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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 January 2008 (03.01.2008)

PCT

(10) International Publication Number
WO 2008/000415 A1

(51) International Patent Classification:
C12N 15/54 (2006.01) *C12P 1/04* (2006.01)
C12N 9/10 (2006.01) *C12N 1/20* (2006.01)
C12P 17/04 (2006.01)

(21) International Application Number:
PCT/EP2007/005591

(22) International Filing Date: 25 June 2007 (25.06.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
06013091.1 26 June 2006 (26.06.2006) EP

(71) Applicant *(for all designated States except US)*: **DSM IP ASSETS B.V.** [NL/NL]; Het Overloon 1, NL-16411 TE Heerlen (NL).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BREMUS, Christoph** [DE/DE]; Fuchsweg 36, 50859 Köln-Lövenich (DE). **BRINGER-MEYER, Stephanie** [DE/DE]; Josef-Rahier-Strasse 13, 52428 Jülich (DE). **HERMANN, Ute** [DE/DE]; Birnbaumweg 72, 88326 Aulendorf (DE). **MOUNCEY, Nigel John** [GB/CH]; Curt-Goetz Strasse 18, CH-4102 Binningen (CH). **SAHM, Hermann** [DE/DE]; Wendelinusstrasse 71, 52428 Jülich (DE).

(74) Agent: **SEIBEL-THOMSON, Nadja .**; c/o DSM Nutritional Products Ltd, Wurmisweg 576, CH-4303 Kaiseraugst (CH).

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

Published:
— *with international search report*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

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

WO 2008/000415 A1

(54) Title: BIOLOGICAL PROCESS USING A TRANSALDOLASE

(57) Abstract: The present invention relates to microorganisms genetically engineered to reduce carbon source diversion throughout the bioconversion of a carbon source, such as e.g. sorbitol, into biomass. Processes for generating such microorganisms are also provided by the present invention. The invention also relates to polynucleotide sequences comprising genes that encode proteins that are involved in the bioconversion of a carbon source such as e.g. sorbitol, into biomass. The invention also features polynucleotides comprising the full-length polynucleotide sequences of the novel genes and fragments thereof, the novel polypeptides encoded by the polynucleotides and fragments thereof, as well as their functional equivalents. Also included are processes of using the polynucleotides and modified polynucleotide sequences to transform host microorganisms leading to a microorganism with reduced carbon source diversion, i.e. higher yield and/or efficiency of biomass production from a carbon source such as e.g. sorbitol.

Novel biological process

The present invention relates to microorganisms genetically engineered to increase yield and/or efficiency of biomass production from a carbon source such as *e.g.* sorbitol. Processes for generating such microorganisms are also provided by the present invention. The invention also relates to polynucleotide sequences comprising genes that encode proteins that are involved in the bioconversion of a carbon source such as *e.g.* sorbitol into biomass. The invention also features polynucleotides comprising the full-length polynucleotide sequences of the novel genes and fragments thereof, the novel polypeptides encoded by the polynucleotides and fragments thereof, as well as their functional equivalents. Also included are processes of using the polynucleotides and modified polynucleotide sequences to transform host microorganisms leading to a microorganism with higher yield and/or efficiency of biomass production from a carbon source such as *e.g.* sorbitol.

The bioconversion of a carbon source may involve many different metabolic routes, and involve several enzymatic steps to generate biomass, wherein the enzymes may be located in the cytosol, on the membrane or in the periplasmic space of a host cell. Furthermore, transporters may also play an important role in the efficient conversion of a carbon source into biomass.

For instance, in the case of acetic acid bacteria, which are obligate aerobe, gram-negative microorganisms belonging to the genus *Acetobacter, Gluconobacter*, and *Gluconacetobacter*, these microorganisms are able to rapidly perform incomplete oxidation of sugars and sugar alcohols at the periplasmic membrane level. The resulting products (*e.g.* organic acids, aldehydes or ketones) are predominantly secreted in the medium. In addition to that, sugars and sugar alcohols can be directly transported into the

cytosol and converted into metabolites and biomass. Acetic acid bacteria lack the Embden-Meyerhof-Parnas pathway of glycolysis. Carbon sources can be converted to glucose-6-phosphate or 6-phosphogluconate and metabolized via the pentose phosphate pathway or the Entner-Doudoroff pathway to produce reducing power in the form of NAD(P)H, energy in form of ATP and tricarboxylic compounds necessary for growth and maintenance.

One disadvantage of such bioconversion processes, however, is the diversion of carbon sources or intermediates throughout said bioconversion process such that for instance the next (enzymatic) step cannot be performed in an optimal way, leading to the loss of available carbon substrate material for conversion into biomass and resultant energy losses, in form of *e.g.* ATP or NADPH. In the case of for instance the bioconversion of glucose into biomass, this loss may be due to transporting the glucose or an intermediate thereof out of the cytosol or by using the glucose or the resulting intermediates as substrates for other pathways not leading to the production of biomass.

It is an object of the present invention to provide microorganisms which are engineered in such a way that the carbon source flow throughout the bioconversion of carbon sources, such as *e.g.* sorbitol, is altered, *e.g.* via a reduction of carbon source diversion, leading to higher production and/or yield of biomass produced from such carbon sources.

Surprisingly, it has been found that proteins or subunits of such proteins having activity towards or which are involved in the metabolization of substrates such as *e.g.* sorbitol, which are selected from transferases [EC 2], preferably transferases transferring aldehyde or ketonic groups [EC 2.2], more preferably transketolases and transaldolases [EC 2.2.1], and most preferably selected from transaldolases [EC 2.2.1.2] referred herein as to tal, play an important role in the formation of biomass.

In particular, it has now been found that proteins encoded by polynucleotides having a nucleotide sequence that hybridizes preferably under highly stringent conditions to a sequence shown in SEQ ID NO:1 play an important role in the bioconversion of a carbon source such as *e.g.* sorbitol to biomass. It has also been found, that by genetically altering such nucleotides in a microorganism, such as for example *Gluconobacter*, the efficiency of said bioconversion within said microorganism can be even greatly improved leading *e.g.* to higher production and/or yield of biomass from a carbon source such as *e.g.* sorbitol.

Consequently, the invention relates to a polynucleotide selected from the group consisting of:

(a) polynucleotides encoding a polypeptide comprising the amino acid sequence according to SEQ ID NO:2;

(b) polynucleotides comprising the nucleotide sequence according to SEQ ID NO:1;

(c) polynucleotides comprising a nucleotide sequence obtainable by nucleic acid amplification such as polymerase chain reaction, using genomic DNA from a microorganism as a template and a primer set according to SEQ ID NO:3 and SEQ ID NO:4;

(d) polynucleotides comprising a nucleotide sequence encoding a fragment or derivative of a polypeptide encoded by a polynucleotide of any of (a) to (c) wherein in said derivative one or more amino acid residues are conservatively substituted compared to said polypeptide, and said fragment or derivative has the activity of a transferase [EC 2], preferably a transferase transferring aldehyde or ketonic groups [EC 2.2], most preferably tal;

(e) polynucleotides the complementary strand of which hybridizes under stringent conditions to a polynucleotide as defined in any one of (a) to (d) and which encode a transferase [EC 2], preferably a transferase transferring aldehyde or ketonic groups [EC 2.2], most preferably tal;

(f) polynucleotides which are at least 70%, such as 85, 90 or 95% identical to a polynucleotide as defined in any one of (a) to (d) and which encode a transferase [EC 2], preferably a transferase transferring aldehyde or ketonic groups [EC 2.2], most preferably tal;

or

the complementary strand of such a polynucleotide.

The polypeptide as isolated from *Gluconobacter oxydans* DSM 2343 (ATCC 621H) shown in SEQ ID NO:2 and described herein was found to be particularly useful, since it appeared that it performs a crucial function in the bioconversion of a carbon source such as *e.g.* sorbitol to biomass in microorganisms, in particular in bacteria, such as acetic acid bacteria, such as *Gluconobacter*, *Acetobacter* and *Gluconacetobacter*. Accordingly, the invention relates to a polynucleotide encoding a polypeptide according to SEQ ID NO:2. This protein may be encoded by a nucleotide sequence as shown in SEQ ID NO:1. The

invention therefore also relates to polynucleotides comprising the nucleotide sequence according to SEQ ID NO:1.

The nucleotide and amino acid sequences determined above were used as a "query sequence" to perform a search with Blast2 program (version 2 or BLAST from National Center for Biotechnology [NCBI] against the database PRO SW-SwissProt (full release plus incremental updates). From the searches, the polynucleotide according to SEQ ID NO:1 was annotated as encoding a protein having transaldolase (tal) activity.

A nucleic acid according to the invention may be obtained by nucleic acid amplification using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers such as the nucleotide primers according to SEQ ID NO:3 and SEQ ID NO:4 according to standard PCR amplification techniques. The nucleic acid thus amplified may be cloned into an appropriate vector and characterized by DNA sequence analysis.

The template for the reaction may be cDNA obtained by reverse transcription of mRNA prepared from strains known or suspected to comprise a polynucleotide according to the invention. The PCR product may be subcloned and sequenced to ensure that the amplified sequences represent the sequences of a new nucleic acid sequence as described herein, or a functional equivalent thereof.

The PCR fragment may then be used to isolate a full length cDNA clone by a variety of known methods. For example, the amplified fragment may be labeled and used to screen a bacteriophage or cosmid cDNA library. Alternatively, the labeled fragment may be used to screen a genomic library.

Accordingly, the invention relates to polynucleotides comprising a nucleotide sequence obtainable by nucleic acid amplification such as polymerase chain reaction, using DNA such as genomic DNA from a microorganism as a template and a primer set according to SEQ ID NO:3 and SEQ ID NO:4.

The invention also relates to polynucleotides comprising a nucleotide sequence encoding a fragment or derivative of a polypeptide encoded by a polynucleotide as described herein wherein in said derivative one or more amino acid residues are conservatively substituted compared to said polypeptide, and said fragment or derivative has the activity of a transferase, preferably tal.

The invention also relates to polynucleotides the complementary strand of which hybridizes under stringent conditions to a polynucleotide as defined herein and which encode a transferase, preferably tal.

The invention also relates to polynucleotides which are at least 70% identical to a polynucleotide as defined herein and which encode a transferase; and the invention also relates to polynucleotides being the complementary strand of a polynucleotide as defined herein above.

Such a polynucleotide may be for instance isolated from *Gluconobacter oxydans* DSM 17078 as shown in SEQ ID NO:5 and encoding a polypeptide according to SEQ ID NO:6. The invention therefore also relates to polynucleotides selected from the group consisting of

(a) comprising the nucleotide sequence according to SEQ ID NO:5;

(b) polynucleotides encoding a polypeptide comprising the amino acid sequence according to SEQ ID NO:6;

(c) polynucleotides comprising a nucleotide sequence obtainable by nucleic acid amplification such as polymerase chain reaction, using genomic DNA from a microorganism as a template and a primer set according to SEQ ID NO:7 and SEQ ID NO:8;

(d) polynucleotides comprising a nucleotide sequence encoding a fragment or derivative of a polypeptide encoded by a polynucleotide of any of (a) to (c) wherein in said derivative one or more amino acid residues are conservatively substituted compared to said polypeptide, and said fragment or derivative has the activity of a transferase [EC 2], preferably a transferase transferring aldehyde or ketonic groups [EC 2.2], most preferably tal;

(e) polynucleotides the complementary strand of which hybridizes under stringent conditions to a polynucleotide as defined in any one of (a) to (d) and which encode a transferase [EC 2], preferably a transferase transferring aldehyde or ketonic groups [EC 2.2], most preferably tal;

(f) polynucleotides which are at least 70%, such as 85, 90 or 95% identical to a polynucleotide as defined in any one of (a) to (d) and which encode a transferase [EC 2],

preferably a transferase transferring aldehyde or ketonic groups [EC 2.2], most preferably tal;

or

the complementary strand of such a polynucleotide.

In one embodiment the polypeptides as of the present invention may be present in the form of a fusion protein, such as *e.g.* tal fused to a further protein, such as *e.g.* an isomerase [EC 5], in particular a glucose-6-phosphate isomerase [EC 5.3.1.9], referred hereafter as to pgi. A polynucleotide expressing such a fusion of tal and pgi as isolated from *Gluconobacter oxydans* DSM 2343 is shown in SEQ ID NO:9, the respective polypeptide is shown in SEQ ID NO:10.

The invention also relates to microorganisms wherein the activity of a transferase, preferably tal, is enhanced and/or improved so that the yield and/or production of biomass which is produced through bioconversion of a carbon source such as such as *e.g.* sorbitol is increased.

A suitable carbon source that can be converted into biomass may be for instance sorbitol, glucose, fructose, mannitol, sorbose, glycerol, or may be selected from carbon sources the assimilation of which results in the formation of said substances, such as sucrose, maltose or starch. Mixtures of the above carbon sources may be also used for the purpose of the present invention, such as glucose and sorbitol, or fructose and sorbitol. Preferably, the carbon source is selected from glucose, fructose, sorbitol, mannitol, sorbose, glycerol, sucrose, and mixtures thereof, wherein sorbitol or mixtures containing sorbitol are most preferred.

The skilled person will know how to enhance and/or improve the activity of a transferase, preferably tal. Such may be for instance accomplished by either genetically modifying the host organism in such a way that it produces more or more stable copies of such protein, preferably tal, than the wild type organism or by increasing the specific activity of such protein, preferably the tal.

In the following description, procedures are detailed to achieve this goal, *i.e.* the increase in the yield and/or production of biomass which is produced through bioconversion of a carbon source such as *e.g.* sorbitol by increasing the activity of tal. These procedures apply mutatis mutandis for other transferases or fusion proteins comprising tal and pgi.

Modifications in order to have the organism produce more copies of the tal gene and/or protein may include the use of a strong promoter, or the mutation (*e.g.* insertion, deletion or point mutation) of (parts of) the tal gene or its regulatory elements. It may also involve the insertion of multiple copies of the gene into a suitable microorganism. An increase in the specific activity of a tal may also be accomplished by methods known in the art. Such methods may include the mutation (*e.g.* insertion, deletion or point mutation) of (parts of) the tal gene. A gene is said to be "overexpressed" if the level of transcription of said gene is enhanced in comparison to the wild type gene. This may be measured by for instance Northern blot analysis quantifying the amount of mRNA as an indication for gene expression. As used herein, a gene is overexpressed if the amount of generated mRNA is increased by at least 1%, 2%, 5% 10%, 25%, 50%, 75%, 100%, 200% or even more than 500%, compared to the amount of mRNA generated from a wild-type gene.

Also known in the art are methods of increasing the activity of a given protein by contacting tal with specific enhancers or other substances that specifically interact with tal. In order to identify such specific enhancers, tal may be expressed and tested for activity in the presence of compounds suspected to enhance the activity of tal. The activity of tal may also be increased by stabilizing the messenger RNA encoding tal. Such methods are also known in the art, see for example, in Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, N. Y.; and Ausubel et al. (eds.), 1995, Current Protocols in Molecular Biology, (John Wiley & Sons, N.Y.).

The invention may be performed in any microorganism carrying a tal gene or homologue thereof. Suitable microorganisms may be selected from the group consisting of yeast, algae and bacteria, either as wild type strains, mutant strains derived by classic mutagenesis and selection methods or as recombinant strains. Examples of such yeast may be, *e.g.*, *Candida*, *Saccharomyces*, *Zygosaccharomyces*, *Schizosaccharomyces*, or *Kluyveromyces*. An example of such algae may be, *e.g.*, *Chlorella*. Examples of such bacteria may be, *e.g.*, *Gluconobacter*, *Acetobacter*, *Gluconacetobacter*, *Ketogulonicigenium*, *Pantoea*, *Pseudomonas*, such as, *e.g.*, *Pseudomonas putida*, and *Escherichia*, such as, *e.g.*, *Escherichia coli*. Preferred are *Gluconobacter* or *Acetobacter aceti*, such as for instance *G. oxydans*, *G. cerinus*, *G. frateurii*, *A. aceti subsp. xylinum* or *A. aceti subsp. orleanus*, preferably *G. oxydans* DSM 17078 or *G. oxydans* DSM 2343 (ATCC 621H). *Gluconobacter oxydans* DSM 17078 (formerly known as *Gluconobacter oxydans* N44-1) has been deposited at Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ), Mascheroder Weg 1B, D-38124 Braunschweig, Germany according to the Budapest Treaty on 26. January 2005.

Microorganisms which can be used for the present invention may be publicly available from different sources, *e.g.*, Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ), Mascheroder Weg 1B, D-38124 Braunschweig, Germany, American Type Culture Collection (ATCC), P.O. Box 1549, Manassas, VA 20108 USA or Culture Collection Division, NITE Biological Resource Center, 2-5-8, Kazusakamatari, Kisarazu-shi, Chiba, 292-0818, Japan (formerly: Institute for Fermentation, Osaka (IFO), 17-85, Juso-honmachi 2-chome,Yodogawa-ku, Osaka 532-8686, Japan). Examples of preferred bacteria deposited with IFO are for instance *Gluconobacter oxydans* (formerly known as *G. melanogenus*) IFO 3293, *Gluconobacter oxydans* (formerly known as *G. melanogenus*) IFO 3292, *Gluconobacter oxydans* (formerly known as *G. rubiginosus*) IFO 3244, *Gluconobacter frateurii* (formerly known as *G. industrius*) IFO 3260, *Gluconobacter cerinus* IFO 3266, *Gluconobacter oxydans* IFO 3287, and *Acetobacter aceti subsp. orleanus* IFO 3259, which were all deposited on April 05, 1954; *Acetobacter aceti subsp. xylinum* IFO 13693 deposited on October 22, 1975, and *Acetobacter aceti subsp. xylinum* IFO 13773 deposited on December 08, 1977. Strain *Acetobacter sp.* ATCC 15164, which is also an example of a preferred bacterium, was deposited with ATCC. Strain *Gluconobacter oxydans* (formerly known as *G. melanogenus*) N 44-1 as another example of a preferred bacterium is a derivative of the strain IFO 3293 and is described in Sugisawa et al., Agric. Biol. Chem. 54: 1201-1209, 1990.

A microorganism as of the present invention may carry further modifications either on the DNA or protein level (see above), as long as such modification(s) has/have a direct impact on the yield, production and/or efficiency of biomass from a substrate such as *e.g.* sorbitol. Such further modifications may for instance affect other genes encoding transferases as described above or further genes encoding enzymes involved in the pentose phosphate pathway, such as for instance glucose-6-phosphate dehydrogenase, 6-phosphogluconate dehydrogenase, 6-phosphogluconolactonase, ribulose-5-phosphate 3-epimerase or transketolase, or further genes encoding enzymes linking the pentose phosphate pathway to other pathways of carbon metabolism, such as for instance 6-phosphogluconate dehydratase, fructose -1,6-bisphosphate or phosphofructokinase. Methods of performing such modifications are known in the art, with some examples further described herein.

In accordance with a further object of the present invention there is provided the use of a polynucleotide as defined above or a microorganism which is genetically engineered using such polynucleotides in the efficient production of biomass from a carbon source.

The invention also relates to processes for the expression of endogenous genes in a microorganism, to processes for the production of polypeptides as defined above in a microorganism and to processes for the production of microorganisms capable of growth on a given carbon source such as *e.g.* sorbitol. All these processes comprise the step of altering a microorganism, wherein "altering" as used herein encompasses the process for "genetically altering" or "altering the composition of the cell culture media and/or methods used for culturing" in such a way that the yield and/or productivity of biomass can be improved compared to the wild-type organism. As used herein, "improved yield of biomass" means an increase of at least 5%, 10%, 25%, 30%, 40%, 50%, 75%, 100%, 200% or even more than 500%, depending on the carbon source used for growth, compared to a wild-type microorganism, *i.e.* a microorganism which is not genetically altered.

The term "genetically engineered" or "genetically altered" means the scientific alteration of the structure of genetic material in a living organism. It involves the production and use of recombinant DNA. More in particular it is used to delineate the genetically engineered or modified organism from the naturally occurring organism. Genetic engineering may be done by a number of techniques known in the art, such as *e.g.* gene replacement, gene amplification, gene disruption, transfection, transformation using plasmids, viruses, or other vectors. A genetically modified organism, *e.g.* genetically modified microorganism, is also often referred to as a recombinant organism, *e.g.* recombinant microorganism.

Advantageous embodiments of the invention become evident from the dependent claims. These and other aspects and embodiments of the present invention should now be apparent to those skilled in the art based on the teachings herein.

The sequence of the gene comprising a nucleotide sequence according to SEQ ID NO:1 encoding tal was determined by sequencing a genomic clone obtained from *Gluconobacter oxydans* DSM 2343.

The invention also relates to a polynucleotide encoding at least a biologically active fragment or derivative of tal as shown in SEQ ID NO:2.

As used herein, "biologically active fragment or derivative" means a polypeptide which retains essentially the same biological function or activity as the polypeptide shown in SEQ ID NO:2. Examples of biological activity may for instance be enzymatic activity, signaling activity or antibody reactivity. The term "same biological function" or "functional equivalent" as used herein means that the protein has essentially the same

biological activity, *e.g.* enzymatic, signaling or antibody reactivity, as a polypeptide shown in SEQ ID NO:2.

The polypeptides and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

The term "isolated" means that the material is removed from its original environment (*e.g.*, the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or polypeptide present in a living microorganism is not isolated, but the same polynucleotide or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or polypeptides could be part of a composition and still be isolated in that such vector or composition is not part of its natural environment.

An isolated polynucleotide or nucleic acid as used herein may be a DNA or RNA that is not immediately contiguous with both of the coding sequences with which it is immediately contiguous (one on the 5'-end and one on the 3'-end) in the naturally occurring genome of the organism from which it is derived. Thus, in one embodiment, a nucleic acid includes some or all of the 5'-non-coding (*e.g.*, promoter) sequences that are immediately contiguous to the coding sequence. The term "isolated polynucleotide" therefore includes, for example, a recombinant DNA that is incorporated into a vector, into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote, or which exists as a separate molecule (*e.g.*, a cDNA or a genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other sequences. It also includes a recombinant DNA that is part of a hybrid gene encoding an additional polypeptide that is substantially free of cellular material, viral material, or culture medium (when produced by recombinant DNA techniques), or chemical precursors or other chemicals (when chemically synthesized). Moreover, an "isolated nucleic acid fragment" is a nucleic acid fragment that is not naturally occurring as a fragment and would not be found in the natural state.

As used herein, the terms "polynucleotide", "gene" and "recombinant gene" refer to nucleic acid molecules which may be isolated from chromosomal DNA, which include an open reading frame encoding a protein, *e.g. G. oxydans* DSM 2343 transferases. A polynucleotide may include a polynucleotide sequence as shown in SEQ ID NO:1 or fragments thereof and regions upstream and downstream of the gene sequences which may

include, for example, promoter regions, regulator regions and terminator regions important for the appropriate expression and stabilization of the polypeptide derived thereof.

A gene may include coding sequences, non-coding sequences such as for instance untranslated sequences located at the 3'- and 5'-ends of the coding region of a gene, and regulatory sequences. Moreover, a gene refers to an isolated nucleic acid molecule as defined herein. It is furthermore appreciated by the skilled person that DNA sequence polymorphisms that lead to changes in the amino acid sequences of transferases may exist within a population, *e.g.*, the *Gluconobacter oxydans* population. Such genetic polymorphism in the tal gene may exist among individuals within a population due to natural variation or in cells from different populations. Such natural variations can typically result in 1-5% variance in the nucleotide sequence of the tal gene. Any and all such nucleotide variations and the resulting amino acid polymorphism in tal are the result of natural variation and that do not alter the function or biological activity of transferases are intended to be within the scope of the invention.

As used herein, the terms "polynucleotide" or "nucleic acid molecule" are intended to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule may be single-stranded or double-stranded, but preferably is double-stranded DNA. The nucleic acid may be synthesized using oligonucleotide analogs or derivatives (*e.g.*, inosine or phosphorothioate nucleotides). Such oligonucleotides may be used, for example, to prepare nucleic acids that have altered base-pairing abilities or increased resistance to nucleases.

The sequence information as provided herein should not be so narrowly construed as to require inclusion of erroneously identified bases. The specific sequences disclosed herein may be readily used to isolate the complete gene from a microorganism capable of converting a given carbon source such as *e.g.* sorbitol into biomass, in particular *Gluconobacter oxydans*, preferably *Gluconobacter oxydans* DSM 17078 or *Gluconobacter oxydans* DSM 2343 which in turn may easily be subjected to further sequence analyses thereby identifying sequencing errors.

Unless otherwise indicated, all nucleotide sequences determined by sequencing a DNA molecule herein were determined using an automated DNA sequencer and all amino acid sequences of polypeptides encoded by DNA molecules determined herein were predicted by translation of a DNA sequence determined as above. Therefore, as is known in the art

for any DNA sequence determined by this automated approach, any nucleotide sequence determined herein may contain some errors. Nucleotide sequences determined by automation are typically at least about 90% identical, more typically at least about 95% to at least about 99.9% identical to the actual nucleotide sequence of the sequenced DNA molecule. The actual sequence may be more precisely determined by other approaches including manual DNA sequencing methods well known in the art. As is also known in the art, a single insertion or deletion in a determined nucleotide sequence compared to the actual sequence will cause a frame shift in translation of the nucleotide sequence such that the predicted amino acid sequence encoded by a determined nucleotide sequence will be completely different from the amino acid sequence actually encoded by the sequenced DNA molecule, beginning at the point of such an insertion or deletion.

The person skilled in the art is capable of identifying such erroneously identified bases and knows how to correct for such errors.

A nucleic acid molecule according to the invention may comprise only a portion or a fragment of the nucleic acid sequence provided by the present invention, such as for instance the sequence shown in SEQ ID NO:1, for example a fragment which may be used as a probe or primer such as for instance SEQ ID NO:3 or SEQ ID NO:4 or a fragment encoding a portion of a protein according to the invention. The nucleotide sequence determined from the cloning of the tal gene allows for the generation of probes and primers designed for use in identifying and/or cloning other tal family members, as well as tal homologues from other species. The probe/primer typically comprises substantially purified oligonucleotides which typically comprises a region of nucleotide sequence that hybridizes preferably under highly stringent conditions to at least about 12 or 15, preferably about 18 or 20, more preferably about 22 or 25, even more preferably about 30, 35, 40, 45, 50, 55, 60, 65, or 75 or more consecutive nucleotides of a nucleotide sequence shown in SEQ ID NO:1 or a fragment or derivative thereof.

A nucleic acid molecule encompassing all or a portion of the nucleic acid sequence of SEQ ID NO:1 may be isolated by the polymerase chain reaction (PCR) using synthetic oligonucleotide primers designed based upon the sequence information contained herein.

A nucleic acid of the invention may be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid thus amplified may be cloned into an appropriate vector and characterized by DNA sequence analysis.

Fragments of a polynucleotide according to the invention may also comprise polynucleotides not encoding functional polypeptides. Such polynucleotides may function as probes or primers for a PCR reaction.

Nucleic acids according to the invention irrespective of whether they encode functional or non-functional polypeptides, may be used as hybridization probes or polymerase chain reaction (PCR) primers. Uses of the nucleic acid molecules of the present invention that do not encode a polypeptide having a tal activity include, inter alia, (1) isolating the gene encoding the protein of the present invention, or allelic variants thereof from a cDNA library, *e.g.*, from other organisms than *Gluconobacter oxydans* and (2) Northern blot analysis for detecting expression of mRNA of said protein in specific cells or (3) use in abolishing or altering the function or activity of homologous tal genes in said other organisms.

Probes based on the nucleotide sequences provided herein may be used to detect transcripts or genomic sequences encoding the same or homologous proteins for instance in other organisms. Nucleic acid molecules corresponding to natural variants and non-*G. oxydans* homologues of the *G. oxydans* tal DNA of the invention which are also embraced by the present invention may be isolated based on their homology to the *G. oxydans* tal nucleic acid disclosed herein using the *G. oxydans* DNA, or a portion thereof, as a hybridization probe according to standard hybridization techniques, preferably under highly stringent hybridization conditions.

In preferred embodiments, the probe further comprises a label group attached thereto, *e.g.*, the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme cofactor.

Homologous gene sequences may be isolated, for example, by performing PCR using two degenerate oligonucleotide primer pools designed on the basis of nucleotide sequences as taught herein.

The template for the reaction may be cDNA obtained by reverse transcription of mRNA prepared from strains known or suspected to express a polynucleotide according to the invention. The PCR product may be subcloned and sequenced to ensure that the amplified sequences represent the sequences of a new nucleic acid sequence as described herein, or a functional equivalent thereof.

The PCR fragment may then be used to isolate a full length cDNA clone by a variety of known methods. For example, the amplified fragment may be labeled and used to screen a bacteriophage or cosmid cDNA library. Alternatively, the labeled fragment may be used to screen a genomic library.

PCR technology can also be used to isolate full-length cDNA sequences from other organisms. For example, RNA may be isolated, following standard procedures, from an appropriate cellular or tissue source. A reverse transcription reaction may be performed on the RNA using an oligonucleotide primer specific for the most 5'-end of the amplified fragment for the priming of first strand synthesis.

The resulting RNA/DNA hybrid may then be "tailed" (*e.g.*, with guanines) using a standard terminal transferase reaction, the hybrid may be digested with RNaseH, and second strand synthesis may then be primed (*e.g.*, with a poly-C primer). Thus, cDNA sequences upstream of the amplified fragment may easily be isolated. For a review of useful cloning strategies, see *e.g.*, Sambrook et al., supra; and Ausubel et al., supra.

Also, nucleic acids encoding other tal family members, which thus have a nucleotide sequence that differs from a nucleotide sequence according to SEQ ID NO:1, are within the scope of the invention. Moreover, nucleic acids encoding tal proteins from different species which thus may have a nucleotide sequence which differs from a nucleotide sequence shown in SEQ ID NO:1 are within the scope of the invention.

The invention also relates to an isolated polynucleotide hybridisable under stringent conditions, preferably under highly stringent conditions, to a polynucleotide according to the present invention, such as for instance a polynucleotide shown in SEQ ID NO:1. Advantageously, such polynucleotide may be obtained from a microorganism capable of converting a given carbon source such as *e.g.* sorbitol into biomass, in particular *Gluconobacter oxydans*, preferably *Gluconobacter oxydans* DSM 17078 or *Gluconobacter oxydans* DSM 2343.

As used herein, the term "hybridizing" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least about 50%, at least about 60%, at least about 70%, more preferably at least about 80%, even more preferably at least about 85% to 90%, most preferably at least 95% homologous to each other typically remain hybridized to each other.

In one embodiment, a nucleic acid of the invention is at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more homologous to a nucleic acid sequence shown in SEQ ID NO:1 or the complement thereof.

A preferred, non-limiting example of such hybridization conditions are hybridization in 6x sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 1x SSC, 0.1% SDS at 50°C, preferably at 55°C, more preferably at 60°C and even more preferably at 65°C.

Highly stringent conditions include incubations at 42°C for a period of several days, such as 2-4 days, using a labeled DNA probe, such as a digoxigenin (DIG)-labeled DNA probe, followed by one or more washes in 2x SSC, 0.1% SDS at room temperature and one or more washes in 0.5x SSC, 0.1% SDS or 0.1x SSC, 0.1% SDS at 65-68°C. In particular, highly stringent conditions include, for example, 2 h to 4 days incubation at 42°C using a DIG-labeled DNA probe (prepared by *e.g.* using a DIG labeling system; Roche Diagnostics GmbH, 68298 Mannheim, Germany) in a solution such as DigEasyHyb solution (Roche Diagnostics GmbH) with or without 100 µg/ml salmon sperm DNA, or a solution comprising 50% formamide, 5x SSC (150 mM NaCl, 15 mM trisodium citrate), 0.02% sodium dodecyl sulfate, 0.1% N-lauroylsarcosine, and 2% blocking reagent (Roche Diagnostics GmbH), followed by washing the filters twice for 5 to 15 minutes in 2x SSC and 0.1% SDS at room temperature and then washing twice for 15-30 minutes in 0.5x SSC and 0.1% SDS or 0.1x SSC and 0.1% SDS at 65-68°C.

Preferably, an isolated nucleic acid molecule of the invention that hybridizes under preferably highly stringent conditions to a nucleotide sequence of the invention corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein). In one embodiment, the nucleic acid encodes a natural *G. oxydans* tal.

The skilled artisan will know which conditions to apply for stringent and highly stringent hybridization conditions. Additional guidance regarding such conditions is readily available in the art, for example, in Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, N. Y.; and Ausubel et al. (eds.), 1995, Current Protocols in Molecular Biology, (John Wiley & Sons, N. Y.). Of course, a polynucleotide which hybridizes only to a poly (A) sequence (such as the 3'-terminal poly

(A) tract of mRNAs), or to a complementary stretch of T (or U) residues, would not be included in a polynucleotide of the invention used to specifically hybridize to a portion of a nucleic acid of the invention, since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (*e.g.*, practically any double-stranded cDNA clone).

In a typical approach, genomic DNA or cDNA libraries constructed from other organisms, *e.g.* microorganisms capable of converting of a given carbon source such as *e.g.* sorbitol into biomass, in particular other *Gluconobacter* species may be screened.

For example, *Gluconobacter* strains may be screened for homologous polynucleotides by Northern blot analysis. Upon detection of transcripts homologous to polynucleotides according to the invention, DNA libraries may be constructed from RNA isolated from the appropriate strain, utilizing standard techniques well known to those of skill in the art. Alternatively, a total genomic DNA library may be screened using a probe hybridisable to a polynucleotide according to the invention.

A nucleic acid molecule of the present invention, such as for instance a nucleic acid molecule shown in SEQ ID NO:1 or a fragment or derivative thereof, may be isolated using standard molecular biology techniques and the sequence information provided herein. For example, using all or portion of the nucleic acid sequence shown in SEQ ID NO:1 as a hybridization probe, nucleic acid molecules according to the invention may be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook et al., Molecular Cloning : A Laboratory Manual. 2nd, ed. , Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989).

Furthermore, oligonucleotides corresponding to or hybridisable to nucleotide sequences according to the invention may be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

The terms "homology" or "percent identity" are used interchangeably herein. For the purpose of this invention, it is defined here that in order to determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps may be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in

the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (*i.e.*, % identity = number of identical positions/total number of positions (*i.e.*, overlapping positions) x 100). Preferably, the two sequences are the same length.

The skilled person will be aware of the fact that several different computer programs are available to determine the homology between two sequences. For instance, a comparison of sequences and determination of percent identity between two sequences may be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (J. Mol. Biol. (48): 444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at http://www.accelrys.com), using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6 or 4 and a length weight of 1, 2, 3, 4, 5 or 6. The skilled person will appreciate that all these different parameters will yield slightly different results but that the overall percentage identity of two sequences is not significantly altered when using different algorithms.

In yet another embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at http://www.accelrys.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70 or 80 and a length weight of 1, 2, 3, 4, 5 or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Meyers and W. Miller (CABIOS, 4: 11-17 (1989) which has been incorporated into the ALIGN program (version 2.0) (available at http://vega.igh.cnrs.fr/bin/align-guess.cgi) using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention may further be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches may be performed using the BLASTN and BLASTX programs (version 2.0) of Altschul, et al. (1990) J. Mol. Biol. 215:403-10. BLAST nucleotide searches may be performed with the BLASTN program, score = 100, word length = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the present invention. BLAST protein searches may be performed with the BLASTX program, score = 50, word length = 3 to obtain amino acid sequences homologous to the protein molecules of the present invention. To obtain gapped

alignments for comparison purposes, Gapped BLAST may be utilized as described in Altschul et al., (1997) Nucleic Acids Res. 25 (17): 3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (*e.g.*, BLASTX and BLASTN) may be used. See http://www.ncbi.nlm.nih.gov.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is the complement of a nucleotide sequence as of the present invention, such as for instance the sequence shown in SEQ ID NO:1. A nucleic acid molecule, which is complementary to a nucleotide sequence disclosed herein, is one that is sufficiently complementary to a nucleotide sequence shown in SEQ ID NO:1 such that it may hybridize to said nucleotide sequence thereby forming a stable duplex.

In a further preferred embodiment, a nucleic acid of the invention as shown in SEQ ID NO:1 or the complement thereof contains at least one mutation leading to a gene product with modified function/activity. The at least one mutation may be introduced by methods described herein. In one aspect, the at least one mutation leads to tal whose function and/or activity compared to the wild type counterpart is enhanced or improved. Methods for introducing such mutations are well known in the art.

The term "increase" of activity as used herein encompasses increasing activity of one or more polypeptides in the producing organism, which in turn are encoded by the corresponding polynucleotides described herein. There are a number of methods available in the art to accomplish increase of activity of a given protein, in this case tal. In general, the specific activity of a protein may be increased or the copy number of the protein may be increased. The term increase of activity or equivalent expressions also encompasses the situation wherein tal activity is introduced in a cell that did not contain this activity before, *e.g.* by introducing a gene encoding tal in a cell that did not contain an equivalent of this gene before, or that could not express an active form of the corresponding protein before.

To facilitate such an increase, the copy number of the genes corresponding to the polynucleotides described herein may be increased. Alternatively, a strong promoter may be used to direct the expression of the polynucleotide. In another embodiment, the promoter, regulatory region and/or the ribosome binding site upstream of the gene can be altered to increase the expression. The expression may also be enhanced or increased by increasing the relative half-life of the messenger RNA. In another embodiment, the activity of the polypeptide itself may be increased by employing one or more mutations in the polypeptide amino acid sequence, which increases the activity. For example, altering

the affinity of the polypeptide for its corresponding substrate may result in improved activity. Likewise, the relative half-life of the polypeptide may be increased. In either scenario, that being enhanced gene expression or increased activity, the improvement may be achieved by altering the composition of the cell culture media and/or methods used for culturing. "Enhanced expression" or "improved activity" as used herein means an increase of at least 5%, 10%, 25%, 50%, 75%, or even 100%, compared to a wild-type protein, polynucleotide, gene; or the activity and/or the concentration of the protein present before the polynucleotides or polypeptides are enhanced and/or improved. The activity of tal may also be enhanced by contacting the protein with a specific or general enhancer of its activity.

Another aspect of the invention pertains to vectors, containing a nucleic acid encoding a protein according to the invention or a functional equivalent or portion thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments may be ligated. Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication). Other vectors are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome.

Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. The terms "plasmid" and "vector" can be used interchangeably herein as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vector includes one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operatively linked" is intended to mean that the nucleotide sequence of interest is linked to the

regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, attenuator). Such regulatory sequences are described, for example, in Goeddel; Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive or inducible expression of a nucleotide sequence in many types of host cells and those which direct expression of the nucleotide sequence only in a certain host cell (*e.g.* tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention may be introduced into host cells to thereby produce proteins or peptides, encoded by nucleic acids as described herein, including, but not limited to, mutant proteins, fragments thereof, variants or functional equivalents thereof, and fusion proteins, encoded by a nucleic acid as described herein, *e.g.*, tal proteins, mutant forms of tal proteins, fusion proteins and the like.

The recombinant expression vectors of the invention may be designed for expression of tal proteins in a suitable microorganism. For example, a protein according to the invention may be expressed in bacterial cells such as strains belonging to the genera *Gluconobacter, Gluconacetobacter or Acetobacter*, such as *e.g.* in *G. oxydans* DSM 17078 or *G. oxydans* DSM 2343. Expression vectors useful in the present invention include chromosomal-, episomal- and virus-derived vectors *e.g.*, vectors derived from bacterial plasmids, bacteriophage, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids.

The DNA insert may be operatively linked to an appropriate promoter, which may be either a constitutive or inducible promoter. The skilled person will know how to select suitable promoters. The expression constructs may contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the mature transcripts expressed by the constructs may preferably include an initiation codon at the beginning and a termination codon appropriately positioned at the end of the polypeptide to be translated.

Vector DNA may be introduced into suitable host cells via conventional transformation or transfection techniques. As used herein, the terms "transformation", "transconjugation"

and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, transduction, infection, lipofection, cationic lipidmediated transfection or electroporation. Suitable methods for transforming or transfecting host cells may be found in Sambrook, et al. (supra), Davis et al., Basic Methods in Molecular Biology (1986) and other laboratory manuals.

In order to identify and select cells which have integrated the foreign DNA into their genome, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those that confer resistance to drugs, such as kanamycin, tetracycline, ampicillin and streptomycin. A nucleic acid encoding a selectable marker is preferably introduced into a host cell on the same vector as that encoding a protein according to the invention or can be introduced on a separate vector such as, for example, a suicide vector, which cannot replicate in the host cells. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

The invention provides also an isolated polypeptide having the amino acid sequence shown in SEQ ID NO:2 or an amino acid sequence obtainable by expressing a polynucleotide of the present invention, such as for instance a polynucleotide sequence shown in SEQ ID NO:1 in an appropriate host.

Polypeptides according to the invention may contain only conservative substitutions of one or more amino acids in the amino acid sequence represented by SEQ ID NO:2 or substitutions, insertions or deletions of non-essential amino acids. Accordingly, a non-essential amino acid is a residue that may be altered in the amino acid sequences shown in SEQ ID NO:2 without substantially altering the biological function. For example, amino acid residues that are conserved among the proteins of the present invention, are predicted to be particularly unamenable to alteration. Furthermore, amino acids conserved among the proteins according to the present invention and other tal proteins are not likely to be amenable to alteration.

The term "conservative substitution" is intended to mean that a substitution in which the amino acid residue is replaced with an amino acid residue having a similar side chain. These families are known in the art and include amino acids with basic side chains (*e.g.*,

lysine, arginine and histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), non-polar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine).

As mentioned above, the polynucleotides of the invention may be utilized in the genetic engineering of a suitable host cell to make it better and more efficient in the bioconversion of a carbon source such as *e.g.* sorbitol to biomass, *i.e.* reduction of carbon source diversion throughout the bioconversion process.

According to the invention a genetically engineered/recombinant host cell (also referred to as recombinant cell or transformed cell) may be produced carrying such a modified polynucleotide wherein the function of the linked protein is significantly modified in comparison to a wild-type cell such that the yield and/or productivity of biomass from a carbon source such as *e.g.* sorbitol is improved. The host cell may be selected from a microorganism capable of converting a given carbon source such as *e.g.* sorbitol into biomass, in particular *Gluconobacter oxydans*, preferably *G. oxydans* DSM 17078 or *G. oxydans* DSM 2343.

A "transformed cell" or "recombinant cell" is a cell into which (or into an ancestor of which) has been introduced, by means of recombinant DNA techniques, a nucleic acid according to the invention, or wherein the activity of tal has been increased and/or enhanced. Suitable host cells include cells of microorganisms capable of converting a given carbon source such as *e.g.* sorbitol into biomass. In particular, these include strains from the genera *Pseudomonas, Pantoea, Escherichia, Corynebacterium, Ketogulonicigenium* and acetic acid bacteria like *e.g.*, *Gluconobacter*, *Acetobacter* or *Gluconacetobacter*, preferably *Acetobacter* sp., *Acetobacter aceti*, *Gluconobacter frateurii, Gluconobacter cerinus, Gluconobacter thailandicus, Gluconobacter oxydans*, more preferably *G. oxydans*, most preferably selected from *G. oxydans* DSM 17078, *G. oxydans* DSM 2343 or *G. oxydans* IFO 3293.

Improved gene expression may also be achieved by modifying the tal gene, *e.g.*, by introducing one or more mutations into the tal gene wherein said modification leads to tal with a function which is significantly improved in comparison to the wild-type protein.

Therefore, in one other embodiment, the polynucleotide carrying the at least one mutation is derived from a polynucleotide as represented by SEQ ID NO:1 or equivalents thereof.

A mutation as used herein may be any mutation leading to a more functional or more stable polypeptide, *e.g.* more functional or more stable tal gene products. This may include for instance an alteration in the genome of a microorganism, which improves the synthesis of tal or leads to the expression of tal with an altered amino acid sequence whose function compared with the wild type counterpart having a non-altered amino acid sequence is improved and/or enhanced. The improvement may occur at the transcriptional, translational or post-translational level.

The alteration in the genome of the microorganism may be obtained *e.g.* by replacing through a single or double crossover recombination a wild type DNA sequence by a DNA sequence containing the alteration. For convenient selection of transformants of the microorganism with the alteration in its genome the alteration may, *e.g.* be a DNA sequence encoding an antibiotic resistance marker or a gene complementing a possible auxotrophy of the microorganism. Mutations include, but are not limited to, deletion-insertion mutations.

An alteration in the genome of the microorganism leading to a more functional polypeptide may also be obtained by randomly mutagenizing the genome of the microorganism using *e.g.* chemical mutagens, radiation or transposons and selecting or screening for mutants which are better or more efficient producers of biomass. Standard methods for screening and selection are known to the skilled person.

In a specific embodiment, it is desired to knockout or suppress a repressor of the tal gene of the present invention, *i.e.*, wherein its repressor gene expression is artificially suppressed in order to improve the yield, productivity, and/or efficiency of production of the fermentation product when introduced into a suitable host cell. Methods of providing knockouts as well as microorganisms carrying such suppressed genes are well known in the art. The suppression of the repressor gene may be induced by deleting at least a part of the repressor gene or the regulatory region thereof. As used herein, "suppression of the gene expression" includes complete and partial suppression, as well as suppression under specific conditions and also suppression of the expression of either one of the two alleles.

The aforementioned mutagenesis strategies for tal proteins may result in increased yield and/or production of biomass. This list is not meant to be limiting; variations on these mutagenesis strategies will be readily apparent to one of ordinary skill in the art. By these mechanisms, the nucleic acid and protein molecules of the invention may be utilized to generate microorganisms such as *Gluconobacter oxydans* or related strains of bacteria

expressing mutated tal nucleic acid and protein molecules such that the yield, efficiency and/or productivity of biomass production from a carbon source such as *e.g.* sorbitol is improved.

Biomass concentration can be measured using several methods known to the person skilled in the art, such as *e.g.* measuring the optical density of the respective cell suspensions at for instance a wavelength of 600 nm (OD_{600}), or measuring either the wet- or dry-cell mass concentration or by counting the numbers of cells of the respective cell suspension. Such methods of measurement are known in the art. Thus, the cell dry weight (CDW) [g/l] may be for instance measured on a dried and tared nitrocellulose filter with a pore size of *e.g.* 0.45 μm on which the culture broth is filtrated. After washing the filter with water, the weight of the dried filter is again determined and the CDW calculated as follows:

CDW = m(dried filter after broth filtration) – m(dried filter before broth filtration)

In one aspect of the invention, microorganisms (in particular from the genera of *Gluconobacter*, *Gluconacetobacter* and *Acetobacter*) are provided that are able to perform such improved bioconversion. When measured by a method as described herein these organisms were found to have an improved capability to produce biomass from a carbon source such as *e.g.* sorbitol. Such may be achieved by increasing the activity of a transferase, preferably tal. The yield of biomass produced from a carbon source, such as for instance sorbitol, when measured according to this method after an incubation period of 60-72 hours may be about 0.1 g or more dry biomass/g carbon source, preferably about 0.2 g or more, or even about 0.3 g or more dry biomass/g carbon source, such as for instance about 0.4 g, 0.5 g, 0.6 g, 0.7 g, 0.8 g or 1.0 g or more dry biomass/g carbon source.

The recombinant microorganism carrying *e.g.* a modified tal gene and which is able to produce biomass from a carbon source such as *e.g.* sorbitol in significantly higher yield, productivity, and/or efficiency may be cultured in an aqueous medium supplemented with appropriate nutrients under aerobic conditions. The cultivation may be conducted in batch, fed-batch, semi-continuous or continuous mode. The cultivation period may vary depending on for instance the host, pH, temperature and nutrient medium to be used, and is preferably about 1 to about 10 days when run in batch or fed-batch mode. The cultivation may be conducted at for instance a pH of about 4.0 to about 9.0, preferably about 5.0 to about 8.0. The preferred temperature range for carrying out the cultivation is from about 13°C to about 36°C, preferably from about 18°C to about 33°C. Usually, the culture

medium may contain besides *e.g.* sorbitol as main carbon source other assimilable carbon sources, *e.g.*, glucose, glycerol, mannitol, sorbose, erythritol, ribitol, xylitol, arabitol, inositol, dulcitol, ribose, fructose, sucrose and ethanol, preferably glucose, mannitol, fructose, glycerol and ethanol; and digestible nitrogen sources such as organic substances, *e.g.*, peptone, yeast extract, baker's yeast, urea, amino acids, and corn steep liquor. Various inorganic substances may also be used as nitrogen sources, *e.g.*, nitrates and ammonium salts. Furthermore, the culture/incubation medium usually may contain inorganic salts, *e.g.*, magnesium sulfate, manganese sulfate, potassium phosphate, and calcium carbonate.

The nucleic acid molecules, polypeptides, vectors, primers, and recombinant microorganisms described herein may be used in one or more of the following methods: identification of *Gluconobacter oxydans* and related organisms; mapping of genomes of organisms related to *Gluconobacter oxydans*; identification and localization of *Gluconobacter oxydans* sequences of interest; evolutionary studies; determination of tal protein regions required for function; modulation of tal activity or function; modulation of the activity of transferase pathway; and modulation of cellular production of biomass from a carbon source such as *e.g.* sorbitol.

The invention provides methods for screening molecules which modulate the activity of tal, either by interacting with the protein itself or a substrate or binding partner of tal, or by modulating the transcription or translation of a tal nucleic acid molecule of the invention. In such methods, a microorganism expressing one or more tal proteins of the invention is contacted with one or more test compounds, and the effect of each test compound on the activity or level of expression of tal is assessed.

The biological, enzymatic or other activity of transferases can be measured by methods well known to a skilled person, such as, for example, by incubating a cell fraction containing tal in the presence of its substrate, fructose-6-phosphate and erythrose-4-phosphate, whereby the substrate-dependent decrease of NADH may be measured spectrophotometrically at 340 nm according to the method described by *e.g.* Sugiyama et al., Biosci. Biotechnol. Biochem., 67 (12), 2524-2532, 2003.

Thus, the present invention is directed to the use of a polynucleotide, polypeptide, vector, primer and recombinant microorganism as described herein in the production of biomass, *i.e.*, the bioconversion of a carbon source such as *e.g.* sorbitol into biomass. In a preferred embodiment, a modified polynucleotide, polypeptide, vector and recombinant

microorganism as described herein is used for improving the yield, productivity, and/or efficiency of said biomass production.

The terms "production" or "productivity" are art-recognized and include the amount of biomass formed within a given time and a given cultivation volume (*e.g.*, kg product per hour per liter) from a given amount or concentration of a carbon source such as *e.g.* sorbitol. The term "efficiency of production" includes the time required for a particular level of production to be achieved (for example, how long it takes for the cell to attain a particular rate of output of a product). The term "yield" is art-recognized and includes the efficiency of the conversion of the carbon source into biomass. This is generally written as, for example, kg biomass per kg carbon source. By "increasing the yield and/or production of biomass" it is meant that in a given amount of culture over a given amount of time the biomass is increased. By measuring the increase in biomass- and decrease in carbon source concentration in a growing culture, one can calculate the yield of biomass on consumed carbon source; *i.e.* the weight of biomass obtained by bioconverting a determined weight of carbon source such as *e.g.* sorbitol, defined as kg biomass per kg carbon source consumed. To determine the consumption of carbon source, the residual amount of carbon source in the growth medium can be measured by methods well known in the art (*e.g.* HPLC). The increase of biomass can be measured by methods well known in the art (*e.g.* optical density at 600 nm, or measurement of dry weight of a sample).

The terms "biosynthesis" or a "biosynthetic pathway" are art-recognized and include the synthesis of a compound/product, preferably an organic compound, by a cell from intermediate compounds in what may be a multistep and highly regulated process. The term "bioconversion" is art-recognized and includes the conversion of a carbon source such as *e.g.* sorbitol into a product, *i.e.* biomass, by means of one or more biosynthetic step(s) which involve one or more enzyme(s) and/or transporter(s). The language "metabolism" is art-recognized and includes the totality of the biochemical reactions that take place in an organism. The metabolism of a particular compound, then, (*e.g.*, the metabolism of an amino acid such as glycine) comprises the overall biosynthetic, modification, and degradation pathways in the cell related to this compound. The language "transport" or "import" is art-recognized and includes the facilitated movement of one or more molecules across a cellular membrane through which the molecule would otherwise either be unable to pass or be passed inefficiently.

In one preferred embodiment, the present invention is related to a process for the production of biomass from a carbon source such as *e.g.* sorbitol wherein a modified

polynucleotide sequence as described above is introduced into a suitable microorganism and the recombinant microorganism is cultured under conditions that allow the production of biomass in high productivity, yield, and/or efficiency.

Recombinant microorganisms according to the present invention carrying *e.g.* a modified transferase gene, in particular a modified tal gene, and which are able to produce biomass from a carbon source such as *e.g.* sorbitol in significantly higher yield, productivity, and/or efficiency may advantageously be used in a number of applications. One particularly suited application is a method for the production of Vitamin C and/or intermediates thereof such like 2-keto-L-gulonic acid (2-KGA).

Thus, it is an aspect of the present invention to provide a process for the production of Vitamin C and/or 2-KGA wherein a nucleotide according to the invention or a modified polynucleotide sequence as described above is introduced into a suitable microorganism as described above wherein a suitable carbon source such as *e.g.* sorbitol is efficiently converted into biomass, the recombinant microorganism is cultured under conditions that allow the production of Vitamin C and/or 2-KGA in high productivity, yield, and/or efficiency, the produced fermentation product is isolated from the culture medium and optionally further purified.

The carbon sources or mixtures thereof used for the production of (1) biomass and (2) Vitamin C and/or 2-KGA may be the same or may differ. In one preferred embodiment, the carbon source(s) used for the production of biomass is different from the carbon source(s) used for the direct production of Vitamin C and/or 2-KGA. These two different carbon sources may be used simultaneously or sequentially, wherein a first carbon source would be used for the production of biomass, and this biomass would then be used to convert a second carbon source into Vitamin C and/or 2-KGA.

A suitable carbon source that can be converted directly into Vitamin C and/or 2-KGA may be selected from the D-glucose or D-sorbitol metabolization pathway such as, for example, D-glucose, D-sorbitol, L-sorbose, L-sorbosone, D-gluconate, 2-keto-D-gluconate, 2,5-diketo-gluconate or mixtures thereof. Preferably, the substrate is selected from for instance D-glucose, D-sorbitol, L-sorbose, L-sorbosone or mixtures thereof, more preferably from D-glucose, D-sorbitol, L-sorbose or mixtures thereof, and most preferably from D-sorbitol, L-sorbose or mixtures thereof.

Vitamin C as used herein may be any chemical form of L-ascorbic acid found in aqueous solutions, such as for instance undissociated, in its free acid form or dissociated as an

anion. The solubilized salt form of L-ascorbic acid may be characterized as the anion in the presence of any kind of cations usually found in fermentation supernatants, such as for instance potassium, sodium, ammonium, or calcium. Also included may be isolated crystals of the free acid form of L-ascorbic acid. On the other hand, isolated crystals of a salt form of L-ascorbic acid are called by their corresponding salt name, *i.e.* sodium ascorbate, potassium ascorbate, calcium ascorbate and the like.

Conversion of a suitable carbon source into Vitamin C and/or 2-KGA in connection with the above process using a microorganism means that the conversion of the carbon source resulting in Vitamin C and/or 2-KGA is performed by the microorganism, *i.e.* the carbon source may be directly converted into Vitamin C and/or 2-KGA. Said microorganism is cultured under conditions which allow such conversion from said carbon source as it is known for the skilled person.

A medium as used herein for the above process using a microorganism may be any suitable medium for the production of Vitamin C and/or 2-KGA. Typically, the medium is an aqueous medium comprising for instance salts, substrate(s), and a certain pH.

The term "direct conversion" and the like is intended to mean that a microorganism is capable of the conversion of a certain carbon source into the specified product by means of one or more biological conversion steps, without the need of any additional chemical conversion step. For instance, the term "direct conversion of D-sorbitol into Vitamin C" is intended to describe a process wherein a microorganism is producing Vitamin C and wherein D-sorbitol is offered as a carbon source without the need of an intermediate chemical conversion step. A single microorganism capable of directly fermenting Vitamin C is preferred.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patent applications, patents and published patent applications, cited throughout this application are hereby incorporated by reference.

Examples

Example 1: Preparation of chromosomal DNA and amplification of DNA fragment by PCR

Chromosomal DNA of *Gluconobacter oxydans* DSM 2343 was prepared from the cells cultivated at 30°C for 1 day in VM liquid medium consisting of 5 g/l yeast extract, 3 g/l tryptone, 10 g/l mannitol, pH 6.0 using the DNeasy Tissue Kit (Qiagen, Hilden, Germany).

A DNA fragment is prepared by PCR with the chromosomal DNA prepared above and a set of primers, Pf (SEQ ID NO:3) and Pr (SEQ ID NO:4). For the reaction, the Herculase PCR kit (Stratagene, Heidelberg, Germany) and 20 ng of the chromosomal DNA is used in total volume of 50 µl according to the supplier's instruction to have the PCR product containing the tal DNA sequence (SEQ ID NO:1). The PCR product is recovered from the reaction and its correct sequence confirmed.

Furthermore, a DNA fragment carrying both genes of tal and pgi fused together was prepared according to the method described above using the PCR primers P1 (SEQ ID NO:11) and P2 (SEQ ID NO:12) and chromosomal DNA of *Gluconobacter oxydans* DSM 2343 as template. PCR was performed as described above using the Herculase PCR kit (Stratagene, Heidelberg, Germany) and 20 ng of the chromosomal DNA resulting in a PCR product containing the pgi/tal gene (SEQ ID NO:9).

Example 2: Overexpression of the tal gene in *G. oxydans* DSM 17078 and *G. oxydans* IFO 3293

The PCR product obtained in Example 1 containing the pgi/tal gene according to SEQ ID NO:9 was phosphorylated and cloned downstream of the strong *G. oxydans tufB* promoter into the dephosphorylated vector pEXGOX (SEQ ID NO:13) using *Swa*I restriction site, resulting in the vector pEXGOX-pgi/tal-fw (SEQ ID NO:14). Identity and correct orientation of the pgi/tal insert was confirmed by DNA sequence analysis.

Using *Kpn*I and *Spe*I as restriction sites, the vector fragment including *tufB* promoter and tal gene was cut out of pEXGOX-pgi/tal-fw and introduced into pBBR1MCS2 (Kovach et al., Gene 166:175-176, 1995) resulting in pBBR1MCS2-pgi/tal-fw (SEQ ID NO:15). The plasmid was used for transformation of *G. oxydans* DSM 17078 and *G. oxydans* IFO 3293 resulting in *G. oxydans* DSM 17078 - pBBR1MCS2-pgi/tal-fw and *G. oxydans* IFO 3293 - pBBR1MCS2-pgi/tal-fw, respectively.

Overexpression of tal was confirmed by determination of the transaldolase activity in *G. oxydans* DSM 17078 - pBBR1MCS2-pgi/tal-fw and *G. oxydans* IFO 3293 - pBBR1MCS2-pgi/tal-fw compared to *G. oxydans* DSM 17078 and *G. oxydans* IFO 3293 with or without plasmid pBBR1MCS2, respectively. Individual colonies of *G. oxydans* were cultivated in 50 ml medium containing 80 g/l D-sorbitol, 20 g/l L-sorbose, 1.5 g/l $CaCl_2$, 15 g/l yeast, 2.5 g/l $MgSO_4$, 0.5 g/l glycerol, 50 µg/l cefoxitine and 50 µg/l kanamycine. The cells were grown at 30°C and harvested between OD 2.5 and 3.5, *i.e.* during mid-exponential growth phase. Activity of tal was measured in a 1 ml sample containing 100 µl 0.5 M potassium phosphate buffer at pH 7 (Roche Diagnostics, Mannheim, Germany), 75 µl 5 mM NADH, 20 µl (100 U/ml) triose phosphate isomerase, 20 µl (100 U/ml) glycerol-3-phosphate dehydrogenase, 50 µl (100 mM) erythrose-4-phosphate and 50 µl of cell-free extract from culture medium at OD 2.5-3.5. The reaction was started by addition of 50 µl (0.1 M) fructose-6-phosphate. A decrease of NADH formation was determined at 340 nm indicating overexpression of tal.

In comparison to *G. oxydans* DSM 17078 and *G. oxydans* DSM 17078- pBBR1MCS2 measurement of tal activity from *G. oxydans* DSM 17078 - pBBR1MCS2-pgi/tal-fw results in an increase of ca. 3 fold. When using *G. oxydans* IFO 3293 as host system, the increase in tal activity is also in the range of 3 fold when using *G. oxydans* IFO 3293 - pBBR1MCS2-pgi/tal-fw compared to the control.

Strains containing the tal gene alone, *i.e. G. oxydans* DSM 17078 - pBBR1MCS2-tal-fw and *G. oxydans* IFO 3293 - pBBR1MCS2-tal-fw, respectively, are prepared accordingly and tested for tal overexpression against *G. oxydans* DSM 17078, *G. oxydans* DSM 17078 - pBBR1MCS2, *G. oxydans* IFO 3293 and *G. oxydans* IFO 3293 - pBBR1MCS2, respectively. As a result, an increase of at least 3 fold is detected.

Example 3: Biomass production upon overexpression of tal in *G. oxydans* DSM 17078 and *G. oxydans* IFO 3293 in liquid culture

Biomass formation from D-sorbitol was tested in shake flasks experiments. Growth of *G. oxydans* DSM 17078 - pBBR1MCS2 and *G. oxydans* IFO 3293 - pBBR1MCS2, respectively, was compared to a strain overexpressing tal according to Examples 1 and 2. Biomass formation was determined via measuring the OD. The following medium was used: 15 g/l yeast extract, 0.5 g/l glycerol, 2.5 g/l $MgSO_4$, 1.5g/l $CaCl_2$, 80g/l D-sorbitol, 20g/l L-sorbose, 50 µg/l cefoxitine, 50 µg/l kanamycine. 50 ml medium in 500ml non-

beveled Erlenmeyer flasks were inoculated to an OD of 0.2 from a 24h pre-culture in the same medium. The culture was incubated at 30°C at 200 rpm shaking.

The OD after 68 h improved from 6.01 for DSM 17078 - pBBR1MCS2 to 9.51 for DSM 17078 - pBBR1MCS2-pgi/tal-fw, corresponding to an increase of 58% in biomass formation. *G. oxydans* IFO 3293 as host, the OD after 68 h improved from 5.12 for *G. oxydans* IFO 3293 - pBBR1MCS2 to 7.18 for *G. oxydans* IFO 3293 - pBBR1MCS2-pgi/tal-fw, corresponding to an increase of 40% in biomass formation.

Improvement of yield and productivity of biomass formation in liquid culture by overexpression of tal in *G. oxydans* DSM 17078 was confirmed by comparing growth on D-sorbitol for *G. oxydans* DSM 17078 - pBBR1MCS2-pgi/tal-fw and *G. oxydans* DSM 17078 - pBBR1MCS2 in a fermenter. Following medium was used: 15 g/l yeast; 2,5 g/l $MgSO_4$; 0.5 g/l glycerol; 100 g/l D-sorbitol; 1.5 g/l $CaCl_2$; 50 µg/l cefoxitine; 50 µg/l kanamycine; one drop antifoam. Fermentations conditions were: 200 ml medium per fermenter, N = 500 rpm, pO_2 = 20%, aeration rate: 15 sl/h, X_{CO2} n the air: 0,0 %, pH = 6.0, inocculation from seed to OD=0.5. Two independent fermentation runs were performed from two different seeds for both *G. oxydans* DSM 17078 - pBBR1MCS2 and *G. oxydans* DSM 17078 - pBBR1MCS2-tal-fw. Growth of biomass was followed over 102 h by determination of the optical density at 600 nm (OD_{600}).

Overexpression of tal in *G. oxydans* DSM 17078 results in an improved productivity as well in an increase of the yield of biomass formation of *G. oxydans* DSM 17078 from D-sorbitol by ca. 47%: The OD_{600} after 102 hours was 33.0 for *G. oxydans* DSM 17078 - pBBR1MCS2-pgi/tal-fw compared to 22.5 for *G. oxydans* DSM 17078 - pBBR1MCS2. In all fermentations for *G. oxydans* DSM 17078 - pBBR1MCS2-pgi/tal-fw and DSM 17078 - pBBR1MCS2, the carbon source D-sorbitol was completely consumed after 102 hours.

An increase of about 40, 50, 60% in yield of biomass production from D-sorbitol is also obtained in experiments according to the method described above using *G. oxydans* DSM 17078 - pBBR1MCS2-tal-fw compared to strain *G. oxydans* DSM 17078 - pBBR1MCS2.

Example 4: Vitamin C production upon overexpression of tal in *G. oxydans* DSM 17078 using resting cells

Cells of *G. oxydans* DSM 17078, *G. oxydans DSM 17078* - pBBR1MCS2-tal-fw are grown at 27°C for 3 days on No. 3BD agar medium containing 70 g/l D-sorbitol, 0.5 g/l

glycerol, 7.5 g/l yeast extract (Difco), 2.5 g/l $MgSO_4 \cdot 7H_2O$, 10 g/l $CaCO_3$ and 18 g/l agar (Difco).

Cells are scraped from the agar plates, suspended in distilled water and used for resting cell reactions conducted at 30°C with shaking at 220 rpm. A series of reactions (0.5 ml reaction mixture in 5 ml reaction tubes) are carried out with 2% D-sorbitol in reaction mixtures further containing 0.3% NaCl, and 1% $CaCO_3$ and is incubated with cells at a final concentration of $OD_{600} = 10$. After an incubation period of 20 hours, samples of the reaction mixtures are analyzed by high performance liquid chromatography (HPLC) using an Agilent 1100 HPLC system (Agilent Technologies, Wilmington, USA) with a LiChrospher-100-RP18 (125 x 4.6 mm) column (Merck, Darmstadt, Germany) attached to an Aminex-HPX-78H (300 x 7.8 mm) column (Biorad, Reinach, Switzerland). The mobile phase is 0.004 M sulfuric acid, and the flow rate was 0.6 ml/min. Two signals are recorded using an UV detector (wavelength 254 nm) in combination with a refractive index detector. In addition, the identification of the L-ascorbic acid is done using an amino-column (YMC-Pack Polyamine-II, YMC, Inc., Kyoto, Japan) with UV detection at 254 nm. The mobile phase is 50 mM $NH_4H_2PO_4$ and acetonitrile (40:60).

An Agilent Series 1100 HPLC-mass spectrometry (MS) system is used to identify L-ascorbic acid. The MS is operated in positive ion mode using the electrospray interface. The separation is carried out using a LUNA-C8(2) column (100 x 4.6 mm) (Phenomenex, Torrance, USA). The mobile phase is a mixture of 0.1% formic acid and methanol (96:4). L-Ascorbic acid elutes with a retention time of 3.1 minutes. The identity of the L-ascorbic acid is confirmed by retention time and the molecular mass of the compound.

The supernatants of the reaction mixtures incubated with cells of *G. oxydans* DSM 17078 - pBBR1MCS2-tal-fw contain at least 20% more Vitamin C than the supernatant of the reaction mixture incubated with cells of *G. oxydans* DSM 17078.

Claims

1. A process for the production of biomass from a carbon source wherein a microorganism which is cultivated in an aqueous nutrient medium under conditions that allow the growth on said carbon source, said microorganism being genetically engineered by a polynucleotide selected from the group consisting of:

(a) polynucleotides encoding a polypeptide comprising the amino acid sequence according to SEQ ID NO:2;

(b) polynucleotides comprising the nucleotide sequence according to SEQ ID NO:1;

(c) polynucleotides comprising a nucleotide sequence obtainable by nucleic acid amplification such as polymerase chain reaction, using genomic DNA from a microorganism as a template and a primer set according to SEQ ID NO:3 and SEQ ID NO:4;

(d) polynucleotides comprising a nucleotide sequence encoding a fragment or derivative of a polypeptide encoded by a polynucleotide of any of (a) to (c) wherein in said derivative one or more amino acid residues are conservatively substituted compared to said polypeptide, and said fragment or derivative has the activity of a transferase [EC 2], preferably a transferase transferring aldehyde or ketonic groups [EC 2.2], most preferably tal;

(e) polynucleotides the complementary strand of which hybridizes under stringent conditions to a polynucleotide as defined in any one of (a) to (d) and which encode a transferase [EC 2], preferably a transferase transferring aldehyde or ketonic groups [EC 2.2], most preferably tal;

(f) polynucleotides which are at least 70%, such as 85, 90 or 95% identical to a polynucleotide as defined in any one of (a) to (d) and which encode a transferase [EC 2], preferably a transferase transferring aldehyde or ketonic groups [EC 2.2], most preferably tal;

or

the complementary strand of such a polynucleotide.

2. Process for the production of a microorganism capable of growth on a given carbon source, comprising the step of altering said microorganism so that the microorganism produces a polypeptide with increased and/or improved transferase [EC 2], preferably transferase transferring aldehyde or ketonic groups [EC 2.2], most preferably tal activity leading to an improved yield and/or production of biomass from said carbon source by said microorganism.

3. Process for the production of a microorganism containing an endogenous gene comprising a polynucleotide as described in claim 1, comprising the step of altering said microorganism so that the endogenous gene is overexpressed, leading to an improved yield and/or production of biomass from a carbon source produced by said microorganism.

4. A microorganism genetically engineered with a polynucleotide selected from the group consisting of:

(a) polynucleotides encoding a polypeptide comprising the amino acid sequence according to SEQ ID NO:2;

(b) polynucleotides comprising the nucleotide sequence according to SEQ ID NO:1;

(c) polynucleotides comprising a nucleotide sequence obtainable by nucleic acid amplification such as polymerase chain reaction, using genomic DNA from a microorganism as a template and a primer set according to SEQ ID NO:3 and SEQ ID NO:4;

(d) polynucleotides comprising a nucleotide sequence encoding a fragment or derivative of a polypeptide encoded by a polynucleotide of any of (a) to (c) wherein in said derivative one or more amino acid residues are conservatively substituted compared to said polypeptide, and said fragment or derivative has the activity of a transferase [EC 2], preferably a transferase transferring aldehyde or ketonic groups [EC 2.2], most preferably tal;

(e) polynucleotides the complementary strand of which hybridizes under stringent conditions to a polynucleotide as defined in any one of (a) to (d) and which encode a transferase [EC 2], preferably a transferase transferring aldehyde or ketonic groups [EC 2.2], most preferably tal;

(f) polynucleotides which are at least 70%, such as 85, 90 or 95% identical to a polynucleotide as defined in any one of (a) to (d) and which encode a transferase [EC 2],

preferably a transferase transferring aldehyde or ketonic groups [EC 2.2], most preferably tal;

or

the complementary strand of such a polynucleotide.

5. The microorganism according to claim 4 wherein said polynucleotide sequence is present in form of a vector.

6. The microorganism according to claim 4, said microorganism being genetically altered in such a way that it leads to an improved yield and/or efficiency of biomass from a carbon source produced by said microorganism.

7. The microorganism according to any one of claims 4 to 6 capable of producing biomass from sorbitol in a yield of at least about 0.5 g biomass/g sorbitol.

8. The microorganism according to claim 7 wherein the yield is at least about 0.2 g biomass/g sorbitol.

9. A polynucleotide selected from the group consisting of:

(a) polynucleotides encoding a polypeptide comprising the amino acid sequence according to SEQ ID NO:2;

(b) polynucleotides comprising the nucleotide sequence according to SEQ ID NO:1;

(c) polynucleotides comprising a nucleotide sequence obtainable by nucleic acid amplification such as polymerase chain reaction, using genomic DNA from a microorganism as a template and a primer set according to SEQ ID NO:3 and SEQ ID NO:4;

(d) polynucleotides comprising a nucleotide sequence encoding a fragment or derivative of a polypeptide encoded by a polynucleotide of any of (a) to (c) wherein in said derivative one or more amino acid residues are conservatively substituted compared to said polypeptide, and said fragment or derivative has the activity of a transferase [EC 2], preferably a transferase transferring aldehyde or ketonic groups [EC 2.2], most preferably tal;

(e) polynucleotides the complementary strand of which hybridizes under stringent conditions to a polynucleotide as defined in any one of (a) to (d) and which encode a transferase [EC 2], preferably a transferase transferring aldehyde or ketonic groups [EC 2.2], most preferably tal;

(f) polynucleotides which are at least 70%, such as 85, 90 or 95% identical to a polynucleotide as defined in any one of (a) to (d) and which encode a transferase [EC 2], preferably a transferase transferring aldehyde or ketonic groups [EC 2.2], most preferably tal;

or

the complementary strand of such a polynucleotide.

10. A vector containing the polynucleotide according to claim 9.

11. A polypeptide encoded by a polynucleotide according to claim 9.

12. A process for producing cells capable of expressing a polypeptide according to claim 11, comprising the step of genetically engineering cells with the vector of claim 10 or with a polynucleotide according to claim 9.

13. Use of a polynucleotide according to claim 9 or a vector according to claim 10 for the production of biomass from a carbon source.

14. Use of a polynucleotide according to claim 9 or a vector according to claim 10 for the production of Vitamin C from a carbon source.

15. Use according to claim 13 or 14, wherein the polynucleotide is operatively linked to expression control sequences and transferred into a microorganism.

16. Use according to claim 15, wherein the expression control sequences comprise a regulation-, and/or promoter-, and/or terminator sequence and wherein at least one of these sequences is altered in such a way that it leads to an improved yield and/or production of biomass from a carbon source produced by said microorganism.

17. Use according to claim 15, wherein the expression control sequences comprise a regulation-, and/or promoter-, and/or terminator sequence and wherein at least one of these sequences is altered in such a way that it leads to an improved yield and/or production of Vitamin C from a carbon source produced by said microorganism.

18. Use according to claim 16 or 17, wherein the expression control sequences comprise a regulation-, and/or promoter-, and/or terminator sequence and wherein at least one of these sequences is altered in such a way that it leads to an increased and/or improved activity of a transferase [EC 2], preferably a transferase transferring aldehyde or ketonic groups [EC 2.2], most preferably tal.

19. The microorganism according to any one of claims 4 to 8 producing a polypeptide according to claim 11 with increased and/or improved transferase [EC 2], preferably transferase transferring aldehyde or ketonic groups [EC 2.2], most preferably tal activity.

20. The microorganism according to claim 19 wherein the polynucleotide according to claim 9 is overexpressed.

21. The microorganism according to any one of claims 4 to 8, 19 or 20 selected from the group consisting of *Pseudomonas, Pantoea, Escherichia, Corynebacterium, Ketogulonicigenium* and acetic acid bacteria like *e.g.*, *Gluconobacter*, *Acetobacter* or *Gluconacetobacter*, preferably *Acetobacter* sp., *Acetobacter aceti*, *Gluconobacter frateurii, Gluconobacter cerinus, Gluconobacter thailandicus, Gluconobacter oxydans*, preferably *Gluconobacter oxydans*, more preferably *Gluconobacter oxydans* DSM 17078.

22. Process for the production of an enhanced endogenous transferase [EC 2], preferably transferase transferring aldehyde or ketonic groups [EC 2.2], most preferably tal gene in a microorganism, said microorganism comprising a polynucleotide according to claim 9, said process comprising the step of altering said polynucleotide in such a way that it leads to an improved yield and/or efficiency of biomass production produced from a carbon source by said microorganism.

23. Process according to any one of claims 1 to 3, 12 or 22 or use according to any one of claims 13 to 18 wherein the carbon source is selected from the group consisting of sorbitol, glucose, fructose, mannitol, sorbose, glycerol, sucrose, maltose, starch and mixtures thereof.

24. The microorganism according to any one of claims 6 to 8, 19 or 21 wherein the carbon source is selected from the group consisting of glucose, fructose, sorbitol, mannitol, sorbose, glycerol, sucrose, maltose, starch, and mixtures thereof.

***

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2007/005591

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/54 C12N9/10 C12P17/04 C12P1/04 C12N1/20

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 practical, search terms used)
EPO-Internal, Sequence Search, WPI Data, BIOSIS

C. 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] 1 March 2005 (2005-03-01), PRUST C ET AL: "Transaldolase (EC 2.2.1.2) Gluconobacter oxydans (Gluconobacter suboxydans)" XP002409917 retrieved from EBI Database accession no. Q5FQA2 the whole document -/--	11

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

[X] See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
31 August 2007	11/09/2007

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	van de Kamp, Mart

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2007/005591

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	-& PRUST C ET AL: "Complete genome sequence of the acetic acid bacterium Gluconobacter oxydans" NATURE BIOTECHNOLOGY, vol. 23, no. 2, February 2005 (2005-02), pages 195-200, XP002334906 ISSN: 1087-0156 abstract page 195, left-hand column, line 1 - right-hand column, line 6 figure 1	7-9,11, 14,17, 23,24
X	DATABASE Geneseq [Online] 20 May 2004 (2004-05-20), "Gluconobacter oxydans glucose-6-phosphoric acid isomerase protein." XP002409918 retrieved from EBI accession no. GSP:ADL90006 Database accession no. ADL90006 the whole document	11
X	-& DATABASE WPI Week 200413 Derwent Publications Ltd., London, GB; AN 2004-127093 XP002409924 & JP 2004 024140 A (AJINOMOTO KK) 29 January 2004 (2004-01-29) abstract	1-6, 9-13,15, 16,18-24
X	DATABASE UniProt [Online] 5 July 2004 (2004-07-05), SUGIYAMA M ET AL: "Transaldolase and glucose-6-phosphate isomerase bifunctional protein Gluconobacter oxydans (Gluconobacter suboxydans)" XP002409919 retrieved from EBI Database accession no. Q76EM6 the whole document	11
X	DATABASE EMBL [Online] 24 January 2005 (2005-01-24), XP002409920 retrieved from EBI accession no. EM_PRO:CP000009_3407 Database accession no. CP000009 the whole document	9,11

-/--

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2007/005591

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 Geneseq [Online] 20 May 2004 (2004-05-20), "Gluconobacter oxydans NADH production related coding sequence." XP002409922 retrieved from EBI accession no. GSN:ADL90004 Database accession no. ADL90004 the whole document	9
X	DATABASE EMBL [Online] 3 February 2004 (2004-02-03), SUGIYAMA M ET AL: "Gluconobacter oxydans tkt, tal-pgi, gnd genes for transketolase, transaldolase and glucose-6-phosphate isomerase bifunctional protein, 6-phosphogluconate dehydrogenase, complete cds." XP002409921 retrieved from EBI accession no. EM_PRO:AB106333 Database accession no. AB106333 the whole document	9,11
X	JP 2004 024140 A (AJINOMOTO KK) 29 January 2004 (2004-01-29) the whole document	1-6, 9-13,15, 16,18-24
X	SUGIYAMA M ET AL: "Transaldolase/glucose-6-phosphate isomerase bifunctional enzyme and ribulokinase as factors to increase xylitol production from D-arabitol in Gluconobacter oxydans." BIOSCIENCE BIOTECHNOLOGY AND BIOCHEMISTRY, vol. 67, no. 12, December 2003 (2003-12), pages 2524-2532, XP002409916 ISSN: 0916-8451 the whole document page 2526, right-hand column, line 31 - page 2527, left-hand column, line 8 page 2528, right-hand column, line 4 - page 2529, right-hand column, line 11 page 2531, right-hand column, lines 29-32	1-6, 9-13,15, 16,18-24

-/--

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2007/005591

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	KARHUMAA KAISA ET AL: "Investigation of limiting metabolic steps in the utilization of xylose by recombinant Saccharomyces cerevisiae using metabolic engineering" YEAST, vol. 22, no. 5, April 2005 (2005-04), pages 359-368, XP002347259 ISSN: 0749-503X abstract figure 2	2
X	WO 2006/009434 A (UNIV DELFT TECH [NL]; WINKLER AARON ADRIAAN [NL]; KUYPER SIPKO MAARTEN) 26 January 2006 (2006-01-26) example 1 claims 2-4	2
X	WO 2005/113759 A2 (BIOTECHNOLOGY RES & DEV [US]; ARGRICULTURAL RES SERVICE UNIT [US]; FOT) 1 December 2005 (2005-12-01) examples 1,2 claims 1-6,10,16	2
X	WO 2005/121349 A2 (BASF AG [DE]; ZELDER OSKAR [DE]; KLOPPROGGE CORINNA [DE]; SCHROEDER HA) 22 December 2005 (2005-12-22) page 15, lines 3-7 page 16, lines 3-6 claims 1,13-22,29-37	2
X	LU J-L ET AL: "Metabolic engineering and control analysis for production of aromatics: Role of transaldolase" BIOTECHNOLOGY AND BIOENGINEERING, vol. 53, no. 2, 20 January 1997 (1997-01-20), pages 132-138, XP002053307 ISSN: 0006-3592 abstract	2
A	DEPPENMEIER U ET AL: "Biochemistry and biotechnological applications of Gluconobacter strains" APPLIED MICROBIOLOGY AND BIOTECHNOLOGY, vol. 60, no. 3, November 2002 (2002-11), pages 233-242, XP002334995 ISSN: 0175-7598 the whole document figure 2 page 237, left-hand column, line 3 - page 238, right-hand column, line 32	

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2007/005591

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
JP 2004024140 A	29-01-2004	NONE	
WO 2006009434 A	26-01-2006	CA 2573483 A1	26-01-2006
WO 2005113759 A2	01-12-2005	US 2006110805 A1	25-05-2006
WO 2005121349 A2	22-12-2005	AR 047058 A1	04-01-2006
		AU 2004320598 A1	22-12-2005
		CA 2546847 A1	22-12-2005
		EP 1704241 A2	27-09-2006
		JP 2007514439 T	07-06-2007
		KR 20070026354 A	08-03-2007