

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
17 August 2006 (17.08.2006)

PCT

(10) International Publication Number
WO 2006/084646 A2

(51) International Patent Classification:
C07K 14/195 (2006.01) **C12P 17/04** (2006.01)

(21) International Application Number:
PCT/EP2006/001012

(22) International Filing Date: 6 February 2006 (06.02.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
05405128.9 11 February 2005 (11.02.2005) EP

(71) Applicant (for all designated States except US): **DSM IP ASSETS B.V.** [NL/NL]; Het Overloon 1, NL-6411 TE Heerlen (NL).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **SHINJOH, Masako** [JP/JP]; 5-5-30 Dai, Kamakura, Kanagana-247-0061 (JP).

(74) Agents: **SCHWANDER, Kuno, Josef** et al.; 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, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, 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, 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:

— without international search report and to be republished upon receipt of that report

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.

(54) Title: NOVEL GENE SMS 22

(57) Abstract: The present invention relates to newly identified microorganisms capable of direct production of L-ascorbic acid (hereinafter also referred to as Vitamin C). The invention also relates to polynucleotide sequences comprising genes that encode proteins which are involved in the synthesis of Vitamin C. 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. The present invention also relates to the use of said polynucleotides and polypeptides as biotechnological tools in the production of Vitamin C from microorganisms, whereby a modification of said polynucleotides and/or encoded polypeptides has a direct or indirect impact on yield, production, and/or efficiency of production of the fermentation product in said microorganism. Also included are methods/processes of using the polynucleotides and modified polynucleotide sequences to transform host microorganisms. The invention also relates to genetically engineered microorganisms and their use for the direct production of Vitamin C.



WO 2006/084646 A2

Novel gene SMS 22

The present invention relates to newly identified microorganisms capable of direct production of L-ascorbic acid (hereinafter also referred to as Vitamin C). The invention also relates to polynucleotide sequences comprising genes that encode proteins that are
5 involved in the synthesis of Vitamin C. 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. The present invention also relates to the use of said polynucleotides and polypeptides as biotechnological tools in the production of Vitamin C
10 from microorganisms, whereby a modification of said polynucleotides and/or encoded polypeptides has a direct or indirect impact on yield, production, and/or efficiency of production of the fermentation product in said microorganism. Also included are methods/processes of using the polynucleotides and modified polynucleotide sequences to transform host microorganisms. The invention also relates to genetically engineered
15 microorganisms and their use for the direct production of Vitamin C.

Vitamin C is one of very important and indispensable nutrient factor for human beings. Vitamin C is also used in animal feed even though farm animals can synthesize it in their own body.

For the past 70 years, Vitamin C has been produced industrially from D-glucose by the
20 well-known Reichstein method. All steps in this process are chemical except for one (the conversion of D-sorbitol to L-sorbose), which is carried out by microbial transformation. Since its initial implementation for industrial production of Vitamin C, several chemical and technical modifications have been used to improve the efficiency of the Reichstein

method. Recent developments of Vitamin C production are summarized in Ullmann's Encyclopedia of Industrial Chemistry, 5th Edition, Vol. A27 (1996), pp. 547ff.

Different intermediate steps of Vitamin C production have been performed with the help of microorganisms or enzymes isolated therefrom. Thus, 2-keto-L-gulonic acid (2-KGA), an intermediate compound that can be chemically converted into Vitamin C by means of an alkaline rearrangement reaction, may be produced by a fermentation process starting from L-sorbose, by means of strains belonging *e.g.* to the *Ketogulonicigenium* or *Gluconobacter* genera, or by an alternative fermentation process starting from D-glucose, by means of recombinant strains belonging to the *Gluconobacter* or *Pantoea* genera.

- 10 Current production methods for Vitamin C have some undesirable characteristics such as high-energy consumption and use of large quantities of organic and inorganic solvents. Therefore, over the past decades, other approaches to manufacture Vitamin C using microbial conversions, which would be more economical as well as ecological, have been investigated.
- 15 Direct Vitamin C production has been reported in several microorganisms, such as algae and yeast, using different cultivation methods. The disadvantage of using these microorganisms, however, is the low yield of Vitamin C produced due to the instability of the product. Using, for instance, microorganisms which are known to be both capable of the production of 2-keto-L-gulonic acid and Vitamin C, the yield of microbiologically produced Vitamin C is further limited by the relatively high production of 2-KGA which is more readily synthesized by said microorganism, leading, for instance, to ratios between the concentration of Vitamin C and 2-KGA which are less than 0.1. Thus, it is an object of the present invention to improve the microbiological production of Vitamin C to get higher yields as with the processes described in the prior art.
- 25 Surprisingly, we now identified acetic acid bacteria able to directly produce Vitamin C in high yields from a number of substrates including D-sorbitol, L-sorbose and L-sorbose. Strains from the genera of *Gluconobacter*, *Gluconacetobacter* and *Acetobacter* were found to be able to directly produce Vitamin C from L-sorbose, whereas at least *Gluconobacter oxydans* DSM 17078 was found to be able to produce Vitamin C directly from D-sorbitol, L-sorbose or L-sorbose. *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.

- 35 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.

In particular, the present invention provides a process for the direct production of Vitamin C comprising converting a substrate into Vitamin C. This may for instance be done in a medium comprising a microorganism, which may be a resting or a growing microorganism, preferably a resting microorganism.

Several substrates may be used as a carbon source in the above-mentioned process. Particularly suited carbon sources are those that are easily obtainable from the D-glucose or D-sorbitol metabolization pathway such as, for example, D-glucose, D-sorbitol, L-sorbose, L-sorbose, 2-keto-L-gulonate, D-gluconate, 2-keto-D-gluconate or 2,5-diketo-gluconate. Preferably, the substrate is selected from for instance D-glucose, D-sorbitol, L-sorbose or L-sorbose, more preferably from D-glucose, D-sorbitol or L-sorbose, and most preferably from D-sorbitol or L-sorbose. The term "substrate" and "production substrate" in connection with the above process using a microorganism is used interchangeably herein.

Conversion of the substrate into Vitamin C in connection with the above process using a microorganism means that the conversion of the substrate resulting in Vitamin C is performed by the microorganism, i.e. the substrate may be directly converted into Vitamin C. Said microorganism is cultured under conditions which allow such conversion from the substrate as defined above.

A medium as used herein for the above process using a microorganism may be any suitable medium for the production of Vitamin C. Typically, the medium is an aqueous medium comprising for instance salts, substrate(s), and a certain pH. The medium in which the substrate is converted into Vitamin C is also referred to as the production medium.

5 In connection with the above process using a microorganism, any microorganism capable of performing the conversion of the substrate to Vitamin C may be used, such as for instance, yeast, algae or 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*,
10 or *Kluyveromyces*. An example of such algae may be, e.g., *Chlorella*. Examples of such bacteria may be, e.g., *Gluconobacter*, *Acetobacter*, *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*,
15 preferably *G. oxydans* DSM 17078.

"Fermentation" or "production" or "fermentation process" as used herein may be the use of growing cells using media, conditions and procedures known to the skilled person, or the use of non-growing so-called resting cells, after they have been cultivated by using media, conditions and procedures known to the skilled person, under appropriate conditions for
20 the conversion of suitable substrates into desired products such as Vitamin C.

The term "direct fermentation", "direct production", "direct conversion" and the like is intended to mean that a microorganism is capable of the conversion of a certain substrate 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
25 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.

30 In connection with the above process using a microorganism it is understood that the above-mentioned microorganisms also include synonyms or basonyms of such species having the same physiological properties, as defined by the International Code of Nomenclature of Prokaryotes. The nomenclature of the microorganisms as used herein is the one officially accepted (at the filing date of the priority application) by the International
35 Committee on Systematics of Prokaryotes and the Bacteriology and Applied Microbiology

Division of the International Union of Microbiological Societies, and published by its official publication vehicle International Journal of Systematic and Evolutionary Microbiology (IJSEM). A particular reference is made to Urbance et al., IJSEM (2001) vol 51:1059-1070, with a corrective notification on IJSEM (2001) vol 51:1231-1233, describing the taxonomically reclassification of *G. oxydans* DSM 4025 as *Ketogulonicigenium vulgare*.

As used herein, resting cells refer to cells of a microorganism which are for instance viable but not actively growing, or which are growing at low specific growth rates [μ], for instance, growth rates that are lower than 0.02 h^{-1} , preferably lower than 0.01 h^{-1} . Cells which show the above growth rates are said to be in a "resting cell mode".

The process of the present invention as above using a microorganism may be performed in different steps or phases: preferably, the microorganism is cultured in a first step (also referred to as step (a) or growth phase) under conditions which enable growth. This phase is terminated by changing of the conditions such that the growth rate of the microorganism is reduced leading to resting cells, also referred to as step (b), followed by the production of Vitamin C from the substrate using the (b), also referred to as production phase.

Growth and production phase as performed in the above process using a microorganism may be performed in the same vessel, *i.e.*, only one vessel, or in two or more different vessels, with an optional cell separation step between the two phases. The produced Vitamin C can be recovered from the cells by any suitable means. Recovering means for instance that the produced Vitamin C may be separated from the production medium. Optionally, the thus produced Vitamin C may be further processed.

For the purpose of the present invention relating to the above process using a microorganism, the terms "growth phase", "growing step", "growth step" and "growth period" are used interchangeably herein. The same applies for the terms "production phase", "production step", "production period".

One way of performing the above process using a microorganism as of the present invention may be a process wherein the microorganism is grown in a first vessel, the so-called growth vessel, as a source for the resting cells, and at least part of the cells are transferred to a second vessel, the so-called production vessel. The conditions in the production vessel may be such that the cells transferred from the growth vessel become resting cells as defined above. Vitamin C is produced in the second vessel and recovered therefrom.

In connection with the above process using a microorganism, in one aspect, the growing step can be performed in an aqueous medium, *i.e.* the growth medium, supplemented with

appropriate nutrients for growth under aerobic conditions. The cultivation may be conducted, for instance, in batch, fed-batch, semi-continuous or continuous mode. The cultivation period may vary depending on the kind of cells, pH, temperature and nutrient medium to be used, and may be for instance about 10 h to about 10 days, preferably about 1 to about 10 days, more preferably about 1 to about 5 days when run in batch or fed-batch mode, depending on the microorganism. If the cells are grown in continuous mode, the residence time may be for instance from about 2 to about 100 h, preferably from about 2 to about 50 h, depending on the microorganism. If the microorganism is selected from bacteria, the cultivation may be conducted for instance at a pH of about 3.0 to about 9.0, preferably about 4.0 to about 9.0, more preferably about 4.0 to about 8.0, even more preferably about 5.0 to about 8.0. If algae or yeast are used, the cultivation may be conducted, for instance, at a pH below about 7.0, preferably below about 6.0, more preferably below about 5.5, and most preferably below about 5.0. A suitable temperature range for carrying out the cultivation using bacteria may be for instance from about 13°C to about 40°C, preferably from about 18°C to about 37°C, more preferably from about 13°C to about 36°C, and most preferably from about 18°C to about 33°C. If algae or yeast are used, a suitable temperature range for carrying out the cultivation may be for instance from about 15°C to about 40°C, preferably from about 20°C to about 45°C, more preferably from about 25°C to about 40°C, even more preferably from about 25°C to about 38°C, and most preferably from about 30°C to about 38°C. The culture medium for growth usually may contain such nutrients as assimilable carbon sources, *e.g.*, glycerol, D-mannitol, D-sorbitol, L-sorbose, erythritol, ribitol, xylitol, arabitol, inositol, dulcitol, D-ribose, D-fructose, D-glucose, and sucrose, preferably L-sorbose, D-glucose, D-sorbitol, D-mannitol, and glycerol; and digestible nitrogen sources such as organic substances, *e.g.*, peptone, yeast extract and amino acids. The media may be with or without urea and/or corn steep liquor and/or baker's yeast. Various inorganic substances may also be used as nitrogen sources, *e.g.*, nitrates and ammonium salts. Furthermore, the growth medium usually may contain inorganic salts, *e.g.*, magnesium sulfate, manganese sulfate, potassium phosphate, and calcium carbonate.

In connection with the above process using a microorganism, in the growth phase the specific growth rates are for instance at least 0.02 h^{-1} . For cells growing in batch, fed-batch or semi-continuous mode, the growth rate depends on for instance the composition of the growth medium, pH, temperature, and the like. In general, the growth rates may be for instance in a range from about 0.05 h^{-1} to about 0.2 h^{-1} , preferably from about 0.06 h^{-1} to about 0.15 h^{-1} , and most preferably from about 0.07 h^{-1} to about 0.13 h^{-1} .

In another aspect of the above process using a microorganism, resting cells may be provided by cultivation of the respective microorganism on agar plates thus serving as

growth vessel, using essentially the same conditions, *e.g.*, cultivation period, pH, temperature, nutrient medium as described above, with the addition of agar agar.

In connection with the above process using a microorganism, if the growth and production phase are performed in two separate vessels, then the cells from the growth phase may be
5 harvested or concentrated and transferred to a second vessel, the so-called production vessel. This vessel may contain an aqueous medium supplemented with any applicable production substrate that can be converted to L-ascorbic acid by the cells. Cells from the growth vessel can be harvested or concentrated by any suitable operation, such as for instance centrifugation, membrane crossflow ultrafiltration or microfiltration, filtration,
10 decantation, flocculation. The cells thus obtained may also be transferred to the production vessel in the form of the original broth from the growth vessel, without being harvested, concentrated or washed, *i.e.* in the form of a cell suspension. In a preferred embodiment, the cells are transferred from the growth vessel to the production vessel in the form of a cell suspension without any washing or isolating step in-between.

15 Thus, in a preferred embodiment of the above process using a microorganism step (a) and (c) of the process of the present invention as described above are not separated by any washing and/or separation step.

In connection with the above process using a microorganism, if the growth and production phase are performed in the same vessel, cells may be grown under appropriate conditions
20 to the desired cell density followed by a replacement of the growth medium with the production medium containing the production substrate. Such replacement may be, for instance, the feeding of production medium to the vessel at the same time and rate as the withdrawal or harvesting of supernatant from the vessel. To keep the resting cells in the vessel, operations for cell recycling or retention may be used, such as for instance cell
25 recycling steps. Such recycling steps, for instance, include but are not limited to methods using centrifuges, filters, membrane crossflow microfiltration or ultrafiltration steps, membrane reactors, flocculation, or cell immobilization in appropriate porous, non-porous or polymeric matrixes. After a transition phase, the vessel is brought to process conditions under which the cells are in a resting cell mode as defined above, and the production
30 substrate is efficiently converted into Vitamin C.

The aqueous medium in the production vessel as used for the production step in connection with the above process using a microorganism, hereinafter called production medium, may contain only the production substrate(s) to be converted into L-ascorbic acid, or may contain for instance additional inorganic salts, *e.g.*, sodium chloride, calcium
35 chloride, magnesium sulfate, manganese sulfate, potassium phosphate, calcium phosphate, and calcium carbonate. The production medium may also contain digestible nitrogen

- sources such as for instance organic substances, *e.g.*, peptone, yeast extract, urea, amino acids, and corn steep liquor, and inorganic substances, *e.g.* ammonia, ammonium sulfate, and sodium nitrate, at such concentrations that the cells are kept in a resting cell mode as defined above. The medium may be with or without urea and/or corn steep liquor and/or
- 5 baker's yeast. The production step may be conducted for instance in batch, fed-batch, semi-continuous or continuous mode. In case of fed-batch, semi-continuous or continuous mode, both cells from the growth vessel and production medium can be fed continuously or intermittently to the production vessel at appropriate feed rates. Alternatively, only production medium may be fed continuously or intermittently to the production vessel,
- 10 while the cells coming from the growth vessel are transferred at once to the production vessel. The cells coming from the growth vessel may be used as a cell suspension within the production vessel or may be used as for instance flocculated or immobilized cells in any solid phase such as porous or polymeric matrixes. The production period, defined as the period elapsed between the entrance of the substrate into the production vessel and the
- 15 harvest of the supernatant containing Vitamin C, the so-called harvest stream, can vary depending for instance on the kind and concentration of cells, pH, temperature and nutrient medium to be used, and is preferably about 2 to about 100 h. The pH and temperature can be different from the pH and temperature of the growth step, but is essentially the same as for the growth step.
- 20 In a preferred embodiment of the above process using a microorganism, the production step is conducted in continuous mode, meaning that a first feed stream containing the cells from the growth vessel and a second feed stream containing the substrate is fed continuously or intermittently to the production vessel. The first stream may either contain only the cells isolated/separated from the growth medium or a cell suspension, coming
- 25 directly from the growth step, *i.e.* cells suspended in growth medium, without any intermediate step of cell separation, washing and/or isolating. The second feed stream as herein defined may include all other feed streams necessary for the operation of the production step, *e.g.* the production medium comprising the substrate in the form of one or several different streams, water for dilution, and base for pH control.
- 30 In connection with the above process using a microorganism, when both streams are fed continuously, the ratio of the feed rate of the first stream to feed rate of the second stream may vary between about 0.01 and about 10, preferably between about 0.01 and about 5, most preferably between about 0.02 and about 2. This ratio is dependent on the concentration of cells and substrate in the first and second stream, respectively.
- 35 Another way of performing the process as above using a microorganism of the present invention may be a process using a certain cell density of resting cells in the production vessel. The cell density is measured as absorbance units (optical density) at 600 nm by

methods known to the skilled person. In a preferred embodiment, the cell density in the production step is at least about 10, more preferably between about 10 and about 200, even more preferably between about 15 and about 200, even more preferably between about 15 to about 120, and most preferably between about 20 and about 120.

- 5 In connection with the above process using a microorganism, in order to keep the cells in the production vessel at the desired cell density during the production phase as performed, for instance, in continuous or semi-continuous mode, any means known in the art may be used, such as for instance cell recycling by centrifugation, filtration, membrane crossflow ultrafiltration of microfiltration, decantation, flocculation, cell retention in the vessel by
- 10 membrane devices or cell immobilization. Further, in case the production step is performed in continuous or semi-continuous mode and cells are continuously or intermittently fed from the growth vessel, the cell density in the production vessel may be kept at a constant level by, for instance, harvesting an amount of cells from the production vessel corresponding to the amount of cells being fed from the growth vessel.
- 15 In connection with the above process using a microorganism, the produced Vitamin C contained in the so-called harvest stream is recovered/harvested from the production vessel. The harvest stream may include, for instance, cell-free or cell-containing aqueous solution coming from the production vessel, which contains Vitamin C as a result of the conversion of production substrate by the resting cells in the production vessel. Cells still
- 20 present in the harvest stream may be separated from the Vitamin C by any operations known in the art, such as for instance filtration, centrifugation, decantation, membrane crossflow ultrafiltration or microfiltration, tangential flow ultrafiltration or microfiltration or dead end filtration. After this cell separation operation, the harvest stream is essentially free of cells.
- 25 In connection with the above process using a microorganism, in one aspect, the process of the present invention leads to yields of Vitamin C which are at least about 1.8 g/l, preferably at least about 2.5 g/l, more preferably at least about 4.0 g/l, and most preferably at least about 5.7 g/l, such as 10 g/l, 20 g/l, 50 g/l, 100 g/l, 200 g/l or more than 300 g/l. In one embodiment, the yield of Vitamin C produced by the process of the present invention
- 30 is in the range of from about 1.8 to about 600 g/l. The yield of Vitamin C refers to the concentration of Vitamin C in the harvest stream coming directly out of the production vessel, *i.e.* the cell-free supernatant comprising the Vitamin C.

In a further aspect, the process of the present invention may be combined with further steps of separation and/or purification of the produced Vitamin C from other components

35 contained in the harvest stream, *i.e.*, so-called downstream processing steps. These steps may include any means known to a skilled person, such as, for instance, concentration,

crystallization, precipitation, adsorption, ion exchange, electrodialysis, bipolar membrane electrodialysis and/or reverse osmosis. Vitamin C may be further purified as the free acid form or any of its known salt forms by means of operations such as for instance treatment with activated carbon, ion exchange, adsorption and elution, concentration, crystallization, filtration and drying. Specifically, a first separation of Vitamin C from other components in the harvest stream might be performed by any suitable combination or repetition of, for instance, the following methods: two- or three-compartment electrodialysis, bipolar membrane electrodialysis, reverse osmosis or adsorption on, for instance, ion exchange resins or non-ionic resins. If the resulting form of Vitamin C is a salt of L-ascorbic acid, conversion of the salt form into the free acid form may be performed by for instance bipolar membrane electrodialysis, ion exchange, simulated moving bed chromatographic techniques, and the like. Combination of the mentioned steps, *e.g.*, electrodialysis and bipolar membrane electrodialysis into one step might be also used as well as combination of the mentioned steps *e.g.* several steps of ion exchange by using simulated moving bed chromatographic methods. Any of these procedures alone or in combination constitute a convenient means for isolating and purifying the product, *i.e.* Vitamin C. The product thus obtained may further be isolated in a manner such as, *e.g.* by concentration, crystallization, precipitation, washing and drying of the crystals and/or further purified by, for instance, treatment with activated carbon, ion exchange and/or re-crystallization.

In a preferred embodiment, Vitamin C is purified from the harvest stream by a series of downstream processing steps as described above without having to be transferred to a non-aqueous solution at any time of this processing, *i.e.* all steps are performed in an aqueous environment. Such preferred downstream processing procedure may include for instance the concentration of the harvest stream coming from the production vessel by means of two- or three-compartment electrodialysis, conversion of Vitamin C in its salt form present in the concentrated solution into its acid form by means of bipolar membrane electrodialysis and/or ion exchange, purification by methods such as for instance treatment with activated carbon, ion exchange or non-ionic resins, followed by a further concentration step and crystallization. These crystals can be separated, washed and dried. If necessary, the crystals may be again re-solubilized in water, treated with activated carbon and/or ion exchange resins and recrystallized. These crystals can then be separated, washed and dried.

Surprisingly, it has now also 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 biotechnological production of Vitamin C. It has also been found, that by genetically altering the expression level of nucleotides according to the invention in a microorganism, such as for example

Gluconobacter, the direct fermentation of Vitamin C by said microorganism can be even greatly improved.

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

5 (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
10 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
15 polypeptide, and said fragment or derivative has the activity of a SMS 22 polypeptide;

(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 SMS 22 polypeptide;

(f) polynucleotides which are at least 70%, such as 85, 90 or 95% homologous to a
20 polynucleotide as defined in any one of (a) to (d) and which encode a SMS 22 polypeptide
or

the complementary strand of such a polynucleotide.

Furthermore, it has also been found that the novel polynucleotides encode proteins which are involved in the Sorbitol/Sorbose Metabolization System. Polynucleotides according to
25 the invention and proteins encoded by these polynucleotides are herein abbreviated by SMS. SMS proteins are herein defined as proteins that function in the direct metabolization of D-sorbitol or L-sorbose.

Metabolisation of D-sorbitol or L-sorbose includes on one side the assimilation of these compounds into the cytosol and further conversion into metabolites useful for assimilation
30 pathways such as the Embden-Meyerhof-Parnas pathway, the pentose phosphate pathway, the Entner-Doudoroff pathway, and the tricarboxylic acid cycle, all of them involved in

all vital energy-forming and anabolic reactions necessary for growth and maintenance of living cells. On the other side, metabolisation of D-sorbitol or L-sorbose also includes the conversion of these compounds into further oxidized products such as L-sorbose, 2KGA and Vitamin C by so-called incomplete oxidation processes. SMS proteins include but are not limited to, for example, enzymes or enzyme subunits having activity towards or involved in the assimilation or conversion of D-sorbitol such as membrane-bound PQQ-dependent D-sorbitol dehydrogenase, membrane-bound FAD-dependent D-sorbitol dehydrogenase, cytosolic NAD-dependent D-sorbitol dehydrogenase, NAD(P)-dependent D-sorbitol dehydrogenase (also called as NADPH-dependent sorbose reductase), NAD-dependent xylitol dehydrogenase, NAD-dependent enzyme annotated as alcohol dehydrogenase, proteins, enzymes or enzyme subunits having activity towards or involved in the assimilation of L-sorbose such as membrane-bound L-sorbose dehydrogenase, NAD(P)H-dependent L-sorbose reductase, enzymes or enzyme subunits having activity towards or involved in the assimilation of L-sorbose, such as cytosolic NADP-dependent sorbose dehydrogenase, cytosolic NAD(P)H-dependent L-sorbose reductase, membrane-bound aldehyde dehydrogenase, cytosolic aldehyde dehydrogenase, and others.

Microorganisms able to assimilate D-sorbitol for growth usually possess enzymes able to oxidize this compound into a universal assimilation substrate such as D-fructose. Also microorganisms able to grow on L-sorbose possess an enzyme, NAD(P)H-dependent L-sorbose reductase, which is able to reduce this compound to D-sorbitol, which is then further oxidized into D-fructose. D-fructose is an excellent substrate for the growth of many microorganisms, after it has been phosphorylated by means of a D-fructose kinase.

As an example, acetic acid bacteria, which are obligate aerobe, gram-negative microorganisms belonging to the genus *Acetobacter*, *Gluconobacter*, and *Gluconacetobacter*, are able to transport D-sorbitol into the cytosol and convert it into D-fructose by means of a cytosolic NAD-dependent D-sorbitol dehydrogenase. Some individual strains, such as *Gluconobacter oxydans* IFO 3292, and IFO 3293, are able as well to transport L-sorbose into the cytosol and reduce it to D-sorbitol by means of a cytosolic NAD(P)H-dependent L-sorbose reductase, which then is further oxidized into D-fructose. In these bacteria, the Embden-Meyerhof-Parnas pathway, as well as the tricarboxylic acid cycle are not fully active, and the main pathway channeling sugars into the central metabolism is the pentose phosphate pathway. D-fructose-6-phosphate, obtained from D-fructose by a phosphorylation reaction enters the pentose phosphate pathway, being further metabolised and producing reducing power in the form of NAD(P)H and tricarboxylic compounds necessary for growth and maintenance.

Acetic acid bacteria are well known for their ability to incompletely oxidize different substrates such as alcohols, sugars, sugar alcohols and aldehydes. These processes are generally known as oxidative fermentations or incomplete oxidations, and they have been well established for a long time in the food and chemical industry, especially in vinegar and in L-sorbose production. A useful product known to be obtained from incomplete oxidations of D-sorbitol or L-sorbose using strains belonging to the *Gluconobacter* genus is 2-keto-L-gulonic acid (2-KGA).

Acetic acid bacteria accomplish these incomplete oxidation reactions by means of different dehydrogenases located either in the periplasmic space, on the periplasmic membrane as well as in the cytoplasm. Different co-factors are employed by the different dehydrogenases, the most common being PQQ and FAD for membrane-bound or periplasmic enzymes, and NAD/NADP for cytoplasmic enzymes.

While all products of these oxidation reactions diffuse back to the external aqueous environment through the outer membrane, some of them can be passively or actively transported into the cell and be further used in metabolic pathways responsible for growth and energy formation. Inside the cell, oxidized products can many times be reduced back to their original substrate by means of reductases, and then be channelled back to the central metabolism.

The SMS 22 protein as isolated from *Gluconobacter oxydans* DSM 17078 shown in SEQ ID NO:2 and described herein was found to be a particularly useful SMS protein, since it appeared that it performs a crucial function in the direct Vitamin C production 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 SMS 22 polynucleotide according to SEQ ID NO:1 was annotated as encoding a putative membrane protein.

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 SMS polypeptide.

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 SMS polypeptide.

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

The invention also relates to primers, probes and fragments that may be used to amplify or detect a DNA according to the invention and to identify related species or families of microorganisms also carrying such genes.

The present invention also relates to vectors which include polynucleotides of the invention and microorganisms which are genetically engineered with the polynucleotides or said vectors.

The invention also relates to processes for producing microorganisms capable of expressing a polypeptide encoded by the above defined polynucleotide and a polypeptide encoded by a polynucleotide as defined above.

5 The invention also relates to microorganisms wherein the activity of SMS 22 is enhanced and/or improved so that the yield of Vitamin C which is directly produced from D-sorbitol or L-sorbose is increased. This may be accomplished, for example, by transferring a polynucleotide according to the invention into a recombinant or non-recombinant microorganism that may or may not contain an endogenous equivalent of the SMS 22 gene.

10 The skilled person will know how to enhance and/or improve the activity of an SMS 22 protein. Such may be accomplished by either genetically modifying the host organism in such a way that it produces more or more stable copies of the SMS 22 protein than the wild type organism. It may also be accomplished by increasing the specific activity of the SMS 22 protein.

15 Modifications in order to have the organism produce more copies of the SMS 22 protein or messenger RNA may include the use of a strong promoter, or the mutation (e.g. insertion, deletion or point mutation) of (parts of) the SMS 22 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 an SMS 22 protein may also be accomplished by
20 methods known in the art. Such methods may include the mutation (e.g. insertion, deletion or point mutation) of (parts of) the SMS 22 gene.

Also known in the art are methods of increasing the activity of a given protein by contacting the SMS 22 protein with a specific enhancer or other substances that specifically interact with the SMS 22 protein. In order to identify such specific enhancers,
25 the SMS 22 protein may be expressed and tested for activity in the presence of compounds suspected to enhance the activity of the SMS 22 protein. The activity of the SMS 22 protein may also be increased by stabilizing the messenger RNA encoding SMS 22. 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.
30), 1995, Current Protocols in Molecular Biology, (JohnWiley & Sons, N. Y.).

In connection with the above process of directly producing Vitamin C using a microorganism, any microorganism capable of performing the conversion of the substrate to Vitamin C may be used, such as for instance, yeast, algae or bacteria, either as wild type strains, mutant strains derived by classic mutagenesis and selection methods or as
35 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*, *Ketogulonicigenium*, *Pantoea*, *Pseudomonas*, such as, e.g., *Pseudomonas putida*, and *Escherichia*, such as, e.g., *Escherichia coli*. Preferred are *Gluconobacter* or
5 *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.

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 production of Vitamin C .

10 The invention also relates to processes for the expression of SMS 22 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 producing Vitamin C. All these processes may comprise the step of altering a microorganism, wherein "altering" as used herein encompasses the process for "genetically
15 altering" or "altering the composition of the cell culture media and/or methods used for culturing" in such a way that the yield of the fermentation product can be improved compared to the wild-type organism.

In accordance with still another aspect of the invention there is provided a process for the production of Vitamin C by direct fermentation.

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

The sequence of the gene comprising a nucleotide sequence according to SEQ ID NO:1 encoding a SMS 22 protein was determined by sequencing a genomic clone obtained from
25 *Gluconobacter oxydans* DSM 17078.

The invention also relates to a polynucleotide encoding at least a biologically active fragment or derivative of a SMS 22 polypeptide as 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.

30 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.

5 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, 10 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 15 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.

20 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 17078 SMS proteins. 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 25 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 30 defined herein. It is furthermore appreciated by the skilled person that DNA sequence polymorphisms that lead to changes in the amino acid sequences of SMS 22 proteins may exist within a population, *e.g.*, the *Gluconobacter oxydans* population. Such genetic polymorphism in the SMS 22 gene may exist among individuals within a population due to natural variation or in cells from different populations. Such natural variations can 35 typically result in 1-5% variance in the nucleotide sequence of the SMS 22 gene. Any and all such nucleotide variations and the resulting amino acid polymorphism in SMS 22 are

the result of natural variation and that do not alter the functional activity of SMS 22 proteins 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 recombinant or non-recombinant microorganism capable of converting a given carbon source directly into Vitamin C, in particular *Gluconobacter oxydans*, preferably *Gluconobacter oxydans* DSM 17078 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
5 encoding a portion of a protein according to the invention. The nucleotide sequence determined from the cloning of the SMS 22 gene allows for the generation of probes and primers designed for use in identifying and/or cloning other SMS 22 family members, as well as SMS 22 homologues from other species. The probe/primer typically comprises substantially purified oligonucleotides which typically comprises a region of nucleotide
10 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
15 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
20 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
25 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 SMS 22 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
30 analysis for detecting expression of mRNA of said protein in specific cells or (3) use in enhancing and/or improving the function or activity of homologous SMS 22 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
35 organisms. Nucleic acid molecules corresponding to natural variants and non-*G. oxydans*

homologues of the *G. oxydans* SMS 22 DNA of the invention which are also embraced by the present invention may be isolated based on their homology to the *G. oxydans* SMS 22 nucleic acid disclosed herein using the *G. oxydans* DNA, or a portion thereof, as a hybridization probe according to standard hybridization techniques, preferably under
5 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
10 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
15 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
20 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
25 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
30 strategies, see *e.g.*, Sambrook et al., *supra*; and Ausubel et al., *supra*.

Also, nucleic acids encoding other SMS 22 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 SMS 22 proteins from

different species which thus 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 as of the
5 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 directly into Vitamin C, in particular *Gluconobacter oxydans*, preferably *Gluconobacter oxydans* DSM 17078.

As used herein, the term "hybridizing" is intended to describe conditions for hybridization
10 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%,
15 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
20 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, for example, 2 h to 4 days incubation at 42°C using a digoxigenin (DIG)-labeled DNA probe (prepared by using a DIG labeling system; Roche
Diagnostics GmbH, 68298 Mannheim, Germany) in a solution such as DigEasyHyb
25 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
30 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* SMS 22 protein.

The skilled artisan will know which conditions to apply for stringent and highly stringent hybridization conditions. Additional guidance regarding such conditions is readily
5 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
10 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,
15 *e.g.* microorganisms capable of converting a given carbon source directly into Vitamin C, 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
20 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
25 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, J., Fritsh, E. F. , and Maniatis, T. *Molecular Cloning : A Laboratory Manual*.
30 2nd, ed. , Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 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 Blossum 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
5 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
10 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
15 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 a SMS 22 protein whose function compared to the wild type counterpart is enhanced or improved. The activity of the SMS 22 protein is thereby increased. Methods for introducing such mutations are well
20 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 the increase of activity of a given protein, in this case the SMS 22
25 protein. 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 SMS 22 protein activity is introduced in a cell that did not contain this activity before, *e.g.* by introducing a gene encoding SMS 22 in a cell that did not contain an equivalent of this gene before, or that could not express an active
30 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
35 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 increase the activity. For example, altering the relative K_m of the polypeptide with its corresponding substrate will result in improved activity. Likewise, the relative half-life of the polypeptide may be increased. In either
5 scenario, that being enhanced gene expression or increased specific 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%, 100%, 200% or even more than 500%, compared to a wild-type protein, polynucleotide, gene; or the activity and/or the
10 concentration of the protein present before the polynucleotides or polypeptides are enhanced and/or improved. The activity of the SMS 22 protein 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
15 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
20 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
25 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
30 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
35 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., SMS 22 proteins, mutant forms of SMS 22 proteins, fusion proteins and the like.

The recombinant expression vectors of the invention may be designed for expression of SMS 22 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*. 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
5 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
10 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
15 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
20 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
25 the proteins according to the present invention and other SMS 22 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.*,
30 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 fermentation, for example in a direct fermentation process for Vitamin C.

According to the invention a genetically engineered/recombinantly produced host cell (also referred to as recombinant cell or transformed cell) 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, production and/or efficiency of production of one or more fermentation products such as Vitamin C is improved. The host cell may be selected from a microorganism capable of directly producing one or more fermentation products such as for instance Vitamin C from a given carbon source, in particular *Gluconobacter oxydans*, preferably *G. oxydans* DSM 17078.

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 the SMS 22 protein has been increased and/or enhanced. Suitable host cells include cells of microorganisms capable of producing a given fermentation product, e.g., converting a given carbon source directly into Vitamin C. 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 *G. oxydans* DSM 17078.

Improved gene expression may also be achieved by modifying the SMS 22 gene, e.g., by introducing one or more mutations into the SMS 22 gene wherein said modification leads to a SMS 22 protein 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 SMS 22 gene products. This may include for instance an alteration in the genome of a microorganism, which improves the synthesis of SMS 22 or leads to the expression of a SMS 22 protein 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 interference 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
5 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
10 *e.g.* chemical mutagens, radiation or transposons and selecting or screening for mutants which are better or more efficient producers of one or more fermentation products. 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 SMS 22 gene of the present invention, *i.e.*, wherein its repressor gene expression is artificially
15 suppressed in order to improve the yield, production, 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. 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
20 expression of either one of the two alleles.

The aforementioned mutagenesis strategies for SMS 22 proteins may result in increased yields of a desired compound in particular Vitamin C. 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
25 may be utilized to generate microorganisms such as *Gluconobacter oxydans* or related strains of bacteria expressing mutated SMS 22 nucleic acid and protein molecules such that the yield, production, and/or efficiency of production of a desired compound such as Vitamin C is improved.

In one aspect of the invention, microorganisms (in particular from the genera of
30 *Gluconobacter*, *Gluconacetobacter* and *Acetobacter*) are provided that are able to directly produce Vitamin C from a suitable carbon source like D-sorbitol and/or L-sorbose. When measured in a resting cell method after an incubation period of 20 hours, these organisms were found to be able to produce Vitamin C directly from D-sorbitol or L-sorbose, even up to a level of 280 mg/l and 670 mg/l respectively. In another aspect of the invention, a
35 microorganism is provided capable of directly producing Vitamin C from D-sorbitol and/or L-sorbose in quantities of 300 mg/l or more or 800 mg/l or more from D-sorbitol or L-

sorbose, respectively when measured in a resting cell method after an incubation period of 20 hours. Such may be achieved by increasing the activity of the SMS 22 protein. The yield of Vitamin C produced from D-sorbitol when measured in a resting cell method after an incubation period of 20 hours may even be as high as 400, 600, 1000 mg/l or even exceed 1.5, 2, 4, 10, 20, 50 g/l. The yield of Vitamin C produced from L-sorbose when measured in a resting cell method after an incubation period of 20 hours may even be as high as 1000 mg/l or even exceed 1.5, 2, 4, 10, 20, 50 g/l.

The recombinant microorganism carrying *e.g.* a modified SMS 22 gene and which is able to produce the fermentation product 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 such nutrients as assimilable carbon sources, *e.g.*, glycerol, D-mannitol, D-glucose, D-sorbitol, L-sorbose, erythritol, ribitol, xylitol, arabitol, inositol, dulcitol, D-ribose, D-fructose, and sucrose, preferably D-sorbitol, D-mannitol, D-glucose and glycerol; 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 medium usually may contain inorganic salts, *e.g.*, magnesium sulfate, manganese sulfate, potassium phosphate, and calcium carbonate. Cells obtained using the procedures described above can then be further incubated at essentially the same modes, temperature and pH conditions as described above, in the presence of substrates such as D-sorbitol, L-sorbose, or D-glucose, in such a way that they convert these substrates directly into Vitamin C. Incubation can be done in a nitrogen-rich medium, containing, for example, organic nitrogen sources, *e.g.*, peptone, yeast extract, baker's yeast, urea, amino acids, and corn steep liquor, or inorganic nitrogen sources, *e.g.*, nitrates and ammonium salts, in which case cells will be able to further grow while producing Vitamin C. Alternatively, incubation can be done in a nitrogen-poor medium, in which case cells will not grow substantially, and will be in a resting cell mode, or biotransformation mode. In all cases, the incubation medium may also contain inorganic salts, *e.g.*, magnesium sulfate, manganese sulfate, potassium phosphate, and calcium chloride.

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

- 5 *Gluconobacter oxydans* sequences of interest; evolutionary studies; determination of SMS 22 protein regions required for function; modulation of a SMS 22 protein activity or function; modulation of the activity of a SMS 22 pathway; and modulation of cellular production of a desired compound, such as Vitamin C.

- 10 The invention provides methods for screening molecules which modulate the activity of a SMS 22 protein, either by interacting with the protein itself or a substrate or binding partner of the SMS 22 protein, or by modulating the transcription or translation of a SMS 22 nucleic acid molecule of the invention. In such methods, a microorganism expressing one or more SMS 22 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
- 15 the SMS 22 protein is assessed.

- The activity of SMS proteins can be measured by methods well known to a skilled person, such as, for example, by incubating a cell fraction containing the SMS protein in the presence of its substrate, electron acceptor(s) or donor(s) including phenazine methosulfate (PMS), dichlorophenol-indophenol (DCIP), NAD, NADH, NADP, NADPH, which
- 20 consumption can be directly or indirectly measured by photometric, colorimetric or fluorimetric methods, and other inorganic components which might be relevant for the development of the activity. Thus, for example, the activity of membrane-bound D-sorbitol dehydrogenase can be measured in an assay where membrane fractions containing this enzyme are incubated in the presence of phosphate buffer at pH 6, D-sorbitol and the
- 25 artificial electron acceptors DCIP and PMS. The rate of consumption of DCIP can be measured at 600 nm, and is directly proportional to the D-sorbitol dehydrogenase activity present in the membrane fraction.

- It may be evident from the above description that the fermentation product of the methods according to the invention may not be limited to Vitamin C alone. The "desired
- 30 compound" or "fermentation product" as used herein may be any natural product of *Gluconobacter oxydans*, which includes the final products and intermediates of biosynthesis pathways, such as for example L-sorbose, L-sorbose, D-gluconate, 2-keto-D-gluconate, 5-keto-D-gluconate, 2,5-diketo-D-gluconate and 2-keto-L-gulonate (2-KGA), in particular the biosynthetic generation of Vitamin C.

- 35 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 Vitamin

C, *i.e.*, the direct conversion of a carbon source into Vitamin C. In a preferred embodiment, a modified polynucleotide, polypeptide, vector and recombinant microorganism as described herein is used for improving the yield, production, and/or efficiency of the production of Vitamin C.

- 5 The terms "production" or "productivity" are art-recognized and include the concentration of the fermentation product (for example, Vitamin C) formed within a given time and a given fermentation volume (*e.g.*, kg product per hour per liter). 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
- 10 fermentation product). The term "yield" or is art-recognized and includes the efficiency of the conversion of the carbon source into the product (*i.e.*, Vitamin C). This is generally written as, for example, kg product per kg carbon source. By increasing the yield or production of the compound, the quantity of recovered molecules, or of useful recovered molecules of that compound in a given amount of culture over a given amount of time is
- 15 increased. The terms "biosynthesis" or a "biosynthetic pathway" are art-recognized and include the synthesis of a compound, preferably an organic compound, by a cell from intermediate compounds in what may be a multistep and highly regulated process. 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,
- 20 (*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 be unable to pass.
- 25 A suitable carbon source that can be converted directly into L-ascorbic acid may be selected from the D-glucose or D-sorbitol metabolization pathway such as, for example, D-glucose, D-sorbitol, L-sorbose, L-sorbose, D-gluconate, 2-keto-D-gluconate or 2,5-diketo-gluconate. Preferably, the substrate is selected from for instance D-glucose, D-sorbitol, L-sorbose or L-sorbose, more preferably from D-glucose, D-sorbitol or L-sorbose,
- 30 sorbose, and most preferably from D-sorbitol or L-sorbose.

L-ascorbic acid or Vitamin C as used interchangeably 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

35 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.

In one preferred embodiment, the present invention is related to a process for the production of Vitamin C wherein a nucleotide according to the invention or a modified
5 polynucleotide sequence as described above is introduced into a suitable microorganism, the recombinant microorganism is cultured under conditions that allow the production of Vitamin C in high productivity, yield, and/or efficiency, the produced fermentation product is isolated from the culture medium and optionally further purified.

This invention is further illustrated by the following examples which should not be
10 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 15 by PCR

Chromosomal DNA of *Gluconobacter oxydans* DSM 17078 were prepared from the cells cultivated at 30°C for 1 day in mannitol broth (MB) liquid medium consisting of 25 g/l mannitol, 5 g/l of yeast extract (Difco), and 3 g/l of Bactopeptone (Difco) by the method described by Sambrook et al (1989) "Molecular Cloning: A Laboratory Manual/Second
20 Edition", Cold Spring Harbor Laboratory Press).

A DNA fragment was 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 Expand High Fidelity PCR kit (Roche Diagnostics) and 10 ng of the chromosomal DNA was used in total volume of 100 µl according to the supplier's instruction to have the PCR product
25 containing SMS 22 DNA sequence (SEQ ID NO:1). The PCR product was recovered from the reaction and its correct sequence confirmed.

Example 2: Production of L-ascorbic acid from L-sorbose using resting cells grown on mannitol broth agar medium

IFO strains 3293, 3292, 3244, 3260, 3266, 3287, 3259, 13693, and 13773 as well as
30 *Acetobacter sp.* ATCC 15164 and *Gluconobacter oxydans* DSM 17078, a derivative of the strain IFO 3293, were used for the production of L-ascorbic acid from L-sorbose.

Strains IFO 13693 and IFO 13773 were grown at 27°C for 3 days on No. 350 medium containing 5 g/l Bactopeptone (Difco), 5 g/l yeast extract (Difco), 5 g/l glucose, 5 g/l

mannitol, 1 g/l $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 5 ml/l ethanol, and 15 g/l agar. All other *Acetobacter* strains and all *Gluconobacter* strains were grown at 27°C for 3 days on mannitol broth (MB) agar medium containing 25 g/l mannitol, 5 g/l yeast extract (Difco Laboratories, Detroit, Mich., USA), 3 g/l Bactopeptone (Difco), and 18 g/l of agar (Difco).

- 5 Cells were scraped from the agar plates, suspended in distilled water and used for resting cell reactions conducted at 30°C for 20 h in 5 ml tubes with shaking at 230 rpm. The reaction mixtures (0.5 ml) contained 1% L-sorbose, 0.3% NaCl, 1% CaCO_3 and cells at a final concentration of 10 absorbance units at 600 nanometers (OD_{600}). At the conclusion of the incubation period, the reaction mixtures were analyzed by high performance liquid chromatography (HPLC) using an Agilent 1100 HPLC system (Agilent Technologies, 10 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 was 0.004 M sulfuric acid, and the flow rate was 0.6 ml/min. Two signals were recorded using an UV detector (wavelength 254 nm) in 15 combination with a refractive index detector. In addition, the identification of the L-ascorbic acid was done using an amino-column (YMC-Pack Polyamine-II, YMC, Inc., Kyoto, Japan) with UV detection at 254 nm. The mobile phase was 50 mM $\text{NH}_4\text{H}_2\text{PO}_4$ and acetonitrile (40:60).

- An Agilent Series 1100 HPLC-mass spectrometry (MS) system was used to identify L- 20 ascorbic acid. The MS was operated in positive ion mode using the electrospray interface. The separation was carried out using a LUNA-C8(2) column (100 x 4.6 mm) (Phenomenex, Torrance, USA). The mobile phase was a mixture of 0.1% formic acid and methanol (96:4). L-Ascorbic acid eluted with a retention time of 3.1 minutes. The identity of the L-ascorbic acid was confirmed by retention time and the molecular mass of the 25 compound.

To exclude the presence of D-isoascorbic acid, the identification of L-ascorbic acid was additionally done by retention time using an amino-column (YMC-Pack Polyamine-II, YMC, Inc., Kyoto, Japan) with UV detection at 254 nm. The mobile phase was 50 mM $\text{NH}_4\text{H}_2\text{PO}_4$ and acetonitrile (40:60).

- 30 The *Gluconobacter* and *Acetobacter* strains produced L-ascorbic acid from L-sorbose as shown in Table 1.

Table 1. Production of L-ascorbic acid from L-sorbose

Strain	L-ascorbic acid (mg/L)
<i>G. oxydans</i> IFO 3293	1740

<i>G. oxydans</i> DSM 17078	570
<i>G. oxydans</i> IFO 3292	410
<i>G. oxydans</i> IFO 3244	1280
<i>G. frateurii</i> IFO 3260	50
<i>G. cerinus</i> IFO 3266	140
<i>G. oxydans</i> IFO 3287	60
<i>A. aceti</i> subsp. <i>Orleanus</i> IFO 3259	30
<i>A. aceti</i> subsp. <i>Xylinum</i> IFO 13693	40
<i>A. aceti</i> subsp. <i>Xylinum</i> IFO 13693	120
<i>Acetobacter</i> sp. ATCC 15164	310
Blank	Not detected

Blank; reaction was done in the reaction mixture without cells.

Example 3: L-Ascorbic acid production from L-sorbose and D-sorbitol in tube and flask fermentations

Cells of *G. oxydans* DSM 17078 were used to inoculate 4 ml of No. 3BD liquid medium and cultivated in a tube (18 mm diameter) at 30°C for 3 days with shaking at 220 rpm. 20 mg/l of L-ascorbic acid had accumulated at the end of the incubation period.

Cells of strain DSM 17078 were cultivated (in triplicate) in 50 ml of No. 5 medium containing 100 g/l D-sorbitol, 0.5 g/l glycerol, 15 g/l yeast extract (Difco), 2.5 g/l of $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, and 15 g/L of CaCO_3 in a 500 ml baffled shake flask at 30°C with shaking at 200 rpm. After 72 h of cultivation, the amounts of L-ascorbic acid measured by HPLC in the three flasks were 400, 500 and 680 mg/l.

Example 4: L-Ascorbic acid production from D-sorbitol in fed-batch fermentation

Cells of *G. oxydans* DSM 17078 were grown in 200 ml No. 5 medium containing 100 g/l D-sorbitol, 0.5 g/l glycerol, 15 g/l yeast extract (Fluka BioChemika, Buchs, Switzerland), 2.5 g/l $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ and 15 g/l CaCO_3 in a 2-l baffled shake flask at 30°C with shaking at 180 rpm. After 48 h, 150 ml of this culture was used to inoculate a 10-l bioreactor (B.

Braun ED10, Melsungen, Germany) previously prepared with 5.3 l medium containing 20 g/l D-sorbitol, 0.5 g/l glycerol, 15 g/l yeast extract (Fluka BioChemika, Buchs, Switzerland) and 2.5 g/l $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ and equipped with temperature, pH and dissolved oxygen sensors and control loops. Temperature was controlled at 30°C, pH was controlled at 6.0 by adding a 28% ammonia solution, airflow was 4.5 l/min and dissolved oxygen was controlled at 30% by a cascade with stirring speed (minimum 300 rpm). After 6 h process time, a 500 g/l sorbitol solution was fed at a rate of 25 g/h for a period of 44 h. After 96 h process time, about 1% substrate was left in the supernatant, and 950 mg/l L-ascorbic acid had been produced.

10 **Example 5: L-Ascorbic acid production from L-sorbose or L-sorbose with a cell membrane fraction**

Cells of *G. oxydans* DSM 17078 were cultivated in 100 ml of No. 3BD liquid medium in a 500 ml baffled shake flask at 30°C with shaking at 220 rpm for 3 days. The resulting culture was centrifuged at 500 rpm to remove CaCO_3 . The supernatant from this step was then centrifuged at 5,000 rpm to pellet the cells. The collected cells were suspended in 3 ml of 50 mM potassium phosphate buffer (pH 7.0) and the cells were disrupted by two passages through a French Pressure cell (SIM-AMINCO Spetronic Instruments, USA) at 900 psi. The resulting homogenate was first centrifuged at 5,000 rpm to remove cell debris. Then the supernatant was diluted to a final protein concentration of 3 mg of protein/ml. This diluted sample is designated as cell-free extract (CFE). The CFE was centrifuged at 100,000 x g for 60 min. The supernatant was discarded and the pellet was collected as the membrane fraction.

The reaction (200 μl) with the membrane fraction (100 μl) was carried out in 50 mM potassium phosphate buffer (pH 7.0), 30°C with shaking at 220 rpm for 15 h. The substrates tested were L-sorbose (1% final concentration) and L-sorbose (2% final concentration). The final protein concentration used in the reaction was 1.5 mg/ml. At the end of the incubation period, 680 mg/l and 10 mg/l of L-ascorbic acid had been produced from 1% L-sorbose and 2% L-sorbose, respectively.

30 **Example 6: Production of L-ascorbic acid from D-sorbitol, L-sorbose or L-sorbose using resting cells grown on 3BD agar medium**

Cells of *G. oxydans* DSM 17078 were grown at 27°C for 3 days on No. 3BD agar medium containing 70 g/l L-sorbose, 0.5 g/l glycerol, 7.5 g/l yeast extract (Difco), 2.5 g/l $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 10 g/l CaCO_3 and 18 g/l agar (Difco). The resting cell reactions (1 ml reaction mixture in 10 ml tube) were carried out with 2% D-sorbitol, 2% L-sorbose, or 1% L-sorbose at 30°C for 24 h as described in Example 2. Strain DSM 17078 produced

280, 400 and 1780 mg/l of L-ascorbic acid from D-sorbitol, L-sorbose, and L-sorbose, respectively.

Other reactions (0.5 ml reaction mixture in 10 ml tube) were carried out with DSM 17078 cells grown on No. 3BD agar medium in reaction mixtures containing 2% D-sorbitol, 2%
5 L-sorbose or 2% L-sorbose for 2 days as described in Example 2. Strain DSM 17078 produced 1.8, 2.0 and 5.1 g/l of L-ascorbic acid from D-sorbitol, L-sorbose, and L-sorbose, respectively.

A reaction using cells of *G. oxydans* IFO 3293 was carried out with 2% L-sorbose as described above. Strain IFO 3293 produced 5.7 g/l of L-ascorbic acid in 2 days.

10 **Example 7: Production of L-ascorbic acid from D-sorbitol using resting cells grown in liquid medium**

Cells of *G. oxydans* DSM 17078 were grown in 200 ml of No. 5 medium containing 100 g/l D-sorbitol, 0.5 g/l glycerol, 15 g/l yeast extract (Fluka BioChemika, Buchs, Switzerland), 2.5 g/l $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ and 15 g/l CaCO_3 in a 2-l baffled shake flask at 30°C
15 with shaking at 180 rpm. After 24 h, the culture was centrifuged at 3220 g (Eppendorf 5810R, Hamburg, Germany), and the cells were resuspended in 0.9% NaCl solution, centrifuged again at 3220 g and the cell pellet was used to inoculate one baffled 500 ml shake flask containing 50 ml of full growth medium (100 g/l D-sorbitol, 0.5 g/l glycerol, 15 g/l yeast extract, 2.5 g/l $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 15 g/l CaCO_3) and another baffled 500 ml shake
20 flask containing 50 ml production medium (100 g/l D-sorbitol, 3 g/l NaCl, 10 g/l CaCO_3). The initial cell density, measured as optical density at 600 nm (OD_{600}), in both flasks was 10. Both flasks were incubated at 30°C with shaking at 180 rpm. After 48 h, the cell suspension in growth medium and production medium had accumulated 1.06 and 1.18 g/l L-ascorbic acid, respectively. No additional growth was observed in full medium during
25 the incubation period time.

Example 8. Overexpression of the SMS 22 gene in *G. oxydans* DSM 17078

The SMS 22 gene obtained in Example 1 above was cloned in an *E. coli* vector pCR2.1-TOPO and used to transform *E. coli* TG1 to have a Apr transformant carrying pCR2.1TOPO-SMS 22. Then, the insert was amplified with a set of primers, P_fNdeI
30 [SEQIDNO: 3 with CCCAT at the 5'-end] and P_rHindIII [SEQIDNO: 4 with CCAAGCTT at the 5'-end], by PCR. Resulting PCR product was digested with NdeI and HindIII and the fragment was inserted together with P_{ct}E-SD (Shine-Dalgarno) fragment (WO 02/099095) as XhoI-NdeI fragment after EcoRI-NdeI P_{ct}E-SD fragment being inserted in pCR2.1-TOPO to have XhoI site upstream of the EcoRI site into pVK100 (ATCC37156)

digested with XhoI and HindIII to yield pVK-PcrtE-SDH. *E. coli* TG1 was transformed with the ligation product to have Tcr transformant carrying plasmid pVK-PcrtE-SMS 22, which was then used to transform *G. oxydans* DSM 17078 by electroporation to have Tc^r *G. oxydans* DSM 17078/pVK-PcrtE-SMS 22.

5 **Example 9: Production of L-ascorbic acid from D-sorbitol, L-sorbose or L-sorbosone using resting cells grown on 3BD agar medium containing 7% L-sorbose**

Cells of *G. oxydans* DSM 17078 and Tcr *G. oxydans* DSM 17078/pVK-PcrtE-SMS 22 were grown at 27°C for 3 days on No. 3BD agar medium containing 70 g/l D-sorbose, 0.5
10 g/l glycerol, 7.5 g/l yeast extract (Difco), 2.5 g/l MgSO₄·7H₂O, 10 g/l CaCO₃ and 18 g/l agar (Difco).

Cells were scraped from the agar plates, suspended in distilled water and used for resting cell reactions conducted at 30°C with shaking at 220 rpm. At the conclusion of the incubation period, the reaction mixtures were analyzed by high performance liquid
15 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 was 0.004 M sulfuric acid, and the flow rate was 0.6 ml/min. Two signals were recorded using an UV detector (wavelength 254 nm) in
20 combination with a refractive index detector. In addition, the identification of the L-ascorbic acid was done using an amino-column (YMC-Pack Polyamine-II, YMC, Inc., Kyoto, Japan) with UV detection at 254 nm. The mobile phase was 50 mM NH₄H₂PO₄ and acetonitrile (40:60).

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

A series of resting cell reactions (0.2 ml reaction mixture in 1.5 ml reaction tube) was carried out with 2% D-sorbitol, 2% L-sorbose, or 1% L-sorbosone, and all reaction mixtures further contained 0.3% NaCl, 1% CaCO₃ and cells at a final concentration of 10 absorbance units at 600 nanometers (OD₆₀₀). After 20 h incubation time, *G. oxydans* DSM

- 39 -

17078 produced 145, 212 and 1130 mg/l of L-ascorbic acid from D-sorbitol, L-sorbose, and L-sorbose, respectively. In comparison, strain Tcr *G. oxydans* DSM 17078/pVK-PcrtE-SMS 22 produced more than 300, 800 and 1200 mg/l of L-ascorbic acid, respectively.

Claims

1. A polynucleotide selected from the group consisting of:
 - (a) polynucleotides encoding a polypeptide comprising the amino acid sequence according to SEQ ID NO:2;
 - 5 (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;
 - 10 (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 SMS 22 polypeptide;
 - (e) polynucleotides the complementary strand of which hybridizes under stringent
15 conditions to a polynucleotide as defined in any one of (a) to (d) and which encode a SMS 22 polypeptide;
 - (f) polynucleotides which are at least 70%, such as 85, 90 or 95% homologous to a polynucleotide as defined in any one of (a) to (d) and which encode a SMS 22 polypeptide
or
20 the complementary strand of such a polynucleotide.
2. A vector containing the polynucleotide according to claim 1.
3. The vector of claim 2 in which the polynucleotide is operatively linked to expression control sequences allowing the expression in prokaryotic or eukaryotic host cells.
4. A microorganism genetically engineered with a polynucleotide according to claim 1
25 or with the vector of claim 2 or 3.

5. A microorganism capable of directly producing Vitamin C from D-sorbitol in quantities of 300 mg/l or more when measured in a resting cell method after an incubation period of 20 hours.
6. A microorganism capable of directly producing Vitamin C from L-sorbose in quantities of 800 mg/l or more when measured in a resting cell method after an incubation period of 20 hours.
7. A polypeptide encoded by a polynucleotide according to claim 1.
8. Process for producing cells capable of expressing a polypeptide according to claim 7, comprising the step of genetically engineering cells with the vector of claim 2 or 3 or with a polynucleotide according to claim 1.
9. Use of a polynucleotide according to claim 1 or a vector according to claims 2 or 3 for the production of Vitamin C.
10. Use according to claim 9, wherein the polynucleotide is operatively linked to expression control sequences and transferred into a microorganism.
11. Use according to claim 10, 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 efficiency of production of Vitamin C produced by said microorganism.
12. Use according to claim 11, 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 the SMS 22 protein.
13. Process for the production of an enhanced endogenous SMS 22 gene in a microorganism, said microorganism comprising a polynucleotide according to claim 1, said process comprising the step of altering said polynucleotide in such a way that it leads to an improved yield and/or efficiency of production of Vitamin C produced by said microorganism.
14. Process for the production of a polypeptide according to claim 7 in a microorganism, comprising the step of altering said microorganism so that the microorganism produces said polypeptide with increased SMS 22 activity leading to an improved yield and/or efficiency of production of Vitamin C produced by said microorganism.

15. Process for the production of a polypeptide according to claim 7 in a microorganism, comprising the step of altering said microorganism so that the microorganism produces said polypeptide with increased SMS 22 activity leading to an improved yield and/or efficiency of production of 2-KGA produced by said microorganism.
- 5 16 Process for the production of a microorganism capable of producing Vitamin C, comprising the step of altering said microorganism so that the microorganism produces a polypeptide with increased and/or improved SMS 22 activity leading to an improved yield and/or efficiency of production of Vitamin C produced by said microorganism.
- 10 17. Process for the production of a microorganism capable of producing Vitamin C comprising the step of altering said microorganism so that the microorganism produces a polypeptide with increased and/or improved SMS 22 activity leading to an improved yield and/or efficiency of production of 2-KGA produced by said microorganism.
- 15 18. Process for the production of a microorganism containing an endogenous gene comprising a polynucleotide according to claim 1, comprising the step of altering said microorganism so that the endogenous gene is overexpressed, leading to an improved yield and/or efficiency of production of Vitamin C produced by said microorganism.
19. Microorganism obtainable by a process according to any one of claims 13 to 18.
20. Microorganism according to claim 19 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.
- 25 21. Process for the direct production of Vitamin C from D-sorbitol or L-sorbose wherein a microorganism according to any one of claims 19, 20 or claims 4 to 6 is cultivated in a aqueous nutrient medium under conditions that allow the direct production of Vitamin C and wherein Vitamin C is isolated as the fermentation product.
- 30 22. Process for the production of Vitamin C with a microorganism according to claims 19, 20 or claims 4 to 6 wherein said microorganism is cultivated in a aqueous nutrient medium under conditions that allow the direct production of Vitamin C from D-sorbitol or L-sorbose and wherein Vitamin C and/or 2KGA is isolated as the fermentation product.

- 1 -

SEQUENCE LISTING

5 <110> DSM IP Assets B.V.
 <120> Novel Gene SMS 22
 <130> 24642
 10 <160> 4
 <170> PatentIn version 3.2.
 <210> 1
 15 <211> 1194
 <212> DNA
 <213> Gluconobacter oxydans DSM 17078
 <400> 1
 20 atgaccgcgc aagacacgcc ggccggactca ccgacgaaac gcgttcttgc cgtcgtcctg 60
 ttcaatcttc tcgtctatat cgtgatcggc ctgcccacgc ccgtcctgac gattttcgtc 120
 catcagggcc tgggctacag cacgaccgtc gcgggcatca cattctcgtc ccagtatttc 180
 25 gcgacttttg cagcccggcc ctccgccgga cggctgggtgg accggatcgg ccccaaacc 240
 gtggtcacgc gcgggctcgt cacctgtctg ctgtccggcc ttctggtgac cggatcggga 300
 30 ctgctcgcgg ccacaccgct gctgtcgtg ggctgtctgc tgtccagccg cctgctgctg 360
 ggatgggcgg aaagctgggc cgcgaccggg gtgatcgtct ggaacatccg gcgtgccggg 420
 gcggaaaata ccgccatcgc catttccttg aacggcatct gctcctatgg cgggattgctg 480
 35 ctgggtgcac cgatcggggc caagctgtcc atgtggccg caccctatgg cgggctgatt 540
 tcggtgggca tcctctcgtt ggccctgccg ctgttcggcc tggcgtgat tggcttctac 600
 40 cctgccgtca aaccatttc ctcgaccagc cagcccatc ctttctcccg cgcgctcaga 660
 ctgctcctcc cgcattggtc ggccctggcg gcagggtcca tcggattcgg ggcatctcg 720
 tcgttcctgg cgctgtatta ttccgtgctg cactggcagg gcgcggcgac ggcgctggcg 780
 45 gttttcggat tcactttcgt cgtcatccgg tttttcttcg cctcgcagat cgcaaagcgc 840
 ggcggcgtgt tcgtggcgat cgtctcgtg ctggtggaag cgctcggact cgccacgctc 900
 50 agcgtgttcc agacaccctc cgcggccctt gtgggcgtg cactgacggg agccggcttc 960
 tccctgctct tccctgccct cggcgttctg gccgttaaca gtgcgggtcc ggagaaccgc 1020
 ggggcggctc ttggggctta ttccatcttc attgaccttg ccatcgcgtg ctccgggctg 1080
 55 tttcttggac aggtcatcga gttcgccgtt tacaggacga tgttcgccac caccgccatc 1140

- 2 -

tgctgtctgg tcggtgtgct gatcagccgt tccctggggc gccggaccgc atga 1194

<210> 2
 5 <211> 397
 <212> PRT
 <213> Gluconobacter oxydans DSM 17078
 <400> 2
 10 Met Thr Ala Gln Asp Thr Pro Ala Asp Ser Pro Thr Lys Arg Val Leu
 1 5 10 15
 15 Ala Val Val Leu Phe Asn Leu Leu Val Tyr Ile Val Ile Gly Leu Pro
 20 25 30
 20 Asn Ala Val Leu Thr Ile Phe Val His Gln Gly Leu Gly Tyr Ser Thr
 35 40 45
 Thr Val Ala Gly Ile Thr Phe Ser Leu Gln Tyr Phe Ala Thr Phe Ala
 50 55 60
 25 Ala Arg Pro Ser Ala Gly Arg Leu Val Asp Arg Ile Gly Pro Lys Pro
 65 70 75 80
 30 Val Val Ile Gly Gly Leu Val Thr Cys Leu Leu Ser Gly Leu Leu Val
 85 90 95
 35 Thr Gly Ser Gly Leu Leu Ala Ala Thr Pro Leu Leu Ser Leu Gly Val
 100 105 110
 40 Leu Leu Ser Ser Arg Leu Leu Leu Gly Trp Ala Glu Ser Trp Ala Ala
 115 120 125
 Thr Gly Val Ile Val Trp Asn Ile Arg Arg Ala Gly Ala Glu Asn Thr
 130 135 140
 45 Ala Ile Ala Ile Ser Trp Asn Gly Ile Cys Ser Tyr Gly Gly Ile Ala
 145 150 155 160
 50 Leu Gly Ala Pro Ile Gly Ala Lys Leu Ser Met Leu Ala Ala Pro Tyr
 165 170 175
 55 Gly Gly Leu Ile Ser Val Gly Ile Leu Ser Val Ala Leu Pro Leu Phe
 180 185 190

- 3 -

Gly Leu Ala Leu Ile Gly Phe Tyr Pro Ala Val Lys Pro Ile Pro Ser
 195 200 205
 5
 Thr Ser Gln Pro Ile Pro Phe Ser Arg Ala Leu Arg Leu Val Leu Pro
 210 215 220
 10 His Gly Ser Ala Leu Ala Ala Gly Ser Ile Gly Phe Gly Ala Ile Ser
 225 230 235 240
 15 Ser Phe Leu Ala Leu Tyr Tyr Ser Val Leu His Trp Gln Gly Ala Ala
 245 250 255
 20 Thr Ala Leu Ala Val Phe Gly Phe Thr Phe Val Val Ile Arg Phe Phe
 260 265 270
 Phe Ala Ser Gln Ile Ala Lys Arg Gly Gly Val Phe Val Ala Ile Val
 275 280 285
 25 Ser Leu Leu Val Glu Ala Leu Gly Leu Ala Thr Leu Ser Val Phe Gln
 290 295 300
 30 Thr Pro Ser Ala Ala Leu Val Gly Ala Ala Leu Thr Gly Ala Gly Phe
 305 310 315 320
 35 Ser Leu Leu Phe Pro Ala Leu Gly Val Leu Ala Val Asn Ser Ala Gly
 325 330 335
 40 Pro Glu Asn Arg Gly Ala Ala Leu Gly Ala Tyr Ser Ile Phe Ile Asp
 340 345 350
 Leu Ala Ile Ala Cys Ser Gly Leu Phe Leu Gly Gln Val Ile Glu Phe
 355 360 365
 45 Ala Gly Tyr Arg Thr Met Phe Ala Thr Thr Ala Ile Cys Cys Leu Val
 370 375 380
 50 Gly Val Leu Ile Ser Arg Ser Leu Gly Arg Arg Thr Ala
 385 390 395
 <210> 3
 55 <211> 20
 <212> DNA
 <213> Artificial

- 4 -

<220>
<223> primer

5 <400> 3
atgaccgcgc aagacacgcc 20

<210> 4
10 <211> 20
<212> DNA
<213> Artificial

<220>
15 <223> primer

<400> 4
tcatgcggtc cggcggccca 20

20