and *A. woodii* mutants.

Aw = *A. woodii* wild type, P = *A. woodii* transformant harboring plasmid pUC19-Ery-pAM β 1, dLDH = double cross-over LDH knockout. SR = Single cross-over LDH knockout.

15

Figure 15. Utilization of Lactate and Acetate production *A. woodii* wildtype and *A. woodii* mutants.

Aw = *A. woodii* wild type, P = *A. woodii* transformant harboring plasmid pUC19-Ery-pAM β 1, dLDH = double cross-over LDH knockout. SR = Single cross-over LDH knockout.

20

Figure 16. Representative mass spectrometry data for the product 1,3-butanediol produced from various pathway combinations incorporating DERA enzymes

25

Figure 17. Examples of downstream products obtainable from 3-hydroxybutanal.

Figure 18 *E.coli* DERA with amino acids within 4Å of the lysine 137 3-hydroxybutanal Schiff-base complex. Residues in positions particularly influencing binding of the acetaldehyde as acceptor aldehyde are clustered in example hotspots A and B. Example hotspot C is involved with binding the phosphate group of the natural substrate. Example hotspots B and D are regions able to influence formation of crotonaldehyde by the dehydration of 3-hydroxybutanal.

30

Figure 19 Alignment of the consensus sequence (SEQ ID NO1) with amino acid sequence of *E. coli* DERA (SEQ ID NO.2) as a query. 12 Amino acid residues that are found in Hotspots A, B, C and D and serve as target sites for mutation are highlighted in green on the consensus sequence (top). By performing a pairwise alignment, all those amino acids in the *E. coli* DERA sequence (bottom) aligning to those highlighted in green on the consensus (i.e. T18, L20, C47, V73, F76, G171, A203, G204, G205, G236, A237 and S238 of the *E. coli* DERA) are those which may be targeted in order to improve the *E. coli* DERA's ability to perform the aldol coupling or condensation. Amino acid positions (e.g. T18) are counted from the N terminal ignoring gaps (-).

Figure 20 Alignment of the consensus sequence (SEQ ID NO1) with amino acid sequence of *Homo sapiens* DERA (SEQ ID NO.3) as a query. 12 Amino acid residues that are found in Hotspots A, B, C and D and serve as target sites for mutation are highlighted in green on the consensus sequence (top). By performing a pairwise alignment, all those amino acids in the *H. sapiens* DERA sequence (bottom) aligning to those highlighted in green on the consensus (i.e. T57, L59, C100, V125, F129, G179, A213, G214, G215, G246, A247 and S248 of the *H. sapiens* DERA) are those which should be targeted in order to improve the *H. sapiens* DERA's ability to perform the aldol coupling and condensation. Amino acid positions (e.g. T57) are counted from the N terminal ignoring gaps (-).

Figure 21 Alignment of the consensus sequence (SEQ ID NO1) with amino acid sequence of *Plasmodium falciparum* DERA (SEQ ID NO.4) as a query. 12 Amino acid residues that are found in hotspots A, B, C and D and serve as target sites for mutation are highlighted in green on the consensus sequence (top). By performing a pairwise alignment, all those amino acids in the *P. falciparum* DERA sequence (bottom) aligning to those highlighted in green on the consensus (i.e. T20, L22, C49, V76, F79, G175, A212, G213, G214, G245, A246 and S247 of the *P. falciparum* DERA) are those which should be targeted in order to improve the *P. falciparum* DERA's ability to perform the aldol coupling and condensation. Amino acid positions (e.g. T20) are counted from the N terminal ignoring gaps (-).

Figure 22 Alignment of the consensus sequence (SEQ ID NO.1) with amino acid sequence of *Pyrobaculum aerophilum* DERA (SEQ ID NO.5) as a query. 12 Amino acid residues that are found in hotspots A, B, C and D and serve as target sites for mutation are highlighted in green on the consensus sequence (top). By performing a pairwise alignment, all those amino acids in the *P. aerophilum* DERA sequence (bottom) aligning to those highlighted in green on the consensus (i.e. A8, L10, C35, V56, F59, G150, A190, G191, G192, G218, T219 and S220 of the *P. aerophilum* DERA) are those which should be targeted in order to improve the *P. aerophilum* DERA's ability to perform the aldol coupling or condensation. Amino acid positions (e.g. T20) are counted from the N terminal ignoring gaps (-).

Figure 23 Alignment of the consensus sequence (SEQ ID NO.1) with amino acid sequence of *Geobacillus thermoglucosidasius*. DERA (SEQ ID NO.6) as a query. 12 Amino acid residues that are found in hotspots A, B, C and D and serve as target sites for mutation are highlighted in green on the consensus sequence (top). By performing a pairwise alignment, all those amino acids in the *Geobacillus*. DERA sequence (bottom) aligning to those highlighted in green on the consensus (i.e. T12, L14, C39, V62, F65, G158, S185, G186, G187, G206, T207 and S208 of the *Geobacillus* DERA) are those which should be targeted in order to improve the *Geobacillus* DERA's ability to perform the aldol coupling or condensation. Amino acid positions (e.g. T20) are counted from the N terminal ignoring gaps (-).

Figure 24 Alignment of the consensus sequence (SEQ ID NO.1) with amino acid sequence of *Acetobacterium woodii* DERA (SEQ ID NO.7) as a query. 12 Amino acid residues that are found in hotspots A, B, C and D and serve as target sites for mutation are highlighted in green on the consensus sequence (top). By performing a pairwise alignment, all those amino acids in the *A. woodii* DERA sequence (bottom) aligning to those highlighted in green on the consensus (i.e. T13, L15, C40, V62, F65, G157, A184, G185, G186, G205, T206 and S207 of the *A. woodii* DERA) are those which should be targeted in order to improve the *A. woodii* DERA's ability to perform the aldol coupling or condensation. Amino acid positions (e.g. T20) are counted from the N terminal ignoring gaps (-).

Figure 25 shows the DERA consensus sequence.

Figure 26 shows a schematic of the screen for identifying DERA variants with improved production of 3-hydroxybutanal.

5

Examples

Methods and materials - cloning, expression and activity assay for gene(s) for engineering into Acetogens to produce 1,3-butanediol

10

The approach to construction of the 1,3-butanediol pathway in a chosen host will depend on the pathway genes already present in the host organism. Those endogenous genes considered suitable for pathway construction may be overexpressed to ensure adequate flux through the pathway to 1,3-butanediol.

15

Metabolic engineering steps required to generate a 1,3-butanediol production strain will depend on whether pyruvate or acetyl CoA or both are selected as the source of acetaldehyde. Subsequent conversion of acetaldehyde is common to all routes. For example, for Route 1, acetaldehyde is derived from acetyl CoA via acetate. Acetate is a natural acetogen product which can accumulate to 10s grams per litre. For example 44g/l was obtained from the acetogen *Acetobacterium woodii* growing on CO₂ and H₂ (Demlar, M. et al. Biotech. Bioeng. 2011, 108, 470). Overexpression of a carboxylic acid reductase, aldehyde ferredoxin oxidoreductase or other enzyme capable of acetate reduction (exemplary sequences given in Table 1) to acetaldehyde in the presence of sufficient reducing equivalents and ATP (if appropriate), allows conversion to acetaldehyde. Other than production of biomass for the fermentation, in this example it is desirable to optimise all carbon flux to acetate or acetyl CoA. Accumulation of by-products which are not required for biosynthesis, such as lactate is avoided by knockout of the respective genes e.g. lactate dehydrogenase (Example 5) overproduction of metabolites required for cell synthesis such as malate or fumarate is avoided by adequate, balanced, carbon flux to avoid bottle necks.

20

25

30

Direct conversion of acetyl CoA to acetaldehyde using acetaldehyde dehydrogenase (overexpression of an endogenous enzyme, or introduction of, for

example eutE, Table 2) can operate in the absence of acetate accumulation (Route 2) or alongside acetate accumulation where flux is directed to acetaldehyde directly or via acetate. The route chosen may be influenced by the energetics requirement of organism which can be related to the feedstock provided. It is most preferable to convert a primary central metabolic intermediate to acetaldehyde directly. If the bioenergetics allow loss of ATP synthesis from acetyl CoA conversion to acetate, acetate accumulation can be prevented in an acetogen by knockout of one or more phosphotransacetylase (pta) or acetate kinase (ack) genes (Example 6 and 8). Furthermore, acetate accumulation may be prevented by natural regulation, or by mutation which directs flux away from acetate synthesis while maintaining Wood Ljungdahl pathway activity. For example growth of the acetogen *Moorella thermoacetica* (renamed from *C. thermoaceticum*) on CO and methanol in the presence of nitrate led to no acetate accumulation due to repression of key Wood Ljungdahl related gene expression (Seifritz, C. et al. J. Bacteriol. 1993, 175, 8008). In that example, sufficient ATP appeared to be provided from nitrate respiration. Acetyl CoA can also be converted to acetaldehyde via pyruvate (Route 5) using pyruvate synthase (EC 1.2.7.1, Table 3). In this example it is particularly desirable to avoid loss of carbon flux to products derived from pyruvate other than acetaldehyde (for example targeting of LDH may be desired), Example 5). If pyruvate is the primary central metabolic intermediate, it is preferable to convert pyruvate to acetaldehyde directly (Route 4) via decarboxylation using example sequences in Table 4 and to optimise the flux by targeting of undesired pathways (for example LDH or pyruvate formate lyase). However, it may alternatively be preferred to allow conversion of pyruvate to acetyl CoA, Route 3 (the natural metabolic route prior to entry to the TCA cycle) or by using gene sequence examples shown in Table 3. It is desirable that the maximum amount of acetaldehyde be converted to 3-hydroxybutanal via an overexpressed endogenous or heterologous DERA (example sequences are shown in Table 6). Hence, loss to oxidation or reduction products (acetate or ethanol) should be avoided by knockout of undesired genes, for example, short chain alcohol dehydrogenases highly active on acetaldehyde, or non-acetylating acetaldehyde dehydrogenase (e.g. EC 1.2.1.5). Reduction of 3-hydroxybutanal is achieved by overexpression of an endogenous, or introduction of a heterologous alcohol dehydrogenase or aldehyde reductase which shows

preference for C4 aldehydes (3-hydroxybutanal) relative to C2 aldehydes (acetaldehyde) e.g Example 9. Such examples are discussed above and example sequences shown in Table 7.

5 The introduction of a heterologous gene into an acetogen is described in Example 7, this method can be cross applied to the introduction of any heterologous gene, for example, a gene within a 1,3-butanediol pathway.

Example 1 – Routes for acetaldehyde and 1,3-BDO synthesis from central metabolites

10

The overall conversion of acetyl CoA to 1,3-butanediol is accomplished in either 3 or 5 steps depending on the route taken (Figure 3) and in 1 or 3 steps to the common pathway intermediate acetaldehyde. Other products obtainable via acetaldehyde and 3-hydroxybutanal are described above (see also Figure 17).

15

The overall conversion of pyruvate to 1,3-butanediol is accomplished in 3 or 4 steps depending on the route taken (Figure 3) and in 1 or 2 steps to the common pathway intermediate acetaldehyde.

20

The two steps from acetaldehyde to 1,3-butanediol are common to all 1,3-butanediol synthetic routes.

25

The description of the pathways is provided as routes for acetaldehyde synthesis (Route 1,2, 3, 4, 5 and 6) and the subsequent conversion of acetaldehyde to 1,3-butanediol via the aldol condensation catalysed by DERA.

Route 1 – Conversion of acetyl CoA to acetaldehyde via acetate

30

Acetogens naturally produce acetate in high yield from sugars, or C1 feedstocks (syngas, CO₂/H₂, CO₂ and methanol) via conversion of acetyl CoA derived from the Wood Ljungdahl pathway. Yields are typically approximately 80% of theoretical or greater, for example, A.E. Bainotti et al., 1988. Journal of fermentation and bioengineering, 85(2), 223-229. Although it is anticipated that even higher yields may be achievable, for example, via modification of the Wood Ljungdahl pathway

which converts CO₂, H₂, CO, or methanol to acetyl CoA or via optimisation of the growth medium. Fundamentally, in acetogens the general fate of acetyl CoA is either to go towards formation or maintenance of biomass, or synthesis of acetate which generates ATP. As the Wood Ljungdahl pathway requires an ATP, in most cases (depending on the growth conditions), acetate synthesis is required in order to balance the energy needs of the system. Acetate is a major natural product of most acetogens.

Acetate can be reduced to acetaldehyde using a carboxylic acid reductase enzyme. Such enzyme activity mainly uses either reduced ferredoxin (aldehyde ferredoxin oxidoreductase) or ATP to drive the thermodynamically unfavourable reduction of a carboxylic acid moiety and tend to be classified in EC 1.2.7.5, EC 1.2.1.30, EC 1.2.99.6. or EC 1.2.1.3. The term carboxylic acid reductase and aldehyde oxidoreductase are used interchangeably in the literature. Aldehyde dehydrogenase is also used to describe enzymes capable of carboxylic acid reduction.

An example of a well-studied carboxylic acid reductase can be found in *Nocardia iowensis* which catalyzes the magnesium, ATP and NADPH-dependent reduction of carboxylic acids to their corresponding aldehydes (Venkitasubramanian et al., J. Biol. Chem. 282:478-485 (2007)). This enzyme is encoded by the car gene and was cloned and functionally expressed in *E. coli* (Venkitasubramanian et al., J. Biol. Chem. 282:478-485 (2007)). Expression of the npt gene product improved activity of the enzyme via post-translational modification. The npt gene encodes a specific phosphopantetheine transferase (PPTase) that converts the inactive apo-enzyme to the active holo-enzyme. The natural substrate of this enzyme is vanillic acid, and the enzyme exhibits broad acceptance of aromatic and aliphatic substrates as small as lactic acid (Venkitasubramanian *et al.*, in *Biocatalysis in the Pharmaceutical and Biotechnology Industries*, ed. R. N. Patel, Chapter 15, pp. 425-440, CRC Press LLC, Boca Raton, Fla. (2006)). Activity towards acetate was not discussed. However, high activity towards lactate suggests that the enzyme is capable of accepting molecules containing as few as three carbons. Hence, this enzyme may potentially be used for acetate reduction in either its native form or as an evolved enzyme.

A further well studied enzyme is the example from *Mycobacterium marinum* which has a wild type substrate preference for C6 to C18 acids (Kalim Akhtar, M. et al. PNAS, 2013, 110, 87). Enzymes capable of carboxylic acid reduction may be evolved or mutated as described above to increase activity towards acetate using enzyme evolution techniques common in the art. The griC and griD genes from *Streptomyces* also code for a carboxylic acid reductase with diverse capability for acid reduction Suzuki et al. 2007. J. Antibiot. 60 (6) 380.

Aldehyde ferredoxin oxidoreductase enzymes use ferredoxin not ATP to drive the carboxylate reduction and are present in many acetogens and other organisms (White, H et al. Biol. Chem Hoppe Seler 1991, 372 (11) 999; White, H and Simon, H. Arch. Microbiol, 1992, 158, 81; Fraisse. L and Simon, H. Arch. Microbiol. 1988, 150,381; (Basen et. al. 2014. PNAS, 111 (49), 17618). The carboxylic acid reducing enzyme from *Moorella thermoacetica* has been purified and characterised ,White, H. et al. Eur. J Biochem, 1989, 184, 89. Further, using propionate reduction to propionaldehyde, the specific activity was shown to increase when the corresponding aldehyde was removed during the reaction. In the case of application of such an enzyme to this invention, the product acetaldehyde would be continuously removed by the DERA enzyme and would not be expected to accumulate significantly. Huber, C. et al. Arch. Microbiol, 1995, 64, 110.

Example genes for acetate reduction are shown in Table 1. The aldehyde oxidoreductase (AOR) genes CLJU_20110 and CLJU_20210 from *Clostridium ljungdahlii* are reported to reduce acetate to acetaldehyde, Kopke, M. et al. PNAS, 2010, 107, 15305. Hence, demonstrating the activity of a wild type enzyme towards the target reduction. Various authors have also described conditions under which AOR enzymes are induced in ethanologenic acetogens for synthesis of ethanol from acetate via acetaldehyde, (Mock et al. 2015, Energy conservation associated with ethanol formation from H₂ and CO₂ in *Clostridium autoethanogenum* involving electron bifurcation, J. Bacteriol. 197 (18) 2965; Nalakath, H. et al. 2015, Bioresource Technology 186, 122.). As described above, in organisms of the invention, it may be desired to target or knockout alcohol dehydrogenases responsible for ethanol production from the intermediate acetaldehyde to thereby

promote synthesis of 3-hydroxybutanal from acetaldehyde catalysed by a DERA enzyme.

- 5 A further source of aldehyde ferredoxin oxidoreductase are the hyperthermophiles, Thermococcus sp. (Kesen, J.H. J. Bacteriol. 1995, 177, 4757 and Pyrococcus sp. (Basen et. al. 2014. PNAS, 111 (49), 17618 where this enzyme has been used to effectively synthesise ethanol from acetate via acetaldehyde driven by carbon monoxide. Although described mainly for oxidation of aldehydes to the corresponding acids, reduction of acetate is also mentioned. The Km values for
- 10 acids appear higher than for the aldehydes, standard enzyme evolution techniques known in the art could be used to improve the enzyme's efficiency for acetate reduction. The use of aldehyde ferredoxin oxidoreductase in the aldehyde oxidation direction is further described by Kletzin, A., et al. J. Bacteriol. 1995,177, 4817.
- 15 An aldehyde dehydrogenase (*aldH*) from E.coli has been shown to reduce 3-hydroxypropionic acid to the corresponding aldehyde as well as the preferred oxidation of 3-hydroxypropionaldehyde, Ji-Eun, J. *et al.*, Appl. Microbiol. Biotechnol 2008. 81, 51. This enzyme was also shown to oxidise acetaldehyde to acetate. Hence, as these authors have shown the enzyme to be reversible, activity towards
- 20 reduction of acetate would be expected.

Table 1. Examples of genes expressing enzymes for application to the reduction of acetate to acetaldehyde (Activity A).

UniProt Entry	NCBI Gene ID	EC number	Gene names	Protein names	Organism
P23883	1293453 4 947003	1.2.1.5	puuC aldH b1300 JW1293	Aldehyde dehydrogenase PuuC (EC 1.2.1.5) (3-hydroxypropionaldehyde dehydrogenase) (Gamma-glutamyl-gamma-aminobutyraldehyde	Escherichia coli (strain K12)

				dehydrogenase) (Gamma-Glu-gamma-aminobutyraldehyde dehydrogenase)	
D8GIZ8	9445627		CLJU_c 20110	Predicted tungsten-containing aldehyde ferredoxin oxidoreductase	Clostridium ljungdahlii (strain ATCC 55383 / DSM 13528 / PETC)
D8GJ08	9445637		CLJU_c 20210	Predicted tungsten-containing aldehyde ferredoxin oxidoreductase	Clostridium ljungdahlii (strain ATCC 55383 / DSM 13528 / PETC)
Q2RG52	3831332	1.2.7.5	Moth_2 300	Aldehyde ferredoxin oxidoreductase (EC 1.2.7.5)	Moorella thermoacetica (strain ATCC 39073)
Q2RKJ9	3830998	1.2.7.5	Moth_0 722	Aldehyde ferredoxin oxidoreductase (EC 1.2.7.5)	Moorella thermoacetica (strain ATCC 39073)
Q2RM4 7	3831866	1.2.7.5	Moth_0 154	Aldehyde ferredoxin oxidoreductase (EC 1.2.7.5)	Moorella thermoacetica (strain ATCC 39073)
C9QU34	1270557 0;128730 31	1.2.7.5	ydhV EcDH1_ 1969	Aldehyde ferredoxin oxidoreductase (EC 1.2.7.5) (Putative	Escherichia coli (strain ATCC 33849)

			ECDH1 ME8569 _1617	oxidoreductase)	/ DSM 4235 / NCIB 12045 / K12 / DH1)
E8Y7H0	1177694 2;127633 67	1.2.7.5	ydhV EKO11_ 2102 KO11_1 4380	Aldehyde ferredoxin oxidoreductase (EC 1.2.7.5) (Putative oxidoreductase)	Escherichia coli (strain ATCC 55124 / KO11)
B1IQ83	6068384;	1.2.7.5	EcolC_1 958	Aldehyde ferredoxin oxidoreductase (EC 1.2.7.5)	Escherichia coli (strain ATCC 8739 / DSM 1576 / Crooks)
E0IXM3	1269559 9;127538 70	1.2.7.5	ydhV ECW_m 1840 WFL_09 015 EschW DRAFT _0881	Aldehyde ferredoxin oxidoreductase (EC 1.2.7.5) (Predicted oxidoreductase) (Putative oxidoreductase)	Escherichia coli (strain ATCC 9637 / CCM 2024 / DSM 1116 / NCIMB 8666 / NRRL B- 766 / W)
C6EA37	8114754; 8160069; 8181416	1.2.7.5	yagT B21_00 248 ECBD_ 3371 ECD_00 245	(2Fe-2S)-binding domain protein (Aldehyde ferredoxin oxidoreductase, Fe-S subunit, subunit of aldehyde ferredoxin oxidoreductase) (EC 1.2.7.5) (Predicted xanthine dehydrogenase, 2Fe- 2S subunit)	Escherichia coli (strain B / BL21-DE3)
C6EA38	8114753; 8160070;	1.2.7.5	yagS B21_00	Aldehyde ferredoxin oxidoreductase, FAD-	Escherichia coli (strain B

	8181415		247 ECBD_ 3372 ECD_00 244	binding subunit, subunit of aldehyde ferredoxin oxidoreductase (EC 1.2.7.5) (Molybdopterin dehydrogenase FAD-binding) (Predicted oxidoreductase with FAD-binding domain)	/ BL21-DE3)
C6EA39	8114752; 8160071; 8181414	1.2.7.5	yagR B21_00 246 ECBD_ 3373 ECD_00 243	Aldehyde ferredoxin oxidoreductase: molybdenum cofactor-binding subunit, subunit of aldehyde ferredoxin oxidoreductase (EC 1.2.7.5) (Aldehyde oxidase and xanthine dehydrogenase molybdopterin binding) (Predicted oxidoreductase with molybdenum-binding domain)	Escherichia coli (strain B / BL21-DE3)
C6ECT2	8113808; 8157240; 8183188	1.2.7.5	ydhV B21_01 632 ECBD_ 1972 ECD_01 642	Aldehyde ferredoxin oxidoreductase (EC 1.2.7.5) (Predicted oxidoreductase)	Escherichia coli (strain B / BL21-DE3)
K3JY12		1.2.7.5	yagR EC3006 _0366	Aldehyde oxidase and xanthine dehydrogenase (EC 1.2.7.5)	Escherichia coli 3006
A0A024		1.2.7.5	PGA_03	Aldehyde ferredoxin	Escherichia

KWW2			435	oxidoreductase (EC 1.2.7.5)	coli D6-113.11
A0A024 LHN1		1.2.7.5	PGC_20250	Aldehyde ferredoxin oxidoreductase (EC 1.2.7.5)	Escherichia coli D6-117.29
Q56303	16548761	1.2.7.5	for OCC_05029	Tungsten-containing formaldehyde ferredoxin oxidoreductase (EC 1.2.7.5)	Thermococcus litoralis (strain ATCC 51850 / DSM 5473 / JCM 8560 / NS-C)
Q6RKB1		1.2.1.-; 1.2.1.30	car	Carboxylic acid reductase (CAR) (EC 1.2.1.-) (ATP/NADPH-dependent carboxylic acid reductase) (Aryl aldehyde oxidoreductase) (EC 1.2.1.30)	Nocardia iowensis
A1YCA5		2.7.8.7	npt	4'-phosphopantetheinyl transferase Npt (PPTase) (EC 2.7.8.7)	Nocardia iowensis
Q5YY80	3108003		NFA_20150	Putative carboxylic acid reductase	Nocardia farcinica (strain IFM 10152)
Q5YSD9	3109498		NFA_40540	Putative phosphopantetheinyl transferase	Nocardia farcinica (strain IFM 10152)
B1VMZ4	6209683		SGR_6790	Putative carboxylic acid reductase	Streptomyces griseus subsp. griseus

					(strain JCM 4626 / NBRC 13350)
B1VRS6	6214265		SGR_665	Putative phosphopantetheinyl transferase	Streptomyces griseus subsp. griseus (strain JCM 4626 / NBRC 13350)
B1VTI3	6210972;		griD SGR_4244	Arylcarboxylate reductase component	Streptomyces griseus subsp. griseus (strain JCM 4626 / NBRC 13350)
B1VTI2	6215140;		griC SGR_4243	Arylcarboxylate reductase component	Streptomyces griseus subsp. griseus (strain JCM 4626 / NBRC 13350)
Q51739	1468181	1.2.7.5	aor, AOR_P YRFU	Tungsten-containing aldehyde ferredoxin oxidoreductase	Pyrococcus furiosus (strain ATCC 43587 / DSM 3638 / JCM 8422 / Vc1)

Additional car and npt genes and other genes coding for enzymes capable of (or involved with) carboxylic acid reduction (Activity A) can be identified based on sequence homology to those examples in Table 1.

5 *Route 2 – Conversion of acetyl CoA to acetaldehyde directly*

Acetaldehyde can be synthesised from acetyl CoA via the reversible enzyme acetaldehyde dehydrogenase EC 1.2.1.10.

- 10 The gene coding for this enzyme can be found in a wide range of different organisms such as: *Acinetobacter* sp.; *Burkholderia xenovorans*; *E. coli*; *Clostridium beijerinckii*, (Run-Tao, Y and Jiann-Shin, C. 1990, Appl. Environ. Microbiol. 56, 2591; Appl. Environ Microbiol, 1999, 65 (11) 4973); *Clostridium kluyveri*; *Pseudomonas* sp. (Platt, A *et al.* 1995, Microbiol., 141, 2223; Soonyoung, H. *et al.* 1999, Biochem. Biophys. Res. Comm. 256, 469) *Propionibacterium* sp. and *Thermoanaerobacter ethanolicus*.

- 20 Many acetogens also have annotated acetaldehyde dehydrogenase genes e.g. *Moorella thermoacetica* (Moth_1776). *Acetobacterium woodii* (Arch. Microbiol, 1992, 158, 132). *Clostridium ljungdahlii* CLJU_c11960.

- 25 The *eutE* gene from the *eut* operon also encodes for an acetaldehyde dehydrogenase. The *eutE* gene from *Salmonella enterica* has been cloned into *E.coli* and shown to efficiently produce acetaldehyde from growth on glucose via acetyl CoA reduction (Huilin, Z. *et al.* 2011. Appl. Environ. Microbiol. 77, 6441). This is an excellent demonstration of an enzyme capable of efficiently providing acetaldehyde substrate for a DERA type enzyme catalysed aldol condensation in a 1,3-butanediol pathway from acetyl CoA. 1,3-Butanediol production using *eutE* to deliver acetaldehyde to DERA from acetyl CoA in a 1,3-BDO pathway, is shown in Example 10.
- 30

Table 2. Examples of genes expressing enzymes for the conversion of acetyl CoA to acetaldehyde.(Activity B).

UniProt Entry	NCBI Gene ID	EC number	Gene names	Protein names	Organism
H6LJM8	11871155	1.1.1.1; 1.2.1.10	adhE Awo_c06310	Bifunctional acetaldehyde- CoA/alcohol dehydrogenase (EC 1.1.1.1) (EC 1.2.1.10)	Acetobacterium woodii (strain ATCC 29683 / DSM 1030 / JCM 2381 / KCTC 1655)
Q79AF6	4010700	1.2.1.10; 1.2.1.87	bphJ Bxeno_C1122 Bxe_C1188	Acetaldehyde dehydrogenase 4 (EC 1.2.1.10) (Acetaldehyde dehydrogenase [acetylating] 4) (Propanal dehydrogenase (CoA- propanoylating)) (EC 1.2.1.87)	Burkholderia xenovorans (strain LB400)
Q143P8	4004910	1.2.1.10	Bxeno_A0903 Bxe_A3547	Acetaldehyde dehydrogenase 1 (EC 1.2.1.10) (Acetaldehyde dehydrogenase [acetylating] 1)	Burkholderia xenovorans (strain LB400)
Q13VU2	4002974	1.2.1.10	amnH Bxeno_A3259 Bxe_A1151	Acetaldehyde dehydrogenase 2 (EC 1.2.1.10) (Acetaldehyde dehydrogenase [acetylating] 2)	Burkholderia xenovorans (strain LB400)
Q13QH7	4007178	1.2.1.10	Bxeno_B0694 Bxe_B2326	Acetaldehyde dehydrogenase 3 (EC 1.2.1.10)	Burkholderia xenovorans (strain LB400)

				(Acetaldehyde dehydrogenase [acetylating] 3)	
Q716S8	5294993	1.2.1.10	ald	Aldehyde dehydrogenase (EC 1.2.1.10) (Coenzyme A acylating aldehyde dehydrogenase) (EC 1.2.1.10) (Coenzyme A-acylating aldehyde dehydrogenase) (EC 1.2.1.10)	Clostridium beijerinckii (Clostridium MP)
D8GIC3	9444813; 9447589	1.2.1.-	CLJU_c11960 CLJU_c39730	Predicted acetaldehyde dehydrogenase (EC 1.2.1.-)	Clostridium ljungdahlii (strain ATCC 55383 / DSM 13528 / PETC)
D8GID4	9447600	1.2.1.-	CLJU_c39840	Predicted acetaldehyde dehydrogenase (EC 1.2.1.-)	Clostridium ljungdahlii (strain ATCC 55383 / DSM 13528 / PETC)
P77580	12932628; 945008	1.2.1.10	mhpF mhpE b0351 JW0342	Acetaldehyde dehydrogenase (EC 1.2.1.10) (Acetaldehyde dehydrogenase [acetylating])	Escherichia coli (strain K12)
A4IT43	4968078	1.2.1.10	nbaJ GTNG_3152	Acetaldehyde dehydrogenase (EC 1.2.1.10) (Acetaldehyde dehydrogenase [acetylating])	Geobacillus thermodenitrificans (strain NG80-2)

Q2RHL2	3832442	1.2.1.10	Moth_1776	Acetaldehyde dehydrogenase (EC 1.2.1.10) (Acetaldehyde dehydrogenase [acetylating])	Moorella thermoacetica (strain ATCC 39073)
C8CEC3		1.2.1.10	nahO	Acetaldehyde dehydrogenase (EC 1.2.1.10) (Acetaldehyde dehydrogenase [acetylating])	Pseudomonas aeruginosa
P41793	1253985		eutE STM2463	Ethanolamine utilization protein EutE	Salmonella typhimurium (strain LT2 / SGSC1412 / ATCC 700720)

Additional genes coding for enzymes capable of acetyl CoA conversion to acetaldehyde (Activity B) can be identified based on sequence homology to those examples in Table 2

5

Route 3. Conversion of pyruvate to acetaldehyde via acetyl CoA

The conversion of pyruvate to acetyl CoA can be carried out using an enzyme such as EC 1.2.7.1 (pyruvate synthase, pyruvate:ferredoxin oxidoreductase). These ferredoxin linked enzymes are particularly common in anaerobes such as the acetogens, but are also present in other aerobic or facultatively anaerobic organisms such as *Hydrogenobacter thermophilus*.

10

The pyruvate dehydrogenase complex is also a central metabolic enzyme well understood in the art which is responsible for conversion of pyruvate (for example, generated from glycolysis) to acetyl CoA for entry into the TCA cycle.

15

The subsequent conversion of acetyl CoA to acetaldehyde is described in Route 2.

Table 3. Examples of genes expressing enzymes for the conversion of pyruvate to acetyl CoA (Activity C).

UniProt Entry	NCBI Gene ID	Gene names	Protein names	Organism
H6LJ55	11873437	porB Awo_c06200	Pyruvate:ferredoxin oxidoreductase, beta subunit PorB (EC 1.2.7.1)	Acetobacterium woodii (strain ATCC 29683 / DSM 1030 / JCM 2381 / KCTC 1655)
H6LJ56	11873438	porA Awo_c06210	Pyruvate:ferredoxin oxidoreductase, alpha subunit PorA (EC 1.2.7.1)	Acetobacterium woodii (strain ATCC 29683 / DSM 1030 / JCM 2381 / KCTC 1655)
U1VQ53		BTCBT_00551 7	Pyruvate synthase subunit porA (EC 1.2.7.1)	Bacillus thuringiensis T01-328
U1W7V0		BTCBT_00551 6	Pyruvate synthase subunit porB (EC 1.2.7.1)	Bacillus thuringiensis T01-328
E5ZLB4		nifJ CSU_1742	Pyruvate synthase (EC 1.2.7.1)	Campylobacter jejuni subsp. jejuni 327
A5I7E8	5187682; 5400580	CBO3423 CLC_3367	Putative subunit of pyruvate:flavodoxin oxidoreductase (EC 1.2.7.1)	Clostridium botulinum (strain Hall / ATCC 3502 / NCTC 13319 / Type A)
A5I7E6	16691482; 5186833	CBO3421	Putative subunit of pyruvate:flavodoxin oxidoreductase (EC 1.2.7.1)	Clostridium botulinum (strain Hall / ATCC 3502 / NCTC 13319 / Type

				A)
A5N1K8	5390957	porB CKL_2996	PorB (EC 1.2.7.1)	Clostridium kluyveri (strain ATCC 8527 / DSM 555 / NCIMB 10680)
A5N1L1	5393792	porC CKL_2999	PorC (EC 1.2.7.1)	Clostridium kluyveri (strain ATCC 8527 / DSM 555 / NCIMB 10680)
A5N1K9	5390958	porA CKL_2997	PorA (EC 1.2.7.1)	Clostridium kluyveri (strain ATCC 8527 / DSM 555 / NCIMB 10680)
A5N1L0	5393791	porD CKL_2998	PorD (EC 1.2.7.1)	Clostridium kluyveri (strain ATCC 8527 / DSM 555 / NCIMB 10680)
Q9LBG1	12100419; 8773721	PorE	Pyruvate:ferredoxin oxidoreductase epsilon subunit (EC 1.2.7.1)	Hydrogenobacter thermophilus
Q9LBF7	12100415; 8773666	porG	Pyruvate:ferredoxin oxidoreductase gamma subunit (EC 1.2.7.1)	Hydrogenobacter thermophilus
Q9LBF8	12100416; 8773723	porB	Pyruvate:ferredoxin oxidoreductase beta subunit (EC 1.2.7.1)	Hydrogenobacter thermophilus
Q9LBF9	12100417; 8773720	porA	Pyruvate:ferredoxin oxidoreductase alpha subunit (EC 1.2.7.1)	Hydrogenobacter thermophilus
Q9LBG0	12100418; 8773719	porD	Pyruvate:ferredoxin oxidoreductase delta subunit (EC 1.2.7.1)	Hydrogenobacter thermophilus

P80900	9704020	porA MTBMA_c031 40	Pyruvate synthase subunit PorA (EC 1.2.7.1) (Pyruvate oxidoreductase alpha chain) (POR) (Pyruvic-ferredoxin oxidoreductase subunit alpha)	Methanothermobacter marburgensis (strain DSM 2133 / 14651 / NBRC 100331 / OCM 82 / Marburg) (Methanobacterium thermoautotrophicum)
P80901	9704019	porB MTBMA_c031 30	Pyruvate synthase subunit PorB (EC 1.2.7.1) (Pyruvate oxidoreductase beta chain) (POR) (Pyruvic-ferredoxin oxidoreductase subunit beta)	Methanothermobacter marburgensis (strain DSM 2133 / 14651 / NBRC 100331 / OCM 82 / Marburg) (Methanobacterium thermoautotrophicum)
P80902	9704022	porC MTBMA_c031 60	Pyruvate synthase subunit PorC (EC 1.2.7.1) (Pyruvate oxidoreductase gamma chain) (POR) (Pyruvic-ferredoxin oxidoreductase subunit gamma)	Methanothermobacter marburgensis (strain DSM 2133 / 14651 / NBRC 100331 / OCM 82 / Marburg) (Methanobacterium thermoautotrophicum)
Q2RH65	3830848	Moth_1924	Pyruvate ferredoxin oxidoreductase, gamma subunit (EC 1.2.7.1)	Moorella thermoacetica (strain ATCC 39073)
Q2RLH9	3832620	Moth_0376	Pyruvate ferredoxin oxidoreductase, gamma subunit (EC	Moorella thermoacetica (strain ATCC

			1.2.7.1)	39073)
Q2RI42	3832737	Moth_1591	Pyruvate ferredoxin oxidoreductase, beta subunit (EC 1.2.7.1)	Moorella thermoacetica (strain ATCC 39073)
Q2RH67	3830846	Moth_1922	Pyruvate ferredoxin oxidoreductase, alpha subunit (EC 1.2.7.1)	Moorella thermoacetica (strain ATCC 39073)
Q2RLH7	3832622	Moth_0378	Pyruvate ferredoxin oxidoreductase, alpha subunit (EC 1.2.7.1)	Moorella thermoacetica (strain ATCC 39073)
Q2RH68	3830845	Moth_1921	Pyruvate ferredoxin oxidoreductase, beta subunit (EC 1.2.7.1)	Moorella thermoacetica (strain ATCC 39073)
Q51804	1468831	porA PF0966	Pyruvate synthase subunit PorA (EC 1.2.7.1) (Pyruvate oxidoreductase alpha chain) (POR) (Pyruvic-ferredoxin oxidoreductase subunit alpha)	Pyrococcus furiosus (strain ATCC 43587 / DSM 3638 / JCM 8422 / Vc1)
Q51805	1468830	porB PF0965	Pyruvate synthase subunit PorB (EC 1.2.7.1) (Pyruvate oxidoreductase beta chain) (POR) (Pyruvic-ferredoxin oxidoreductase subunit beta)	Pyrococcus furiosus (strain ATCC 43587 / DSM 3638 / JCM 8422 / Vc1)
O05651	896831	porA TM_0017	Pyruvate synthase	Thermotoga

			subunit PorA (EC 1.2.7.1) (Pyruvate oxidoreductase alpha chain) (POR) (Pyruvic-ferredoxin oxidoreductase subunit alpha)	maritima (strain ATCC 43589 / MSB8 / DSM 3109 / JCM 10099)
O05650	896829	porC porG TM_0015	Pyruvate synthase subunit PorC (EC 1.2.7.1) (Pyruvate oxidoreductase gamma chain) (POR) (Pyruvic-ferredoxin oxidoreductase subunit gamma)	Thermotoga maritima (strain ATCC 43589 / MSB8 / DSM 3109 / JCM 10099)
Q56317	896832	porB TM_0018	Pyruvate synthase subunit PorB (EC 1.2.7.1) (Pyruvate oxidoreductase beta chain) (POR) (Pyruvic-ferredoxin oxidoreductase subunit beta)	Thermotoga maritima (strain ATCC 43589 / MSB8 / DSM 3109 / JCM 10099)

Additional genes coding for enzymes capable of the conversion of pyruvate to acetyl CoA (Activity C) can be identified based on sequence homology to those examples in Table 3, or to common sequences for the pyruvate dehydrogenase complex.

Route 4. Conversion of pyruvate to acetaldehyde directly

The conversion of pyruvate to acetaldehyde is well known in the art. Pyruvate decarboxylase is a homotetrameric enzyme (EC 4.1.1.1) that catalyses the decarboxylation of pyruvic acid to acetaldehyde and carbon dioxide in the cytoplasm

of prokaryotes, and in the mitochondria of eukaryotes. It is also called 2-oxo-acid carboxylase, alpha-ketoacid carboxylase, and pyruvic decarboxylase. Under anaerobic conditions, this enzyme is part of the fermentation process that occurs in yeast, especially of the *Saccharomyces* genus, to produce ethanol by fermentation.

5 Pyruvate decarboxylase starts this process by converting pyruvate into acetaldehyde and carbon dioxide.

The pyruvate ferredoxin oxidoreductase from *Pyrococcus furiosus* (Table 3) has also been shown to catalyse pyruvate decarboxylation to acetaldehyde (Ma, K. *et al.*
10 1997. PNAS, 94, 9608).

Examples 12,13 and 14 show the production of 1,3-butanediol using pyruvate decarboxylase to deliver acetaldehyde to DERA from pyruvate, in a novel, unnatural 1,3-BDO pathway. Table 4 . Examples of genes expressing enzymes for application
15 to the decarboxylation of pyruvate to acetaldehyde (Activity D).

Uniprot Entry	NCBI GeneID	Gene names	Protein names	Organism
O82647	829444	PDC1 At4g33070 F4I10.4	Pyruvate decarboxylase 1 (AtPDC1)	Arabidopsis thaliana (Mouse-ear cress)
Q9FFT4	835587	PDC2 At5g54960 MBG8.23	Pyruvate decarboxylase 2 (AtPDC2)	Arabidopsis thaliana (Mouse-ear cress)
Q9M039	831414	PDC3 At5g01330 T10O8.40	Pyruvate decarboxylase 3 (AtPDC3)	Arabidopsis thaliana (Mouse-ear cress)
Q9M040	830867	PDC4 At5g01320 T10O8.30	Pyruvate decarboxylase 4 (AtPDC4)	Arabidopsis thaliana (Mouse-ear cress)
Q2UKV4	5991796	pdca AO090003000661	Pyruvate decarboxylase	Aspergillus oryzae (strain ATCC 42149 / RIB 40) (Yellow koji mold)
P51844		pdca pdc	Pyruvate	Aspergillus parasiticus

			decarboxylase	
Q0CNV1	4320296	pdca ATEG_04633	Pyruvate decarboxylase	Aspergillus terreus (strain NIH 2624 / FGSC A1156)
P83779	3642780	PDC11 PDC1 CaO19.10395 CaO19.2877	Pyruvate decarboxylase	Candida albicans (strain SC5314 / ATCC MYA-2876) (Yeast)
Q6FJA3	2891742	PDC1 PDC CAGL0M07920g	Pyruvate decarboxylase	Candida glabrata (strain ATCC 2001 / CBS 138 / JCM 3761 / NBRC 0622 / NRRL Y-65) (Yeast) (Torulopsis glabrata)
P87208	2872690	pdca AN4888	Pyruvate decarboxylase	Emmericella nidulans (strain FGSC A4 / ATCC 38163 / CBS 112.46 / NRRL 194 / M139) (Aspergillus nidulans)
P34734		PDC	Pyruvate decarboxylase	Hanseniaspora uvarum (Yeast) (Kloeckera apiculata)
Q12629	2894295	PDC1 KLLA0E16357g	Pyruvate decarboxylase	Kluyveromyces lactis (strain ATCC 8585 / CBS 2359 / DSM 70799 / NBRC 1267 / NRRL Y-1140 / WM37) (Yeast) (Candida sphaerica)
P33149		PDC1	Pyruvate decarboxylase	Kluyveromyces marxianus (Yeast) (Candida kefyr)
Q4WXX9	3511715	pdca AFUA_3G11070	Pyruvate decarboxylase	Neosartorya fumigata (strain ATCC MYA-4609 / Af293 / CBS 101355 / FGSC A1100) (Aspergillus fumigatus)
P33287	3875734	cfp pdc-1	Pyruvate	Neurospora crassa (strain

		NCU02193	decarboxylase (8-10 nm cytoplasmic filament- associated protein) (P59NC)	ATCC 24698 / 74-OR23- 1A / CBS 708.71 / DSM 1257 / FGSC 987)
A2Y5L9		PDC1 OsI_019612	Pyruvate decarboxylase 1 (PDC)	Oryza sativa subsp. indica (Rice)
A2XFI3		PDC2 OsI_010826	Pyruvate decarboxylase 2 (PDC)	Oryza sativa subsp. indica (Rice)
A2YQ76		PDC3 OsI_026469	Pyruvate decarboxylase 3 (PDC)	Oryza sativa subsp. indica (Rice)
Q0DHF6	4339066	PDC1 Os05g0469600 LOC_Os05g39310 OsJ_018109 OSJNBa0052E20. 2	Pyruvate decarboxylase 1 (PDC)	Oryza sativa subsp. japonica (Rice)
Q10MW3	4332519	PDC2 Os03g0293500 LOC_Os03g18220	Pyruvate decarboxylase 2 (PDC)	Oryza sativa subsp. japonica (Rice)
Q0D3D2	4344382	PDC3 Os07g0693100 LOC_Os07g49250 OsJ_024667	Pyruvate decarboxylase 3 (PDC)	Oryza sativa subsp. japonica (Rice)
P51850		PDC1	Pyruvate decarboxylase 1 (PDC)	Pisum sativum (Garden pea)
P06169	850733	PDC1 YLR044C L2104	Pyruvate decarboxylase isozyme 1 (EC	Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's

			4.1.1.-)	yeast)
P16467	850825	PDC5 YLR134W L3133 L9606.7	Pyruvate decarboxylase isozyme 2 (EC 4.1.1.-)	Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast)
P26263	852978	PDC6 YGR087C	Pyruvate decarboxylase isozyme 3	Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast)
O42873	2543400	SPAC3G9.11c	Putative pyruvate decarboxylase C3G9.11c	Schizosaccharomyces pombe (strain 972 / ATCC 24843) (Fission yeast)
Q09737	3361478	SPAC13A11.06 SPAC3H8.01	Putative pyruvate decarboxylase C13A11.06	Schizosaccharomyces pombe (strain 972 / ATCC 24843) (Fission yeast)
Q9P7P6	2542602	SPAC186.09	Probable pyruvate decarboxylase C186.09	Schizosaccharomyces pombe (strain 972 / ATCC 24843) (Fission yeast)
P28516	542376	PDC1 PDC	Pyruvate decarboxylase 1 (PDC)	Zea mays (Maize)
P06672	3188496	pdv ZMO1360	Pyruvate decarboxylase (PDC)	Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4)

Additional genes coding for enzymes capable of pyruvate conversion to acetaldehyde (Activity D) can be identified based on sequence homology to those examples in Table 4.

The conversion of acetyl CoA to pyruvate (Activity E) can be achieved using the reversible enzyme pyruvate ferredoxin oxidoreductase (pyruvate synthase, EC 1.2.7.1). Gene sequences coding for this enzyme are listed in Table 3.

- 5 The subsequent conversion of pyruvate to acetaldehyde is described in Route 4.

Route 6. Conversion of acetate to acetaldehyde via acetyl CoA.

- 10 The conversion of acetate to acetyl CoA can be achieved using acetyl CoA synthetase or a CoA transferase for example, EC 6.2.1.1 or EC 2.8.3.8 and subsequently converted to acetaldehyde via EC 1.2.1.10 (Route 2.). Examples of gene sequences coding for enzymes capable of the conversion of acetate to acetyl CoA are shown in Table 5.

- 15 Table 5 .Examples of genes expressing enzymes for application to the synthesis of acetyl CoA from acetate (Activity F).

UniProt Entry	NCBI Gene ID	Gene names	Protein names	Organism
A4SJM6	4995560	acsA ASA_09 67	Acetyl-coenzyme A synthetase (AcCoA synthetase) (Acs) (EC 6.2.1.1) (Acetate--CoA ligase) (Acyl-activating enzyme)	Aeromonas salmonicida (strain A449)
Q9KWA3	874783	acsA riorf81	Acetyl-coenzyme A synthetase (AcCoA synthetase) (Acs) (EC 6.2.1.1) (Acetate--CoA ligase) (Acyl-activating enzyme)	Agrobacterium rhizogenes
Q8UBV5	1134783	acsA acs Atu2745 AGR_C_4980	Acetyl-coenzyme A synthetase (AcCoA synthetase) (Acs) (EC 6.2.1.1) (Acetate--CoA ligase) (Acyl-activating	Agrobacterium tumefaciens (strain C58 / ATCC 33970)

			enzyme)	
Q758X0	4620668	ACS1 ADR408 W	Acetyl-coenzyme A synthetase 1 (EC 6.2.1.1) (Acetate--CoA ligase 1) (Acyl-activating enzyme 1)	Ashbya gossypii (strain ATCC 10895 / CBS 109.51 / FGSC 9923 / NRRL Y- 1056) (Yeast) (Eremotheciu m gossypii)
Q750T7	4622812	ACS2 AGL148 C	Acetyl-coenzyme A synthetase 2 (EC 6.2.1.1) (Acetate--CoA ligase 2) (Acyl-activating enzyme 2)	Ashbya gossypii (strain ATCC 10895 / CBS 109.51 / FGSC 9923 / NRRL Y- 1056) (Yeast) (Eremotheciu m gossypii)
P39062	937324	acsA BSU296 80	Acetyl-coenzyme A synthetase (AcCoA synthetase) (Acs) (EC 6.2.1.1) (Acetate--CoA ligase) (Acyl-activating enzyme)	Bacillus subtilis (strain 168)
Q89WV 5	1049589	acsA blr0573	Acetyl-coenzyme A synthetase (AcCoA synthetase) (Acs) (EC 6.2.1.1) (Acetate--CoA ligase) (Acyl-activating enzyme)	Bradyrhizobiu m diazoefficiens (strain JCM 10833 / IAM 13628 / NBRC 14792

				/ USDA 110)
Q8FYQ3	1167504 ; 1213757 5	acsA BR1811 BS1330 _I1805	Acetyl-coenzyme A synthetase (AcCoA synthetase) (Acs) (EC 6.2.1.1) (Acetate--CoA ligase) (Acyl-activating enzyme)	Brucella suis biovar 1 (strain 1330)
Q8NJJ3	3644652 ; 3644710	ACS2 CaO19.1 064 CaO19.8 666	Acetyl-coenzyme A synthetase 2 (EC 6.2.1.1) (Acetate--CoA ligase 2) (Acyl-activating enzyme 2)	Candida albicans (strain SC5314 / ATCC MYA- 2876) (Yeast)
Q8KBY0	1006138	acsA acs CT1652	Acetyl-coenzyme A synthetase (AcCoA synthetase) (Acs) (EC 6.2.1.1) (Acetate--CoA ligase) (Acyl-activating enzyme)	Chlorobium tepidum (strain ATCC 49652 / DSM 12025 / TLS)
P16928	2871910	facA acuA AN5626	Acetyl-coenzyme A synthetase (EC 6.2.1.1) (Acetate--CoA ligase) (Acyl-activating enzyme)	Emmericella nidulans (strain FGSC A4 / ATCC 38163 / CBS 112.46 / NRRL 194 / M139) (Aspergillus nidulans)
P27550	1293368 1;94857 2	acs yfaC b4069 JW4030	Acetyl-coenzyme A synthetase (AcCoA synthetase) (Acs) (EC 6.2.1.1) (Acetate--CoA ligase) (Acyl-activating	Escherichia coli (strain K12)

			enzyme)	
O60011	2896335	ACS1 KLLA0A 03333g	Acetyl-coenzyme A synthetase 1 (EC 6.2.1.1) (Acetate--CoA ligase 1) (Acyl-activating enzyme 1)	Kluyveromyce s lactis (strain ATCC 8585 / CBS 2359 / DSM 70799 / NBRC 1267 / NRRL Y-1140 / WM37) (Yeast) (Candida sphaerica)
P27095		acsA acs	Acetyl-coenzyme A synthetase (AcCoA synthetase) (Acs) (EC 6.2.1.1) (Acetate--CoA ligase) (Acyl-activating enzyme)	Methanosaet a concillii (Methanotherix soehngeni)
P9WQD 1	1331727 6;88547 9	acsA acs Rv3667 MTV025. 015	Acetyl-coenzyme A synthetase (AcCoA synthetase) (Acs) (EC 6.2.1.1) (Acetate--CoA ligase) (Acyl-activating enzyme)	Mycobacteriu m tuberculosis (strain ATCC 25618 / H37Rv)
O93730	1463659	acsA acs PAE286 7	Acetyl-coenzyme A synthetase (AcCoA synthetase) (Acs) (EC 6.2.1.1) (Acetate--CoA ligase) (Acyl-activating enzyme)	Pyrobaculum aerophilum (strain ATCC 51768 / IM2 / DSM 7523 / JCM 9630 / NBRC 100827)
Q9Z3R3	1232358	acsA1 R00719	Acetyl-coenzyme A synthetase 1 (AcCoA	Rhizobium meliloti (strain

		SMc007 74	synthetase 1) (Acs 1) (EC 6.2.1.1) (Acetate-- CoA ligase 1) (Acyl- activating enzyme 1)	1021) (Ensifer meliloti) (Sinorhizobiu m meliloti)
O68040	9004945	acsA acs RCAP_r cc02126	Acetyl-coenzyme A synthetase (AcCoA synthetase) (Acs) (EC 6.2.1.1) (Acetate--CoA ligase) (Acyl-activating enzyme)	Rhodobacter capsulatus (strain ATCC BAA-309 / NBRC 16581 / SB1003)
Q01574	851245	ACS1 YAL054 C FUN44	Acetyl-coenzyme A synthetase 1 (EC 6.2.1.1) (Acetate--CoA ligase 1) (Acyl-activating enzyme 1)	Saccharomyc es cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast)
Q8ZKF6	1255801	acs STM427 5	Acetyl-coenzyme A synthetase (AcCoA synthetase) (Acs) (EC 6.2.1.1) (Acetate--CoA ligase) (Acyl-activating enzyme)	Salmonella typhimurium (strain LT2 / SGSC1412 / ATCC 700720)
Q82EL5	1211019	acsA SAV_45 99	Acetyl-coenzyme A synthetase (AcCoA synthetase) (Acs) (EC 6.2.1.1) (Acetate--CoA ligase) (Acyl-activating enzyme)	Streptomyces avermitilis (strain ATCC 31267 / DSM 46492 / JCM 5070 / NCIMB 12804 / NRRL 8165 / MA-4680)
Q55404	951871	acsA acs sll0542	Acetyl-coenzyme A synthetase (AcCoA	Synechocysti s sp. (strain

			synthetase) (Acs) (EC 6.2.1.1) (Acetate--CoA ligase) (Acyl-activating enzyme)	PCC 6803 / Kazusa)
--	--	--	---	--------------------

UniProt Entry	NCBI Gene ID	Gene names	Protein names	Organism
F1CYZ5		carA	Acetate CoA-transferase YdiF (EC 2.8.3.8)	Acetobacterium woodii
H6LGM4	11871862	carA2 Awo_c15700	Acetate CoA-transferase YdiF (EC 2.8.3.8)	Acetobacterium woodii (strain ATCC 29683 / DSM 1030 / JCM 2381 / KCTC 1655)
G6XSJ6		ATCR1_08124	Acetate CoA-transferase YdiF (EC 2.8.3.8)	Agrobacterium tumefaciens CCNWGS0286
V5MSC1	17703598	U712_10415	Putative coenzyme A transferase subunit beta (EC 2.8.3.8)	Bacillus subtilis PY79
V5MSQ3	17703599	U712_10420	Putative coenzyme A transferase subunit alpha (EC 2.8.3.8)	Bacillus subtilis PY79
Q8FY42	1167750; 1213783	BR2047 BS1330 _I2041	Acetate CoA-transferase YdiF (EC 2.8.3.8)	Brucella suis biovar 1 (strain 1330)

	0			
C6Q271		Ccar_0559 CcarbDR AFT_5139 CLCAR_0656	Acetate CoA-transferase YdiF (EC 2.8.3.8)	Clostridium carboxidivora ns P7
P76458	12930185;947525	atoD b2221 JW2215	Acetate CoA-transferase subunit alpha (EC 2.8.3.8) (Acetyl-CoA:acetoacetate-CoA transferase subunit alpha)	Escherichia coli (strain K12)
P76459	12933993;946719	atoA b2222 JW2216	Acetate CoA-transferase subunit beta (EC 2.8.3.8) (Acetyl-CoA:acetoacetate CoA-transferase subunit beta)	Escherichia coli (strain K12)
P37766	12931296;946211	ydiF b1694 JW1684	Acetate CoA-transferase YdiF (EC 2.8.3.8) (Short-chain acyl-CoA:acetate CoA-transferase)	Escherichia coli (strain K12)
A9HGB7	5790057;6975653	GDI1530 Gdia_2224	Acetate CoA-transferase YdiF (EC 2.8.3.8)	Gluconacetobacter diazotrophicus (strain ATCC 49037 / DSM 5601 / PAI5)
M7PTY6		G000_0	Acetate CoA-	Klebsiella

		9783 Kpn2146 _2726	transferase YdiF (EC 2.8.3.8)	pneumoniae ATCC BAA- 2146
Q2RJ16	3833054	Moth_12 59	Acetate CoA- transferase YdiF (EC 2.8.3.8)	Moorella thermoacetica (strain ATCC 39073)
A0QZW 0	4534496	MSMEG _4168 MSMEI_ 4070	Acetate CoA- transferase YdiF (EC 2.8.3.8)	Mycobacteriu m smegmatis (strain ATCC 700084 / mc(2)155)
B6VK66	1386106 9	atoD PAU_01 020 PA- RVA1- 4466	Acetate CoA- transferase subunit alpha (EC 2.8.3.8) (Acetate coa- transferase subunit alpha (Ec 2.8.3.8) (Acetyl coa:acetoacetate coa transferase subunit alpha)) (EC 2.8.3.8)	Photorhabdus asymbiotica subsp. asymbiotica (strain ATCC 43949 / 3105- 77) (Xenorhabdus luminescens (strain 2))
B6VK67	1386323 4	atoA PAU_01 019 PA- RVA1- 4467	Acetate coa-transferase beta subunit (Acetyl- coa:acetoacetate co transferase beta subunit) (EC 2.8.3.8)	Photorhabdus asymbiotica subsp. asymbiotica (strain ATCC 43949 / 3105- 77) (Xenorhabdus luminescens (strain 2))
Q8Y265	1219275	RSc047 1	Acetate CoA- transferase YdiF (EC	Ralstonia solanacearum

			2.8.3.8)	(strain GMI1000) (Pseudomona s solanacearum)
F6G5B4	1262703 0	mdcA RSPO_c 02923	Acetate transferase YdiF (EC 2.8.3.8)	Ralstonia solanacearum (strain Po82)
Q92YU3	1235806	SMa140 9	Acetate transferase YdiF (EC 2.8.3.8)	Rhizobium meliloti (strain 1021) (Ensifer meliloti) (Sinorhizobiu m meliloti)
Q92KZ4	1234998	R03304 SMc043 99	Acetate transferase YdiF (EC 2.8.3.8)	Rhizobium meliloti (strain 1021) (Ensifer meliloti) (Sinorhizobiu m meliloti)
Q8ZPR5		ydiF STM135 7	Acetate transferase YdiF (EC 2.8.3.8)	Salmonella typhimurium (strain LT2 / SGSC1412 / ATCC 700720)
E1WFM 4		ydiF SL1344_ 1291	Acetate transferase YdiF (EC 2.8.3.8)	Salmonella typhimurium (strain SL1344)
U2NIA4		L581_27 35	Acetate transferase YdiF (EC 2.8.3.8)	Serratia fonticola AU- AP2C

U2NF66		L581_45 07	Acetate transferase YdiF (EC 2.8.3.8)	CoA-	Serratia fonticola AU- AP2C
U2NBM 1		L580_13 93	Acetate transferase YdiF (EC 2.8.3.8)	CoA-	Serratia fonticola AU- P3(3)
W8UTY 0		CH52_0 4550 DA92_0 0435	Acetate transferase YdiF (EC 2.8.3.8)	CoA-	Staphylococc us aureus
D3ES62	1286367 5	SA2981 _0235	Acetate transferase YdiF (EC 2.8.3.8)	CoA-	Staphylococc us aureus (strain 04- 02981)

Additional genes coding for enzymes for application to the synthesis of acetyl CoA from acetate (Activity F) can be identified based on sequence homology to those examples in Table 5.

5 **Example 2 –Conversion of acetaldehyde to 3-hydroxybutanal**

The syntheses described herein via the intermediate compound 3-hydroxybutanal. Synthesis of 3-hydroxybutanal is achieved using an enzyme capable of the aldol coupling of two molecules of acetaldehyde. This reaction is catalysed by deoxyribose phosphate aldolase (DERA, EC 4.1.2.4). This aldolase can be sourced from a wide range of microorganisms with the example from E.coli having been studied in detail for the synthesis of statin intermediates.

The NIH Genbank® database of publicly available nucleotide sequences (<http://www.ncbi.nlm.nih.gov/gene>) may be used to identify genes encoding proteins classified as EC 4.1.2.4. Bacterial genes annotated with EC 4.1.2.4 number 1137 as of 6th July 2014; by phylum, there are 394 examples in the firmicutes, 387 in proteobacteria, 153 in actinobacteria, 50 in cyanobacteria and 153 in others. There are also 52 archaeal genes annotated with EC 4.1.2.4. These data are summarised in Table 6.

Table 6. Distribution of genes annotated with EC 4.1.2.4 in the Bacteria and Archaea identified using the NIH Genbank® database of publicly available nucleotide sequences; accessed 6th July 2014. Example sequences for conversion of acetaldehyde to 3-hydroxybutanal (Activity G).

Domain	Phylum	Order	Number of genes
Bacteria			1137
	Firmicutes		394
		Bacillales	168
		Lactobacillales	110
		Clostridiales	78
		Thermoanaerobacterales	26
		Halanaerobiales	6
		Selenomonadales	4
		Natranaerobiales	1
		Erysipelotrichales	1
	Proteobacteria		387
		g-proteobacteria	276
		a-proteobacteria	61
		b-proteobacteria	36
		d-proteobacteria	14
	Actinobacteria		153

		Actinomycetales	146
		high GC Gram+	7
	Cyanobacteria		50
Archaea			52

UniProt Entry	NCBI Gene ID	Gene names	Protein names	Organism
H6LE06	11871631;	deoC4 Awo_c12650	Deoxyribose-phosphate aldolase DeoC4 (EC 4.1.2.4)	<i>Acetobacterium woodii</i> (strain ATCC 29683 / DSM 1030 / JCM 2381 / KCTC 1655)
H6LE04	11871629;	deoC2 Awo_c12630	Deoxyribose-phosphate aldolase DeoC2 (EC 4.1.2.4)	<i>Acetobacterium woodii</i> (strain ATCC 29683 / DSM 1030 / JCM 2381 / KCTC 1655)
H6LF13	11870761;	deoC1 deoC Awo_c01090	Deoxyribose-phosphate aldolase (DERA) (EC 4.1.2.4) (2-deoxy-D-ribose 5-phosphate aldolase) (Phosphodeoxyriboal dolase)	<i>Acetobacterium woodii</i> (strain ATCC 29683 / DSM 1030 / JCM 2381 / KCTC 1655)
H6LFY1	11871799;	deoC5 Awo_c14870	Deoxyribose-phosphate aldolase DeoC5 (EC 4.1.2.4)	<i>Acetobacterium woodii</i> (strain ATCC 29683 / DSM 1030 / JCM 2381 / KCTC 1655)
H6LE05	11871630;	deoC3 Awo_c12640	Deoxyribose-phosphate aldolase DeoC3 (EC 4.1.2.4)	<i>Acetobacterium woodii</i> (strain ATCC 29683 / DSM 1030 /

				<i>JCM 2381 / KCTC 1655)</i>
C9RDA8	8491097	deoC Adeg_1109	Deoxyribose-phosphate aldolase (DERA) (EC 4.1.2.4) (2-deoxy-D-ribose 5-phosphate aldolase) (Phosphodeoxyriboaldolase)	<i>Ammonifex degensii</i> (strain DSM 10501 / KC4)
P39121	938608	deoC dra BSU39420	Deoxyribose-phosphate aldolase (DERA) (EC 4.1.2.4) (2-deoxy-D-ribose 5-phosphate aldolase) (Phosphodeoxyriboaldolase) (Deoxyriboaldolase)	<i>Bacillus subtilis</i> (strain 168)
Q97IU5	1117728	deoC CA_C1545	Deoxyribose-phosphate aldolase (DERA) (EC 4.1.2.4) (2-deoxy-D-ribose 5-phosphate aldolase) (Phosphodeoxyriboaldolase) (Deoxyriboaldolase)	<i>Clostridium acetobutylicum</i> (strain ATCC 824 / DSM 792 / JCM 1419 / LMG 5710 / VKM B-1787)
Q8NTC4	1021418	deoC Cgl0383 cg0458	Deoxyribose-phosphate aldolase (DERA) (EC 4.1.2.4) (2-deoxy-D-ribose 5-phosphate aldolase) (Phosphodeoxyriboaldolase) (Deoxyriboaldolase)	<i>Corynebacterium glutamicum</i> (strain ATCC 13032 / DSM 20300 / JCM 1318 / LMG 3730 / NCIMB 10025)
P0A6L0	12934356;	deoC dra thyR	Deoxyribose-	<i>Escherichia coli</i>

	948902	b4381 JW4344	phosphate aldolase (DERA) (EC 4.1.2.4) (2-deoxy-D-ribose 5- phosphate aldolase) (Phosphodeoxyriboal dolase) (Deoxyriboaldolase)	(strain K12)
A4IR26	4967361	deoC GTNG_2435	Deoxyribose- phosphate aldolase (DERA) (EC 4.1.2.4) (2-deoxy-D-ribose 5- phosphate aldolase) (Phosphodeoxyriboal dolase) (Deoxyriboaldolase)	<i>Geobacillus thermodenitrificans</i> (strain NG80-2)
Q8ZXK7	1465578	deoC PAE1231	Deoxyribose- phosphate aldolase (DERA) (EC 4.1.2.4) (2-deoxy-D-ribose 5- phosphate aldolase) (Phosphodeoxyriboal dolase) (Deoxyriboaldolase)	<i>Pyrobaculum aerophilum</i> (strain ATCC 51768 / IM2 / DSM 7523 / JCM 9630 / NBRC 100827)
Q8ZJV8	1256093	deoC STM4567	Deoxyribose- phosphate aldolase (DERA) (EC 4.1.2.4) (2-deoxy-D-ribose 5- phosphate aldolase) (Phosphodeoxyriboal dolase) (Deoxyriboaldolase)	<i>Salmonella typhimurium</i> (strain LT2 / SGSC1412 / ATCC 700720)
P99174	1124840	deoC2 SA1939	Deoxyribose- phosphate aldolase 2 (DERA 2) (EC	<i>Staphylococcus aureus</i> (strain N315)

			4.1.2.4) (2-deoxy-D-ribose 5-phosphate aldolase 2) (Phosphodeoxyriboaldolase 2) (Deoxyriboaldolase 2)	
Q99Y51	3571313; 902077	deoC SPy_1867 M5005_Spy15 85	Deoxyribose-phosphate aldolase (DERA) (EC 4.1.2.4) (2-deoxy-D-ribose 5-phosphate aldolase) (Phosphodeoxyriboaldolase) (Deoxyriboaldolase)	<i>Streptococcus pyogenes</i> serotype M1
Q9X1P5	897566	deoC TM_1559	Deoxyribose-phosphate aldolase (DERA) (EC 4.1.2.4) (2-deoxy-D-ribose 5-phosphate aldolase) (Phosphodeoxyriboaldolase) (Deoxyriboaldolase)	<i>Thermotoga maritima</i> (strain ATCC 43589 / MSB8 / DSM 3109 / JCM 10099)
Q72JE9	2775585	deoC TT_C0823	Deoxyribose-phosphate aldolase (DERA) (EC 4.1.2.4) (2-deoxy-D-ribose 5-phosphate aldolase) (Phosphodeoxyriboaldolase) (Deoxyriboaldolase)	<i>Thermus thermophilus</i> (strain HB27 / ATCC BAA-163 / DSM 7039)
Q8ZGH4	1147807; 1174165; 2764428	deoC1 dra YPO1323 y2860 YP_1269	Deoxyribose-phosphate aldolase 1 (DERA 1) (EC 4.1.2.4) (2-deoxy-D-	<i>Yersinia pestis</i>

			ribose 5-phosphate aldolase 1) (Phosphodeoxyriboal dolase 1) (Deoxyriboaldolase 1)	
A2BLE9	4781378	Hbut_0962	Deoxyribose- phosphate aldolase (DERA) (EC 4.1.2.4)	<i>Hyperthermus butylicus (strain DSM 5456 / JCM 9403)</i>
C0ZUQ6	7712431	deoC, RER_15930	Deoxyribose- phosphate aldolase (DERA) (EC 4.1.2.4)	<i>Rhodococcus erythropolis (strain PR4 / NBRC 100887)</i>
B4A422		deoC SNSL317_A20 05	Deoxyribose- phosphate aldolase (DERA) (EC 4.1.2.4)	<i>Salmonella enterica subsp. enterica serovar Newport str. SL317</i>
Q88264	1061480	deoC Ip_0497	Deoxyribose- phosphate aldolase (DERA) (EC 4.1.2.4)	<i>Lactobacillus plantarum (strain ATCC BAA-793 / NCIMB 8826 / WCFS1)</i>
S8F6M2	7901233	TGME49_318 750	Deoxyribose- phosphate aldolase (DERA) (EC 4.1.2.4)	<i>Toxoplasma gondii ME49</i>
Q87710	3233914	deoC TK2104	Deoxyribose- phosphate aldolase (DERA) (EC 4.1.2.4)	<i>Thermococcus kodakaraensis (strain ATCC BAA-918 / JCM 12380 / KOD1) (Pyrococcus kodakaraensis (strain KOD1))</i>
B9DS93	7393055	deoC SUB 0952	Deoxyribose- phosphate aldolase (DERA) (EC 4.1.2.4)	<i>Streptococcus uberis (strain ATCC BAA- 854 / 0140J)</i>
A4WHP	5054261	deoC	Deoxyribose-	<i>Pyrobaculum</i>

		Pars_0301	phosphate aldolase (DERA) (EC 4.1.2.4)	<i>arsenaticum</i> (strain DSM 13514 / JCM 11321)
A1RU26	4617152	deoC Pisl_1295	Deoxyribose- phosphate aldolase (DERA) (EC 4.1.2.4)	<i>Pyrobaculum</i> <i>islandicum</i> (strain DSM 4184 / JCM 9189)
C4M5C6	3406093	EHl_121800	Deoxyribose- phosphate aldolase (DERA) putative	<i>Entamoeba histolytica</i>
A8A8B0	5593924	deoC EcHS_A4616	Deoxyribose- phosphate aldolase (DERA) (EC 4.1.2.4)	<i>Escherichia coli</i> O9:H4 (strain HS)
Q0SEY5	4218140	deoC RHA1_ro0209 4	Deoxyribose- phosphate aldolase (DERA) (EC 4.1.2.4)	<i>Rhodococcus</i> sp. (strain RHA1)
F8K193	11354892 12650565	Dera, SCAT_3805, SCATT_37940	Deoxyribose- phosphate aldolase	<i>Streptomyces cattleya</i> (strain ATCC 35852 / DSM 46488 / JCM 4925 / NBRC 14057 / NRRL 8057)
A7FU73	5395000	deoC, CLB_1583	Deoxyribose- phosphate aldolase	<i>Clostridium botulinum</i> (strain ATCC 19397 / Type A)
B9E4U5	7273626	deoC, CKR_2469	Deoxyribose- phosphate aldolase	<i>Clostridium kluyveri</i> (strain NBRC 12016)
D8GI14	9445430	deoC CLJU_c18130	Deoxyribose- phosphate aldolase	<i>Clostridium ljungdahlii</i> (strain ATCC 55383 / DSM 13528 / PETC)
E3GHB9	9881953	deoC ELI_0052	Deoxyribose- phosphate aldolase	<i>Eubacterium limosum</i> (strain KIST612)
B5Y277	6936643	deoC KPK_4777	Deoxyribose- phosphate aldolase	<i>Klebsiella</i> <i>pneumoniae</i> (strain 342)

Q5KY02	3184692	deoC GK_2499	Deoxyribose- phosphate aldolase	<i>Geobacillus kaustophilus</i> (strain HTA426)
Q6HK62	2856540	deoC BT9727_1732	Deoxyribose- phosphate aldolase	<i>Bacillus thuringiensis</i> subsp. <i>konkukian</i> (strain 97-27)
I3DVS3		deoC PB1_12319	Deoxyribose- phosphate aldolase	<i>Bacillus methanolicus</i> PB1

Additional genes coding for enzymes capable of the coupling of two molecules of acetaldehyde (Activity G) can be identified based on sequence homology to those examples in Table 6.

5

Example 3 – Reduction of 3-hydroxybutanal to 1,3-butanediol

Genes coding for enzymes capable of the reduction of an aldehyde to the corresponding alcohol (EC 1.1.1.-) are widespread in nature and with respect to this application are generally classified in EC 1.1.1.78; 1.1.1.265; 1.1.1.373; 1.1.1.1; 1.1.1.2; 1.1.1.21; 1.1.1.26; 1.1.1.31; 1.1.1.71; 1.1.1.72; 1.1.1.77 and 1.1.1.283. For this application it is desirable that aldehyde reductase or alcohol dehydrogenases enzymes (both terms refer to enzymes capable of aldehyde reduction) show preference towards a C4 aldehyde relative to a C2 aldehyde such as acetaldehyde. Alcohol dehydrogenases involved in ethanol synthesis for example, preferring acetaldehyde as a substrate would not be preferred for this application, but evolution of these well described short chain dehydrogenase or reductases using techniques well known in the art, could be used to alter the substrate preference towards longer chain aldehydes.

20

This reaction can be catalysed by a medium chain alcohol dehydrogenase which showed preference for alcohols of C4 or greater, for example (gene *alrA*) see Appl. Environ. Microbiol, 2000, 66, 5231. Further, alcohol dehydrogenases showing preference for longer chain alcohols from *Acinebacter calcoaceticus* NCIB 8250 and from *Saccharomyces cerevisiae* D273-10B are described by Wales, M and Fewson, C. Microbiol 1994, 140, 173. Although measured in the oxidative direction,

25

the dehydrogenase also accepts 1,4-butanediol as a substrate. 2,3- butanediol is not a substrate, clearly demonstrating the desired primary alcohol as opposed to secondary alcohol specificity for application to 3-hydroxybutanal reduction.

5 A further excellent candidate enzyme is bcALD, GRE_2 (EC 1.1.1.265 also classified in EC 1.1.1.283) from *S. cerevisiae* var. *uvarum* W34 described by van Iersel, M. F. M et al. Appl. Environ. Microbiol. 1997. 63, 4079. This enzyme shows strong preference for butanal and derivatives with a poor preference for acetaldehyde. Kms are : acetaldehyde 158mM; butanal 2.76mM; 2-methylbutanal
10 1.85mM; 3-methylbutanal 0.21mM. The preference of GRE2 derived dehydrogenase for a C4 aldehyde (butanal) relative to acetaldehyde is shown in Example 9. From these data this dehydrogenase would also be expected to show selective preference for 3-hydroxybutanal relative to acetaldehyde.

15 Another excellent example is the GOX1615 gene from *Gluconobacter oxydans* (Richter, N. et al. Chembiochem. 2009, 10,1888.). This enzyme has been characterised and shown to have very poor preference for acetaldehyde reduction compared to longer chain and hydroxysubstituted substrates. 3-Hydroxybutanal was not specifically tested in the reductive direction. However, 1,3-butanediol was tested
20 in the undesired oxidative direction and poor activity was reported. Hence, based on the data presented it is expected that this enzyme would show the desired preference for 3-hydroxybutanal reduction compared to undesired acetaldehyde reduction also with a poor oxidative activity towards the product 1,3-butanediol. The use of GOX1615 for selective reduction of 3-hydroxybutanal to 1,3-butanediol within
25 a novel, unnatural 1,3-BDO pathway, is shown in Example 12,13 and 14. Example 9 also confirms its predicted preference for the DERA product 3-hydroxybutanal relative to acetaldehyde.

Further examples include yqhD from *E. coli* which is reported as having a
30 preference for alcohols of C3 or greater, Sulzenbacher et al.,2004. J. Mol. Biol. 342:489-502. Alcohol dehydrogenases are understood to be reversible enzymes capable of operating in a reductive or oxidative direction. Genes bdh A and bdh B (proteins bdh I and bdh II) from *C. acetobutylicum* code for enzymes which convert butanal into butanol (Walter et al. 1992. J. Bacteriol. 174:7149-7158. The use of

bdhII (gene bdhB) butanol dehydrogenase for reduction of 3-hydroxybutanal to 1,3-butanediol within a novel, unnatural 1,3-BDO pathway, is shown in Example 12. Further, butanol dehydrogenase examples include bdh from *C. saccharoperbutylacetonicum* and Cbei_1722, Cbei_2181 and Cbei_2421 in *C. Beijerincki* (Gene announce. 2012, 194, (19) 5470.). Other gene products classified as methylglyoxal reductases (EC 1.1.1.283) in addition to GRE_2 described above, may also be candidates (Eur. J. Biochem, 1988, 171,213).

Additional aldehyde reductase gene candidates in *Saccharomyces cerevisiae* include the aldehyde reductases GRE3, ALD2-6 and HFD1, glyoxylate reductases GOR1 and YPL113C and glycerol dehydrogenase GCY1 (Atsumi et al., *Nature* 451:86-89 2008).

Table 7 . Examples of genes expressing enzymes for the conversion of 3-hydroxybutanal to 1,3-butanediol (Activity H).

UniProt Entry	NCBI Gene ID	Gene names	Protein names	Organism
Q9F1R1		alrA	NADPH-dependent alcohol dehydrogenase (EC 1.1.1.2)	Acinetobacter sp. M-1
Q04944	1119481	bdhA CA_C3299	NADH-dependent butanol dehydrogenase A (EC 1.1.1.-) (BDH I)	Clostridium acetobutylicum (strain ATCC 824 / DSM 792 / JCM 1419 / LMG 5710 / VKM B-1787)
Q04945	1119480	bdhB CA_C3298	NADH-dependent butanol dehydrogenase B (EC 1.1.1.-) (BDH II)	Clostridium acetobutylicum (strain ATCC 824 / DSM 792 / JCM 1419 / LMG 5710 / VKM B-

				1787)
Q46856	12933386; 947493	yqhD b3011 JW2978	Alcohol dehydrogenase YqhD (EC 1.1.1.-)	Escherichia coli (strain K12)
Q12068	854014;	GRE2 YOL151W	NADPH-dependent methylglyoxal reductase GRE2 (EC 1.1.1.283) (3- methylbutanal reductase) (EC 1.1.1.265) (Genes de respuesta a estres protein 2) (Isovaleraldehyde reductase)	Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast)
P00331	855349	ADH2 ADR2 YMR303C YM9952.05C	Alcohol dehydrogenase 2 (EC 1.1.1.1) (Alcohol dehydrogenase II) (YADH-2)	Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast)
P0A9S1	12930229; 947273	fucO b2799 JW2770	Lactaldehyde reductase (EC 1.1.1.77) (Propanediol oxidoreductase)	Escherichia coli (strain K12)
P20368	3188393	adhA ZMO1236	Alcohol dehydrogenase 1 (EC 1.1.1.1) (Alcohol dehydrogenase I) (ADH I)	Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4)
A2PYM4		bdh	Butanol dehydrogenase	Clostridium saccharoperbutyl acetonicum
A6LU64	5292938	Cbei_1722	Iron-containing	Clostridium

			alcohol dehydrogenase	beijerinckii (strain ATCC 51743 / NCIMB 8052) (Clostridium acetobutylicum)
A6LVG8	5293392	Cbei_2181	Iron-containing alcohol dehydrogenase	Clostridium beijerinckii (strain ATCC 51743 / NCIMB 8052) (Clostridium acetobutylicum)
A6LW49	5293624	Cbei_2421	Iron-containing alcohol dehydrogenase	Clostridium beijerinckii (strain ATCC 51743 / NCIMB 8052) (Clostridium acetobutylicum)
P38715	856504	GRE3 YHR104W	NADPH-dependent aldose reductase GRE3 (EC 1.1.1.21) (NADPH-dependent methylglyoxal reductase GRE3) (Xylose reductase) (EC 1.1.1.-)	Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast)
Q5FQJ0	3248904	GOX1615	Putative oxidoreductase (EC 1.1.1.-)	Gluconobacter oxydans (strain 621H) (Gluconobacter suboxydans)
P28811	879097	mmsB,	EC 1.1.1.31	Pseudomonas

		PA3569	3-hydroxyisobutyrate dehydrogenase	aeruginosa (strain ATCC 15692 / PAO1 / 1C / PRS 101 / LMG 12228)
Q5SLQ6	3168163	TTHA0237	EC 1.1.1.31 3-hydroxyisobutyrate dehydrogenase	Thermus thermophilus (strain HB8 / ATCC 27634 / DSM 579)
A6LU64	5292938	Cbei_1722	Iron-containing alcohol dehydrogenase.	Clostridium beijerinckii (strain ATCC 51743 / NCIMB 8052) (Clostridium acetobutylicum)
B3LMK7		SCRG_02216	Medium chain alcohol dehydrogenase	Saccharomyces cerevisiae (strain RM11-1a) (Baker's yeast)
C4QWW0	8196620	PAS_chr1-1_0357	Medium chain alcohol dehydrogenase (NADPH)	Komagataella pastoris (strain GS115 / ATCC 20864) (Yeast) (Pichia pastoris)
A6ZTC9		ADH7, SCY_0522	Medium chain alcohol dehydrogenase	Saccharomyces cerevisiae (strain YJM789) (Baker's yeast)
P27250	12934055 948802	Ahr yjgB b4269, JW5761	Aldehyde reductase	Escherichia coli (strain K12)

Additional genes coding for enzymes capable of the conversion of 3-hydroxybutanal to 1,3-butanediol (Activity H) can be identified based on sequence homology to those examples in Table 7.

5 **Example 4 - Culture of acetogen strains for production of 1,3-butanediol or other downstream product**

For the production of downstream products of 3-hydroxybutanal such as 1,3-butanediol the recombinant acetogen strain may be cultured in a defined, semi-
10 defined or undefined medium supplemented with syngas as the only or principle carbon and energy source is well known in the art. Examples of additional sources of energy or carbon may be nitrate, methanol or sugar. It is highly desirable to maintain anaerobic conditions as the acetogen strains of the present Example are strict anaerobes. Initial tests with the wild type organism and with genetically
15 modified organisms before moving to a fermenter can be done in small bottles that are fitted with thick rubber stoppers and aluminium crimps employed to seal the bottles and as those skilled in the art will understand.

Suitable replicates such as triplicate cultures can be grown for each engineered
20 strain and culture supernatants can be tested for products formed. For example, syngas composition in the media, metabolic intermediates, 1,3-butanediol and by-product(s) formed in the engineered production host can be measured as a function of time and can be analysed by methods such as High Performance Liquid Chromatography (HPLC), GC (Gas Chromatography), GC-MS (Gas
25 Chromatography-Mass Spectroscopy) and LC-MS (Liquid Chromatography-Mass Spectroscopy) or other suitable analytical methods using routine procedures well known in the art.

Acetate, pyruvate, acetyl-Co A, 3-hydroxybutanal, 1,3-butanediol and intermediates
30 or other desired products can be quantified by HPLC using as appropriate, a refractive index detector or UV detector or other suitable assay and detection methods well known in the art. The individual enzyme or protein activities expressed from the heterologous DNA sequences or overexpressed endogenous DNA sequences, can also be assayed using methods well known in the art.

5 Fermentations can be performed in continuous cultures, batch or fed-batch. All of these processes are well known in the art. Important process considerations for syngas fermentation are high biomass concentration and good gas-liquid mass transfer Bredwell *et al.*, (1999), Biotechnol. Prog. 15:834-844. As carbon monoxide has a lower solubility in water compared to oxygen, continuously gas-sparged fermentations are recommended and can be performed in controlled fermentors with constant off-gas analysis by mass spectrometry and periodic liquid sampling and analysis discussed above. Other feedstocks such as methanol or sugar can be
10 fed to the fermentor using traditional approaches.

Example 5. Generation of a lactate dehydrogenase gene knockout in *Acetobacterium woodii*.

15 Plasmids and Primers

Primer	Sequence	Rest ricti on site
EryFor	ggGGATCCAATGATACACCAATCAGTGC (SEQ ID NO.8)	Bam HI
EryXbaRev	ggTCTAGATTGAACCCGTCTCCTTACG (SEQ ID NO.9)	XbaI
01For(uplacdehyEco)	ggGAATTCatgtcgatcatattgaagg (SEQ ID NO.10)	Eco RI
01Rev(uplacdehyBam)	aaGGATCCcttaaacgaccatcc (SEQ ID NO.11)	Bam H1
02For(dwnlacdehyXba)	Cctctagagtgagatgatggataac (SEQ ID NO.12)	XbaI
02Rev(dwnlacdehyPstHind)	gcAAGCTTCTGCAGtcaccaaattgtgcttaacaacc (SEQ ID NO.13)	PstI, Hind III
16For	aaTCTAGAttagtcggtgcctgtg (SEQ ID NO.14)	XbaI
16Rev	ggAAGCTTtctttacgttctacg (SEQ ID NO.15)	Hind

		III
13b(LDHantiFor50)	aa <u>ACTAGT</u> ggcgattccaacggcgat (SEQ ID NO.16)	<i>SpeI</i>
13(LDHantiRev)	aa <u>CTCGAG</u> ctacatctgacagactttttcgg (SEQ ID NO.17)	<i>XhoI</i>
pAM β 1For	CGATTTCGATTGATTGCTT (SEQ ID NO.18)	-
pAM β 1Rev	AAT CCC AAA TGA GCC AACAG (SEQ ID NO.19)	-

Plasmid	Description
pUC19	ampicillin resistance , suitable for blue white cloning
pTRKH2	Erythromycin resistance, replicon for gram+ and <i>E. coli</i>
pUC19-Ery(XbaI) (pE)	pUC19 vector containing the Erythromycin resistance gene in the <i>Bam</i> HI and <i>Xba</i> I site
pUC19-Ery-LDHup-LDHdown (pE01-02)	pUC19 vector containing the Erythromycin resistance gene in the <i>Bam</i> HI and <i>Xba</i> I site and the LDH upstream region cloned in <i>Eco</i> RI- <i>Bam</i> HI and LDH downstream region of cloned into <i>Xba</i> I- <i>Hind</i> III
pUC19-Ery-LDHk/o (pE16)	pUC19 vector containing the Erythromycin resistance gene in the <i>Bam</i> HI and <i>Xba</i> I site and a 728 bp fragment of the LDH gene
pUC19-pAM β 1-Ery (pEP)	pUC19 vector containing the gram+ replicon pAM β 1 in the <i>Eco</i> RI and <i>Kpn</i> I site, Erythromycin resistance gene in the <i>Bam</i> HI and <i>Xba</i> I

Media for *Acetobacterium woodii*

- 5 The medium was prepared using anaerobic techniques and contained under an N₂-CO₂ atmosphere (80:20) the following per 1000 ml:

	KH ₂ PO ₄	1.76 g
	K ₂ HPO ₄	8.44 g
10	NH ₄ Cl	1.0 g

	cysteine hydrochloride	0.5 g
	MgSO ₄ x 6H ₂ O	0.33 g
	NaCl	2.9 g
	yeast extract	2.0 g
5	KHCO ₃	6.0 g
	Resazurin	0.001 g
	Trace element solution SL 9	1.0 ml
	selenite-tungstate solution	1.0 ml
	vitamin solution DSMZ 141	2.0 ml
10	Carbon source	20 - 40 mM

Carbon source and magnesium were added after autoclaving from an anaerobe, sterile stock solution (2 M Fructose or Lactate and 0.75 M respectively).

- 15 For solid media 15 g/L agar was added and the media boiled to remove dissolved oxygen and then cooled down under a stream of gas, before autoclaving.

Vitamin solution

	Biotin	2.00 mg
20	Folic acid	2.00 mg
	Pyridoxine-HCl	10.00 mg
	Thiamine-HCl x 2 H ₂ O	5.00 mg
	Riboflavin	5.00 mg
	Nicotinic acid	5.00 mg
25	D-Ca-pantothenate	5.00 mg
	Vitamin B12	0.10 mg
	p-Aminobenzoic acid	5.00 mg
	Lipoic acid	5.00 mg
	1000 ml dH ₂ O	

30

Selenite-tungstate solution

	NaOH	0.5 g
	Na ₂ SeO ₃ x 5 H ₂ O	3 mg
	Na ₂ WO ₄ x 2 H ₂ O	4 mg

1000 ml dH₂O

Trace element solution SL 9

	Nitrilotriacetic acid	12.8 g
5	FeCl ₂ x 4 H ₂ O	2.0 g
	ZnCl ₂	0.070 g
	MnCl ₂ x 4 H ₂ O	0.1 g
	H ₃ BO ₃	0.006 g
	CoCl ₂ x 6 H ₂ O	0.19 g
10	CuCl ₂ x 2 H ₂ O	0.002 g
	NiCl ₂ x 6 H ₂ O	0.024 g
	Na ₂ MoO ₄ x 2 H ₂ O	0.036 g

1000 ml dH₂O, Nitrilotriacetic acid was dissolved first and the pH adjusted to 6.0 with NaOH before all other Trace elements were added.

15

Construction of the LDH knockout mutant

To construct a Lactate-dehydrogenase knockout mutant two strategies were followed.

20

- (1) Involving two cross-over recombination processes where the full length LDH gene is replaced with the Erythromycin cassette (Figure 11)
- (2) A single recombination event leading to integration of the complete cloning plasmid and therefore interrupting the LDH gene. (Figure 12)

25

To construct the plasmids for the LDH knockout mutants of *A. woodii* pUC19 was used as backbone plasmid. First a erythromycin antibiotic cassette was cloned into the *Bam*HI and *Xba*I restriction sites yielding the plasmid pUC19-Ery. The Erythromycin resistance cassette was amplified by PCR using gene specific primers (EryFor, EryXbaRev) and plasmid pTRKH2 as template. Functionality of the antibiotic resistance was confirmed by growth of *E. coli* DH10B harbouring this plasmid in the presence of 150-300 µg/ml Erythromycin.

30

To confirm that the plasmid does not integrate into the *A. woodii* genome randomly an aliquot was used to transform *A. woodii*. As expected no erythromycin resistance bacteria were obtained after several attempts.

5 For strategy (1) approximately 1000 bp flanking region upstream and downstream of the LDH gene were amplified using the gene specific primers (01For, 01Rev) for the upstream region and (02For, 02Rev) for the downstream region with genomic DNA as template. The upstream region was cloned into the *EcoRI* and *BamHI* region leading to plasmid pUC19-Ery-LDHup which was then used to clone the
10 downstream region into *XbaI* and *HindIII* sites yielding plasmid pUC19-Ery-LDHup-LDHdown (Plasmid pE01-02).

In strategy (2) a 728 bp fragment of the LDH gene generated by PCR from genomic DNA using primer 16For/16Rev, This PCR products was cloned into pUC19-Ery into *XbaI* and *HindIII* site, giving plasmid (pE16).

15 Plasmid pE01-02 was used to transform *A. woodii*. All following procedures were carried out under anaerobic conditions. 1 ml of a fresh 5 ml over-night culture was used to inoculate 10 ml media and *A. woodii* grown to an OD₆₀₀ of approximately 0.5. Cells were spun down in hungate tubes for 10 min at 4000 rpm at 4°C and
20 washed twice with 10 ml ice-cold anaerobic 270 mM sucrose solution. The pellet was resuspended in 200 µL sucrose and transferred on ice into the anaerobic chamber. 4 µl of each plasmid was added to 40 µl cells, transferred to a 0.2-cm electroporation cuvette and kept on ice for 5 min. For electroporation the following settings were used: electric pulse of 10 kV, electric resistance of 400 Ω and 25 µF.
25 Following electroporation the cells were kept on ice for another 5 min before 960 µl media was added. The transformed cells were incubated o/n at 30°C and then transferred to 50 ml media containing the required antibiotic (20 µg/ml erythromycin). Single cross-over resistant cells grew within 48 hours. An aliquot of the liquid culture was plated on solid media with the required antibiotic (50 µg/ml
30 erythromycin). Single colonies were obtained after 4-5 days, which were picked and grown up in 5 ml cultures in the presence of erythromycin.

For the double cross-over knockout the plasmid was cut using the restriction enzymes *EcoRI* and *HindIII* to obtain an approximately 3837 bp linear fragment

containing the erythromycin gene flanked by the up and downstream LDH region. The linear fragment was gel extracted and used to transform *A. woodii*. The transformation of *A. woodii* was performed as described above. An erythromycin resistant culture was obtained after a 3-4 days. The culture were plated on solid media and colonies obtained after 5-6 days.

In both cases the colonies were screened for the presence of the erythromycin gene using primers EryFor, EryXbaRev and in the case of the double cross-over for the absence of the LDH gene using primer set 13For/Rev. In the latter case 2 colonies showed the absence of the LDH gene.

Those 2 colonies, as well as 2 colonies for the single cross-over recombination (SR) were picked and analysed for their growth on Fructose and Lactate. Two cultures of *A. woodii* wild type (Aw1, Aw2) and *A. woodii* harbouring plasmid pUC19-pAM β 1-Ery (P1, P2) were used as control strains. Plasmid pUC19-pAM β 1-Ery was constructed by cloning the gram positive replicon pAM β 1 from plasmid pTRKH2 into the *Eco*RI and *Kpn*I sites of pUC19 and then cloning the Erythromycin cassette into the *Xba*I-*Bam*HI site. The plasmid was then transformed into *A. woodii* as described before and resistance growth obtained after 2-3 days. Single colonies were picked and analyzed for the plasmid by PCR using the specific primers for the erythromycin gene (EryFor, EryXbaRev) as well for the pAM β 1 replicon (pAM β 1For, pAM β 1Rev).

For the growth curves 500 μ l of a fresh over-night culture was used to inoculate a 50 ml anaerobe culture. Either 20 mM Fructose or 40 mM DL-Lactate ((Lactic acid) were used as a substrate. Erythromycin was prepared as a stock concentration of 2 mg/ml in water. One ml samples were taken twice a day over a period of 4 days. Of this 1 ml, 500 μ l were used for OD₆₀₀ measurement. The remaining 500 μ l were spun down and the supernatant frozen at -20°C for HPLC analysis.

Figure 13 shows the obtained growth curve in Fructose media. Here, all mutants grew similar to the control strains (Aw1, Aw2, P1, and P2). In contrast no growth was obtained for the mutants when grown in Lactate, confirming the expected phenotype for the Lactate Dehydrogenase knockout (Figure 13).

For HPLC analysis 10 µl samples were injected on the HPLC column Rezex ROA Organic Acid H⁺ (300x 7.8 mm, Phenomex). The used mobile Phase was 100 % 0.01N H₂SO₄. Samples were analyzed for 30 minutes with a flow rate of 0.6 ml/min. The HPLC analysis of those cultures showed that the fructose is consumed and acetate is produced by those mutants at a similar rate and amount (Figure 14). In contrast, no Lactate is consumed by the mutants, while the control strains utilize it and produce acetate (Figure 15).

Additionally, the double cross-over knockouts are stable in contrast to the single cross-over mutants (SR). After 5 days a growth on Lactate is observed, probably due to degradation of the Erythromycin which then allows a second recombination step where the cassette is removed from the genome and a functional LDH gene is obtained. HPLC data also confirmed that Lactate is slowly utilized after 5 days by SR1 and SR2 and acetate is produced (data not shown). Those results confirm that the constructed vector can be used to generate stable knockout mutants.

Example 6. Generation of a phosphotransacetylase (PTA) gene knockout in *Acetobacterium woodii*.

Primers:

EryFor ggGGATCCAATGATACACCAATCAGTGC (SEQ ID NO.8)
 EryXbaRev ggTCTAGATTGAACCCGTCTCCTTACG (SEQ ID NO.9)
 Pr1_cisFor ggACTAGTTGTTATTTGGCGATCAGC (SEQ ID NO.20)
 Pr4_iorRev ggCTGCAGCGCACCCATACAAAGC (SEQ ID NO.21)

Pr5_PTAfrag1Rev

AACATCAACATGCGGCCGCACTTACCAAATTATCTGCGTGC (SEQ ID NO.22)

Pr6_PTAfrag3For

AATTTGGTAAGTGCGGCCGCGATGTTGATGTTATTCTCATGC ((SEQ ID NO.23)

Plasmids:

pUC19 Ampicillin resistance

pTRKH2 Erythromycin resistance, replicon for gram+ and *E. coli*

pUC19-Ery

pUC19-Ery- Δ PTA2

Strategy

5 Erythromycin resistance gene was amplified from plasmid pTRKH2 using primer Ery for and EryXabRev and cloned into pUC19 yielding a non-replicative plasmid, pUC19-Ery.

The PTA knock-out cassette was constructed as following: Primers Pr1_cisFor and Pr5_PTAfrag1Rev were used to amplify the upstream region of the PTA gene as well as the N-terminal part of the PTA gene from genomic *A. woodii* DNA. Pr4_iorRev and Pr6_PTAfrag3For were used to amplify the C-terminal part of the PTA gene as well as the downstream region. This two fragments were cloned together using SOE PCR and primers Pr1_cisFor and Pr4_iorRev. The so constructed knockout cassette harbors a modified PTA gene sequence, consisting of the N and C-terminal part of the PTA gene only. Both parts are separated by a *NotI* restriction site which was introduced by previous PCR round. The knockout cassette was cloned into *XbaI* and *PstI* site of pUC19-Ery, yielding plasmid pUC19-Ery- Δ PTA2

20 Plasmid pUC19-Ery- Δ PTA2 was used to transform *A. woodii*. All following procedures were carried out under anaerobic conditions. A 10 ml culture was inoculated with *A. woodii* and grown to an OD600 of approximately 0.5. Cells were spun down in hungate tubes for 10 min at 4000 rpm at 4°C and washed twice with 10 ml 10 ice-cold anaerobic 270 mM sucrose solution. The pellet was resuspended in 200 μ L sucrose. 4 μ L of plasmid was added to 40 μ L cells, transferred to a 0.2-cm electroporation cuvette and kept on ice for 5 min. For electroporation the following settings were used: electric pulse of 10 kV, electric resistance of 400 Ω and 25 μ F. Following electroporation the cells were kept on ice for another 5 min. The transformed cells were recovered in media, incubated anaerobe for 6h at 30°C and then transferred to 50 ml of medium containing the required antibiotic (20 μ g/ml erythromycin). The culture was incubated at 30°C until growth was obtained. An aliquot of the culture was plated on solid medium. Single colonies were obtained after 5-7 days, which were picked and grown up in 10 ml cultures in the presence of erythromycin.

The cultures were genetically analyzed by specific primers to confirm the integration of the plasmid. The culture was passaged into liquid media without erythromycin to allow looping out of the plasmid and generation of a stable PTA knockout via a second recombination event. Passages were plated on solid media until such an event occurred. PTA knockout clones were screened by replica plating in the presence and absence of Erythromycin. Clones not capable of growing in the presence of Erythromycin were picked and analysed for the PTA genotype by PCR, which was confirmed.

Example 7. Heterologues gene expression and protein production in *A. woodii*

Primers

Pr55

gaGTCGACGCAGTATCTTAAATTTGTATAATAGGAATTGAAGTTAAATTAGAT
GCTAAAAATTTGTAATTAAGAAGGAGTGATTACATGTTACGTCCTGTAGAAACC
(SEQ ID NO.24)

Pr54

TTGCATGCTCATTGTTTGCCTCCC (SEQ ID NO.25)

Strategy

The uidA (GUS) from *E. coli* BL21star(DE3) was amplified using gene specific primers, which included the sequence for constitutive promoters. Primer Pr55 includes the sequence from the Enterococcus faecalis Erythromycin resistance gene promoter, while primer Pr56 include the promoter sequence of the *C. ljungdhalii* PTA gene.

The amplified fragments were cloned into a plasmid capable of replicating in *A. woodii*. The replicative plasmid (pEP) carries the Erythromycin gene (described in Example 6) as well as the replicon pAM β 1. However, any other replicon suitable for *A. woodii* can be used. In this fashion two plasmids were generated, pEP55

carrying the uidA gene under control of the Enterococcus faecalis promoter and pEP56, carrying the uidA gene under control of the Clj promoter.

5 The generated plasmids were used to transform *A. woodii*. All following procedures were carried out under anaerobic conditions. A 10 ml culture was inoculate with *A. woodii* and grown to an OD600 of approximately 0.5. Cells were spun down in hungate tubes for 10 min at 4000 rpm at 4C and washed twice with 10 ml 10 ice-cold anaerobic 270 mM sucrose solution. The pellet was resuspended in 200 μ L sucrose. 4 μ L of plasmid was added to 40 μ L cells, transferred to a 0.2-cm
10 electroporation cuvette and kept on ice for 5 min. For electroporation the following settings were used: electric pulse of 10 kV, electric resistance of 400 Ω and 25 μ F. Following electroporation the cells were kept on ice for another 5 min. The transformed cells were recovered in media, incubated anaerobe for 6h at 30C and then transferred to 50 ml of medium containing the required antibiotic (20 μ g/ml
15 erythromycin). The cultures were incubated at 30C until growth was obtained. An aliquot of the cultures was plated on solid medium containing 20 μ g/ml erythromycin. Single colonies were obtained after 5-7 days. For each transformation 2 independent colonies were picked and grown up in 10 ml cultures in the presence of erythromycin.

20 Functionality of uidA was established by restreaking the cultures on anaerobe selection media containing MUG (4-Methylumbelliferyl- β -D-glucopyranosiduronic acid, final concentration of 0.1g/L) leading to a fluorescent product visible under UV-light as seen in Figure 10.

25 The same strategy described above may be applied to the introduction any other heterologous gene (such as DERA, eutE etc) during construction of a 1,3-butanediol pathway and expressing it under a strong constitutive promoter on a replicative plasmid into *A. woodii*. The above described reporter gene uidA can be used to
30 confirm the expression of any other gene, when cloned in an operon. Further, the expression of uidA can be used to determine promotor strength and hence promotor selection, as the efficiency of expression is related to fluorescence intensity. To enhance genetic stability heterologous genes may be introduced into the genome.

Example 8. Generation of a phosphotransacetylase (PTA) mutant of *Moorella thermoacetica* ATCC39073 by homologous recombination.

Introduction

5 The *M. thermoacetica* ATCC39073 genome sequence has been published (Pierce et al., 2008) and is available at the NCBI with accession number NC_007644. The KEGG map of central carbon metabolism for *M. thermoacetica* ATCC39073 (http://www.genome.jp/kegg-bin/show_pathway?mta01200) was used to identify two putative phosphotransacetylases (PTAs), Moth_0864 and Moth_1181 (EC 2.3.1.8);
10 which appear to be isoenzymes and are identified as being members of the PduL superfamily of bacterial propanediol utilisation proteins, based on sequence homology. Members of the phosphate acetyl/butaryl transferase (PTA/PTB) superfamily were not identified in the *M. thermoacetica* ATCC39073 genome; a BLASTP search of the partial PTA from *Clostridium tyrobutyricum* (Zhu et al., 2005)
15 returned no significant alignments.

Construction of Moth_0864 and Moth_1181 knockout plasmids

Construction of the knockout plasmid backbone

The mobilisable shuttle vector pS797 is used as the backbone for construction of *M.*
20 *thermoacetica* ATCC39073 knockout plasmids, since it already contains three of the desired genetic elements comprising the final construct; a pMB1 origin of replication for *E. coli*, an antibiotic selection marker (*bla*) and an RK4-derived conjugal origin of transfer (*oriT*) (Yakobson and Guiney, 1984). A thermostable (pJH1-derived) kanamycin resistance gene for *M. thermoacetica* ATCC39073 has previously been
25 described in the literature (Iwasaki et al., 2013), and was synthesised without further modification using the gene sequence from the *S. faecalis* pJH1 kanamycin resistance gene (Genbank accession number V01547) fused to the native G3PDH promoter. Note that the knockout plasmid backbone does not include a replicon for *M. thermoacetica* ATCC39073, to ensure that kanamycin resistance can only be
30 maintained in *Moorella* following a chromosome recombination event. Kanamycin-resistant transconjugants of *M. thermoacetica* ATC39073 are therefore all presumptive single crossover (SCO) chromosome mutants.
Primers APB57-65 were designed to generate knockout cassettes for Moth_0864 and Moth_1181.

APB #	Sequence (5'-3')	Description
57	AGCTTTTCGAGCGCGGAAC (SEQ ID NO.26)	5' phosphorylated. Universal EMP R2 to splice knockout cassettes into knockout plasmids.
58	GAGTTCCATGTGGTCTAC CATAC (SEQ ID NO.27)	SOE F1 to clone upstream region of chromosomal homology for Moth_0864.
59	CATGGAGGTTAAGGCTGA GTTGACGATACTGTC (SEQ ID NO.28)	SOE R2; pair with APB58. 945bp product. Includes 13bp overhang for assembly with APB60/61 PCR product.
60	GTCAACTCAGCCTTAACCT CCATGACGACCAG (SEQ ID NO.29)	SOE 3F; to clone downstream region of chromosomal homology for Moth_0864 knockout. Includes 11bp overhang for assembly with APB58/59 PCR product.
61	GACGAGCAAGGCAAGACC GGGATCCGACAGTAACCG TAGGTACCTTCG (SEQ ID NO.30)	SOE 4R; pair with APB60. 1015bp product. Includes 25bp 3' overhang for EMP splicing into plasmid backbone.
62	CCAGTGATCTCTTTATCGA CCTCC (SEQ ID NO.31)	SOE F1 to clone upstream region of chromosomal homology for Moth_1181
63	GGTGTGCATGTGCAAGGA CACGCACCTTTTCTAG (SEQ ID NO.32)	SOE F2, pair with APB62. 994bp product. 3' overhang for SOE splicing with APB64/65.
64	TGCGTGCCTTGACATG CACACCGATGAGG (SEQ ID NO.33)	SOE 3F. Clone downstream region of chromosomal homology for Moth_1181. 5' overhang to splice with APB62/63.
65	GACGAGCAAGGCAAGACC GGGATCCGCTTCAACCCA AGCTTGTAGC (SEQ ID NO.34)	SOE 4R. Pair with APB64. 994bp product. Includes 25bp 3' overhang to EMP PCR splice into knockout plasmid backbone.

Upstream and downstream regions of approximately 1kb flanking a 282bp and 176bp internal region of Moth_0864 and Moth_1189, respectively were PCR-amplified from *M. thermoacetica* genomic DNA with compatible overhanging ends.

The two flanking regions for each gene were then assembled into a single molecule of approximately 2kb using SOE (splicing by overlap extension) PCR. These assembled knockout cassettes were independently spliced into the knockout plasmid backbone pDH160 by EMP PCR (Ulrich et al., 2012) to generate two new constructs; pDH177 and pDH180 (Moth_0864 and Moth_1189 knockout plasmids, respectively).

Generation of independent SCO mutant strains of Moth_0864 and Moth_1189 in *M. thermoacetica* ATCC39073

Knockout plasmids are used to independently transform *E. coli* conjugal donor strain S17-1 and resulting strains maintained on selective agar media containing 100µg/ml carbenicillin. For each gene knockout, biomass equivalent to a 10µl inoculation “loopful” from overnight growth of the conjugal donor strain and the conjugal recipient strain (the latter being wild-type *M. thermoacetica* ATCC39073 grown on brain-heart infusion agar (BHIA; Oxoid) supplemented with 2% (w/v) fructose (BHIAF) and incubated at 55°C) are emulsified and spread onto BHIA. The conjugation mix is incubated for 8 hours at 37°C and is then re-suspended in 1ml of pre-reduced ATCC medium 1754 using a sterile spreader. The emulsified conjugation mix is diluted 10^{-1} to 10^{-6} in ATCC medium 1754 and 200µl of each dilution is spread onto selective agar (BHIAF plus kanamycin 150µg/ml) and incubated at 55°C in anaerobic jars. Transconjugant *Moorella* colonies (presumptive SCOs) are typically recovered within 8-10 days. This is believed to be the first account of genetic transformation of *Moorella* sp. using conjugation.

Generation of double crossover (DCO) stable mutants of Moth_0864 and Moth_1189 in *M. thermoacetica* ATCC39073

The following method can be used to isolate chromosomal deletion mutants, generated from SCOs by homologous recombination following sequential passage. Single, isolated transconjugant colonies of *M. thermoacetica* ATCC39073 can be used to independently inoculate 20mL aliquots of pre-reduced ATCC medium 1754 in sealed Hungate tubes and are incubated for at least 24 hours, until turbid; this is passage 1. Following incubation, 4mL of passaged culture is added to 4mL of 50% (v/v) pre-reduced glycerol in a sealed serum bottle and is stored at -80°C. In addition, 100µL of passaged culture is diluted 10^{-1} to 10^{-6} in ATCC medium 1754 and 200µl of each dilution are spread onto selective agar (BHIAF plus kanamycin

150µg/ml) and incubated at 55°C in anaerobic jars to isolate single colonies. Finally, 200µl of passaged culture is used to inoculate a 20mL aliquot of pre-reduced ATCC medium 1754 in a sealed Hungate tube and is incubated for at least 24 hours, until turbid (passage 2). Passaging of SCOs proceeds until kanamycin-sensitive colonies are isolated (see below).

Single colonies isolated from each passage (approximately 100) are replica-plated onto BHIAF with and without 150µg/ml kanamycin and are incubated at 55°C. Kanamycin-sensitive colonies are presumptive double-crossover mutants (i.e the knockout plasmid has been lost following a second recombination event). Genomic DNA is prepared from presumptive DCO mutants and the target gene is PCR-cloned and sequenced to check for the designed deletion mutation.

References for Example 8

Iwasaki, Y., Kita, A., Sakai, S., Takaoka, K., Yano, S., Tajima, T., Kato, J., Nishio, N., Murakami, K., and Nakashimada, Y. (2013). Engineering of a functional thermostable kanamycin resistance marker for use in *Moorella thermoacetica* ATCC39073. *FEMS Microbiol. Lett.* *343*, 8–12.

Pierce, E., Xie, G., Barabote, R.D., Saunders, E., Han, C.S., Detter, J.C., Richardson, P., Brettin, T.S., Das, A., Ljungdahl, L.G., et al. (2008). The complete genome sequence of *Moorella thermoacetica* (f. *Clostridium thermoaceticum*). *Environ. Microbiol.* *10*, 2550–2573.

Ulrich, A., Andersen, K.R., and Schwartz, T.U. (2012). Exponential Megaprimer PCR (EMP) Cloning—Seamless DNA Insertion into Any Target Plasmid without Sequence Constraints. *PLoS ONE* *7*, e53360.

Yakobson, E.A., and Guiney, D.G. (1984). Conjugal transfer of bacterial chromosomes mediated by the RK2 plasmid transfer origin cloned into transposon Tn5. *J. Bacteriol.* *160*, 451–453.

Zhu, Y., Liu, X., and Yang, S.-T. (2005). Construction and characterization of pta gene-deleted mutant of *Clostridium tyrobutyricum* for enhanced butyric acid fermentation. *Biotechnol. Bioeng.* *90*, 154–166.

Example 9. Enzymes for the reduction of 3-hydroxybutanal to 1,3-butanediol

Introduction

This example describes the ability for selected reductases to demonstrate a preference for a C4 aldehyde (model substrate butanal and target 3-hydroxybutanal) relative to a C2 aldehyde (acetaldehyde) as discussed in Example 3; GOX1615 from *Gluconobacter oxydans*, BdhB from *Clostridium acetobutylicum* and GRE2 from *Saccharomyces cerevisiae* were selected for demonstration of this required principle. There follows a description of cloning, purification and enzyme assay for these three selected enzymes.

Gene and protein information

Gene Name	UniProt entry	NCBI ID	Organism	Size of his tagged protein
GOX1615	Q5FQJ0	3248904	<i>Gluconobacter oxydans</i>	39376.7
BdhB CA_C3298	Q04945	1119480	<i>Clostridium acetobutylicum</i>	45450
GRE2 YOL151W	Q12068	854014	<i>Saccharomyces cerevisiae</i> (strain ATCC 204508 / S288c)	

a) GOX1615

Construction of a GOX1615 expression plasmid

Primers Pr89 (5'- GCCATATGGCATCCGACACCATCC (SEQ ID NO.35)) and Pr90 (5'- CCGGATCCTCAGTCCCGTGCC (SEQ ID NO.36)) were used to amplify the *G. oxydans* GOX1615 gene; which had previously been obtained by commercial DNA synthesis and delivered on a plasmid. The amplicon was cloned into pET3a and pET14b (Novagen); with the latter construct adding an N-terminal 6-His tag to the GOX1615 coding sequence in order to facilitate purification of the enzyme by nickel-affinity chromatography.

PCR was performed using Q5 proofreading DNA polymerase (New England Biolabs) following the manufacturer's protocol and using an annealing temperature of 55°C. The resulting PCR product (1008 bp) was purified by gel extraction, and was then digested using NdeI and BamHI restriction endonucleases (New England Biolabs). Following heat inactivation of the restriction enzymes (manufacturer's

protocol), the digested PCR product was ligated into pET14b and pET3a and an aliquot of the ligation mix was used to transform *E. coli* DH10B. Transformants were screened for presence of the GOX16515 gene by colony PCR using T7 forward and reverse primers (using a Taq polymerase with annealing at 55°C). Two positive clones from each transformation were picked for plasmid DNA extraction and the correct constructs further confirmed by restriction digest. The positive clones were stored in 15% glycerol at -80°C. (pET14b-GOX1615: pDH358 and pDH359; pET3a-GOX1615 pDH351, pDH353). Expression plasmid pDH358 was subsequently confirmed by sequencing using primers pET3a-F and pET3a-R.

Expression and purification of GOX1615

Plasmid pDH358 (pET14b-GOX) was used to transform *E. coli* BL21 Star (DE3) with the resulting strain (DH369) stored in 15% glycerol at -80°C. A single colony of DH369, and vector control strain DH228, were inoculated into 5 mL auto inducing medium (per litre: 6g Na₂HPO₄, 3g KH₂PO₄, 5g Yeast extract, 5g NaCl with 10mL 60% v/v glycerol, 5mL 10% w/v glucose, 25mL 8% w/v lactose filter sterilised and added post autoclaving; Studier, F. W. 2005) in 50 mL tubes and the cultures grown overnight at 37°C with continuous shaking at 225rpm. 1 mL aliquots of cells were harvested by centrifugation and the cell pellets resuspended in 200 µl of Bugbuster (Novagen). After incubation at room temperature for 20 minutes, the mix was centrifuged for 5 min at 14000 x g and the supernatant retained. The resulting pellet was resuspended in 200µl 50 mM Tris-HCl pH 7.0. Supernatant and pellet samples were mixed 1:1 was 2x SDS buffer, boiled for 10 min, centrifuged for 2 min at 14000 x g and subsequently 5 µl of each sample was loaded on a 12% gel SDS PAGE gel to confirm presence of GOX1615 protein.

For purification of His-tagged GOX1615, 1 mL of an overnight DH369 culture grown in LB containing 100 µg/mL carbenicillin was used to inoculate 100 mL of the same medium. Cells were grown at 37°C to an OD_{600nm} of 0.6-0.9. Gene expression was induced with 0.4 mM IPTG and cells incubated overnight with shaking at 200 rpm at 18°C.

The culture was harvested by centrifugation and samples were maintained at 4°C for subsequent steps. The pellet was resuspended in 5 mL binding buffer (50 mM Na-phosphate pH 8.0, 0.5 M NaCl, 5 mM Imidazole) and sonicated on ice: 5x 30 sec; amplitude 10% with a 30 sec break between pulses. The lysed cell suspension

was clarified by centrifugation at 14000 x g for 15 mins, 4°C and GOX1615 purified from the cleared cell lysate by affinity purification using a HisTrap™ HP 5 mL column and AKTA start chromatography (GE Healthcare Life Sciences) system following manufacturers protocols. Purified protein was stored at -80°C.

5

Assay of GOX1615 reductase activity with C2 and C4 substrates

Assays were performed in a volume of 1 mL directly in 1.5 mL UV cuvettes. Consumption of NAD(P)H was measured at 340 nm. The reaction was started by addition of 100 µL of substrate solution and measured over 2-5 min.

10 3-Hydroxybutanal (technical grade) was purchased from BOC. Activities are therefore minimal activities when in the linear range as the standard was not pure. Crotonaldehyde was purchased from Sigma.

The results shown below indicate the desired preference GOX1615 displays for a C4 aldehyde including 3-hydroxybutanal and crotonaldehyde compared with activity
15 towards acetaldehyde as substrate:

GOX1615, 0.15 mM NADPH, 50 mM Na-phosphate pH 7.0, 25°C			
Concentration	3-hydroxybutanal	Butanal	Acetaldehyde
Substrate (mM)	(units)		
1	0.1	9.63	0.006
5	2.49	17.7	
10	3.75	23.3	0.12
20	5.28		
GOX1615, 0.15 mM NADH, 50 mM Na-phosphate pH 7.0, 25°C			
10		0.37	

The following table comprises recent data demonstrating an enzyme capable of
20 selective crotonaldehyde reduction to crotyl alcohol:

		NADPH		
Conc. (mM)	3-Hydroxybutanal	Crotonaldehyde	Butanal	Acetaldehyde
	(units)			
0	0	0	0	0
1	0.28	0.09	9.76	0.01

5	1.31	0.38	20.71	0.06
10	2.24	0.90	21.54	0.11

1unit = 1μmol/min/mg

b) BdhB

5

Construction of BdhB expression plasmids

The *bdhB* gene of *Clostridium acetobutylicum* was obtained by commercial DNA synthesis and was independently spliced into expression vectors pET3a and pET14b; in-frame with the 3' sequence encoding a 6-His tag of the latter, using EMP PCR. Sequenced clones pDH365, pDH366 (pET14b-*BdhB*) and pDH380, pDH381 (pET3a-*BdhB*) were used to transform *E. coli* DH10B and were stored in 15% Glycerol at -80°C.

Expression and purification of BdhB

15 Plasmid pDH365 (pET14b-BdhB, MP1) was used to transform BL21*(DE3). The generated culture was stored in LB containing 15% glycerol at -80°C (DH372). A 400 mL auto-inducing media (Foremedia) culture was inoculated from glycerol stock and grown for 20-24h at 30°C, 250 rpm shaking. The culture was spun down (4000 rpm, 30 min, 4°C), the pellet was washed twice with 10 mM sodium phosphate buffer pH 7.0 and resuspended in 5 mL Binding buffer with ZnSO₄ and DTT (10 mM Na-phosphate pH 7.0, 5 mM Imidazole, 0.1 mM ZnSO₄, 1 mM DTT and protease inhibitors). Cells were lysed by glass beads (four cycles of 20 s, 5.5 m/s, two minutes on ice between each cycle) cell lysate was cleared by centrifugation at 4000 rpm, 15 min, 4°C and purified using a 5 ml HisTrap column and AKTA 900 system as described previously for purification of GOX1615. All procedures were carried out under anaerobic conditions

Assay of BdhB reductase with C2 and C4 substrates

30 The activity of BdhB against acetaldehyde and butanal was measured at 25°C at 1 ml reaction volume under anaerobic conditions. Enzyme assays were performed in 50 mM MES buffer pH 6.5 containing 1 mM DTT and 0.1 mM ZnSO₄. Reactions were carried out in disposable UV cuvette sealed with a rubber stopper. Consumption of NAD(P)H was measured at 340 nm. The reaction was started by

adding 100 μ l of the substrate. The linear reaction was measured over a range of 10 min. The following data were obtained.

BdhB, 0.15 mM NADH, 50 mM MES pH 6.5, 0.1mM ZnSO ₄ 25°C		
Concentration	Butanal	Acetaldehyde
Substrate (mM)	(units)	
1	0.03	No measured activity
5	0.06	No measured activity

5

BdhB, 0.15 mM NADPH, 50 mM MES pH 6.5, 0.1mM ZnSO ₄ 25°C		
Concentration	Butanal	Acetaldehyde
Substrate (mM)	(units)	
1	0.14	No measured activity
5	0.27	No measured activity

1unit = 1 μ mol/min/mg

In a separate experiment 3-hydroxybutanal was also shown to be a substrate. At 5mM 3-hydroxybutanal with co factor NADH the rate was 0.012 μ mol/min/mg protein

10

c) GRE2

Construction of a GRE2 expression plasmid

15 GRE2 was PCR-amplified from *Saccharomyces cerevisiae* genomic DNA using a proofreading DNA polymerase and primer pair Pr91 (5'-GCCATATGTCAGTTTTCTGTTTCAGG (SEQ ID NO.37)) and Pr92 (5'-CGGATCCTTATATTCTGCCCTC (SEQ ID NO.38)). The 1038bp PCR product was purified by gel extraction and was then restriction-cloned into expression plasmids
20 pET14b and pET3a via 5' NdeI and 3' BamHI enzyme cleavage sites; a method

well-known in the art. The resulting ligation mixes were used to independently transform aliquots of chemically-competent *E. coli* DH10B. Successful clones for each ligation were identified by colony PCR and further confirmed by restriction analysis of plasmid minipreps. Resulting plasmids were assigned the following IDs:

5 pET14b-GRE2: pDH360; pET3a-GRE2: pDH376.

Expression and purification of GRE2

Plasmid pDH360 was used to transform *E. coli* BL21 Star (DE3). For protein production, the resulting strain (DH370) was used to inoculate an LB medium containing 100 µg/mL carbenicillin and incubated at 37°C to an OD600 of 0.6-0.9. GRE2 expression was induced with 0.4 mM IPTG and incubation at 18°C with shaking at 200 rpm for 18 hours.

Induced bacteria were recovered by centrifugation and protein purification was carried out as follows: bacteria were resuspended in 5 ml binding buffer (50 mM Na-phosphate pH 8.0, 0.5 M NaCl, 5 mM Imidazole) and sonicated on ice for 5x30 sec, amplitude 10% with a 30 sec break between pulses. Lysed bacteria were recovered by centrifugation at 15000 rpm and 4°C. An aliquot of the supernatant was kept for analysis by SDS PAGE before the remaining supernatant was loaded on a 3 or 5 mL nickel affinity column (Qiagen, NTA), which had been equilibrated with 10 column volumes of binding buffer. The flow-through was collected for analysis by SDS PAGE. Unbound protein was washed from the column with 15 ml of binding buffer and 15 ml of wash buffer (50 mM Na-phosphate pH 8.0, 0.5 M NaCl, 100 mM Imidazole). Again the flow-through was collected for SDS gel analysis. Finally bound protein was eluted with 10 ml of elution buffer (50 mM Na-phosphate pH 8.0, 0.5 M NaCl, 400 mM Imidazole). The presence of protein in eluted fractions was rapidly confirmed by Bradford assay and were then further analysed by SDS PAGE for the presence of GRE2. Enzyme-containing fractions were buffer-exchanged to 50 mM Na-phosphate pH 7.0 immediately after purification using PD10 columns. Purified enzyme was stored at 4°C until assay.

Assay of GRE2 reductase activity with C2 and C4 substrates

The activity of GRE2 against acetaldehyde and butanal was studied. Reactions were carried out 25°C at 1 mL reaction volume. Enzyme assays were performed in

- disposable UV cuvette sealed with a rubber stopper. Consumption of NAD(P)H was measured at 340 nm. The reaction was started by adding 100 μ l of the substrate. The linear reaction was measured over a range of 2-5 min. The following data were obtained and show the required preference for the longer chain aldehyde butanal relative to acetaldehyde. As shown for the examples above, GRE2 would be expected to be active on 3-hydroxybutanal.

GRE2, 0.15 mM NAD(P)H, 50 mM Na-phosphate pH 7.0, 25 C			
NADPH		NADH	
Butanal (mM)	units	Butanal (mM)	units
1	5.5	1	0.14
2	13.8	5	0.63
5	24	10	0.84
10	21.6	20	1.35
NADPH		NADH	
Acetaldehyde (mM)	units	Acetaldehyde (mM)	units
10	0.57	5	0.029
40	1.64	10	0.032

1unit = 1 μ mol/min/mg

References for Example 9

Studier, F.W. Protein production by auto-induction in high density shaking cultures. Protein Expr Purif. 2005 May;41(1):207-34.

Example 10. Production of 1,3-butanediol from acetyl CoA (Route 2, Figure 3)

- Described below is an in vitro example of a pathway where DERA is supplied with acetaldehyde from an acetaldehyde dehydrogenase and where the DERA product 3-hydroxybutanal is reduced to 1,3-butanediol using GOX1615 reductase. Further methodology is described in Example 11 below. These data demonstrate how DERA can be supplied acetaldehyde from a preceding pathway enzyme to effect synthesis of hydroxybutanal and a downstream product (here: 1,3-butanediol).

Gene Name	UniProt entry	NCBI ID	Organism	Size of his tagged protein

<i>eutE</i>	P41793	1253985	<i>Salmonella</i> <i>enterica</i> subsp. <i>enterica</i> serovar <i>Typhimurium</i> str. LT2	50KDa
-------------	--------	---------	---	-------

Expression and preparation of lysates containing EutE

E. coli BL21 Star (DE3) cells bearing either an empty pET3a vector (DH228) or a DERA:EutE fusion (DH357; Example 11 below) were inoculated as a seed culture in 5 ml LB medium (10 g/l tryptone, 5 g/l yeast extract, 10 g/L NaCl) containing 100 µg/ml carbenicillin. After overnight growth at 37°C, cultures were diluted to OD_{590nm} 0.1 and grown at 37°C in 50 ml of the same medium to an OD_{590nm} of 0.4 to 0.6. Protein expression was then induced by adding 0.4 mM IPTG, followed by incubation at 18°C overnight with shaking.

Following overnight growth, induced bacteria were recovered by centrifugation at 4,000 rpm for 10 min at 4°C. The pellet was then washed twice with 10 mM sodium phosphate buffer, pH 7.0 and resuspended in 2 ml of lysis buffer (10 mM sodium phosphate buffer pH 7.0 containing 1 mM DTT and protease inhibitor cocktail (SigmaFast, Sigma S8820)). Cells were lysed by sonication as described previously. The resulting lysate was clarified by a centrifugation step at 15,000 x g for 5 min at 4°C and the supernatant was recovered.

Production of 1,3-butanediol was carried out for an arbitrary 120 hours at 25°C using the reagents shown below with a lysate volume of 1/10th the assay volume.

Reagents and concentrations used for the production of 1,3-BDO from acetyl-CoA. Cofactor recycle was achieved using glucose and glucose dehydrogenase (GDH)

Reagent	Concentration in assay
Acetyl-coA	10 mM
NADH	0.15 mM

NADPH	0.15 mM
GDH	0.1 mg/ml
GOX1615	0.011 mg/ml
Glucose	10 mM
EcDERA (Sigma 91252), added where indicated	12mg/ml (3.6 units/mg, Sigma)
Buffer	qs 250 μ l

A control reaction comprised 10 mM acetaldehyde, 0.15 mM NADH, 0.15 mM NADPH, 0.1 mg/mL GDH and 0.011 mg/mL GOX1615 in a 10 mM sodium phosphate buffer, pH 7.0 without addition of DERA or lysate. This control confirmed that no 1,3-BDO was produced via abiotic chemical coupling of acetaldehyde to 3-hydroxybutanal

Detection of 1,3-BDO, ethanol and acetaldehyde was carried out using HPLC (Phenomenex Rezex OA column organic acid H+ 300 $\times$ 7.8 mm).

10 Results

DH228 (Empty Vector no eutE) + EcDERA			
Incubation Conditions	[Acetaldehyde] mM	[1,3 Butanediol] mM	[Ethanol] mM
-GOX1615	ND	ND	ND
+GOX1615	ND	ND	ND
DH357 (eutE-EcDERA fusion) + EcDERA			
Incubation Conditions	[Acetaldehyde] mM	[1,3 Butanediol] mM	[Ethanol] mM
-GOX1615	ND	ND	ND
+GOX1615	ND	0.31	ND

ND = None detected

The control reaction (containing 10mM acetaldehyde alone) exhibited no detectable 1,3-butanediol.

15 1,3-Butanediol was confirmed by LC/mass spectrometry. Representative mass spectrometry data are shown in Figure 16.

Example 11: Construction of a EutE:DERA fusion protein**Introduction**

5 A fusion of heterologous pyruvate dehydrogenase (PDC) and alcohol dehydrogenase (ADHE), expressed in *E. coli* has been previously shown to exhibit improved ethanol production when compared to individually expressed enzymes alone, despite the fusion enzyme having a 20-fold less specific activity for ADH (Lewicka et al 2014). Substrate channelling of acetaldehyde (a cytotoxic, volatile intermediate of relevance to this invention) was attributed to the observed
10 improvement in ethanol titre.

A functional enzyme fusion of DERA and an acetaldehyde dehydrogenase (e.g eutE), or DERA and pyruvate decarboxylase or DERA and an enzyme capable of acetate reduction for example, would be expected to work in a comparable way to
15 improve the conversion of acetyl-CoA to 3-hydroxybutanal; (or pyruvate or acetate to 3-hydroxybutanal) either by substrate channelling of the acetaldehyde intermediate, or by locally increasing the substrate concentration around the DERA active site. In principle any component of a complete DERA pathway could be introduced as a fusion protein to optimise pathway performance.

20 Proteins comprising a biosynthetic pathway may also be linked by other approaches whereby the enzymes are not fused, but are retained in close proximity. For example, the localisation to a bacterial microcompartment by the use of an N-terminal targeting peptide in order to generate an “ethanol bioreactor” within the cell
25 (Lawrence et al 2014); or the potential for use of bacterial scaffoldins to position proteins into a complex (Ding et al 2003). These techniques could equally be applied to the current invention.

Construction of a EutE:DERA fusion

30 An enzyme fusion comprising the *E. coli* K12 DERA (GenBank: CAA26974.1) and EutE (*Salmonella typhimurium* LT2; GenBank: AAL21357.1) was constructed by removing the corresponding start and stop codons from an existing polycistronic expression operon using inverse PCR without the addition of a linker, with the resulting fusion enzyme found to be functional. The following method for creating

enzyme fusions may be applied to one or more of the enzymes comprising a metabolic pathway containing DERA for the purpose of the synthesis of 1,3-butanediol or other chemicals.

5 A divergent primer pair APB142F (5'-AATCAACAGGATATTGAACAGGTGGTG (SEQ ID NO.39); 5' phosphorylated) and APB143R (5'-GTAGCTGCTGGCGCTCTTAC (SEQ ID NO.40)) were designed to remove the intergenic region between adjacent DERA and EutE genes (including the stop codon of the DERA coding sequence and the start codon of the downstream EutE coding sequence in plasmid pDH291, pET3a-DERA-EutE-GOX1615) by inverse PCR; such that the two coding sequences would be fused into one continuous open reading frame when the PCR product (comprising the entire expression plasmid) was re-ligated. The linear 7.8 kb APB142/APB143 inverse PCR product was purified, ligated and used to transform chemically-competent *E. coli* JM109. Two
10 carbenicillin-resistant transformants were subcultured on selective media and assigned strain IDs DH337 and DH338, respectively. Strains DH356 (*E. coli* BL21 Star (DE3)/pDH337), DH357 (*E. coli* BL21 Star (DE3)/pDH338), DH301 (*E. coli* BL21 Star (DE3)/pDH291) and DH228 (*E. coli* BL21 Star/pET3a; negative control) were used to independently inoculate 6 mL of auto-inducing media (Studier 2005) with 100 µg/L carbenicillin and were incubated overnight (16-18h) at 37°C, 225 rpm. Following incubation, biomass was recovered by centrifugation and lysed at 4x 5.5 m/s in a FastPrep bead beater, using 0.1mM acid-washed glass beads. The soluble fraction (supernatant) was recovered by centrifugation (13.4 krpm at 4°C in a bench-top centrifuge) and proteins resolved by 10% SDS PAGE to confirm expression of
20 the 76.76 KDa DERAE-EutE fusion protein in both strains DH356 and DH357.

The fusion protein was expressed as described in Example 10. The fusion protein was confirmed to be active with respect to both acetaldehyde dehydrogenase activity (eutE) and deoxyribose-5-P-phosphate aldolase (DERA) activity. Assays
30 were carried out as described in Example 13 using an alcohol dehydrogenase linked assay to detect the product acetaldehyde from either acetyl CoA or deoxyribose-5-P-phosphate respectively. Measured activities were 4.2µmol/min/mg and 5µmol/min/mg.

References for Example 11

- Ding SY, Lamed R, Bayer EA, Himmel ME. The bacterial scaffoldin: structure, function and potential applications in the nanosciences. *Genet Eng (N Y)*. 2003;25:209-25.
- Lawrence AD, Frank S, Newnham S, Lee MJ, Brown IR, Xue WF, Rowe ML, Mulvihill DP, Prentice MB, Howard MJ, Warren MJ. Solution structure of a bacterial microcompartment targeting peptide and its application in the construction of an ethanol bioreactor. *ACS Synth Biol*. 2014 Jul 18;3(7):454-65.
- Lewicka AJ, Lyczakowski JJ, Blackhurst G, Pashkuleva C, Rothschild-Mancinelli K, Tautvaišas D, Thornton H, Villanueva H, Xiao W, Slikas J, Horsfall L, Elfick A, French C. Fusion of pyruvate decarboxylase and alcohol dehydrogenase increases ethanol production in *Escherichia coli*. *ACS Synth Biol*. 2014 Dec 19;3(12):976-8
- Example 12. Production of 1,3-butanediol from pyruvate using pyruvate decarboxylase (isolated enzymes). Route 4, Figure 3.**

Introduction

- Described below are in vitro examples of a pathway where DERA is supplied acetaldehyde from pyruvate decarboxylase and where the DERA product 3-hydroxybutanal is reduced to 1,3-butanediol using GOX1615 reductase or bdhB dehydrogenase. This work provides detailed data regarding the production of 1,3-butanediol from pyruvate. Increasing pyruvate concentration provides increasing amounts of acetaldehyde supply to the DERA enzyme.
- These data demonstrate how DERA can be supplied acetaldehyde from a preceding pathway enzyme to effect synthesis of hydroxybutanal and a downstream product (here: 1,3-butanediol).

1,3-butanediol production with increasing pyruvate concentration

- Acetaldehyde was supplied to the enzyme system (1ml) via decarboxylation of pyruvate *using S. cerevisiae* pyruvate decarboxylase (PDC1) 0.5 U/ ml (Sigma P9474). Pyruvate was added at 5, 10, 15, 20, 30 and 50mM. The reaction also contained 12mg *E. coli* deoxyribose-5-P aldolase (DERA, 3.6 units/mg, Sigma) and 0.011 mg/ml of purified GOX1615). Recycling of the cofactor NADPH (0.15 mM)

was provided by glucose dehydrogenase (GDH) from *Pseudomonas sp.* (Sigma 19359) added at a final concentration of 0.1 mg/ml. The concentration of reactants and products were monitored by HPLC (Phenomenex Rezex OA column organic acid H+ 300×7.8 mm).

- 5 The reaction was incubated at 25°C for an arbitrary 96 hours.

1,3-butanediol production from increasing concentration of pyruvate.

Test Reaction (PDC + DERA + GOX1615)				
Initial [Pyruvate] mM	[1,3 Butanediol] mM	[Ethanol] mM	[Acetaldehyde] mM	
5	0.65	ND	ND	
10	1.28	ND	ND	
15	1.84	ND	3.06	
20	2.28	1.18	3.17	
30	2.56	1.60	7.64	
50	4.17	1.53	14.86	
Control PDC + GOX1615 (No DERA)				
Initial [Pyruvate] mM	[1,3 Butanediol] mM	[Ethanol] mM	[Acetaldehyde] mM	
5	ND	1.01	3.51	
10	ND	2.19	7.23	
15	ND	2.84	9.86	
20	ND	3.53	13.07	
30	ND	5.16	22.81	
50	ND	4.70	33.61	
Control PDC + DERA (No GOX 1615)				
Initial [Pyruvate] mM	[1,3 Butanediol] mM	[Ethanol] mM	[Acetaldehyde] mM	
5	ND	ND	1.37	
10	ND	ND	1.73	
15	ND	ND	2.34	
20	ND	ND	3.66	
30	ND	ND	7.17	
50	ND	ND	14.70	

An example using bdhB dehydrogenase is given below. The same method was used except that pyruvate was provided at 30mM and bdhB added at 0.07mg/ml

96 Incubations	Hour	BdhB Test reaction (PDC + DERA + GOX1615)		
Initial mM	[Pyruvate]	[Acetaldehyde] mM	[1,3 Butanediol] mM	[Ethanol] mM
5		1.35	ND	1.73
30		9.34	0.4	2.46
Control PDC + BdhB (No DERA)				
Initial mM	[Pyruvate]	[Acetaldehyde] mM	[1,3 Butanediol] mM	[Ethanol] mM
5		4.64	ND	1.14
30		32.87	ND	1.45

ND Not detected

1,3-butanediol was confirmed in all cases by LC/mass spectrometry. Representative mass spectrometry data is shown in Figure 16.

1,3-butanediol production from 5mM or 30mM pyruvate at different DERA concentrations

Reactions containing 5mM or 30mM pyruvate as substrate and GDH cofactor recycling for GOX1615 were set up as above except the amount of *E.coli* DERA (3.6 units/mg, Sigma) was varied at , 12, 6, 3, 1.5, 0.75 mg. Reactants and products were monitored as above.

The reaction was incubated at 25°C for an arbitrary 96 hours.

1,3-butanediol production

Pyruvate supplied at 5mM			
DERA mg/ml	[Acetaldehyde] mM	[1,3 Butanediol] mM	[Ethanol] mM
0.75	4.07	0.17	ND
1.5	3.26	0.24	ND
3	1.89	0.36	ND

6	1.64	0.43	ND
12	1.35	0.53	ND
Pyruvate supplied at 30mM			
DERA mg/ml	[Acetaldehyde] mM	[1,3 Butanediol] mM	[Ethanol] mM
0.75	31.32	0.47	2.52
1.5	29.62	0.80	2.11
3	26.62	1.36	1.69
6	22.01	1.93	1.71
12	14.76	2.48	1.25
Control PDC and GOX1615 no DERA			
[Pyruvate] mM	[Acetaldehyde] mM	[1,3 Butanediol] mM	[Ethanol] mM
5	4.61	ND	ND
30	30.68	ND	2.14

ND = none detected

1,3-butanediol was confirmed in all cases by LC/mass spectrometry. A representative mass spectrum is shown in Figure 16.

- 5 Overall, the production of ethanol as a by-product can be improved either by improvement of the DERA enzyme (e.g evolution for better kinetics) or further evolution of a selective reductase towards reduction of 1,3-butanediol.

10 **Example 13. Production of 1,3-butanediol from pyruvate using selected DERAs from different microbial sources**

Introduction

15 This example demonstrates that DERA's from a range of microorganisms can be suitable for coupling of two molecules of acetaldehyde to the intermediate 3-hydroxybutanal as part of a novel, in vivo unnatural metabolic pathway.

Target gene and protein information

Name	UniProt entry	NCBI ID	Source organism	Size of protein
------	------------------	---------	-----------------	--------------------

EcDERA	P0A6L0	948902	<i>Escherichia coli</i>	27.73 KDa
AwDERA	H6LF13	WP_01435452 3.1	<i>Acetobacterium woodii</i>	23.84 KDa
PaDERA	Q8ZXK7	1465578	<i>Pyrobaculum aerophilum</i>	24.54 KDa
GtDERA	A4IR26	WP_00887991 4.1	<i>Geobacillus thermodenitrificans NG80-2</i>	23.3 KDa

Isolation and sequencing of a *Geobacillus thermodenitrificans* strain NG80-2 DERA homolog from *Geobacillus thermodenitrificans* strain K1041

5 A homolog of the gene GTNG_2435 from *G. thermodenitrificans* strain NG80-2 (for which there is a published genome sequence) was identified, PCR cloned and sequenced from *G. thermodenitrificans* K1041. PCR primers APB106F (5'-ATGACGGTGAATATTGCTAAAATGATCG (SEQ ID NO.41)) and APB107R (5'-TTAATAGTCAGCGCCGCCGTTTGTG (SEQ ID NO.42)) were designed based on

10 the GNTG_2435 sequence as a template, and were used along with Q5 High-Fidelity DNA Polymerase (New England Biolabs) and the manufacturer's recommended PCR reaction conditions to PCR-clone an approximate 672bp product from *G. thermodenitrificans* K1041 genomic DNA; confirmed by agarose gel electrophoresis. The PCR product was directly ligated into cloning vector pJET1.2

15 using the CloneJET PCR Cloning Kit (Thermo Scientific) according to the manufacturer's protocol for blunt-ended PCR products, with the resulting ligation mix used to transform chemically-competent *E. coli* DH10B, with transformants selected by incubation on Luria Agar (LA; Sigma) plus carbenicillin at a concentration of 100µg/mL. Transformant *E. coli* DH10B colonies recovered

20 following 16 hours incubation at 37°C were replica-plated onto LA plus carbenicillin 100µg/mL and checked for the presence of cloned APB106F/APB107R PCR product using primer pair APB106F and pJET1.2 reverse sequencing primer (Thermo Scientific) with DreamTaq DNA Polymerase (Thermo Scientific) in a colony PCR reaction; with the presence of an approximate 729bp product in PCR-positive

25 transformant colonies confirmed by agarose gel electrophoresis. Two of these were stored with strain IDs DH208 and DH209, respectively. Plasmids were isolated from

these strains by alkaline lysis and were sequenced using pJET1.2 forward and reverse sequencing primers (Thermo Scientific) to derive the sequence of *G. thermodenitrificans* strain K1041 GNTG_2435 homolog. This gene was subsequently identified as encoding a putative DERA by both nucleotide sequence homology with GNTG_2435 and identification of conserved domains within the translated primary amino acid sequence using the NCBI BLAST web server. The sequence of the K1041 homolog is reproduced below:

ATGACGGTGAATATTGCTAAAATGATCGATCATACGTTGCTTAAGCCAGAAGCG
ACGGAAGAGCAAATCATTCAACTATGCGACGAAGCAAAGCAACACGGCTTCGC
CTCGGTGTGCGTCAACCCAGCGTGGGTGAAAACAGCGGCACGCGAGCTTTCC
GACACTGATGTCCGCGTCTGCACGGTCATCGGCTTTCCGCTTGGGGCGACGA
CGCCGGAACAAAGGCGTTTGAAACGAACAACGCTATCGAAAACGGCGCCCG
CGAAGTCGATATGGTAATCAACATCGGCGCGTTAAAAAGTGGTAACGATGAAC
TCGTTGAGCGCGACATTCGTGCGGTTGTTGAGGCGGCGTCCGGGAAAGCGCT
TGTGAAAGTGATCATCGAAACGGCCTTGTTGACTGATGAGGAAAAAGTGCGCG
CCTGCCAATTGGCGGTGAAAGCGGGCGCCGATTACGTAAAAACGTCGACCGG
ATTCTCAGGCGGCGGAGCGACGGTCTGAAGACGTGGCGCTGATGCGCCGGAC
AGTTGGCGATAAAGCAGGTGTCAAAGCCTCAGGAGGCGTCCGCGACCGAAAA
ACAGCCGAAGCGATGATTGAAGCTGGGGCCACGCGCATTGGGACGAGCTCCG
GGGTGGCGATCGTCAGCGGCCAAACCGGCGGCGCTGACTATTAA (SEQ ID
NO.43)

Construction of expression vectors containing *E. coli*, *A. woodii*, *P. aerophilum* and *G. thermodenitrificans* deoxyribose-5-P aldolases (DERAs)

PaDERA was obtained by commercial DNA synthesis using the published gene sequence as a template (NCBI GID: 1465578) and was supplied on a plasmid. The *G. thermodenitrificans* DERA was isolated as described above. *E. coli* and *A. woodii* were directly isolated from their respective genomic DNA with primers designed using the published genome sequences as templates. Using methods well known in the art in order to generate the final expression constructs; each of the target DERAs was PCR-amplified with 5' NdeI and 3' BamHI restriction sites and then independently subcloned using standard restriction enzyme-based cloning methods into the corresponding sites of the pET3a expression plasmid backbone, such that they were in frame with the plasmid-encoded T7 inducible promoter.

Expression and preparation of lysates containing DERAs for 1,3-butanediol production

- 5 Strains of *E. coli* BL21 Star (DE3) bearing either an empty pET3a vector or a cloned DERA from *E. coli*, *G. thermodenitrificans*, *A. woodii* and *P. aerophilum* were grown in 50 mL of commercial auto-induction medium (Formedium) containing 100µg/ml carbenicillin, at 30°C with shaking at 250rpm. Following overnight growth, bacteria were lysed by bead-beating as described previously. The resulting lysates were
- 10 clarified by centrifugation prior to activity assays. DERA activity for each lysate was determined in the retro aldol direction against deoxyribose-5-phosphate using a NADH linked assay for detection of the product acetaldehyde. The assay was carried out using 0.15mM NADH, 5mM 2-Deoxyribose 5-phosphate (Sigma: D3126) and 10U/ml alcohol dehydrogenase (Sigma: A7011)
- 15 Lysates were diluted and reactions were run at the following volumetric and specific activities:

Cloned DERA	Volumetric activity	Specific Activity
<i>Escherichia coli</i>	2U/ml and 20U/ml	0.09U/mg and 0.9U/mg
<i>Geobacillus thermodenitrificans</i>	2U/ml and 20U/ml	0.09U/mg and 0.9U/mg
<i>Acetobacterium woodii</i>	2U/ml and 20U/ml	0.09U/mg and 0.9U/mg
<i>Pyrobaculum aerophilum</i>	0.3U/ml	0.021U/mg

- Acetaldehyde was supplied to the enzyme system (1mL) via decarboxylation of pyruvate using yeast pyruvate decarboxylase (PDC) 0.5 U/ ml (Sigma: P9474).
- 20 Pyruvate was added at 5 and 30mM. The reaction also contained the cloned DERA at either 2, 20 or 0.3 U/ml as appropriate and 0.033 mg/ml of purified GOX1615

(Example 9). The assays were carried out in 10mM sodium phosphate buffer, pH 7 containing 0.1mM thiamine pyrophosphate, 1mM MgSO₄ and 1mM DTT. Recycling of the cofactor NADPH for GOX1615 (0.15 mM) was provided by glucose dehydrogenase (GDH) from *Pseudomonas sp.* (Sigma: 19359) added at a final concentration of 0.1 mg/ml and 10mM glucose. The reaction was incubated at 25°C, shaking at 250rpm for an arbitrary 96hr and was cooled on ice prior analysis. The concentration of reactants and products were monitored by HPLC (Phenomenex Rezex OA column organic acid H+ 300×7.8 mm).

A control comprised 5 and 30 mM pyruvate, 0.5U/ml PDC, 0.15 mM NADPH, 0.1 mg/mL GDH, 10mM glucose and 0.033 mg/ml GOX 1615 in buffer, pH 7.0 without addition of DERA lysate. This control confirmed that no 1,3-butanediol was produced via abiotic chemical coupling of acetaldehyde to 3-hydroxybutanal.

Results

1,3-butanediol production (*Escherichia coli* DERA)

Pyruvate supplied at 5mM			
DERA U/ml	[Acetaldehyde] mM	[1,3 Butanediol] mM	[Ethanol] mM
2	ND	0.23	1.86
20	ND	0.31	3.67
Pyruvate supplied at 30mM			
DERA mg/ml	[Acetaldehyde] mM	[1,3 Butanediol] mM	[Ethanol] mM
2	23.80	0.54	5.68
20	ND	2.32	7.74
Control PDC and GOX1615 no DERA			
[Pyruvate] mM	[Acetaldehyde] mM	[1,3 Butanediol] mM	[Ethanol] mM
5	3.56	ND	1.37
30	26.11	ND	5.62

1,3-butanediol production (*Acetobacterium woodii* DERA)

Pyruvate supplied at 5mM			
DERA U/ml	[Acetaldehyde] mM	[1,3 Butanediol] mM	[Ethanol] mM
2	ND	0.32	1.54
20	ND	0.37	2.69
Pyruvate supplied at 30mM			
DERA mg/ml	[Acetaldehyde] mM	[1,3 Butanediol] mM	[Ethanol] mM
2	23.4	0.95	5.31
20	ND	3.19	6.32
Control PDC and GOX1615 no DERA			
[Pyruvate] mM	[Acetaldehyde] mM	[1,3 Butanediol] mM	[Ethanol] mM
5	3.56	ND	1.37
30	26.11	ND	5.62

1,3-butanediol production (Geobacillus thermodenitrificans DERA)

Pyruvate supplied at 5mM			
DERA U/ml	[Acetaldehyde] mM	[1,3 Butanediol] mM	[Ethanol] mM
2	ND	0.64	3.00
20	ND	0.24	7.45
Pyruvate supplied at 30mM			
DERA mg/ml	[Acetaldehyde] mM	[1,3 Butanediol] mM	[Ethanol] mM
2	ND	3.39	5.46
20	ND	2.19	14.55
Control PDC and GOX1615 no DERA			
[Pyruvate] mM	[Acetaldehyde] mM	[1,3 Butanediol] mM	[Ethanol] mM
5	3.61	ND	0.82
30	26.83	ND	6.41

1,3-butanediol production (Pyrobaculum aerophilum DERA)

Pyruvate supplied at 5mM			
DERA U/ml	[Acetaldehyde] mM	[1,3 Butanediol] mM	[Ethanol] mM
0.3U/ml	ND	1.03	3.27
Pyruvate supplied at 30mM			
DERA mg/ml	[Acetaldehyde] mM	[1,3 Butanediol] mM	[Ethanol] mM
0.3	ND	4.36	3.25
Control PDC and GOX1615 no DERA			
[Pyruvate] mM	[Acetaldehyde] mM	[1,3 Butanediol] mM	[Ethanol] mM
5	3.63	ND	1.98
30	27.6	ND	6.09

ND = none detected

1,3-butanediol was confirmed in all cases by LC/mass spectrometry. A representative mass spectrum is shown in Figure 16.

Example 14. Production of 1,3-butanediol from pyruvate using a cloned full pathway operon

10 Introduction

For exemplification of a complete 1,3-BDO biosynthetic pathway expressed as an operon, EcDERA, PDC1 and GOX1615 genes were assembled as a single polycistronic operon, under a lactose-inducible T7 promoter (in expression vector pET3a) and were actively expressed to produce 1,3-BDO in *E. coli* BL21 Star (DE3) cell lysate.

The aldehyde oxidoreductase (eutE) from *Salmonella enterica subsp. enterica serovar Typhimurium* strain LT2 (Uniprot ID: P41793, Genbank GID: 1253985) was initially used as an endogenous source of acetaldehyde substrate for EcDERA. In a later embodiment cloned PDC1 from *Saccharomyces cerevisiae* was used to replace eutE in this construct, as described below.

Construction of EcDERA and PDC with GOX1615 expression vectors

The PDC1 gene was PCR cloned from *S. cerevisiae* genomic DNA (SG ID S000004034 and Candy et al. 1991) with 24bp of homology for 5' UTR of GOX1615 in plasmid pDH384 (pET3a-EcDERA-EutE-GOX1615). The purified PCR product was then spliced into pDH384 using EMP PCR; such that PDC1 would replace the eutE coding sequence in the final construct and would also be cloned in-frame with the original ribosome-binding site; creating expression plasmid pDH527 (pET3a-EcDERA-PDC1-GOX1615).

Expression and preparation of lysates containing the expressed operon

Cell lysates from induced strains of *E. coli* BL21 Star (DE3) were prepared as described in Example 13.

The assays were carried out in 10mM sodium phosphate buffer, pH 7 containing 0.1mM thiamine pyrophosphate, 1mM MgSO₄ and 1mM DTT. Pyruvate was added to a final concentration of 5mM and 30mM. The lysate was diluted to contain units of expressed PDC, DERA and GOX 1615 as described below. The reaction was incubated at 25°C, shaking at 250rpm for an arbitrary 96hr and was cooled on ice prior to analysis by HPLC as described in Example 13.

The activity of the cloned DERAs was carried out using a NADH linked assay using 2-deoxyribose-5-phosphate as the substrate. The assay was carried out using 0.15mM NADH, 5mM 2-Deoxyribose 5-phosphate (Sigma: D3126) and 10U/ml alcohol dehydrogenase (Sigma: A7011). The activity of PDC was carried out using a linked assay using 10mM sodium pyruvate, 0.15mM NADH and 10U/ml alcohol dehydrogenase (Sigma: A7011). The activity of GOX 1615 was carried out using 10mM butanal and 0.15mM NADPH.

Activity of each pathway enzyme in the reaction	Volumetric activity in each reaction	Specific Activity
---	--------------------------------------	-------------------

	as appropriate	
DERA measured against 2-deoxyribose-5-phosphate)	2U/ml or 20U/ml	0.22U/mg or 2.2U/mg
PDC measured against pyruvate	0.28U/ml or 2.8U/ml	0.03U/mg or 0.3U/mg
GOX1625 measured against butanal	0.06U/ml and 0.6U/ml	0.007U/mg and 0.07U/mg

Results

Pyruvate supplied at 5mM				
DERA U/ml		[Acetaldehyde] mM	[1,3 Butanediol] mM	[Ethanol] mM
2		ND	0.15	2.00
20		ND	0.45	3.63
Pyruvate supplied at 30mM				
DERA U/ml		[Acetaldehyde] mM	[1,3 Butanediol] mM	[Ethanol] mM
2		ND	0.82	5.21
20		3.24	4.01	7.67
Control PDC and GOX1615 no DERA)				
[Pyruvate] mM		[Acetaldehyde] mM	[1,3 Butanediol] mM	[Ethanol] mM
5		ND	ND	4.05
30		ND	ND	7.94

ND = none detected

1,3-butanediol was confirmed in all cases by LC/mass spectrometry. A representative mass spectrum is shown in Figure 16.

These data successfully demonstrate the ability for synthesis of 1,3-butanediol from a fully cloned novel, unnatural metabolic pathway, containing the key enzyme DERA for coupling of two molecules of acetaldehyde.

References for Example 14

- 5 Candy JM, Duggleby RG, Mattick JS. Expression of active yeast pyruvate decarboxylase in Escherichia coli. J Gen Microbiol. 1991 Dec;137(12):2811-5.

Example 15 - Creation of the novel consensus sequence as a tool for targeted DERA engineering

10

DERA protein sequences for generation of the consensus sequence were obtained from the UniProtKB/Swiss-Prot database (<http://www.uniprot.org>), searching using the Enzyme Commission (EC) number of DERA, EC 4.1.2.4. Results were filtered to include only manually annotated and reviewed prokaryotic and eukaryotic peptide

15

sequences (336 sequences at time of writing, Accession Numbers: A0AKA1, A0KCV5, A0KPE4, A0KU07, A0PVY3, A0PYX5, A0QLL2, A1AJU8, A1JM38, A1KFB2, A1RH87, A1RU26, A1S474, A2RCN4, A2RK45, A3CMP6, A3D7J4, A3N123, A3QGT3, A4IR26, A4SRU4, A4TN86, A4VVB2, A4W1L5, A4W698, A4WHP7, A4XLW0, A4Y9A8, A5F5S6, A5I237, A5IM24, A5N0Z3, A5TZK6,

20

A5UCY8, A6T962, A6UCC1, A6VQW4, A6WRB8, A7FK39, A7FU73, A7GDP8, A7GNS8, A7MGB0, A7MUW7, A7ZAF1, A7ZVS4, A8A8B0, A8AX59, A8FJ14, A8FYQ9, A8G9H6, A8H728, A8MH18, A9KZ80, A9NG35, A9R4U4, A9VQK2, B0BPV4, B0JK91, B0K709, B0KA53, B0R6I6, B0S0N1, B0TQ91, B0UTU0, B1AJM6, B1HTK8, B1IB13, B1ILA1, B1IS38, B1JRK6, B1KCV0, B1KRP8, B1L1S5,

25

B1LB68, B1LEI6, B1WW41, B1XFJ1, B1XI10, B1YB76, B1YKT7, B2A4J4, B2GG36, B2HQU3, B2INW0, B2K9V8, B2KBN0, B2S2L2, B2TL86, B2TZR4, B2V184, B2VH50, B3H1P7, B3PM95, B4ENZ2, B4EWA4, B4TGZ9, B4U2S5, B5E3L5, B5FA98, B5FTC5, B5XID8, B5Y277, B5Y7V3, B5Z4R3, B5ZCA0, B6ENG3, B6I6M8, B7HIT0, B7HMQ9, B7IS80, B7JJK4, B7K8P2, B7LEM7,

30

B7LNS1, B7LXU3, B7MNI8, B7N2V7, B7NH49, B7NW61, B7UR09, B7VJF8, B8CKI4, B8DBT6, B8E0F1, B8E6P4, B8F671, B8FUK9, B8HXS4, B8ZNP4, B8ZT93, B9DMC3, B9DS93, B9E4U5, B9E8I5, B9IXA5, B9LR64, C0M9B0, C0MG52, C0ZB54, C1AKF6, C1C6H7, C1CDJ0, C1CJT2, C1CS48, C1EQH0, C1FN33, C1KWT9, C3KVV7, C3L6K8, C3LQC2, C3P786, C3PI20, C4L415,

- C4Z3P0, C4ZT63, C5A366, C5BHI2, C5C9E5, C5D4T6, C6DEP1, O26909,
 O66540, O83288, P09924, P0A6L0, P0A6L1, P0CH94, P0DA62, P0DA63, P39121,
 P43048, P44430, P47296, P47722, P57937, P61084, P61108, P63930, P63932,
 P73618, P99102, P99174, P9WP02, P9WP03, Q02YD6, Q04L69, Q086G0,
 5 Q0HLF0, Q0HXQ4, Q0I2N7, Q0RRG0, Q0SRC4, Q0SX30, Q0T8T2, Q0TNQ8,
 Q0W6M2, Q18C22, Q18FF3, Q19264, Q1BVA7, Q1CAE0, Q1CG96, Q1J4Z3,
 Q1J9Z9, Q1JF38, Q1JK46, Q24SU9, Q2RRZ2, Q2SRA7, Q2YUL6, Q2YUU4,
 Q31SV8, Q327L5, Q38XI2, Q39NL8, Q3A247, Q3ABV0, Q3IQA5, Q3JYQ4,
 Q3MF82, Q3T0V9, Q3YU12, Q48RH2, Q49Z84, Q4A5W6, Q4A7J6, Q4A9F7,
 10 Q4JSV3, Q4L819, Q4QLH8, Q4ZMV1, Q5E7J7, Q5FLZ2, Q5HE63, Q5HJN0,
 Q5HM84, Q5KX02, Q5N4L8, Q5SJ28, Q5UX95, Q5WHF4, Q5XA31, Q5YP36,
 Q600B2, Q63CR9, Q65D22, Q65H59, Q65WA1, Q66CP8, Q66EW0, Q67RZ3,
 Q6A655, Q6A8F1, Q6ALS3, Q6D3R0, Q6F0H8, Q6F1Z6, Q6G7H3, Q6GCY6,
 Q6GET8, Q6GKG7, Q6HK62, Q6KHA7, Q6LUH4, Q6MGR8, Q6MSE7, Q6NJX0,
 15 Q71Y27, Q72JE9, Q73A11, Q73QJ5, Q73SV2, Q74IA4, Q7MI38, Q7MP37,
 Q7N932, Q7NAQ0, Q7NFI7, Q7NRS9, Q7UPT7, Q7VMS9, Q81EY8, Q81RZ3,
 Q839J1, Q83P02, Q877I0, Q87M22, Q88Z64, Q892U4, Q89ZF2, Q8CNH7,
 Q8DBT2, Q8DJZ1, Q8DQC4, Q8DU34, Q8DWZ0, Q8E2U1, Q8EHK4, Q8EMT9,
 Q8EPW8, Q8EWT4, Q8FSJ0, Q8NTC4, Q8NVF5, Q8NYR1, Q8RB49, Q8UJ09,
 20 Q8XB36, Q8XIR2, Q8Y5R1, Q8YPM0, Q8Z0U3, Q8ZGH4, Q8ZIQ4, Q8ZJV8,
 Q8ZXK7, Q91YP3, Q92A19, Q92MQ3, Q97CC6, Q97IU5, Q97RH2, Q98QP7,
 Q99Y51, Q9CB45, Q9CFM7, Q9HKB7, Q9HP08, Q9KD67, Q9KPL7, Q9PPQ4,
 Q9RV25, Q9X1P5, Q9Y315, Q9Y948).
- 25 Sequences corresponding to the accession numbers were downloaded in FASTA
 format and a multiple sequence alignment generated using CLC Sequence Viewer 7
 (<http://www.clcbio.com/products/clc-sequence-viewer/>). The parameters of the
 alignment were the default settings of the program: Gap open cost = 10.0; Gap
 extension cost = 1.0; End gap cost = As any other; Alignment = Very accurate
 30 (slow); Redo alignments = No; Use fixpoints = No. This alignment allowed for a
 novel consensus sequence encompassing all the above DERAs to be elucidated
 (SEQ ID NO1). DERA consensus sequence is shown in Figure 25.

Example 16 - Alignment of any query sequence with the consensus sequence

Pairwise alignment of any two amino acid sequences allows for evolutionarily conserved regions to be identified independent of the sequence origin. In this example the utility of the consensus sequence to identify complementary residues in any query amino acid sequence is demonstrated. Thus, in the circumstance where residues in the consensus sequence have been identified as targets for substitution by an alternative amino acid (e.g. to improve a desired function), these target sites can be transcribed onto the query sequence and the equivalent substitutions introduced in the query. The utility of query sequence alignment to the consensus sequence in identifying targets for substitution is further enhanced as it does not rely on the query and consensus sequences being of equal length. Indeed, significant gap regions can be inserted by the alignment algorithm and yet the amino acids in the query corresponding to those identified for mutation in the consensus remain perfectly aligned. Therefore, by counting amino acid positions from the N terminus of the query sequence (ignoring gaps) the exact residue in the query corresponding to the 12 hotspot amino acids of the consensus can be identified. The broad applicability is demonstrated by performing a non-exhaustive list of alignments of the consensus sequence against DERA sequences from evolutionarily distant domains of life (Figures 19, 20, 21, 22, 23, and 24).

Further, with respect to conserved regions of interest beyond those labelled as Hotspot A, B, C and D in Figure 18, the consensus sequence may be applied with equal utility. Hence, this invention relates to any mutation at a position that aligns with the consensus sequence and which imparts a performance benefit with respect to the coupling or condensation of two molecules of acetaldehyde to form either 3-hydroxybutanal or crotonaldehyde.

The directed evolution of 2-deoxy-D-ribose 5-phosphate aldolase DERA for industrial purposes has been described in the literature (Jennewein *et al.* 2006, Biotechnol J. 1(5):537; WO2005118794 A3). Here random mutagenesis of the *E. coli* K12 deoC gene encoding DERA was used, in conjunction with a high throughput screen, to generate and identify DERA mutants with enhanced resistance to chloroacetaldehyde and improved production of (3*R*,5*S*)-6-chloro-

2,4,6-trideoxyhexapyranoside. Several improved variants were discovered including those with the following amino acid substitutions: K13C, T19S, M185V, M185T, F200I and S239C.

5 The utility of the consensus more broadly than identifying equivalent residues to just the 12 amino acids in the hotspot regions of the consensus sequence, can be demonstrated using the example above. Despite being performed with a target of improving coupling of two molecules of acetaldehyde with chloroacetaldehyde to form the C6 molecule (3*R*,5*S*)-6-chloro-2,4,6-trideoxyhexapyranoside, rather than
10 coupling of 2 molecules of acetaldehyde to form 3-hydroxybutanal, as is the intention in the current disclosure, the amino acids identified above as playing an important role as targets for substitution could be identified using the consensus sequence in SEQ ID No1 as a tool. In practice, on alignment of the consensus sequence with the *E. coli* DERA sequence (Figure 19), it can be seen that residues
15 K3, L9, M166, V176 and A203 of the consensus sequence align with K13, T19, M185, F200 and S239 of the *E. coli* DERA sequence. Therefore, without any *a priori* knowledge of the *E. coli* DERA enzyme, the consensus would have identified these influential target amino acids for mutagenesis of the underlying DNA. Any amino acids which align with the consensus sequence are considered potentially influential
20 and hence may be targets for substitution with specific regard to improving DERA performance for 3-hydroxybutanal or crotonaldehyde synthesis from acetaldehyde.

Pairwise alignment of query sequences against the consensus sequence

Pairwise alignments to the consensus sequence (SEQ ID NO1) were performed
25 using CLC Sequence Viewer 7 on default settings (as described above). Alternative software for pairwise alignment include EMBOSS Needle (http://www.ebi.ac.uk/Tools/psa/emboss_needle/) on default settings (Matrix = BLOSUM62; Gap Open = 10; Gap Extend = 0.5; Output format = pair; End Gap Penalty = false; End Gap Open = 10; End Gap Extend = 0.5). Example alignments
30 were performed under these conditions with selected examples of DERA peptide sequences from *E. coli* (Figure 19), *Homo sapiens* (Figure 20), *Plasmodium falciparum* (Figure 21), *Pyrobaculum aerophilum* (Figure 22), *Geobacillus thermoglucosidasius* (Figure 23) and *Acetobacterium woodii* (Figure 24).

Example 17 - Identification of deoxyribose-5'-phosphate aldolase enzyme variants with enhanced activity for 3-hydroxybutanal synthesis or crotonaldehyde synthesis

- 5 Following selection of example residues within regions of interest (Hotspots A, B, C, and D - Figure 18), alternative residues may also be selected for introduction. A list of example mutants is presented in the tables 8 and 9 below, encompassing single and multiple mutations of each of these key residues.

10

Table 8 - List of selected residues for mutagenesis

Hotspot	Consensus	Changes (including original consensus amino acid)	Number of variants
A	T8	L,I,F,V,A,Y,W	8
D	L10	,I,F,V,A,Y,W,D,E,H	10
A	C35	L,I,F,V,A,Y,W	8
A	V57	,L,I,F,A,Y,W	7
A	F60	L,I,V,A,Y,W	7
C	G151	,L,I,F,V,A,Y,W,D,E	10
B	S178	H,Y,R,D,E,,L,I,F,V,A,W	12
C	G179	L,I,F,V,A,Y,W,D,,E	10
C	G180	,L,I,F,V,A,Y,W,D,E	10
B	G199	H,Y,R,D,E,L,I,F,V,A,W	12
C	A200	,L,I,F,V,Y,W,D,,E	9
C	S201	,L,I,F,V,A,Y,W,D,,E	10

Table 9 - Reduced list of selected residues for mutagenesis.

Hotspot	Consensus	Changes (including original consensus amino acid)	Number of variants
A	T8	L,I,V,F, A	5
D	L10	H,D,E	3
A	C35	L,I,V,F, A	5
A	V57	L,I,F,A	4
A	F60	L,I,V,A	4
C	G151	, L, I, F, V, A,, W, D	8
B	S178	D,E,H	3
C	G179	L, I, F, V, A, W, D	8
C	G180	L, I, F, V, A, W, D	8
B	G199	D,E,H	3
C	A200	L, I, F, V, W, D	7
C	S201	L, I, F, V, A, W, D	8

Example 18 - Assessment of improved DERA variants producing 3-hydroxybutanal or crotonaldehyde

5

10

DERA variants may be purified or assayed in crude cell lysates as appropriate and assayed for improved performance alongside the parent DERA enzyme from which the variant originates and does not comprise the modifications. Assay of DERA using acetaldehyde supply via decarboxylation of pyruvate using pyruvate decarboxylase is described in examples 12 and 13. Alternatively, acetaldehyde can be directly added to the system. Due to the equilibrium with respect to the aldolase reaction, assessment of DERA performance via a linked assay with co factor recycle is preferred. Hence, 3-hydroxybutanal production efficiency is linked to

reduction of 3-hydroxybutanal to 1,3-BDO in a system where the reductase enzyme is not rate limiting. The reductase requires high selectivity for the target C4 aldehyde relative to the C2 aldehyde acetaldehyde. The reductase GOX1615 (Example 12 and 13) demonstrates such selectivity, further exemplified by Richter, N *et al.* 2009, ChemBioChem 10, 1888.

Activity assessment *in vivo* within a metabolic pathway

DERA variant performance may be tested *in vivo* as part of an unnatural metabolic pathway. For example, DERA variants may be assessed alongside the parent wild type DERA in an unnatural 1,3-butanediol biosynthetic pathway expressed from an operon *in vivo*, using *E. coli* BL21 Star (DE3) strain DH527. The construction of this strain is described in Example 14). In this example, *in vivo* acetaldehyde delivery to the DERA enzyme is achieved by decarboxylation of pyruvate using pyruvate decarboxylase. Reduction of DERA generated 3-hydroxybutanal to 1,3-butanediol is achieved using GOX1615 reductase.

Growth of *E. coli* cultures

Strain DH527 as well as *E. coli* BL21 Star (DE3) bearing an empty pET3a vector (control) were grown overnight in 50 ml of commercial auto-induction medium (Formedium) containing 100µg/ml carbenicillin, at 30°C with shaking at 250rpm. Cells were harvested by centrifugation at 4,000 rpm for 10 min at 4°C. The pellet was then washed with 10 mM sodium phosphate buffer, pH 7.0 and resuspended to a selected final OD_{600nm} in sodium phosphate buffer pH7 containing 0.1mM thymine pyrophosphate (TPP), 1mM MgSO₄ and 10g /l D-glucose. The selected OD_{600nm} depended on expected enzyme performance. The resting cell system was incubated at 25°C with shaking at 250rpm overnight.

Assay of *in vivo* 1,3 butanediol production

A 1ml sample was taken and cells harvested by centrifugation at 14,000rpm for 5 minutes at 4°C. 100µL of supernatant was mixed 1:1 with 1% trifluoroacetic acid and the concentration of reactants and products was monitored by HPLC

(Phenomenex Rezex OA column organic acid H+ 300×7.8 mm). Reactions were assayed at time intervals as appropriate to performance. After 24h incubation, the wild type *E. coli* DERA in this system was sufficient to produce 1,3-butanediol at a concentration of 2.05mM. Confirmation of 1,3-butanediol production was also supported by LC mass spectrometry, and performed as described in previous Examples.

Example 19 - High throughput screening of DERA variants

A high throughput screen is desirable for any enzyme mutation programme which may generate a large number of variants. Figure 26. shows a screen designed to specifically identify the secondary alcohol generated when two molecules of acetaldehyde are combined to produce 3-hydroxybutanal by a DERA enzyme. If the aldehyde moiety is beneficially reduced to form 1,3-butanediol in the screen, the secondary alcohol is still retained and still represents DERA activity. Secondary alcohol oxidase enzymes (EC 1.1.3.18) oxidise a secondary alcohol to a ketone at the expense of oxygen and generate a molecule of hydrogen peroxide. The release of hydrogen peroxide can be captured using a diaminobenzidine assay where the substrate reacts with hydrogen peroxide to produce a red colour. This means of visualising hydrogen peroxide release is well known in the art and its application is well reported. For example, a plate and liquid screen approach using the diaminobenzidine system was used to generate improved mutants of amine oxidases (Alexeeva, M. et al. Org. Biomol. Chem. 2003, 1, 4133). These enzymes carry out a similar reaction to a secondary alcohol oxidase by oxidising an amine to a ketone generating hydrogen peroxide.

Hence, an evolution host such as *E.coli* harbouring DERA variants can be grown on either agar or a liquid medium. The *E. coli* cells can then be provided with acetaldehyde (at a desired concentration) which permeates the cell. The efficiency of a DERA mutant enzyme can be determined by the rate of hydrogen peroxide formation linked to the colorimetric visualisation assay in the presence of a selected secondary alcohol oxidase which is in excess, whereby 3-hydroxybutanal is converted to 3-ketobutanal. If 3-hydroxybutanal is beneficially reduced to 1,3-

butanediol, 1,3-butanediol is converted to 3-ketobutanol. Overexpression of the oxidase within the host cell is preferred.

5 Secondary alcohol oxidases are known to be active towards 1,3-butanediol while showing no activity towards the potential by product ethanol (produced from undesired acetaldehyde reduction in the screen), (Kawagoshi, Y. and Fujita, M. 1997. World J. Microbiol Biotechnol. 13, 273. The substrate selectivity of secondary alcohol oxidases has been thoroughly reviewed by Pickl, M. *et al.* 2015. Appl. Microbiol. Biotechnol, 99, 6617.

10

Claims

1. A modified 2-deoxyribose phosphate aldolase (DERA) enzyme variant comprising one or more mutations that improve the aldolase catalytic performance for synthesis of 3-hydroxybutanal or crotonaldehyde, relative to the parent DERA enzyme from which the variant originates and which does not comprise such a modification.
2. A DERA variant according to claim 1, comprising a polypeptide sequence which when aligned to a consensus sequence having at least 80% sequence identity with the sequence of SEQ ID NO: 1 shows an alignment with amino acid residues at 9 or more of 15 defined positions in the consensus sequence.
3. A DERA variant according to claim 2, wherein at least one of said one or more mutations are mutations of residues that are at equivalent positions to one or more of the amino acid residues at 9 or more of the 15 defined positions in the consensus sequence.
4. A DERA variant according to claim 3, wherein residues at equivalent positions to those of the consensus sequence are identified by alignment of the parent DERA polypeptide sequence with the consensus sequence.
5. A DERA variant according to claims 3 or 4, wherein the one or more mutations are at residues equivalent to positions T8, L10, C35, V57, F60, G151, S178, G179, G180, G199, A200 and/or S201 of the consensus sequence, wherein equivalent positions are determined by alignment to the consensus sequence.
6. A DERA variant according to any of claims 3-5, wherein if the one or more mutations is at a residue equivalent to position:
 - a. 8, the modification introduces a hydrophobic residue;

- b. 10, the modification introduces a positively or negatively charged residue or a hydrophobic residue;
 - c. 35, the modification introduces a hydrophobic residue;
 - d. 57, the modification introduces a hydrophobic residue;
 - 5 e. 60, the modification introduces a hydrophobic residue;
 - f. 151, the modification introduces a hydrophobic, or negatively charged residue;
 - g. 178, the modification introduces a positively or negatively charged residue or hydrophobic residue
 - 10 h. 179, the modification introduces a negatively charged, or a hydrophobic residue. 180 the modification introduces a negatively charged, or a hydrophobic residue.
 - i. 199, the modification introduces a positively or negatively charged residue or a hydrophobic residue;
 - 15 j. 200, the modification introduces a negatively charged, or a hydrophobic; and
 - k. 201, the modification introduces a negatively charged, or a hydrophobic..
7. A DERA variant according to claim 6, wherein
- 20 if the mutation is at position 8, the hydrophobic residue is selected from any of L, I, V, F, or A,
 - if the mutation is at position 10, the positively charged residue is H or the negatively charged residue is selected from D or E,
 - if the mutation is at position 35, the hydrophobic residue is selected from any of
 - 25 L, I, V, F, or A,
 - if the mutation is at position 57, the hydrophobic residue is selected from any of L, I, F, or A,
 - if the mutation is at position 60, the hydrophobic residue is selected from any of L, I, V, or A,
 - 30 if the mutation is at position 151, the hydrophobic residue is selected from any of A, , L, I, V, F, or W or the negatively charged residue is D,
 - if the mutation is at position 178, the positively charged residue is H or the negatively charged residue is selected from either of D and E,

if the mutation is at position 179, the negatively charged residue is D and a hydrophobic residue is selected from any of A, L, I, V, F, or W,

if the mutation is at position 180, the negatively charged residue is D, and the hydrophobic residue is selected from any of A, , L, I, V, F, or W,

5 if the mutation is at position 199, the positively charged residue is H and the negatively charged residue is D or E,

if the mutation is at position 200, the negatively charged residue is D and the hydrophobic residue is selected from any of L, I, V, F, or W, and/or

10 if the mutation is at position 201, a negatively charged residue is D and the hydrophobic residue is selected from any of A, L, I, V, F, or W.

8. A DERA variant according to claim 2, wherein said one or more mutations include mutations of amino acid residues that are at positions that are not equivalent to any of said 15 defined positions in the consensus sequence.

15

9. A DERA variant according to claim 8, wherein none of said mutations are mutations of amino acid residues that are at positions that are equivalent to any of said 15 defined positions in the consensus sequence.

20 10. A DERA variant according to any preceding claim, wherein said one or more mutations improve coordination of a substrate carbonyl group.

11. A DERA variant according to any preceding claim, wherein said one or more mutations improve coordination of a substrate methyl group.

25

12. A DERA variant according to any preceding claim, wherein said one or more mutations reduce the coordination of a substrate phosphate group.

30 13. A DERA variant according to any preceding claim, wherein said one or more mutations increases the negative charge in the active site.

14. A DERA variant according to any preceding claim, wherein said one or more mutations increases the hydrophobicity in the active site.

5 15. A modified 2-deoxyribose phosphate aldolase (DERA) enzyme variant comprising mutations that increase the ability of the enzyme to perform dehydration of 3-hydroxybutanal to form crotonaldehyde, relative to the activity of the parent DERA enzyme from which the variant originates and which does not comprise such a modifications,
10 wherein said DERA enzyme comprises a polypeptide sequence which when aligned to a consensus sequence having at least 80% sequence identity with the sequence of SEQ ID NO: 1 shows an alignment with amino acid residues at 9 or more of 15 defined positions in the consensus sequence, and
wherein said mutations are substitutions of residues at a position equivalent to positions 10, 178 and 199 in the consensus sequence, wherein at positions 10,
15 178 and 199 there is a H or a negatively charged residue selected from D or E.

16. An isolated polypeptide comprising a polypeptide sequence which when aligned to a consensus sequence having at least 80% sequence identity with the sequence of SEQ ID NO: 1 shows an alignment with amino acid residues at 9 or
20 more of 15 defined positions in the consensus sequence.

17. An isolated polypeptide according to claim 16, wherein said 15 defined positions in the consensus sequence are positions 8, 10, 35, 57, 60, 86, 147, 151, 176, 178, 179 180, 199, 200 and 201.

25

18. An isolated polypeptide according to claim 17, wherein said polypeptide sequence comprises one or more of the following residues at positions that are equivalent to one or more of positions 8, 10, 35, 57, 60, 151, 178, 179 180, 199, 200 and/or 201 of the consensus sequence, wherein equivalent positions are
30 determined by alignment to the consensus sequence:

at position 8, there is a hydrophobic residue selected from any of L, I, V, F, or A,
at position 10, there is a H or a negatively charged residue selected from D or E,

- at position 35, there is a hydrophobic residue selected from any of L, I, V, F, or A,
at position 57, there is a hydrophobic residue selected from any of L, I, F, or A,
at position 60, there is a hydrophobic residue selected from any of L, I, V, or A,
5 at position 151, there is a hydrophobic residue selected from any of A, L, I, V, F, or W, or the negatively charged residue is D,
at position 178, there is a H or there is a negatively charged residue selected from either of D and E,
at position 179, there is a negatively charged residue is D and a hydrophobic
10 residue selected from any of A, L, I, V, F, or W, ,
at position 180, there is a negatively charged residue, D and a hydrophobic residue selected from any of A, L, I, V, F, or W, ,
at position 199, there is a H, or a negatively charged residue selected from D or E,
15 at position 200, there is a negatively charged residue, D and a hydrophobic residue selected from any of L, I, V, F, or W, and/or
at position 201, there is a negatively charged residue, D and a hydrophobic residue selected from any of A, L, I, V, F, or W.
- 20 19. Use of an isolated polypeptide according to claims 16-18 to improve the aldolase catalytic performance for synthesis of 3-hydroxybutanal or crotonaldehyde of a 2-deoxyribose phosphate aldolase (DERA) enzyme.
- 25 20. An isolated polynucleotide sequence encoding a polypeptide sequences according to claims 16-18.
21. An expression system comprising the isolated polynucleotide sequence of claim 20 operably linked to suitable control sequences.
- 30 22. A recombinant microorganism transformed with the expression system of claim 21.

23. A non-naturally occurring polypeptide comprising an amino acid sequence having at least 80% identity to the amino acid sequence of SEQ ID NO. 1.

5 24. A non-naturally occurring polypeptide according to claim 23, comprising the amino acid sequence of SEQ ID NO. 1.

25. A non-naturally occurring polypeptide according to claim 24 consisting of the amino acid sequence of SEQ ID NO. 1.

10

26. Use of a non-naturally occurring polypeptide according to any of claims 23-25 to identifying regions within a DERA enzyme that can be mutated to influence the catalytic activity of the DERA enzyme, wherein said catalytic activity is the coupling or condensation of two molecules of acetaldehyde to form 3-hydroxybutanal and/or crotonaldehyde.

15

27. A method for identifying one or more residues within a DERA polypeptide sequence that can influence catalytic activity of the DERA enzyme, comprising:

20 a. aligning the DERA polypeptide sequence to a consensus sequence having at least 80% homology to SEQ ID No 1; and

b. identifying residues within the DERA sequence that are in alignment with residues of the consensus sequence,
wherein residues of the DERA sequence that are in alignment with residues within the consensus sequence are identified as residues that can influence catalytic activity.

25

28. A method according to claim 27, wherein said catalytic activity is:

a. coupling of two aldehyde molecules to produce 3-hydroxybutanal; and/ or
b. dehydration of 3-hydroxybutanal to crotonaldehyde.

30

29. A method according to claims 27 or 28, wherein step (b) of claim 27 comprises identifying residues within the DERA sequence that are in alignment with residues at 9 or more of the following 15 positions within the consensus

sequence: 8, 10, 35, 57, 60, 86, 147, 151, 176, 178, 179 180, 199, 200 and/or 201.

5 30. A method according to claims 27-29, wherein if the DERA sequence comprises at least 9 residues at positions equivalent to 8, 10, 35, 57, 60, 151, 178, 179, 180, 199, 200 and/or 201 of the consensus sequence, the DERA sequence can be modified to improve the catalytic activity of the enzyme, relative to the parent DERA enzyme.

10 31. A method of identifying regions within a DERA enzyme that can be mutated to influence the coupling or condensation of two molecules of acetaldehyde to form 3-hydroxybutanal and/or crotonaldehyde, comprising aligning the polypeptide sequence of the DERA enzyme with a consensus sequence having at least 80% identity to the amino acid sequence of SEQ ID
15 NO. 1, and identifying regions in the DERA enzyme sequence that align with known regions within the consensus sequence.

20 32. A method according to claim 31, wherein the known regions within the consensus sequence comprise are one or more of the following residues: T8, L10, C35, V57, F60, G151, S178, G179 G180, G199, A200 and/or S201.

33. A method of increasing the catalytic activity of a DERA enzyme, comprising:
(i) aligning the polypeptide sequence of the DERA enzyme with a consensus
25 sequence having at least 80% identity to the amino acid sequence of SEQ ID NO. 1,
(ii) identifying regions in the DERA sequence that align with regions within the consensus sequence, and
(iii) mutating the DNA codons encoding one more or amino acid residues within
30 said regions in order to increase the catalytic activity of the enzyme.

34. A method according to claim 33, wherein said regions are single amino acids or groups of amino acids that can be mutated to influence the coupling or

condensation of two molecules of acetaldehyde to form either 3-hydroxybutanal or crotonaldehyde.

5 35. A method according to claims 33 or 34, wherein said mutation of one more or amino acid residues results in increased synthesis of 3-hydroxybutanal and/or synthesis of crotonaldehyde.

10 36. Use of a polypeptide having at least 80% sequence identity with the sequence of SEQ ID NO: 1 in a method for identifying regions a DERA enzyme that can be mutated to influence the coupling or condensation of two molecules of acetaldehyde to form 3-hydroxybutanal and/or crotonaldehyde according to any of claims 27-32 or in a method for increasing the catalytic activity of a DERA enzyme according to any of claims 33-35.

15 37. A method of increasing the catalytic activity of a DERA enzyme, comprising:
(i) aligning the polypeptide sequence of the DERA enzyme with a consensus sequence having at least 80% identity to the amino acid sequence of SEQ ID NO. 1,
20 (ii) identifying regions in the DERA sequence that align with regions within the consensus sequence, and
(iii) mutating the DNA codons encoding one more or amino acid residues within said regions in order to increase the catalytic activity of the enzyme.

25 38. A method of increasing the catalytic activity of a DERA enzyme, comprising:
(i) aligning the polypeptide sequence of the DERA enzyme with a consensus sequence having at least 80% identity to the amino acid sequence of SEQ ID NO. 1,
30 (ii) identifying residues in the DERA sequence that align with positions 10, 178 and 199 within the consensus sequence, and
(iii) mutating the DNA codons encoding these three residues in order to increase the catalytic activity of the enzyme,

wherein said mutations are substitutions of residues at positions equivalent to positions 10, 178 and 199 in the consensus sequence, wherein at positions 10, 178 and 199 there is a H or a negatively charged residue selected from D or E, and wherein said catalytic activity is the dehydration of 3-hydroxybutanal to form crotonaldehyde.

39. A non-naturally occurring microbial organism which includes a DERA variant as defined in claims 1 to 14 and a genetic modification in its genome which enhances production of 3-hydroxybutanal or crotonaldehyde or a downstream product of 3-hydroxybutanal by the microbial organism from at least one endogenous central metabolic intermediate via a 3-hydroxybutanal or crotonaldehyde synthetic pathway in which two molecules of acetaldehyde are either coupled or condensed to form said 3-hydroxybutanal using the DERA variant capable of aldolase activity accepting acetaldehyde as both the acceptor and donor in an aldol coupling or condensation.

40. A non-naturally occurring microbial organism as claimed in claim 39 wherein the genetic modification also increases the production of the acetaldehyde from the at least one endogenous central metabolic intermediate or increases its availability to the DERA variant.

41. A non-naturally occurring microbial organism as claimed in claim 40 wherein the genetic modification:

- (i) introduces a heterologous gene encoding an enzyme having an activity utilised in generation of acetaldehyde from one or more of the central metabolic intermediates;
- (ii) up-regulates at least one endogenous enzyme having an activity utilised in generation of acetaldehyde from one or more of the central metabolic intermediates; and/or
- (iii) down-regulates or inactivates an endogenous enzyme which utilises acetaldehyde as a substrate,

thereby increasing production or availability to the aldolase of the acetaldehyde, thereby increasing production of the 3-hydroxybutanal or crotonaldehyde from the DERA variant.

5 42. A microbial organism as claimed in any one of claims 39 to 41 wherein the genetic modification confers on the microorganism the capability to produce an increased amount of 3-hydroxybutanal or crotonaldehyde or a downstream product of thereof, wherein the downstream product is obtained directly or indirectly from reduction, oxidation and/or acylation with Coenzyme A of 3-
10 hydroxybutanal or crotonaldehyde.

43. A microbial organism as claimed in claim 42 wherein the downstream product is:
(i) selected from 1,3-BDO; 2-hydroxisobutyrate, Crotyl alcohol, Crotonic acid,
15 Butanol, Butyrate, 3-hydroxybutyrate, 3-hydroxybutylamine, Polyhydroxybutyrate, Acetone, Isopropanol, 2-methylsuccinic acid, and/or
(ii) is obtained via an intermediate selected from: 3-hydroxybutyryl CoA, 2-hydroxyisobutyryl CoA, Crotonyl CoA, , Butyryl CoA, Butanal, Acetoacetyl CoA, and acetoacetate.

20 44. A microbial organism as claimed in any one of claims 39 to 43 wherein the microorganism lacks the ability to produce the downstream product in the absence of said modification.

45. A microbial organism as claimed in any one of claims 39 to 44 wherein
25 the DERA variant is
 encoded by at least one heterologous gene

46. A microbial organism as claimed in claim 45 wherein the heterologous gene encodes the DERA variant as a fusion protein encoding also one or more
30 other enzymes involved in the 3-hydroxybutanal or crotonaldehyde pathway.

47. A microbial organism as claimed in claim 46 wherein said one or more other enzymes involved in the 3-hydroxybutanal or crotonaldehyde pathway

provide an activity enhancing the provision of the DERA variant substrate acetaldehyde or provide an activity converting the 3-hydroxybutanal or crotonaldehyde to the downstream product.

5 48. A microbial organism as claimed in any one of claims 39 to 47 wherein the DERA variant is a variant of EC 4.1.2.4 enzyme.

10 49. A microbial organism as claimed in any one of claims 39 to 48 wherein the DERA variant is a variant of an enzyme selected from Table 6 capable of the coupling and/ or condensation of two molecules of acetaldehyde to form 3-hydroxybutanal or crotonaldehyde.

15 50. A microbial organism as claimed in any one of claims 39 to 49 wherein the DERA variant is a variant of deoxyribose phosphate aldolase (DERA).

20 51. A microbial organism as claimed in any one of claims 39 to 50 wherein the DERA from which the variant is derived is optionally obtained from: *E.coli*, *Geobacillus thermodenitrificans*, *Acetobacterium woodii*, *Pyrobaculum* spp., which is optionally *P. aerophilum*.

25 52. A microbial organism as claimed in any one of claims 39 to 51 wherein the downstream product is 1,3-BDO or crotyl alcohol and the genetic modification causes down-regulation or inactivation of an endogenous alcohol dehydrogenase with a preference for reduction of acetaldehyde to ethanol relative to reduction of 3-hydroxybutanal or crotonaldehyde to 1,3-BDO or crotyl alcohol, respectively..

30 53. A microbial organism as claimed in any one of claims 39 to 52 wherein the central metabolic intermediate is selected from one or both of: acetyl CoA or pyruvate

54. A microbial organism as claimed in any one of claims 39 to 53 wherein central metabolic intermediate is converted to acetaldehyde via 1, 2, or 3 enzymatic steps.

5 55. A microbial organism as claimed in any one of claims 39 to 54 wherein the genetic modification comprises the introduction of 1, 2, 3, 4, 5, 6, 7, 8, or 9 heterologous genes each encoding a 3-hydroxybutanal or crotonaldehyde pathway enzyme.

10 56. A microbial organism as claimed in any one of claims 39 to 55 wherein the genetic modification comprises the introduction of a heterologous gene encoding one or more of the following 3-hydroxybutanal or crotonaldehyde pathway enzymes:

15 (i) an enzyme having an activity utilised in generation of the central metabolic intermediates from feedstock;

(ii) an enzyme having an activity utilised in generation of acetaldehyde from the central metabolic intermediates;

(iii) an enzyme having an activity utilised in generation of the downstream product of 3-hydroxy butanal or crotonaldehyde.

20

57. A microbial organism as claimed in any one of claims 39 to 56 wherein acetyl CoA and pyruvate are generated in the organism by one or more of the following metabolic pathways encoded in the organism genome: the Wood-Ljungdahl pathway; the ribulose monophosphate (RuMP) pathway; the reverse
25 TCA cycle; the serine cycle; glycolysis; the pentose phosphate pathway, the Calvin cycle, the 3-hydroxypropionate cycle; the dicarboxylate cycle/4-hydroxybutyrate cycle.

58. A microbial organism as claimed in claim 57 which is an acetogen, more
30 preferably a carboxydophilic acetogen.

59. A microbial organism as claimed in claim 58 which has the Wood-Ljungdahl pathway naturally encoded in its genome, wherein the genetic

modification confers the ability to produce 3-hydroxybutanal or crotonaldehyde or the downstream product thereof or the ability to produce an increased flux of 3-hydroxybutanal or crotonaldehyde or the downstream product thereof from a feedstock selected from: syngas, CO₂, CO, and H₂, methanol, sugar or combinations thereof

60. A microbial organism as claimed in any one of claims 57 to 59 which encodes in its genome one or more methyltransferase enzymes utilised in generation of the central metabolic intermediates from methanol feedstock, wherein said methyltransferase enzyme or enzymes are optionally heterologous to the microbial organism.

61. A microbial organism as claimed in claim 60 wherein the methyltransferase enzyme or enzymes are selected from the list consisting of: methanol methyltransferase (MtaB); Corrinoid protein (MtaC); Methyltetrahydrofolate:corrinoid protein methyltransferase (MtaA); Methyltetrahydrofolate:corrinoid protein methyltransferase (AcsE); Corrinoid iron-sulfur protein (AcsD)

62. A microbial organism as claimed in any one of claims 39 to 57 which is a methylotroph or methanotroph.

63. A microbial organism as claimed in claim 62 which has a RuMP or serine cycle pathway naturally encoded in its genome, wherein the genetic modification confers the ability to produce 3-hydroxybutanal or crotonaldehyde or the downstream products thereof or the ability to produce an increased flux of 3-hydroxybutanal or crotonaldehyde or the downstream products thereof from a feedstock selected from: methanol or methane or both.

64. A microbial organism as claimed in any one of claims 39 to 57 which encodes in its a genome enzymes comprising the Calvin cycle utilised in generation of the central metabolic intermediates.

65. A microbial organism as claimed in claim 59 which has the Wood-Ljungdahl pathway naturally encoded in its genome and wherein the genetic modification causes down-regulation or inactivation of an endogenous enzyme converting acetyl CoA to acetate.

5

66. A microbial organism as claimed in claim 65 wherein the endogenous enzyme activity converting acetyl CoA to acetate is an endogenous phosphotransacetylase or an endogenous acetate kinase.

10 67. A microbial organism as claimed in any one of claims 39 to 64 wherein the genetic modification comprises the introduction of a heterologous gene or up-regulates at least one endogenous gene which encodes an enzyme activity which confers the ability to produce 3-hydroxybutanal or a downstream product thereof or the ability to produce an increased flux of 3-hydroxybutanal or
15 crotonaldehyde or a downstream product thereof from acetate, such that the 3-hydroxybutanal or crotonaldehyde or a downstream product thereof accumulates and can be recovered or further converted enzymatically or chemically.

20 68. A microbial organism as claimed in claim 65 wherein the enzyme activity is utilised in generation of acetaldehyde from acetate via acetyl CoA.

69. A microbial organism as claimed in any one of claims 65 to 68 wherein the enzyme activity is used in the generation of acetyl CoA from acetate and is
25 an acetyl CoA synthetase or a CoA transferase optionally selected from EC 6.2.1.1 or EC 2.8.3.8.

70. A microbial organism as claimed in any one of claims 65 to 69 wherein the acetyl CoA synthesising activity is selected from an enzyme of Table 5 or a
30 variant of an enzyme of Table 5 capable of generation of acetyl CoA from acetate.

71. A microbial organism as claimed in claim 67 wherein the enzyme activity is a carboxylic acid reductase activity

5 72. A microbial organism as claimed in claim 67 or claim 71 wherein the carboxylic acid reductase activity is selected from an enzyme of Table 1 or a variant of an enzyme of Table 1 capable of the reduction of acetate to acetaldehyde.

10 73. A microbial organism as claimed in any one of claims 39 to 73 wherein the genetic modification comprises the introduction of a heterologous gene or up-regulates at least one endogenous gene which encodes an enzyme activity which confers the ability to produce 3-hydroxybutanal or crotonaldehyde or the downstream products thereof or the ability to produce an increased flux of 3-hydroxybutanal or crotonaldehyde or the downstream products thereof from the
15 central metabolic intermediate acetyl-coA, such that the 3-hydroxybutanal or crotonaldehyde or the downstream products thereof accumulate and can be recovered, or can be further converted enzymatically or chemically.

20 74. A microbial organism as claimed in claim 73 wherein the enzyme activity is utilised in generation of acetaldehyde from acetyl CoA.

75. A microbial organism as claimed in claim 74 wherein the enzyme activity is an aldehyde dehydrogenase, optionally an acetaldehyde dehydrogenase.

25 76. A microbial organism as claimed in claim 74 or claim 75 wherein the enzyme activity is selected from: EC 1.2.7.1; EC 4.1.1.1; EC 1.2.1.10.

30 77. A microbial organism as claimed in any one of claims 74 to 76 wherein the enzyme activity is selected from an enzyme of Table 2 or a variant of an enzyme of Table 2 capable of the conversion of acetyl CoA to acetaldehyde.

78. A microbial organism as claimed in claim 74 wherein the enzyme activity is selected from an enzyme of Table 3 or a variant of an enzyme of Table 3 capable of the conversion of acetyl CoA to pyruvate.

5 79. A microbial organism as claimed in any one of claims 39 to 58 or claim 78 wherein the genetic modification comprises the introduction of a heterologous gene or up-regulates at least one endogenous gene which encodes an enzyme activity which confers the ability to produce 3-hydroxybutanal or crotonaldehyde or the downstream products thereof or the ability to produce an increased flux of
10 3-hydroxybutanal or crotonaldehyde or the downstream products thereof from the central metabolic intermediate pyruvate, such that the 3-hydroxybutanal or crotonaldehyde or the downstream products thereof accumulate and can be recovered, or further converted enzymatically or chemically.

15 80. A microbial organism as claimed in claim 79 wherein the enzyme activity is utilised in generation of acetaldehyde from pyruvate.

81. A microbial organism as claimed in claim 79 or claim 80 wherein the enzyme activity is a pyruvate decarboxylase.

20 82. A microbial organism as claimed in claim 80 or claim 81 wherein the enzyme activity is selected from: EC 1.2.7.1; EC 1.2.1.51; EC 1.2.4.1; EC 1.2.1.10; EC 4.1.1.1.

83. A microbial organism as claimed in any one of claims 80 to 82 wherein
25 the enzyme activity is selected from an enzyme of Table 3 or a variant of an enzyme of Table 3 capable of the conversion of pyruvate to acetyl CoA.

84. A microbial organism as claimed in any one of claims 80 to 82 wherein
30 the enzyme activity is selected from an enzyme of Table 4 or a variant of an enzyme of Table 4 capable of the conversion of pyruvate to acetaldehyde.

85. A microbial organism as claimed in any one of claims 79 to 84 wherein the genetic modification causes down-regulation or inactivation of an endogenous enzyme capable of converting pyruvate to lactate or other products.
- 5 86. A microbial organism as claimed in claim 73 wherein the endogenous enzyme capable of converting pyruvate to lactate or other products is an LDH or malate dehydrogenase or pyruvate formate lyase, optionally selected from 1.1.1.27 or 1.1.1.37 or 2.3.1.54.
- 10 87. A microbial organism as claimed in any one of claims 39 to 86 wherein the genetic modification comprises the introduction of a heterologous gene or up-regulates at least one endogenous gene which encodes an enzyme activity which confers the ability to convert 3-hydroxybutanal or crotonaldehyde to 1,3-BDO or crotyl alcohol, respectively.
- 15 88. A microbial organism as claimed in claim 87 wherein the enzyme has alcohol dehydrogenase activity or has aldehyde reductase activity optionally classified as EC 1.1.1.-
- 20 89. A microbial organism as claimed in claim 88 wherein the enzyme activity is a medium chain alcohol dehydrogenase or an aldehyde reductase which shows a preference for 3-hydroxybutanal or crotonaldehyde relative to acetaldehyde.
- 25 90. A microbial organism as claimed in any one of claims 87 to 89 wherein the enzyme activity is selected from an enzyme of Table 7 or a variant of an enzyme of Table 7 capable of the conversion of 3-hydroxybutanal or crotonaldehyde to 1,3-butanediol or crotyl alcohol, respectively.
- 30 91. A microbial organism as claimed in claim 90, wherein the enzyme is selected from GOX 1615, GRE2, BbhB, or variants thereof, preferably wherein the enzyme is GOX 1615 or a variant thereof.

92. A process method for producing 3-hydroxybutanal or crotonaldehyde or downstream products thereof, which process comprises culturing a microbial organism as claimed in any one of claims 39 to 91 for a sufficient period of time to produce said 3-hydroxybutanal or crotonaldehyde or the downstream products thereof.

93. A process as claimed in claim 92, wherein the microbial organism is an acetogen as defined in any claim 58 or 59 or 65 or 66, and the microbial organism is cultured on a feedstock selected from: a sugar, methanol, CO₂, H₂, CO, syngas, or a mixture thereof.

94. A process as claimed in claim 92, wherein the microbial organism is an acetogen as defined in claim 60 or claim 61, and the microbial organism is cultured on a feedstock selected from: methanol and a more oxidised co-substrate which is optionally CO₂.

95. A process as claimed in claim 92, wherein the microbial organism is a methylotroph or methanotroph as defined in claim 62 or claim 63, and the microbial organism is cultured on a feedstock selected from: methanol; methane.

96. A process as claimed in claim 92, wherein the microbial organism is as defined in claim 64, and the microbial organism is an Alga or photosynthetic bacterium which is cultured in sunlight.

97. A process as claimed in any one of claims 92 to 96, further comprising recovering the 3-hydroxybutanal or crotonaldehyde or downstream products thereof.

98. A process as claimed in any one of claims 92 to 97 further comprising converting the 3-hydroxybutanal or crotonaldehyde or downstream products thereof to a further chemical compound, which is optionally selected from 1,3-butadiene and methylethyl ketone.

99. A process for producing a microbial organism according to any one of claims 39 to 91, which comprises introducing said genetic modification into a parent strain.

5

100. A process as claimed in claim 99 wherein the parent strain lacks the capability to produce 3-hydroxybutanal or crotonaldehyde or the downstream products thereof.

101. A process as claimed in claim 99 or claim 100 for increasing the flux of 3-hydroxybutanal or crotonaldehyde or the downstream products thereof.

102. A process as claimed in any one of claims 99 to 101 wherein the genetic modification comprises either:

- 15 (i) introduction at least one heterologous gene encoding an enzyme, or
(ii) up-regulation of at least one endogenous gene encoding an enzyme.

103. A non-naturally occurring microbial organism obtained or obtainable by the process of any one of claims 99 to 102.

20

104. A downstream product obtained or obtainable by the process of any one of claims 92 to 98.

105. Use of an alcohol oxidase for detecting DERA variants with improved aldolase catalytic performance for the synthesis of 3-hydroxybutanal or downstream products thereof relative to the parent DERA enzyme from which the variant originates and which does not comprise such a modification.

106. An assay for screening for DERA variants having improved aldolase catalytic performance for the synthesis of 3-hydroxybutanal or downstream products thereof, comprising determining an increase in the rate of H₂O₂ formation from the oxidation of 3-hydroxybutanal or downstream products thereof in the presence of an alcohol oxidase.

30

107. An assay according to claim 106, wherein the rate of H₂O₂ formation is compared to a control value.

5 108. An assay according to claim 107, wherein the downstream product is 1,3-butanediol formed by the selective reduction of 3-hydroxybutanal.

109. An assay according to claim 108, wherein selective reduction is performed by introducing a heterologous enzyme with alcohol dehydrogenase or aldehyde reductase activity which shows a preference for reduction of 3-hydroxybutanal over acetaldehyde.

110. An assay according to claim 109, wherein the heterologous enzyme is selected from Table 7, or is a variant of an enzyme from Table 7, preferably wherein the heterologous enzyme is GOX 1615, GRE2, or BdhB, or a variant thereof.

111. An assay according to any of claims 106 to 110, wherein the assay is performed in a microbial host, preferably wherein the host is *E.coli*.

112. An assay according to any of claims 106 to 110, wherein the assay is performed in a cell lysate.

113. An assay according to any of claims 106 to 110, wherein the assay is performed *in vitro* using recombinant proteins.

114. An assay according to any of claims 106 to 113, wherein the formation of H₂O₂ is determined using a colorimetric assay or fluorimetric assay, preferably a diaminobenzidine assay.

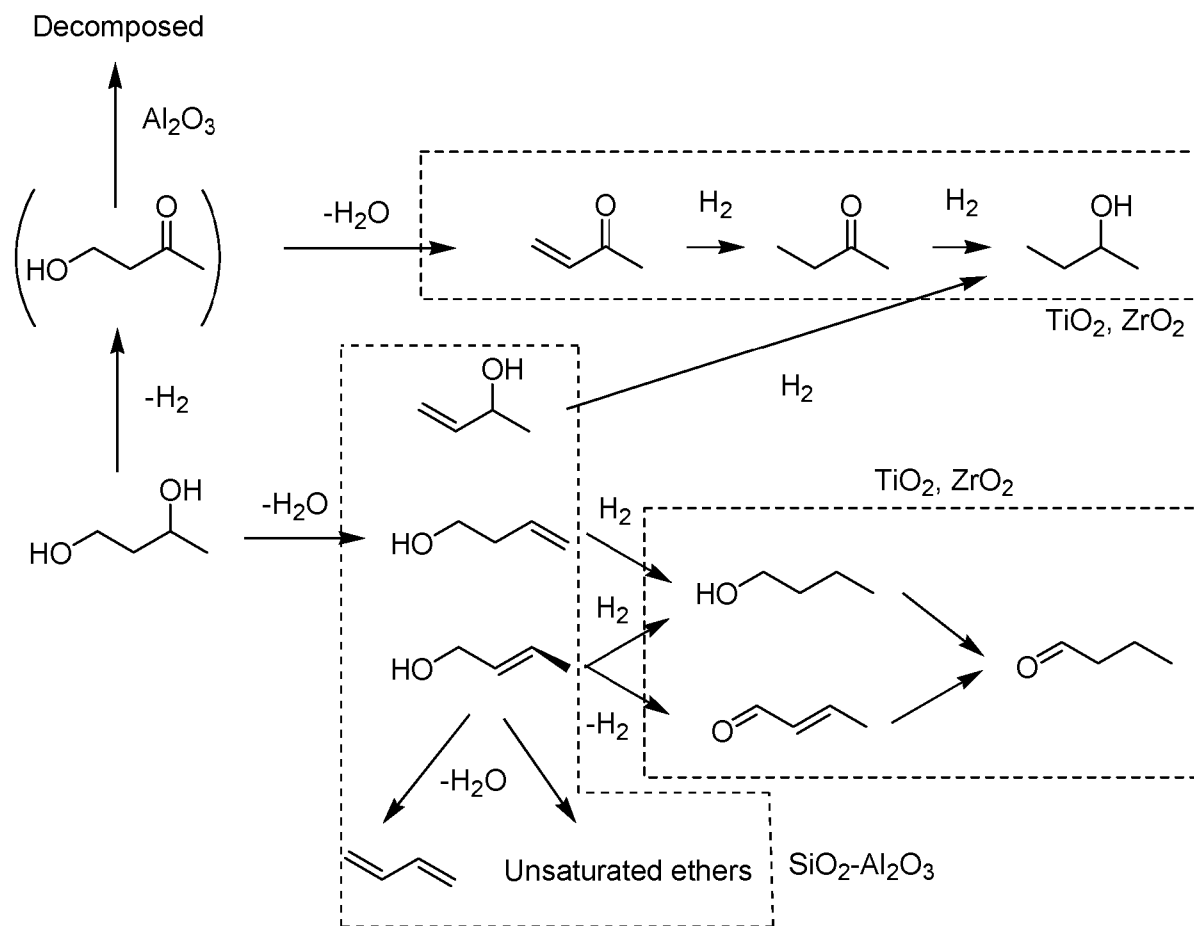


Figure 1

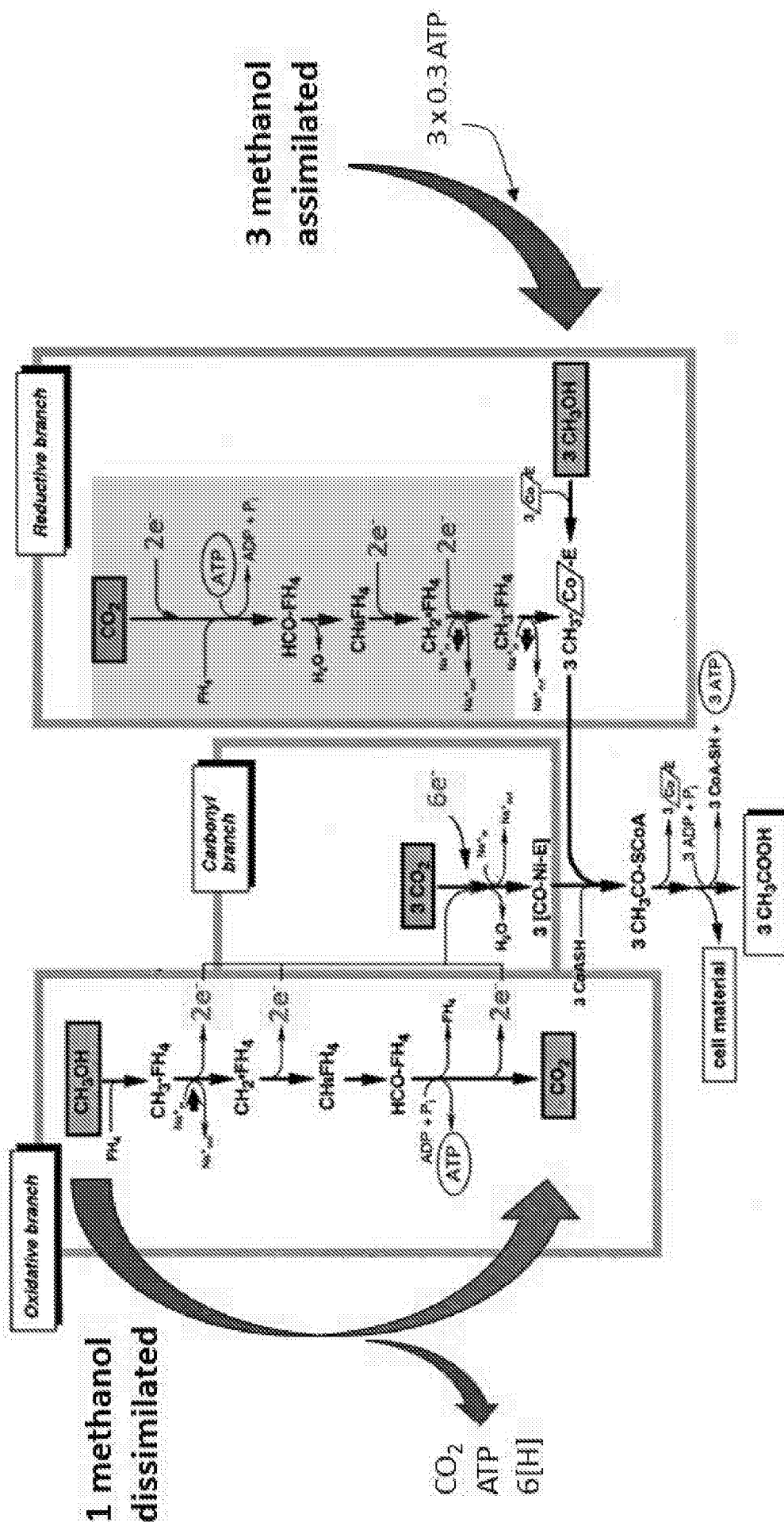


Figure 2

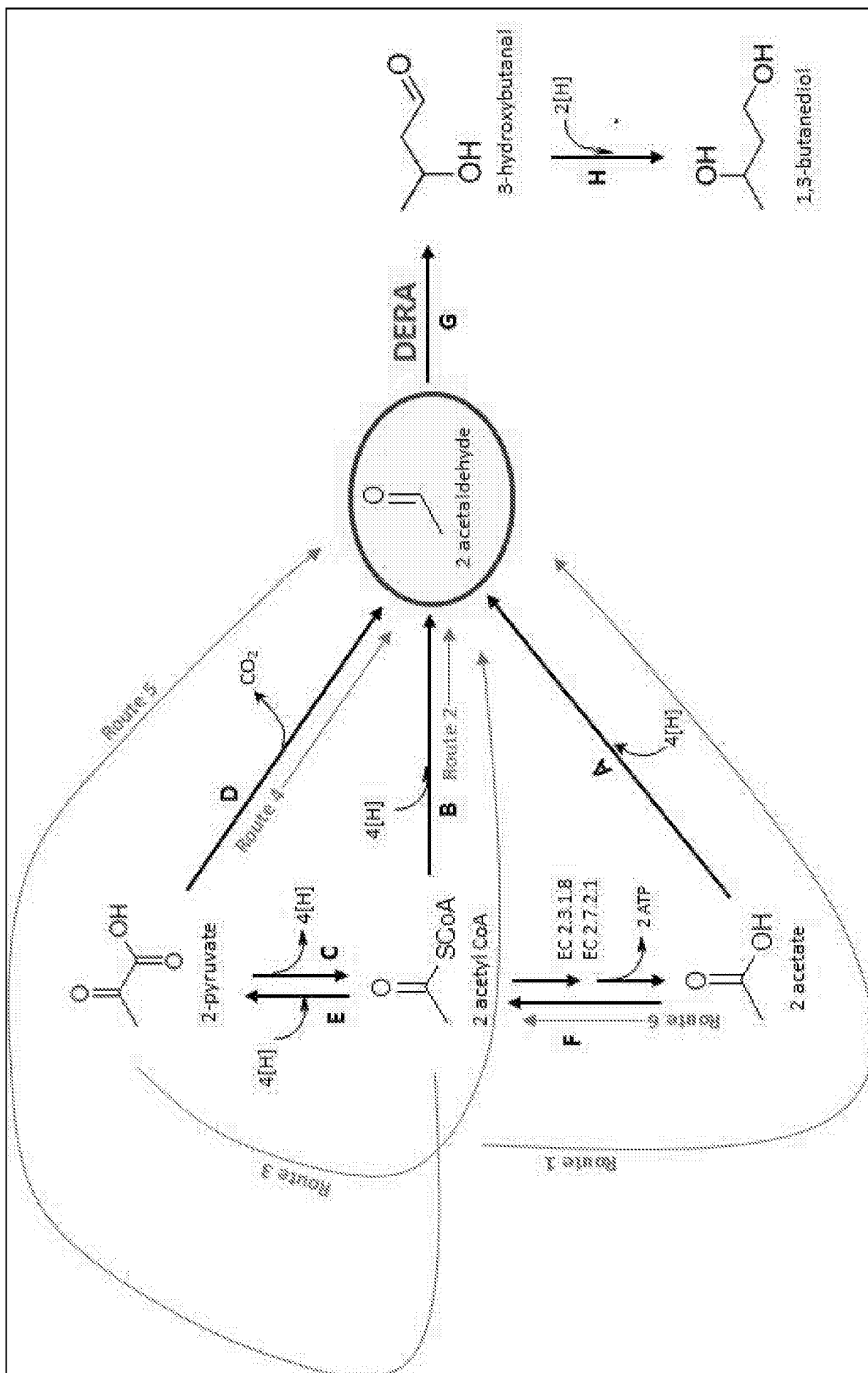


Figure 3

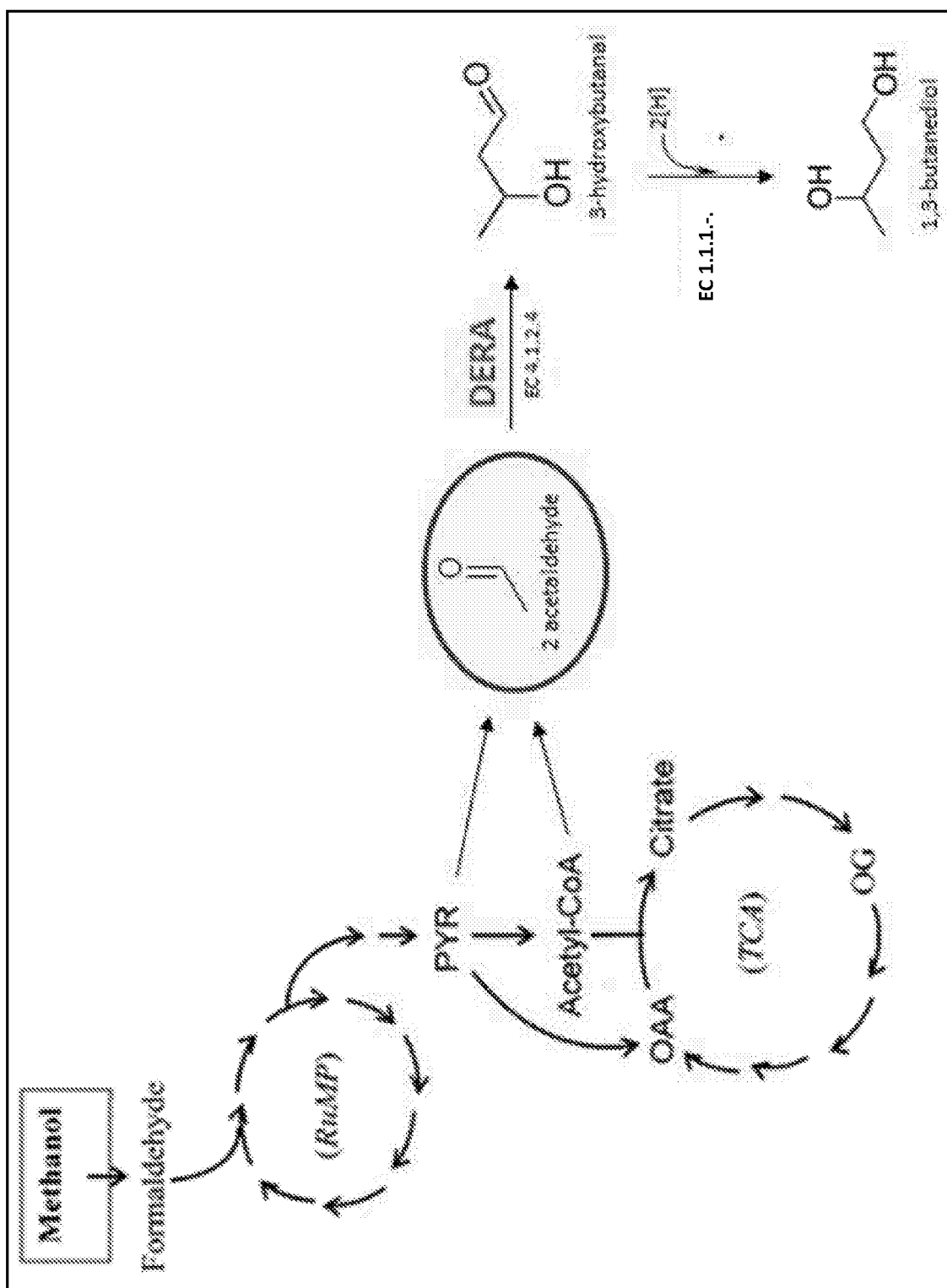


Figure 4

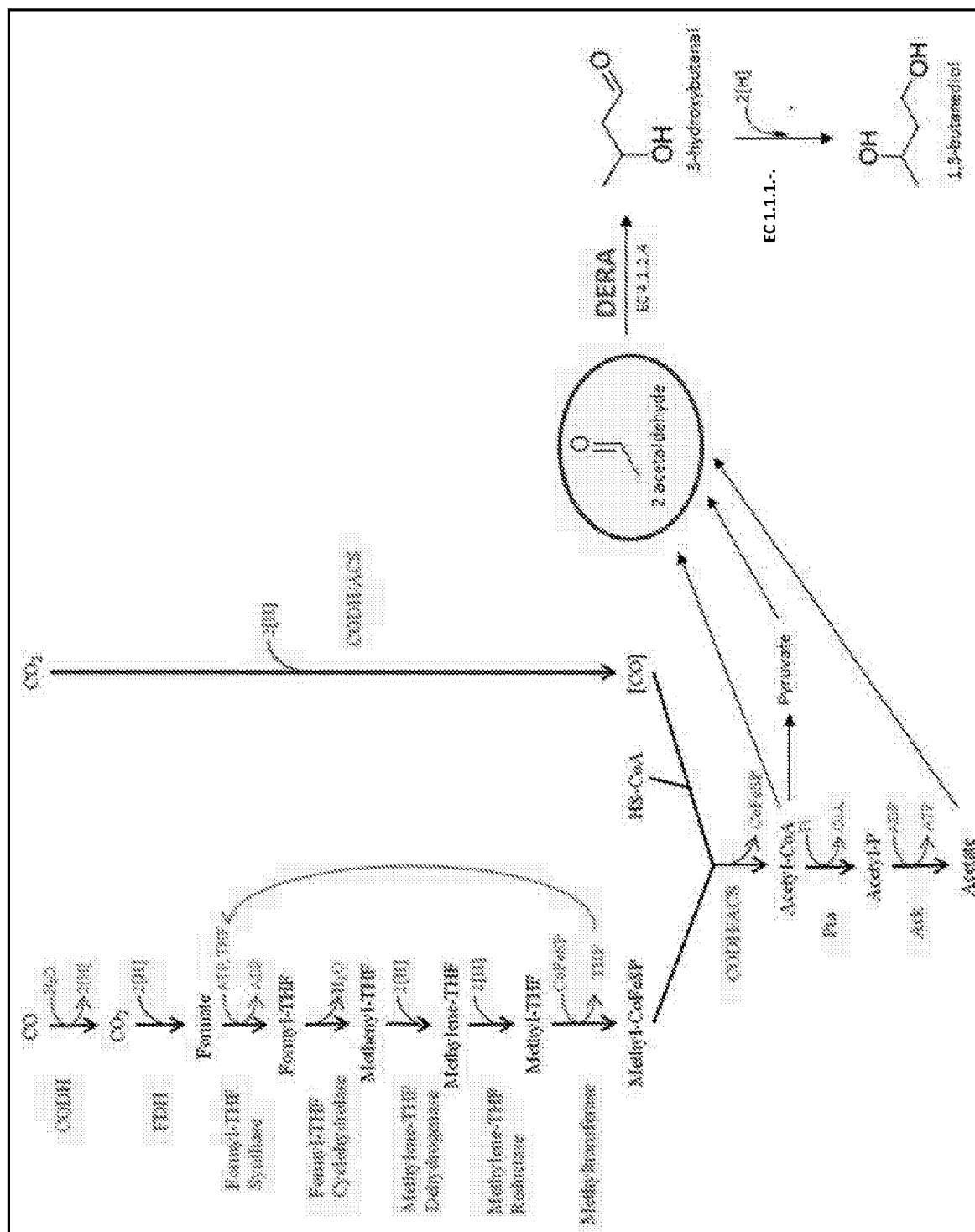


Figure 5

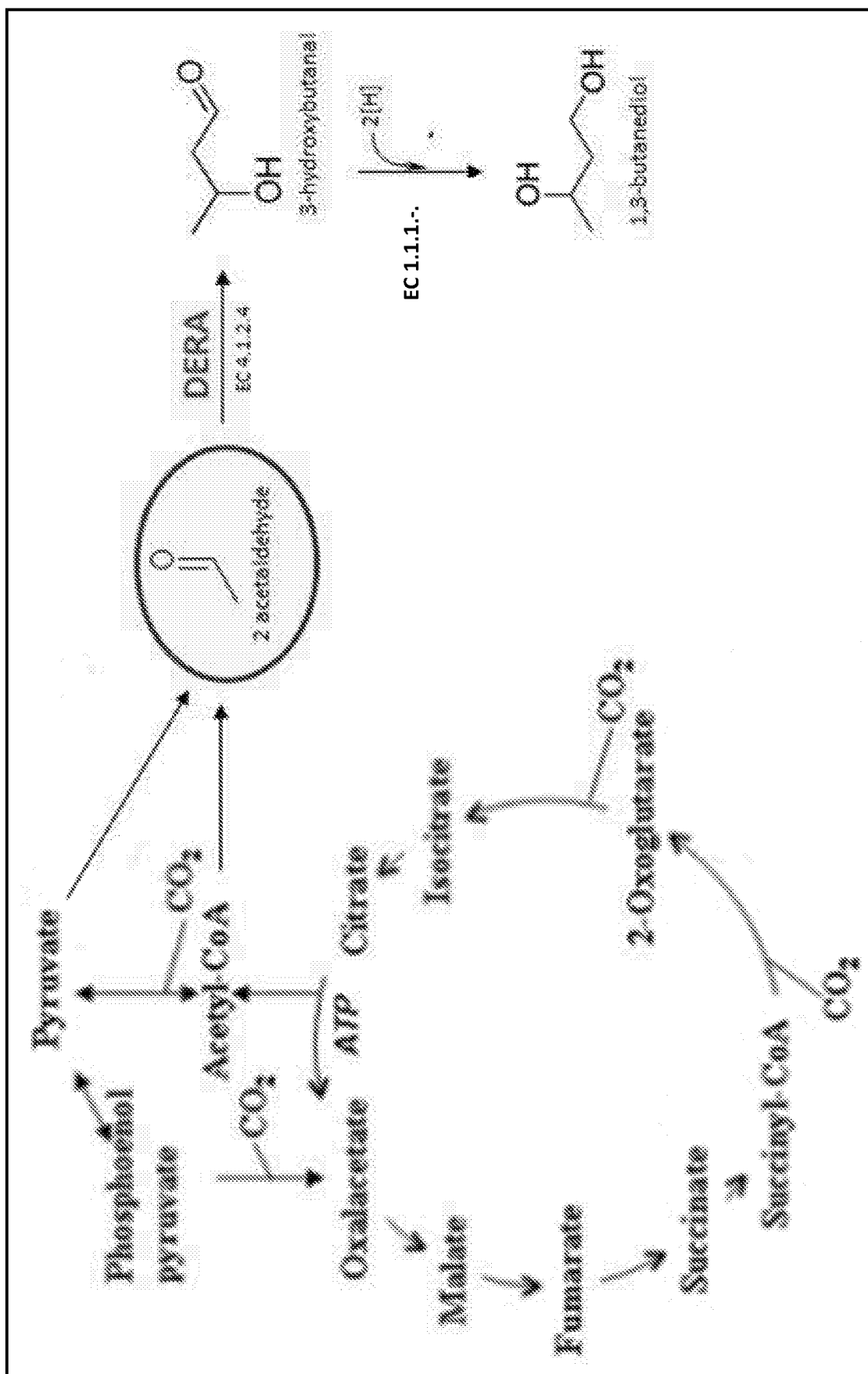
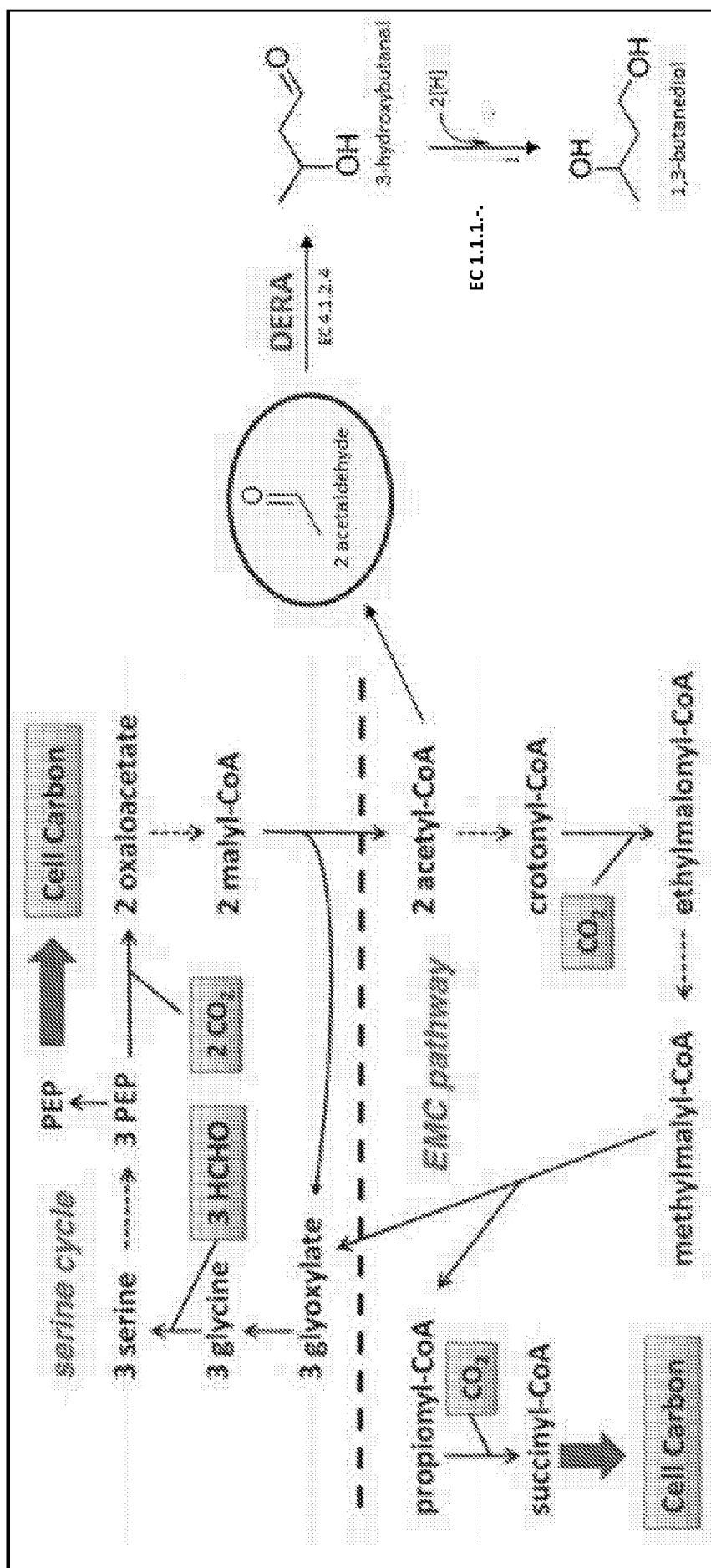


Figure 6



2019

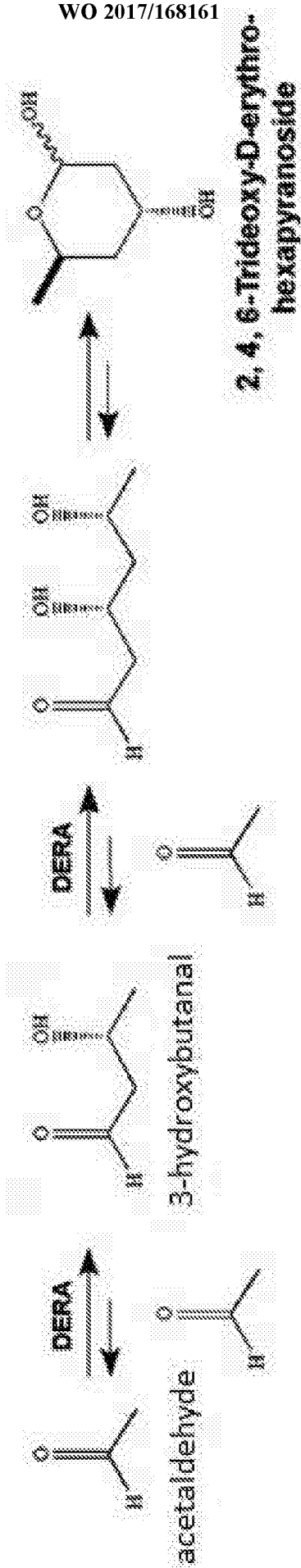


Figure 8

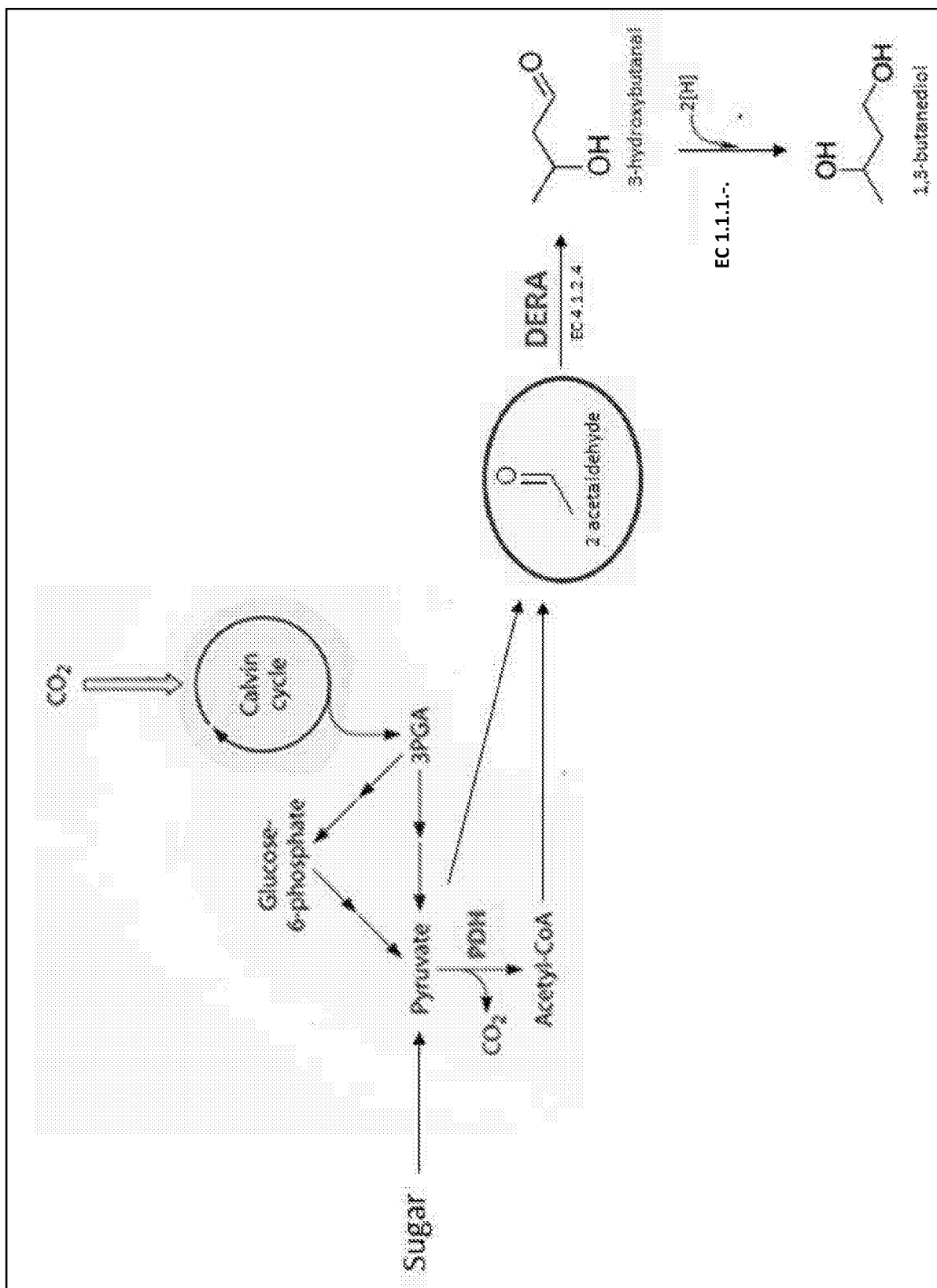


Figure 9

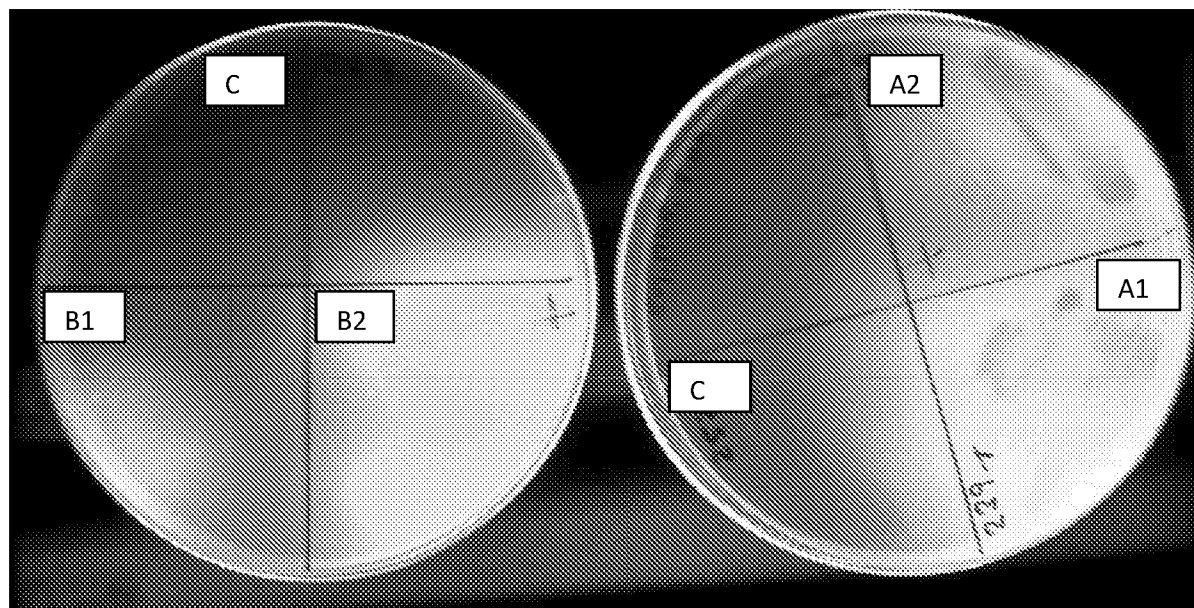


Figure 10

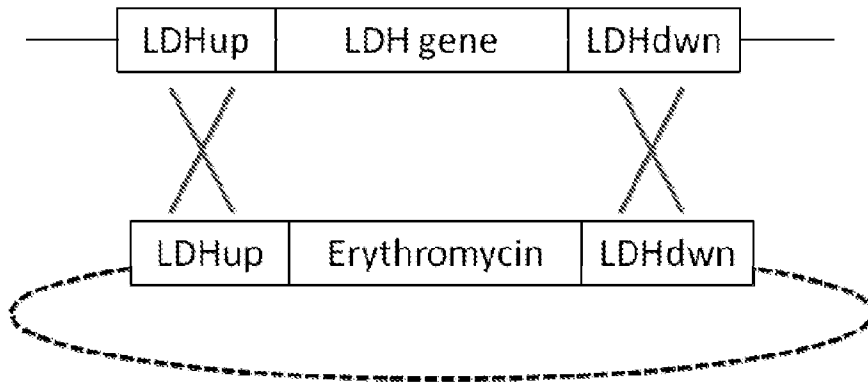


Figure 11

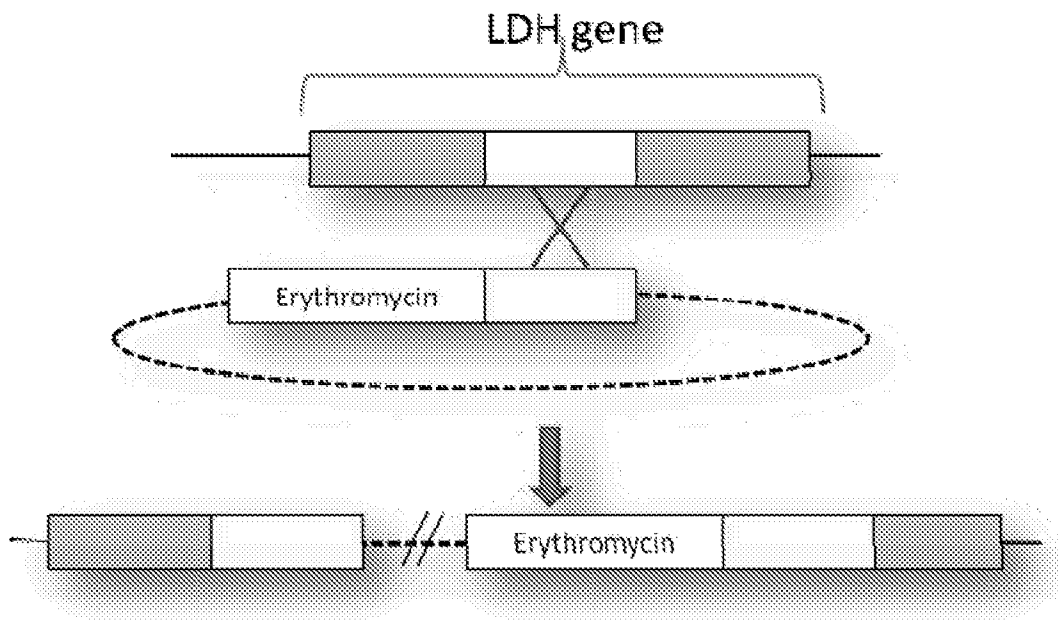


Figure 12

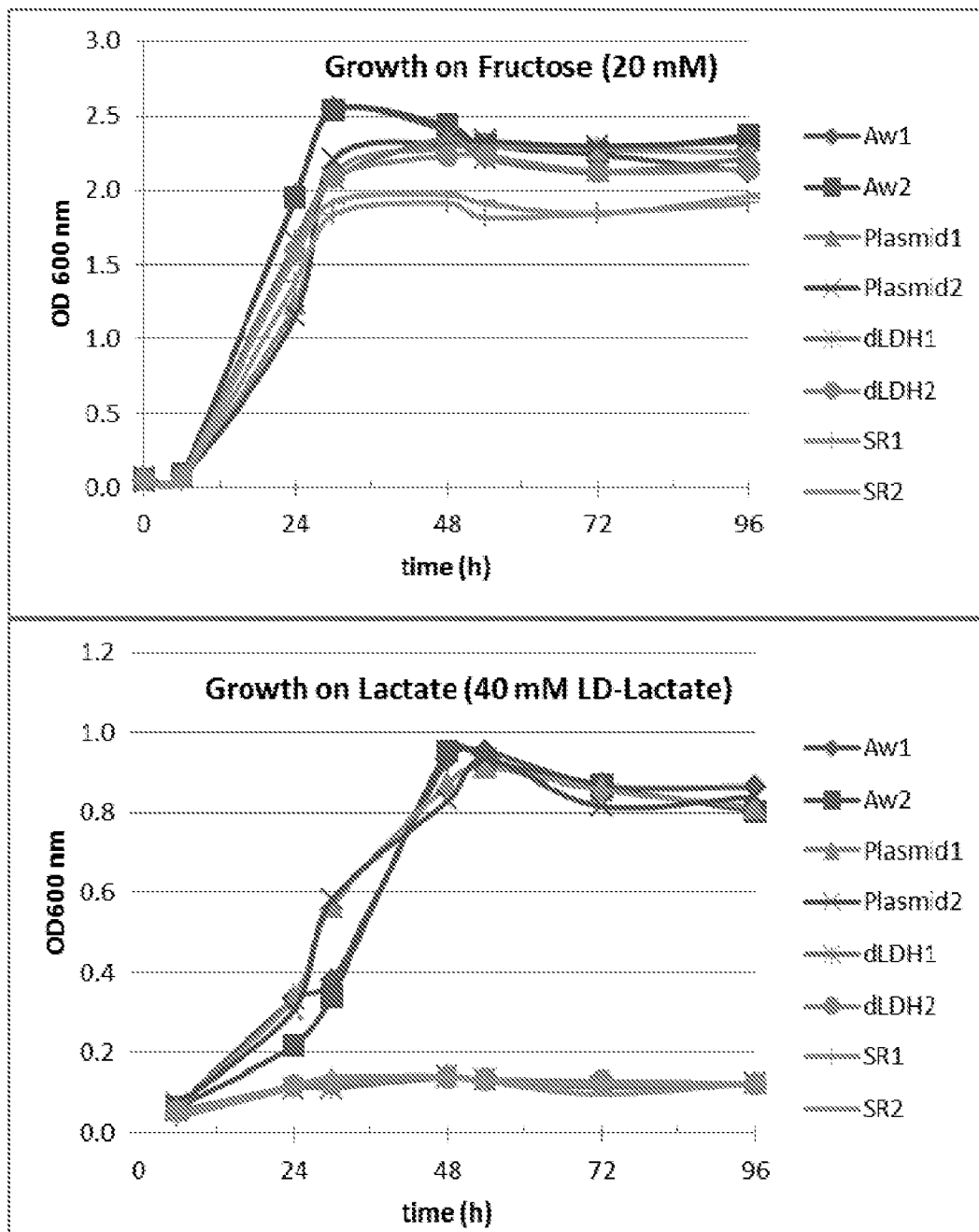


Figure 13

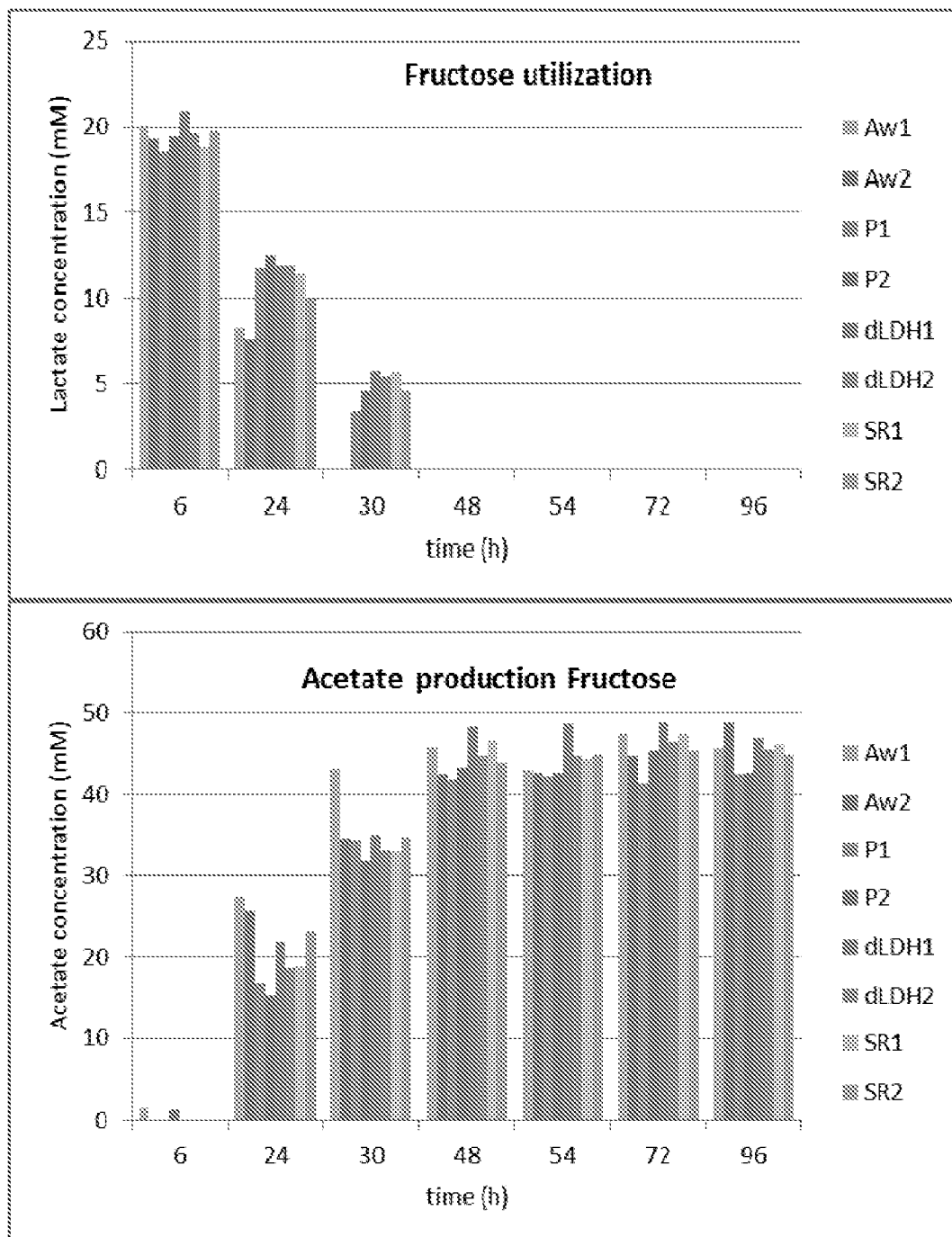


Figure 14

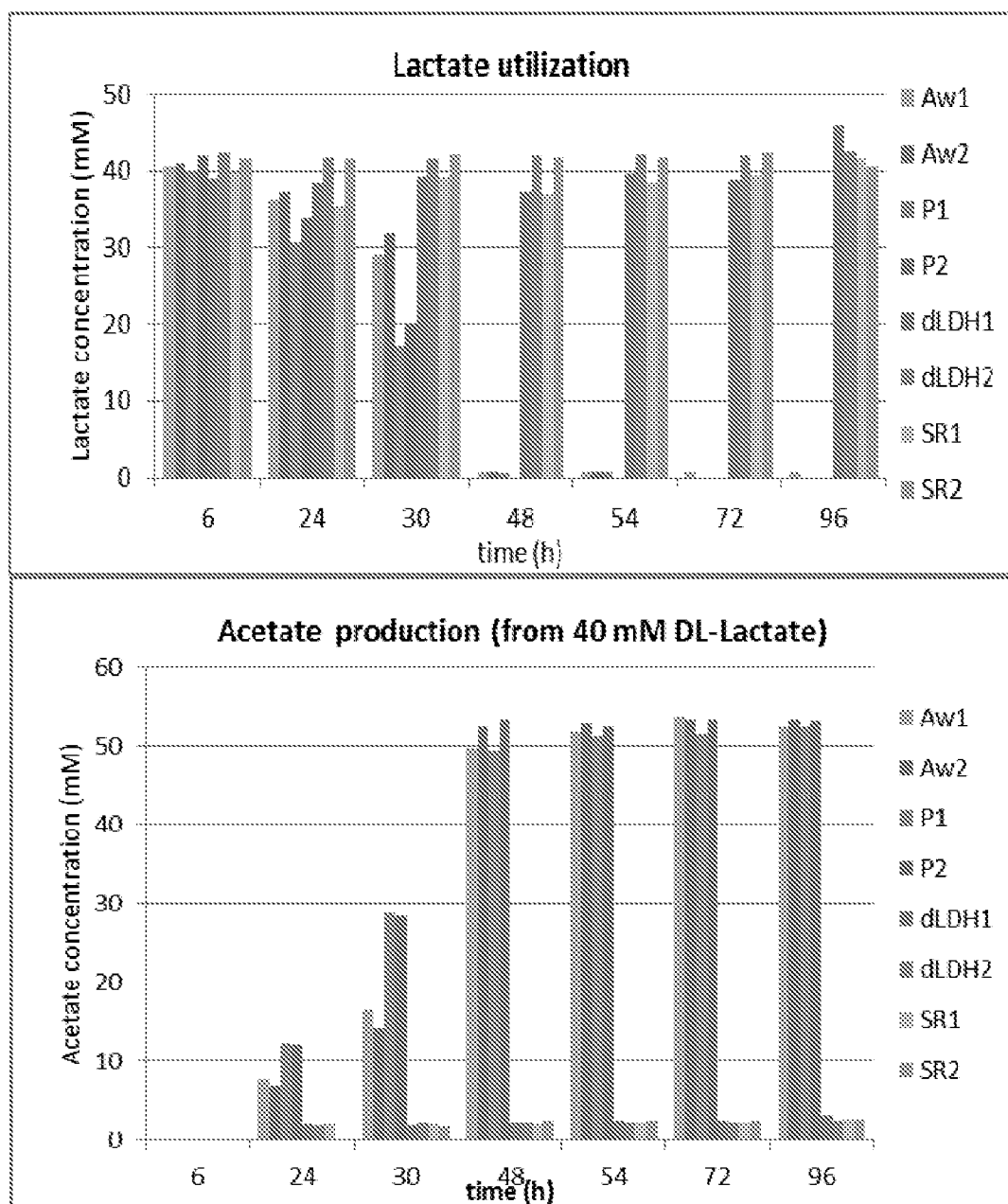


Figure15

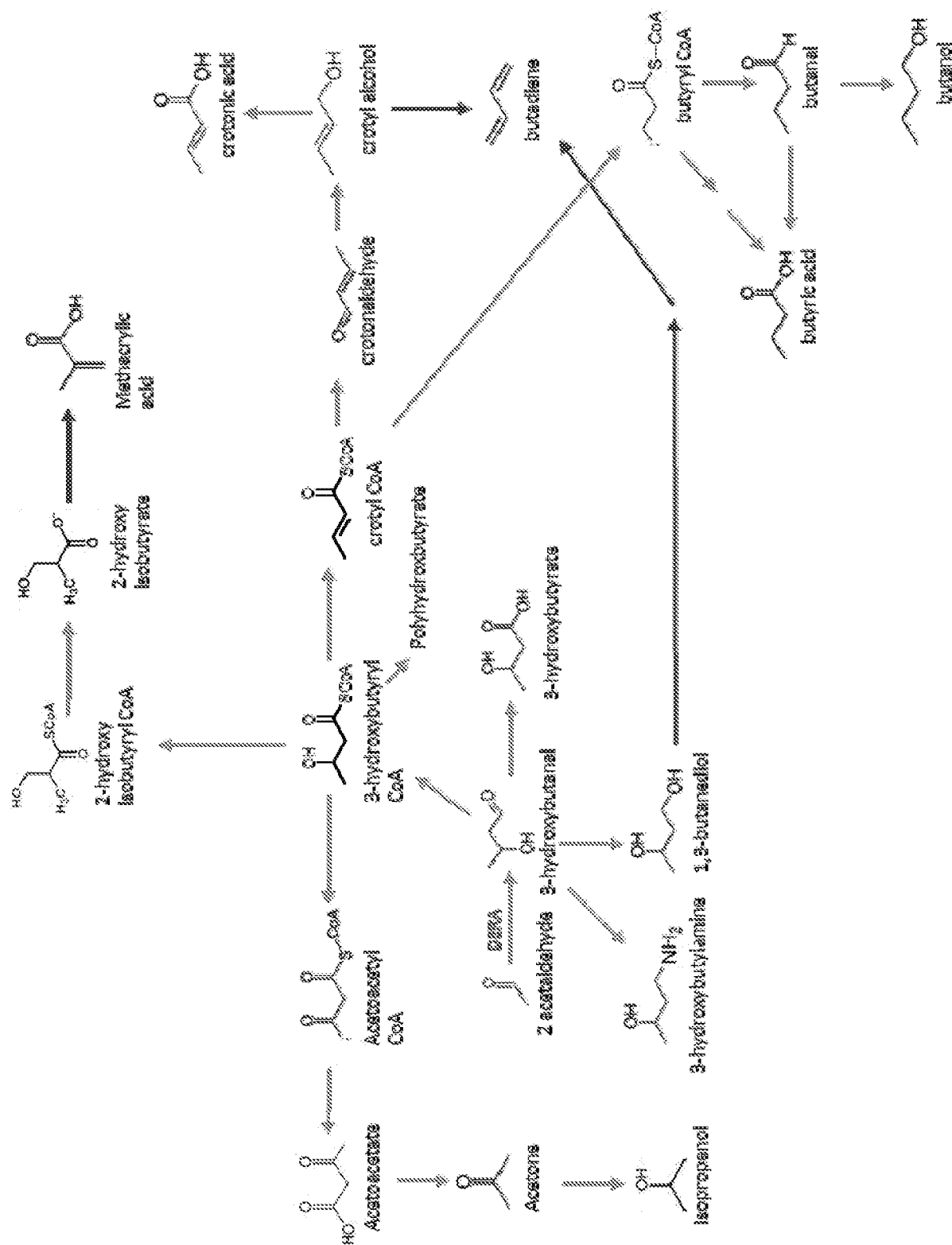


Figure 17

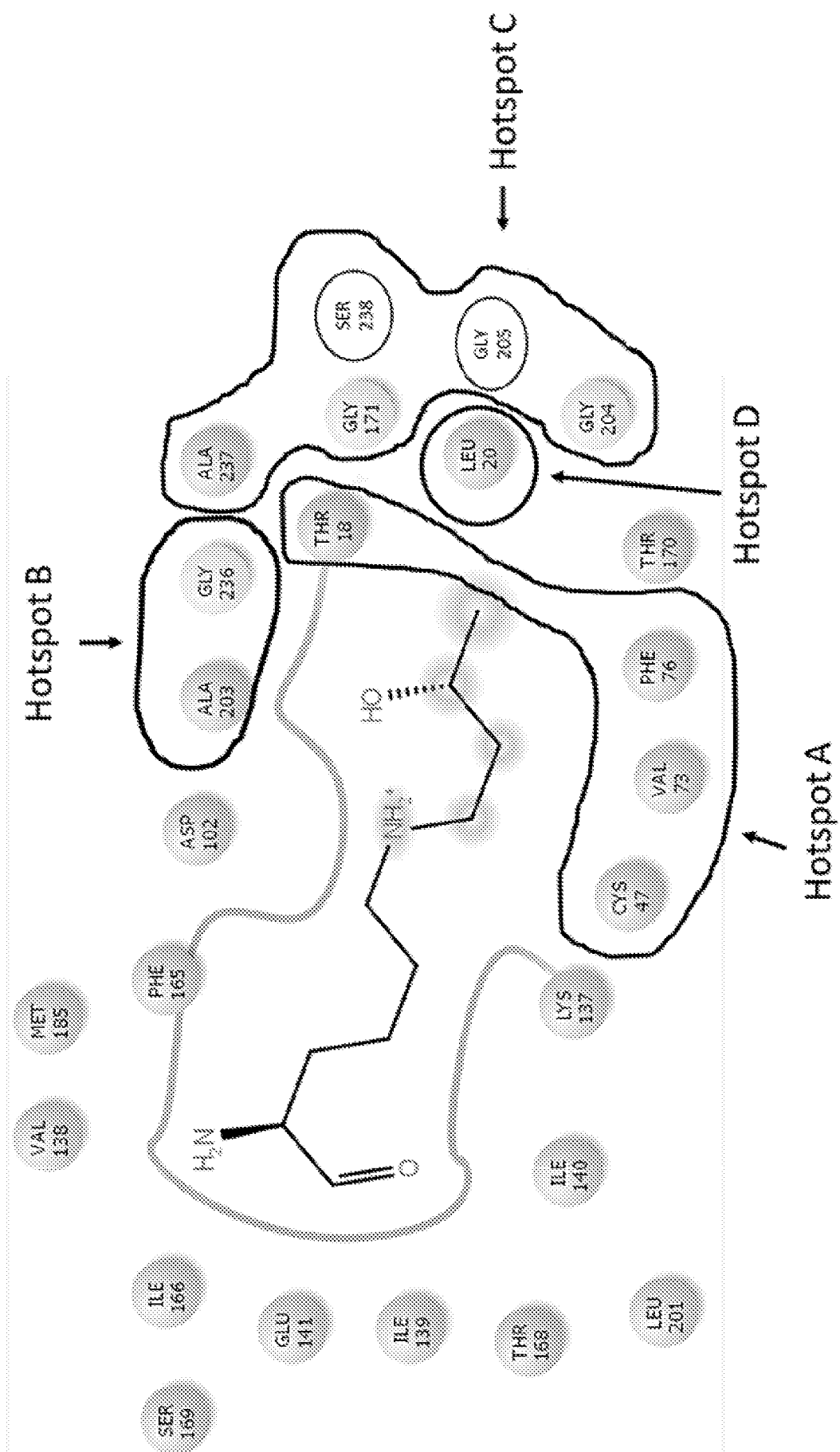


Figure 18

			20		40
DERA Consensus	MA - - - - -	- - KLIDH T L L	KPDATDED I I	KLCHEAKEY -	29
DERA E coli	MTDLKASSLR	ALKLMDLTTL	NDDDTDEKVI	ALCHQAKTPV	40
		60		80	
DERA Consensus	- NFASV Q VNP	RFVPLAAETL	EG - - - - DVKV	CT V I G F PLGA	64
DERA E coli	GNTAAICIYP	RFIPIARKTL	KEQGTPEIRT	ATVTNFPHGN	80
		100		120	
DERA Consensus	DTTAVKAFET	KAAIANGADE	IDMVINIGAL	KAGNEDVVEE	104
DERA E coli	DDIEIALAET	RAAIAYGADE	VDVFPYRAL	MAGNEQVGFD	120
		140		160	
DERA Consensus	DIKAVVEACG	- - - VLLKV I I	ETCLLTDEE I	V - KACE I S I K	140
DERA E coli	LVKACKEACA	AANVLLKV I I	ETGELKDEAL	IRKASE I S I K	160
		180		200	
DERA Consensus	AGAD F V K T S T	G F S T G G A T V E	DVRLMRETV -	- - GPD - - V G V	175
DERA E coli	AGAD F I K T S T	GKVAVNATPE	SARIMMEVIR	DMGVEKTVGF	200
		220		240	
DERA Consensus	K A S G G V R T A E	DA - - - LAMIE	A - - GAT - - - -	- - - R I G A S A G	203
DERA E coli	KPAGGVRTAE	DAQKYLA IAD	ELFGADWADA	RHYRFGASSL	240
DERA Consensus	VA - I V K - - - -	- - - - - SGY			211
DERA E coli	LASLLKALGH	GDGKSASSY			259

Figure 19

		20		40	
DERA Consensus	MAK - - - - -		- - - - -	- - - - -	3
DERA H sapiens	MSAHNRGTEL		DLSWISKIQV	NHPAVLRRAE	QIQARRTVKK 40
		60		80	
DERA Consensus	- - - - -		--LIDHTLLK	PDATDEDI IK	LCHEAKE - - - 28
DERA H sapiens	EWQAAWLLKA		VTFIDLTTLS	GDDTSSNIQR	LCYKAKYPIR 80
		100		120	
DERA Consensus	- - - - -		---YNFASVQ	VNPRFVPLAA	ETLEG - - - DV 52
DERA H sapiens	EDLLKALNMH		DKGITTAAVC	VYPARVCDV	KALKAAGCNI 120
		140		160	
DERA Consensus	KVCTMI - GFP		LGADTTAVKA	FETKAAIANG	ADEIDMVINI 91
DERA H sapiens	PVASVAAGFP		AGQTHLKTRL	EEIRLAVEDG	ATEIDVVINR 160
		180		200	
DERA Consensus	GALKAGNEDV		VEEDIKAVVE	ACGVLLKVII	ETCLLTDEEI 131
DERA H sapiens	SLVLTGQWE -		- - - - -	- - - - -	- - - - - 169
		220		240	
DERA Consensus	VKACEISIKA		GADFVKTSTG	FSTGGAT - - -	-VEDVRLMRE 167
DERA H sapiens	- - - - -		GSDFIKTSTG	KETVNATFPV	AIVMLRAIRD 199
		260		280	
DERA Consensus	- - - TVGPDVG		VKASGGVVRTA	EDALAMI - - -	- - - EAGAT - - - 196
DERA H sapiens	FFWKTGNKIG		FKPAGGIRSA	KDSLAWLSLV	KEELGDEWLK 239
		300			
DERA Consensus	- - - - RIGASA		GVAIVK - - - -	- - - - - SGY	- - - - - 211
DERA H sapiens	PELFRIGAST		LLSDIERQIY	HHVTGRYAAY	HDLPMS 275

Figure 20

		20		40	
DERA Consensus	MAK - - - - -	- - - - - L I D H T	L L K P D A T D E D	I I K L C H E A - K	27
DERA P falciparum	MANYAETFS	WTSICLT DHT	LLGENDT EDD	I RELCEESV K	40
		60		80	
DERA Consensus	EYNFAS - V C V	NPRFVPLAAE	TLEGDV - - - -	- K V C T V I G F P	61
DERA P falciparum	TCPFAAAVCV	KPEFVKFIND	TIKKEICPFK	PKVACVINFP	80
		100		120	
DERA Consensus	LGADTTAVKA	FETKAAIANG	ADEIDMVINI	GALKAGNEDV	101
DERA P falciparum	HGTDVVEKVL	EDTKKVLEDG	VDEIDLVINY	KKVLQNVEEG	120
		140		160	
DERA Consensus	VEEDIKAVVE	ACGVL - - - - L	KV I I E T C - L L	TDEEIVKACE	136
DERA P falciparum	LKEATELTKK	VKNLMKEKVL	KV I I E T G E L K	NDELI I K T T L	160
		180		200	
DERA Consensus	I S I K A G A D F V	K T S T G F - - - -	S T G G A T V E D V	R L M R E T V G P D	172
DERA P falciparum	AVLEGNADFV	KTSSGKVPVN	ATPGAVK A I I	EAMKQYVEKN	200
		220		240	
DERA Consensus	- - - - - V G V K	A S G G V R T A E D	A - - - - - L A - - -	M I E A G A T - - -	196
DERA P falciparum	PQKKDKIGLK	VAGGVGDLNS	ASY Y I L L A R R	FLSALACHPN	240
		260			
DERA Consensus	- - R I G A S A G V	A I V K S - - - - -	- G Y		211
DERA P falciparum	NFRIGSSSLV	PKLRK I I A Q N	PAV		263

Figure 21

		20		40	
DERA Consensus	MAKLIDHTLL	KPDATDEDII	KLCHEAKEYN	FASVQVNPRF	40
DERA P aerophilum	MIHLVDYALL	KPYLTVDEAV	AGARKAEELG	VAAYCVNPIY	40
		60		80	
DERA Consensus	VPLAAETLEG	DVKVCTVIGF	PLGADTTAVK	AFETKAAIAN	80
DERA P aerophilum	APVVRPLLR-	KVKLCVVADF	PFGALPTASR	-IALVSRLAE	78
		100		120	
DERA Consensus	GADEIDMVIN	IGALKAGNED	VVEEDIKAVV	EACGV-LLKV	119
DERA P aerophilum	VADEIDVVAP	IGLVKSRRWA	EVRRLISVV	GAAGGRVVKV	118
		140		160	
DERA Consensus	IIETCLLTDE	EIVKACEISI	KAGADFKVTS	TGF-----	152
DERA P aerophilum	ITEEPYLRDE	ERYTLYDIIA	EAGAHFIKSS	TGFAEEAYAA	158
		180		200	
DERA Consensus	-----STG	GATVEDVRLM	RETVGPDVGV	KASGGVVRTAE	185
DERA P aerophilum	RQGNPVHSTP	ERAAAIARYI	KEK-GYRLGV	KMAGGIRTRE	197
		220			
DERA Consensus	DALAMIEA--	-----GATRI	GASAGVAIVK	SGY	211
DERA P aerophilum	QAKAIVDAIG	WGEDPARVRL	GTSTPEALL-	---	226

Figure 22

				20				40	
DERA Consensus	M - - - AKLID	HTLLKPDATD	EDI I KLCHEA	KEYNFASV	CV	36			
DERA G thermoglucosidasius	MENNI AKMID	HTLLKADTTK	AQIVKLCEEA	KQYGFASV	CV	40			
				60				80	
DERA Consensus	NPRFVPLAAE	TLEG - DVKVC	TVIGFPLGAD	TTAVKAFETK		75			
DERA G thermoglucosidasius	NPTWVATAAE	LLKGT DVKVC	TVIGFPLGAN	TPETKAFETK		80			
				100				120	
DERA Consensus	AAIANGADEI	DMVINIGALK	AGNEDVVEED	IKAVVEACG	-	114			
DERA G thermoglucosidasius	NAIENGATEV	DMVINIGALK	DGND DLVERD	IRAVVEAAKG		120			
				140				160	
DERA Consensus	-VLLKVIIET	CLLTDEEIVK	ACEISIKAGA	DFVKTSTGFS		153			
DERA G thermoglucosidasius	KALVKVIIET	CLLTEEEKVR	ACQLAVKAGA	DYVKTSTGFS		160			
				180				200	
DERA Consensus	TGGATVEDVR	LMRETVGPDV	GVKASGGVRT	AEDALAMIEA		193			
DERA G thermoglucosidasius	TGGATTEDVA	LMRKTVGQNI	GVKASGGVRD	IKSAMAMVEA		200			
				220					
DERA Consensus	GATRI GASAG	VAIVK - - - -	SGY		211				
DERA G thermoglucosidasius	GATRI GTSSG	AAIVAGKTAA	SDY		223				

Figure 23

			20			40	
DERA Consensus	MAKL - - - - I	DHTLLKPDAT	DEDI I KLCHE	AKEYNFASVC	35		
DERA A woodii	MTNKEILSHI	DHTLLKQFAT	WEEIKVICED	AIEYQTASVC	40		
			60			80	
DERA Consensus	VNPRFVPLAA	ETLEGDVKVC	TVIGFPLGAD	TTAVKAFETK	75		
DERA A woodii	IPPAYVKKIK	KTYGDKLNIC	TVIGFPLGYN	TTATKVFEAV	80		
			100			120	
DERA Consensus	AAIANGADEI	DMVINIGALK	AGNEDVVEED	IKAVVEACGV	115		
DERA A woodii	DAVKNGASEI	DMVINITALK	NKDYDLVQNE	IAEIKKAIGA	120		
			140			160	
DERA Consensus	-LLKVIIETC	LLTDEEIVKA	CEISIKAGAD	FVKTSTGFST	154		
DERA A woodii	KILKVIVETC	FLTAAEKVKV	CELVTNAGAD	YIKTSTGFGT	160		
			180			200	
DERA Consensus	GGATVEDVRL	MRETVGPDVG	VKASGGVRTA	EDALAMIEAG	194		
DERA A woodii	AGATIEDIRL	FKAHIGPNVK	MKAAGGVKTV	DDLIAFINEG	200		
			220				
DERA Consensus	ATRIGASAGV	AIVK - - - SG	Y	211			
DERA A woodii	CARIGTSSAV	KLLKGEAVSG	Y	221			

Figure 24

MAKLIDHTLLKPDATDEDIIKLCHEAKEYNFASVCVNPRFVPLAAETLEGDVKVCTVIGFPLGADT
TAVKAFETKAAIANGADEIDMVINIGALKAGNEDVVEEDIKAVVEACGVLLKVIIETCLLTDEEIVKA
CEISIKAGADFVKTSTGFSTGGATVEDVRLMRETVGPDVGVKASGGVRTAEDALAMIEAGATRI
GASAGVAIVKSGY

Figure 25

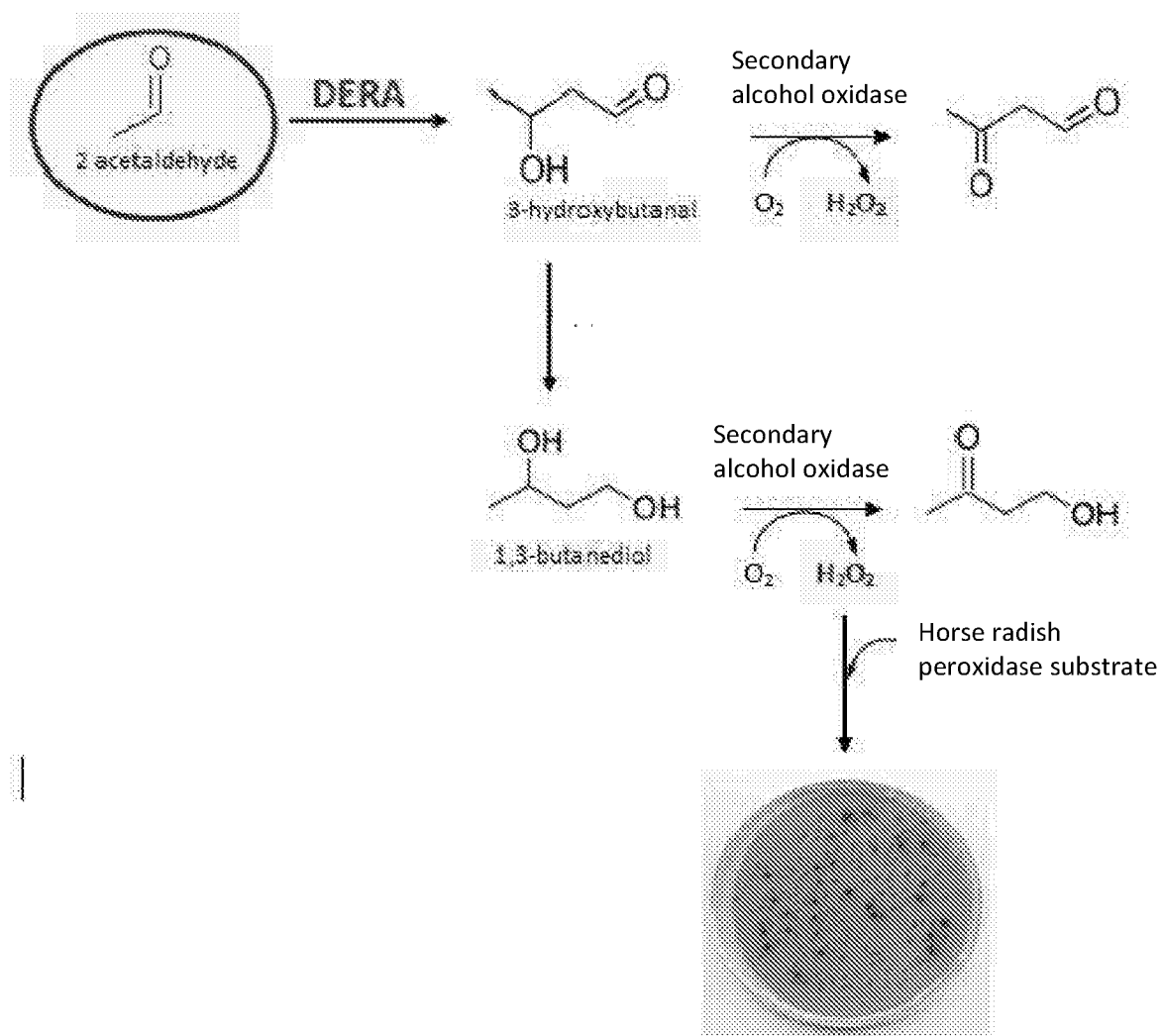


Figure 26

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2017/050901

A. CLASSIFICATION OF SUBJECT MATTER

INV. C12N9/10 C12N15/52 C12P7/02
ADD. C12N9/02

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12N C12P

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

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

EPO-Internal, Sequence Search, EMBASE, FSTA, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2005/118794 A2 (DSM IP ASSETS BV [NL]; JENNEWEIN STEFAN MARTIN [DE]; SCHUERMANN MARTIN) 15 December 2005 (2005-12-15) cited in the application	1-22, 39-42, 44-51, 103,104
Y	p.5 l.9-22, p.32 l.14- p.33 l.19 -/--	43, 52-102



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

5 July 2017

Date of mailing of the international search report

11/09/2017

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Bonello, Steve

INTERNATIONAL SEARCH REPORT

International application No

PCT/GB2017/050901

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	& DATABASE UniProt [Online] 20 May 2008 (2008-05-20), "RecName: Full=Deoxyribose-phosphate aldolase {ECO:0000255 HAMAP-Rule:MF_00114}; Short=DERA {ECO:0000255 HAMAP-Rule:MF_00114}; EC=4.1.2.4 {ECO:0000255 HAMAP-Rule:MF_00114}; AltName: Full=2-deoxy-D-ribose 5-phosphate aldolase {ECO:0000255 HAMAP-Rule:MF_00114}; AltName: Full=Phosphodeoxyriboaldolase {E", retrieved from EBI accession no. UNIPROT:A8AX59 Database accession no. A8AX59 sequence -----	
X	WO 2015/181074 A1 (SCIENTIST OF FORTUNE SA [LU]) 3 December 2015 (2015-12-03)	1-42, 44-51, 103,104
Y	p.10 par.2- p.11 par.1, p.12 par.2-5, p.13 par.2; sequence 10 & DATABASE UniProt [Online] 20 May 2008 (2008-05-20), "RecName: Full=Deoxyribose-phosphate aldolase {ECO:0000255 HAMAP-Rule:MF_00114}; Short=DERA {ECO:0000255 HAMAP-Rule:MF_00114}; EC=4.1.2.4 {ECO:0000255 HAMAP-Rule:MF_00114}; AltName: Full=2-deoxy-D-ribose 5-phosphate aldolase {ECO:0000255 HAMAP-Rule:MF_00114}; AltName: Full=Phosphodeoxyriboaldolase {E", retrieved from EBI accession no. UNIPROT:A8AX59 Database accession no. A8AX59 sequence -----	43, 52-102
Y	WO 2014/063156 A2 (BRASKEM S A AP 09 [BR]; BRASKEM AMERICA INC [US]; GARCEZ LOPES MATEUS) 24 April 2014 (2014-04-24) p.1 par.4, p.2 5, p.15 par.77, p.28 par.81, Tbl.1, 2, p.93 par.17 -----	43, 52-102
A	WO 2013/057194 A1 (SCIENTIST OF FORTUNE SA [LU]) 25 April 2013 (2013-04-25) p.2 par.2, 3 -----	52-102
	----- -/--	

INTERNATIONAL SEARCH REPORT

International application No

PCT/GB2017/050901

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE UniProt [Online] 21 July 1986 (1986-07-21), "RecName: Full=Deoxyribose-phosphate aldolase; Short=DERA; EC=4.1.2.4; AltName: Full=2-deoxy-D-ribose 5-phosphate aldolase; AltName: Full=Phosphodeoxyriboaldolase; Short=Deoxyriboaldolase;", XP002771757, retrieved from EBI accession no. UNIPROT:P0A6L0 Database accession no. P0A6L0 sequence -----	1-18
A	CLAIRE L WINDLE ET AL: "Engineering aldolases as biocatalysts", CURRENT OPINION IN CHEMICAL BIOLOGY, vol. 19, 1 April 2014 (2014-04-01), pages 25-33, XP055387415, GB ISSN: 1367-5931, DOI: 10.1016/j.cbpa.2013.12.010 p.25 col.2 par.3, p.28 col.1 par.3, p.31 col.1 par.2 -----	1-22
A	KATERYNA FESKO ET AL: "Biocatalytic Methods for C?C Bond Formation", CHEMCATCHEM, vol. 5, no. 6, 21 February 2013 (2013-02-21), pages 1248-1272, XP055388022, DE ISSN: 1867-3880, DOI: 10.1002/cctc.201200709 p.1254 col.2 par.3 -p.1255 col.2 par.1 -----	1-22
T	WO 2016/050842 A1 (ZUVASYNTHA LTD [GB]) 7 April 2016 (2016-04-07) -----	
T	WO 2017/011915 A1 (GOVERNING COUNCIL OF THE UNIV OF TORONTO [CA]) 26 January 2017 (2017-01-26) -----	

INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB2017/050901

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-104

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-104

Directed towards a modified 2-deoxyribose phosphate aldolase (DERA) enzyme variant.

2. claims: 105-114

Directed towards use of an alcohol oxidase for detecting DERA variants.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/GB2017/050901

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2005118794 A2	15-12-2005	CA 2568728 A1 CN 1965084 A EP 1751287 A2 JP 2008541693 A TW 200604344 A US 2009209001 A1 WO 2005118794 A2	15-12-2005 16-05-2007 14-02-2007 27-11-2008 01-02-2006 20-08-2009 15-12-2005
WO 2015181074 A1	03-12-2015	AU 2015266132 A1 CA 2947212 A1 CN 106536742 A EP 3149180 A1 JP 2017516475 A US 2017191095 A1 WO 2015181074 A1	17-11-2016 03-12-2015 22-03-2017 05-04-2017 22-06-2017 06-07-2017 03-12-2015
WO 2014063156 A2	24-04-2014	NONE	
WO 2013057194 A1	25-04-2013	AU 2012324935 A1 BR 112014008061 A2 EP 2771473 A1 JP 2014530615 A US 2014256009 A1 WO 2013057194 A1	06-03-2014 11-04-2017 03-09-2014 20-11-2014 11-09-2014 25-04-2013
WO 2016050842 A1	07-04-2016	CA 2962828 A1 EP 3201344 A1 WO 2016050842 A1	07-04-2016 09-08-2017 07-04-2016
WO 2017011915 A1	26-01-2017	NONE	