

NEW MAJOR FACILITATOR SUPERFAMILY (MFS) PROTEIN (FRED) IN HMO PRODUCTION*FIELD*

5 The present invention relates to the field of recombinant production of biological molecules in genetically modified cells. More particularly, it relates to a method for recombinant production of human milk oligosaccharides (HMO) using a genetically modified cell expressing a protein of the major facilitator superfamily (MFS), the protein expressed being Fred.

BACKGROUND

10 Human milk represents a complex mixture of carbohydrates, fats, proteins, vitamins, minerals and trace elements. The by far most predominant fraction is represented by carbohydrates, which can be further divided into lactose and more complex oligosaccharides (Human milk oligosaccharides, HMO). Whereas lactose is used as an energy source, the complex oligosaccharides are not
15 metabolized by the infant. The fraction of complex oligosaccharides accounts for up to 1/10 of the total carbohydrate fraction and consists of probably more than 150 different oligosaccharides. The occurrence and concentration of these complex oligosaccharides are specific to humans and thus cannot be found in large quantities in the milk of other mammals, like for example domesticated dairy animals.

To date, the structures of at least 115 HMOs have been determined (see Urashima et al.: Milk
20 Oligosaccharides, Nova Biomedical Books, *New York*, 2011, ISBN: 978-1-61122-831-1), and considerably more are probably present in human milk. (Kunz C. *et al.*, (2014) Food Oligosaccharides: Production, Analysis and Bioactivity, 1st Edition, p 5-20, Eds. Moreno J. and Luz Sanz M., John Wiley & Sons, Ltd).

HMOs have become of great interest in the last decade, due to the discovery of their important
25 functionality in human development. Besides their prebiotic properties, HMOs have been linked to additional positive effects, which expands their field of application (Kunz C. *et al.*, (2014) Food Oligosaccharides: Production, Analysis and Bioactivity, 1st Edition, p 5-20, Eds. Moreno J. and Luz Sanz M., John Wiley & Sons, Ltd). The health benefits of HMOs have enabled their approval for use in foods, such as infant formulas and foods, and for consumer health products.

30 Due to the limited availability of HMOs, an effective commercial, i.e., large scale production is highly desirable. Chemical routes to some HMOs have been developed. However, production of HMOs by chemical synthesis, enzymatic synthesis or fermentation has proved to be challenging. The manufacturing of large-scale quantities as well as qualities, required for food and medical applications, through chemical synthesis, has yet to be provided. Furthermore, chemical synthetic

routes to HMOs involve several noxious chemicals, which impose a contamination risk to the final product.

To bypass the drawbacks associated with the chemical synthesis of HMOs, several enzymatic methods and fermentative approaches have been developed. Fermentation based processes have
5 been developed for several HMOs such as 2'-fucosyllactose, 3-fucosyllactose, lacto-N-tetraose, lacto-N-neotetraose, 3'-sialyllactose and 6'-sialyllactose. Fermentation based processes typically utilize genetically modified bacterial strains, such as recombinant *Escherichia coli* (*E. coli*).

Biotechnological production, such as a fermentation process, of HMOs is a valuable, cost-efficient and large-scale approach to HMO manufacturing. It relies on genetically modified bacteria
10 constructed so as to express the glycosyltransferases needed for synthesis of the desired oligosaccharides and takes advantage of the bacteria's innate pool of nucleotide sugars as HMO precursors.

Recent developments in biotechnological production of HMOs have made it possible to overcome certain inherent limitations of bacterial expression systems. For example, HMO-producing bacterial
15 cells may be genetically modified to increase the limited intracellular pool of nucleotide sugars in the bacteria (WO2012112777), to improve activity of enzymes involved in the HMO production (WO2016040531), or to facilitate the secretion of synthesized HMOs into the extracellular media (WO2010142305, WO2017042382). Further, expression of genes of interest in recombinant cells may be regulated by using particular promoters or other gene expression regulators, like e.g. what
20 has recently been described in WO2019123324.

The approach described in WO2010142305 and WO2017042382 has an advantage in that it allows to reduce the metabolic burden inflicted on the producing cell by high levels of recombinant gene expression, e.g., using methods of WO2012112777, WO2016040531 or WO2019123324. This approach attracts growing attention in recombinant HMO-producing cells engineering, e.g.,
25 recently there have been described several new sugar transporter genes encoding proteins and fermentation processes that can facilitate efflux of a recombinantly produced 2'-fucosyllactose (2'-FL), the most abundant HMO of human milk (WO2018077892, US201900323053, US201900323052).

However, at present, there is no algorithm that is able to pinpoint the right transporter protein
30 capable of efflux of different recombinantly produced HMO structures among numerous bacterial proteins with predicted transporter function in multiple protein databases, e.g., UniProt, since the structures-function relationship defining substrate specificity of sugar transporters is still not well-studied and remains to be highly unpredictable.

SUMMARY

The present disclosure shows that overexpression in HMO producing strains of the heterologous gene *fred*, which encodes the Fred protein from the Major facilitator superfamily (MFS), increases the amount of HMO produced by the cells. Identification of new efficient sugar efflux transporter
5 proteins having specificity for different recombinantly produced HMOs and development of recombinant cells expressing said protein are advantageous for high scale industrial HMO manufacturing.

In its broadest aspect, the present invention thus relates to a genetically modified cell capable of producing one or more Human Milk Oligosaccharides (HMOs), wherein said genetically modified
10 cell comprises a heterologous nucleic acid sequence encoding a major facilitator superfamily (MFS) polypeptide shown in SEQ ID NO: 1, or a functional homologue thereof which amino acid sequence is more than 95.4 % identical to SEQ ID NO: 1, such as at least 95.5, such as at least 96 % identical to SEQ ID NO: 1.

The genetically modified cell according to the present invention can further comprise a nucleic acid
15 sequence comprising a regulatory element for the regulation of the expression of the heterologous nucleic acid sequence. Said regulatory element in one aspect regulates the expression of the MFS polypeptide shown in SEQ ID NO: 1, or a functional homologue thereof which amino acid sequence is more than 95.4 % identical to SEQ ID NO: 1.

The amino acid sequence identified herein as SEQ ID NO: 1 is the amino acid sequence that is
20 100 % identical to the amino acid sequence having the GenBank accession ID WP_087817556.1. The MFS transporter protein having the amino acid sequence of SEQ ID NO: 1 is identified herein as "Fred protein" or "Fred transporter" or "Fred", interchangeably; a nucleic acid sequence encoding Fred protein is identified herein as SEQ ID NO: 2 "Fred coding nucleic acid/DNA" or "*fred* gene" or "*fred*".

The present invention shows that use of HMO producing recombinant cells that express Fred
25 protein results in very distinct improvements of the HMO manufacturing process related both to fermentation and purification of the HMOs. The disclosed herein recombinant cells and methods for HMO production provide both higher yields of total produced HMOs, lower by-product formation or by-product-to-product ratio and facilitated recovery of the HMOs during downstream processing
30 of the fermentation broth.

Surprisingly, expression of a DNA sequence encoding Fred in different HMO producing cells is found to be associated with accumulation of some particular HMOs in the extracellular media and other HMOs inside of the producing cells, and in an increase in total production of the HMOs. Surprisingly, an increase in the efflux of the produced HMOs is found to be characteristic for HMOs
35 that consist of either tri or tetra units of monosaccharides, i.e. HMOs that are trisaccharides and tetrasaccharides, e. g. 2'-fucosyllactose (2'-FL), 3-fucosyllactose (3-FL), 3'-sialyllactose (3'-SL), 6'-sialyllactose (6'-SL), lacto-N-triose 2, (LNT-2), lacto-N-neotetraose (LNnT) and lacto-N-tetraose

(LNT), especially for 2'-fucosyllactose (2'-FL), 3-fucosyllactose (3-FL), and lacto-N-tetraose (LNT) as seen from the Examples herein, but not for larger oligosaccharide structures, like pentasaccharides and hexasaccharides, which accumulate inside of the producing cells.

Surprisingly, it is also found that the total production of the major HMO, e.g. 2'-fucosyllactose (2'-FL), 3-fucosyllactose (3-FL), and lacto-N-tetraose (LNT), in the corresponding HMO producing cells expressing *fred* gene is also increased, while the by-product production, e.g. di-fucosyllactose (DFL), lacto-N-fucopentaose V (LNFP V) or para-lacto-neo-hexaose-I (pLNH-I), in these cells, correspondingly, is often decreased and said by-product oligosaccharides typically accumulate inside of the production cells.

10 A genetically modified cell according to the present invention typically further comprises a nucleic acid sequence comprising a regulatory element. The regulatory element can regulate the expression of the nucleic acid encoding a polypeptide of the Major facilitator superfamily (MFS), and can be selected from the group consisting of *PglpF*, *PglpF_SD4*, and *PglpF_SD7*.

The invention also relates to a nucleic acid construct comprising a nucleic acid sequence encoding
15 a polypeptide of the Major facilitator superfamily (MFS), wherein the nucleic acid sequence encoding the Major facilitator superfamily (MFS) polypeptide has at least 70 %, more preferably at least 80 %, more preferably at least 90 %, more preferably at least 95 % and even more preferably at least 99 % sequence identity to SEQ ID NO: 2, also the invention relates to a genetically modified cell comprising the nucleic acid construct, which is *Escherichia coli*.

20 In one aspect, the nucleic acid construct comprising a nucleic acid sequence(s) encoding a polypeptide of the Major facilitator superfamily (MFS), wherein the nucleic acid sequence is at least 70% identical to SEQ ID NO: 2.

The invention also provides a method for the biosynthetic production of one or more Human Milk Oligosaccharides (HMOs), the method comprising the steps of:

- 25 (i) providing a genetically modified cell according to the present invention;
- (ii) culturing the genetically modified cell according to (i) in a suitable cell culture medium to express said polypeptide capable of efflux sugar transportation and produce one or more Human Milk Oligosaccharides (HMOs) and;
- 30 (iii) harvesting one or more HMOs produced in step (ii).

The invention also relates to the use of a genetically modified cell or a nucleic acid construct comprising a heterologous nucleic acid sequence encoding a Major facilitator superfamily (MFS) polypeptide, said nucleic acid sequence having at least 70 % sequence identity to SEQ ID NO: 2, for the biosynthetic production of one or more Human Milk Oligosaccharides (HMOs).

As mentioned above, during the culturing of genetically modified cells capable of producing one or more HMOs, which cells comprise a nucleic acid sequence encoding a Fred transporter protein, it has surprisingly been found that the corresponding one or more HMOs are produced in high yields, while by-product and biomass formation is reduced. This facilitates recovery of the HMOs during downstream processes, e.g. the overall recovery and purification procedure may comprise less steps and overall time of purification may be shortened.

The effects of increased product yields and facilitation of the product recovery makes the present invention superior to the disclosures of the prior art.

Other aspects and advantageous features of the present invention are described in detail and illustrated by non-limiting working examples below.

FIGURES

Figure 1 shows the relative 2'-FL production of a modified *E. coli* strain (Strain 1) with and without integration of the *fred* gene according to SEQ ID NO: 2, encoding for the expression of the Fred MFS transporter protein according to SEQ ID NO: 1, Strain 2 and Strain 1, respectively. Data is obtained from a deep-well feed batch assay.

Figure 2 shows the relative distribution of 2'-FL inside and outside the cells of a modified *E. coli* strain (Strain 1) with and without integration of the *fred* gene according to SEQ ID NO: 2, encoding for the expression of the Fred MFS transporter protein according to SEQ ID NO: 1, Strain 2 and Strain 1, respectively. Data is obtained from a deep-well feed batch assay.

Figure 3 shows the relative 3-FL production of a modified *E. coli* strain (Strain 3) with and without integration of the *fred* gene according to SEQ ID NO: 2, encoding for the expression of the Fred MFS transporter protein according to SEQ ID NO: 1, Strain 4 and Strain 3, respectively. Data is obtained from a deep-well feed batch assay.

Figure 4 shows the relative distribution of 3-FL inside and outside the cells of a modified *E. coli* strain (Strain 3) with and without integration of the *fred* gene according to SEQ ID NO: 2, encoding for the expression of the Fred MFS transporter protein according to SEQ ID NO: 1, Strain 4 and Strain 3, respectively. Data is obtained from a deep-well feed batch assay.

Figure 5 shows the relative LNT production of a modified *E. coli* strain (Strain 5) with and without integration of the *fred* gene according to SEQ ID NO: 2, encoding for the expression of the Fred MFS transporter protein according to SEQ ID NO: 1, Strain 6 and 7, versus Strain 5, respectively. Data is obtained from a deep-well feed batch assay.

Figure 6 shows the relative distribution of LNT inside and outside the cells of a modified *E. coli* strain (Strain 5) with and without integration of the *fred* gene according to SEQ ID NO: 2, encoding

for the expression of the Fred MFS transporter protein according to SEQ ID NO: 1, Strain 6 and 7, versus Strain 5, respectively. Data is obtained from a deep-well feed batch assay.

Figure 7 shows the percentage (%) relative LNnT concentrations in total samples for seven strains that express GlcNacT and Gal4T genes and a heterologous transporter gene, *yberC0001_9420*,
5 *bad*, *fred*, *marc*, *vag*, *nec*, or *yabM*, correspondingly. The strain MP3672 expresses the glycosyltransferase genes and no transporter genes;

Figure 8 shows the percentage (%) relative LNnT concentrations in the supernatant of cells expressing *vag* or *yabM*.

10 DETAILED DESCRIPTION

In the following, embodiments of the invention will be described in further detail. Each specific variation of the features can be applied to other embodiments of the invention unless specifically stated otherwise.

Generally, all terms used herein are to be interpreted according to their ordinary meaning in the
15 technical field, and applicable to all aspects and embodiments of the invention, unless explicitly defined or stated otherwise.

All references to "a/an/the [cell, sequence, gene, transporter, step, etc]" are to be interpreted openly as referring to at least one instance of said cell, sequence, gene, transporter, step, etc., unless explicitly stated otherwise.

20 The steps of any method disclosed herein do not have to be performed in the exact order disclosed, unless explicitly stated.

The present invention in general relates to a genetically modified cell for efficient production of oligosaccharides and use of said genetically modified cell in a method of producing the oligosaccharides. In particular, the present invention relates to a genetically modified cell enabled
25 to synthesize an oligosaccharide, preferably a heterologous oligosaccharide, in particular a human milk oligosaccharide (HMO).

Accordingly, a genetically modified cell of the invention is modified to express a set of recombinant nucleic acids that are necessary for synthesis of one or more HMOs by the cells (which enable the host cell to synthesize one or more HMOs), such as genes encoding one or more enzymes with
30 glycosyltransferase activity as described below. The oligosaccharide producing recombinant cell of the invention is further modified to comprise a heterologous recombinant nucleic acid sequence, preferably, a DNA sequence, encoding a putative MFS (major facilitator superfamily) transporter protein, originating from the bacterium *Yersinia frederiksenii*.

Specifically, the invention relates to a genetically modified cell optimized for the production of one or more particular oligosaccharides, in particular one or more particular HMOs, comprising a recombinant nucleic acid encoding a Fred protein.

A nucleic acid sequence encoding a Fred protein having the nucleic acid sequence of SEQ ID NO:

5 2 is herein identified as "Fred coding nucleic acid/DNA" or "*fred* gene" or "*fred*".

The MFS transporter protein identified herein as "Fred protein" or "Fred transporter" or "Fred", interchangeably, has the amino acid sequence of SEQ ID NO: 1; The amino acid sequence identified herein as SEQ ID NO: 1 is an amino acid sequence that has 100 % identity with the amino acid sequence having the GenBank accession ID WP_087817556.1.

10 Accordingly, one aspect of the invention relates to a genetically modified cell capable of producing one or more Human Milk Oligosaccharides (HMOs), wherein said genetically modified cell comprises a heterologous nucleic acid sequence encoding a polypeptide capable of sugar transportation, said nucleic acid sequence having at least 70 %, more preferably at least 80 %, more preferably at least 90 %, more preferably at least 95 % and even more preferably at least 99
15 % sequence identity to SEQ ID NO: 2.

Further, the invention relates to a genetically modified cell optimized for the production of one or more particular oligosaccharides, in particular one or more particular HMOs, comprising a recombinant nucleic acid encoding a protein having more than 95.4%, such as at least 95.5 % sequence identity, preferably at least 96 %, more preferably at least 97 %, more preferably at least
20 98 %, and even more preferably at least 99 % sequence identity to the amino acid sequence of SEQ ID NO: 1.

Accordingly, one aspect of the invention relates to a genetically modified cell capable of producing one or more Human Milk Oligosaccharides (HMOs), wherein said genetically modified cell comprises a heterologous nucleic acid sequence encoding a polypeptide capable of sugar
25 transportation, said nucleic acid sequence having at least 70 %, more preferably at least 75 %, more preferably at least 80 % more preferably at least 90 % and even more preferably at least 95 %, such as 99% sequence identity to SEQ ID NO: 2.

Accordingly, another aspect of the invention relates to a genetically modified cell capable of producing one or more HMOs, wherein said cell comprises a recombinant nucleic acid encoding a
30 protein of SEQ ID NO: 1, or a functional homologue thereof which amino acid sequence is at least 95.4 % identical, preferably at least 96 %, more preferably at least 97 %, more preferably at least 99 % identical to SEQ ID NO: 1.

By the term "functional homolog" in the present context is meant a protein that has an amino acid sequence that is more than 95.4%, such as 95.5 %-99,9 % identical to SEQ ID NO: 1 and has a
35 function that is beneficial to achieve at least one advantageous effect of the invention, e.g. an increase of the total HMO production or of a selection of HMO by the genetically modified cell,

facilitate recovery of the produced HMO(s), HMO production efficiency and/or viability of an HMO producing cell.

By the term "Major Facilitator Superfamily (MFS)" is meant a large and exceptionally diverse family of the secondary active transporter class, which is responsible for transporting a range of different substrates, including sugars, drugs, hydrophobic molecules, peptides, organic ions, etc. The specificity of sugar transporter proteins is highly unpredictable and the identification of novel transporter protein with specificity towards for example oligosaccharides requires unburden laboratory experimentation (for more details see review by Reddy V.S. et al., (2012), FEBS J. 279(11): 2022–2035). The term "MFS transporter" means in the present context a protein that facilitates transport of an oligosaccharide, preferably, an HMO, through the cell membrane, preferably transport of an HMO/oligosaccharide synthesized by the genetically modified cell from the cell cytosol to the cell medium, preferably an HMO/oligosaccharide comprising three or four sugar units, e.g. 2'-FL, 3-FL, LNT-2, LNT, LNT, 3'-SL or 6'-SL. Additionally, or alternatively, the MFS transporter may also facilitate efflux of molecules that are not considered HMO or oligosaccharides according to the present invention, such as lactose, glucose, cell metabolites or toxins.

The term "sequence identity of [a certain] %" in the context of two or more nucleic acid or amino acid sequences means that the two or more sequences have nucleotides or amino acid residues in common in the given percent when compared and aligned for maximum correspondence over a comparison window or designated sequences of nucleic acids or amino acids (i.e. the sequences have at least 90 percent (%) identity). Percent identity of nucleic acid or amino acid sequences can be measured using a BLAST 2.0 sequence comparison algorithm with default parameters, or by manual alignment and visual inspection (see e.g. <http://www.ncbi.nlm.nih.gov/BLAST/>). This definition also applies to the complement of a test sequence and to sequences that have deletions and/or additions, as well as those that have substitutions. An example of an algorithm that is suitable for determining percent identity, sequence identity and for alignment is the BLAST 2.2.20+ algorithm, which is described in Altschul et al. Nucl. Acids Res. 25, 3389 (1997). BLAST 2.2.20+ is used to determine percent sequence identity for the nucleic acids and proteins of the invention. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). Examples of sequence alignment algorithms are CLUSTAL Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>), EMBOSS Needle (http://www.ebi.ac.uk/Tools/psa/emboss_needle/), MAFFT (<http://mafft.cbrc.jp/alignment/server/>) or MUSCLE (<http://www.ebi.ac.uk/Tools/msa/muscle/>).

In the context of the invention, the term "oligosaccharide" means a saccharide polymer containing a number of monosaccharide units. In some embodiments, preferred oligosaccharides are saccharide polymers consisting of three or four monosaccharide units, i.e. trisaccharides or tetrasaccharides. Preferable oligosaccharides of the invention are human milk oligosaccharides (HMOs).

HMO

The term "human milk oligosaccharide" or "HMO" in the present context means a complex carbohydrate found in human breast milk (for reference, see Urashima et al.: Milk

- 5 Oligosaccharides. Nova Science Publisher (2011); or Chen, Adv. Carbohydr. Chem. Biochem. 72, 113 (2015)). The HMOs have a core structure comprising a lactose unit at the reducing end that can be elongated by one or more β -N-acetyl-lactosaminy and/or one or more β -lacto-N-biosyl units, and this core structure can be substituted by an α -L-fucopyranosyl and/or an α -N-acetyl-neuraminy (sialyl) moiety. In this regard, the non-acidic (or neutral) HMOs are devoid of a sialyl
- 10 residue, and the acidic HMOs have at least one sialyl residue in their structure. The non-acidic (or neutral) HMOs can be fucosylated or non-fucosylated. Examples of such neutral non-fucosylated HMOs include lacto-N-triose 2 (LNT-2) lacto-N-tetraose (LNT), lacto-N-neotetraose (LNnT), lacto-N-neohexaose (LNnH), para-lacto-N-neohexaose (pLNnH), para-lacto-N-hexaose (pLNH) and lacto-N-hexaose (LNH). Examples of neutral fucosylated HMOs include 2'-fucosyllactose (2'-FL),
- 15 lacto-N-fucopentaose I (LNFP-I), lacto-N-difucohexaose I (LNDFH-I), 3-fucosyllactose (3-FL), difucosyllactose (DFL), lacto-N-fucopentaose II (LNFP-II), lacto-N-fucopentaose III (LNFP-III), lacto-N-difucohexaose III (LNDFH-III), fucosyl-lacto-N-hexaose II (FLNH-II), lacto-N-fucopentaose V (LNFP-V), lacto-N-difucohexaose II (LNDFH-II), fucosyl-lacto-N-hexaose I (FLNH-I), fucosyl-para-lacto-N-hexaose I (FpLNH-I), fucosyl-para-lacto-N-neohexaose II (F-pLNnH II) and fucosyl-
- 20 lacto-N-neohexaose (FLNnH). Examples of acidic HMOs include 3'-sialyllactose (3'-SL), 6'-sialyllactose (6'-SL), 3-fucosyl-3'-sialyllactose (FSL), 3'-O-sialyllacto-N-tetraose a (LST a), fucosyl-LST a (FLST a), 6'-O-sialyllacto-N-tetraose b (LST b), fucosyl-LST b (FLST b), 6'-O-sialyllacto-N-neotetraose (LST c), fucosyl-LST c (FLST c), 3'-O-sialyllacto-N-neotetraose (LST d), fucosyl-LST d (FLST d), sialyl-lacto-N-hexaose (SLNH), sialyl-lacto-N-neohexaose I (SLNH-I), sialyl-lacto-N-
- 25 neohexaose II (SLNH-II) and disialyl-lacto-N-tetraose (DSLNT). In the context of the present invention lactose is not regarded as an HMO species.

In some embodiments of the invention, tri-HMOs and tetra-HMOs may be a preferred, e.g. trisaccharides 2'-FL, 3-FL, LNT-2, 3'-SL, 6'-SL, and tetrasaccharides DFL, LNT, LNnT, FSL.

- 30 In a presently preferred aspect of the invention, tri-HMOs are preferred, in particular trisaccharides selected from 2'-FL and 3-FL, and tetrasaccharides selected from DFL and LNT.

2'-FL

- 2'-Fucosyllactose (2'-FL or 2'-O-fucosyllactose) is a trisaccharide, more precisely, fucosylated, neutral trisaccharide composed of L-fucose, D-galactose, and D-glucose units (Fuc α 1-2Gal β 1-4Glc). It is the most prevalent human milk oligosaccharide (HMO) naturally present in human
- 35 breast milk, making up about 30% of all of HMOs. In a genetically modified cell or in an enzymatic

reaction, 2'-FL is produced primarily by an α -1,2-fucosyltransferase enzymatic reaction with lactose and a fucosyl donor.

3-FL

- 3-Fucosyllactose (3-FL) is a trisaccharide, more precisely, fucosylated, neutral trisaccharide composed of D-galactose, L-fucose, and D-glucose ($\text{Gal}\beta 1-4(\text{Fuc}\alpha 1-3)\text{Glc}$). It is naturally present in human milk. In a genetically modified cell or in an enzymatic reaction, 3-FL is produced primarily by an α -1,3-fucosyltransferase or α -1,3/4-fucosyltransferase enzymatic reaction with lactose and a fucosyl donor.

LNT

- Lacto-N-tetraose (LNT) is a tetrasaccharide, more precisely, a neutral tetrasaccharide composed of galactose, N-acetylglucosamine, galactose, and glucose ($\text{GlcNAc}\beta 1-3\text{Gal}\beta 1-4\text{Glc}$). It is naturally present in human milk.

DFL

- Difucosyllactose (DFL or 2',3-di-O-fucosyllactose) is an oligosaccharide, more precisely, fucosylated neutral tetrasaccharide composed of L-fucose, D-galactose, L-fucose, and D-glucose ($\text{Fuc}\alpha 1-2\text{Gal}\beta 1-4(\text{Fuc}\alpha 1-3)\text{Glc}$). It is naturally present in human milk. In a genetically modified cell or in an enzymatic reaction, DFL is produced primarily by an α -1,2-fucosyltransferase, α -1,3-fucosyltransferase and/or α -1,3/4-fucosyltransferase enzymatic reaction with lactose and two fucosyl donors.

20 *Functional enzymes*

- To be able to synthesize one or more HMOs, the recombinant cell of the invention comprises at least one recombinant nucleic acid which encodes a functional enzyme with glycosyltransferase activity. The galactosyltransferase gene may be integrated into the genome (by chromosomal integration) of the genetically modified cell, or alternatively, it may be comprised in a plasmid DNA and expressed as plasmid-borne. If two or more glycosyltransferases are needed for the production of an HMO, e.g. LNT or LNnT, two or more recombinant nucleic acids encoding different enzymes with glycosyltransferase activity may be integrated in the genome and/or expressed from a plasmid, e.g. a β -1,3-N-acetylglucosaminyltransferase (a first recombinant nucleic acid encoding a first glycosyltransferase) in combination with a β -1,3-galactosyltransferase (a second recombinant nucleic acid encoding a second glycosyltransferase) for the production of LNT, where the first and second recombinant nucleic acid can independently from each other be integrated chromosomally or on a plasmid.

- In one preferred embodiment, both the first and second recombinant nucleic acids are stably integrated into the chromosome of the production cell; in another embodiment at least one of the first and second glycosyltransferase is plasmid-borne. A protein/enzyme with glycosyltransferase activity (glycosyltransferase) may be selected in different embodiments from enzymes having the

- activity of α -1,2-fucosyltransferase, α -1,3-fucosyltransferase, α -1,3/4-fucosyltransferase, α -1,4-fucosyltransferase, α -2,3-sialyltransferase, α -2,6-sialyltransferase, β -1,3-N-acetylglucosaminyltransferase, β -1,6-N-acetylglucosaminyltransferase, β -1,3-galactosyltransferase and β -1,4-galactosyltransferase. For example, the production of 2'-FL requires that the modified cell expresses an active α -1,2-fucosyltransferase enzyme; for the production of 3-FL the modified cell needs expression of an active α -1,3-fucosyltransferase enzyme; for the production of LNT the modified cell need to express at least two glycosyltransferases, a β -1,3-N-acetylglucosaminyltransferase and a β -1,3-galactosyltransferase; for the production of 6'-SL the modified cell has to express an active α -2,6-sialyltransferase enzyme and a pathway for CMP-sialic acid synthesis; for the production of 3'SL the modified cell has to express an active α -2,3-sialyltransferase enzyme and a pathway for CMP-sialic acid synthesis. Some non-limiting embodiments of proteins having glycosyltransferase activity, which can be encoded by the recombinant genes comprised by the production cell, can be selected from non-limiting examples of Table 1.

15 **Table 1**

Gene	Protein Sequence ID (GenBank)	Description	HMO example
<i>IgtA_Nm</i>	WP_002248149.1	β -1,3-N-acetylglucosaminyltransferase	LNT, LNnT, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, pLNH, F-pLNH I, pLNnH
<i>IgtA_Nm_MC58</i>	AAF42258.1	β -1,3-N-acetylglucosaminyltransferase	LNT, LNnT, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, pLNH, F-pLNH I, pLNnH
<i>IgtA_Hd</i>	AAN05638.1	β -1,3-N-acetylglucosaminyltransferase	LNT, LNnT, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, pLNH, F-pLNH I, pLNnH
<i>IgtA_Ng_PID2</i>	AAK70338.1	β -1,3-N-acetylglucosaminyltransferase	LNT, LNnT, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, pLNH, F-pLNH I, pLNnH
<i>IgtA_Ng_NCCP11945</i>	ACF31229.1	β -1,3-N-acetylglucosaminyltransferase	LNT, LNnT, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, pLNH, F-pLNH I, pLNnH

Gene	Protein Sequence ID (GenBank)	Description	HMO example
<i>lgtA_Past</i>	AAK02595.1	β -1,3-N-acetylglucosaminyl-transferase	LNT, LNnT, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, pLNH, F-pLNH I, pLNnH
<i>lgtA_Nc</i>	EEZ72046.1	β -1,3-N-acetylglucosaminyl-transferase	LNT, LNnT, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, pLNH, F-pLNH I, pLNnH
<i>lgtA_Nm_87255</i>	ELK60643.1	β -1,3-N-acetylglucosaminyl-transferase	LNT, LNnT, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, pLNH, F-pLNH I, pLNnH
<i>galT_Hp/HP0826</i>	NP_207619.1	β -1,4-galactosyltransferase	LNnT, LNFP-III, LNFP-VI, pLNH I, F-pLNH I, pLNnH
<i>galT_Nm/lgtB</i>	AAF42257.1	β -1,4-galactosyltransferase	LNnT, LNFP-III, LNFP-VI, pLNH I, F-pLNH I, pLNnH
<i>wbgO</i>	WP_000582563.1	β -1,3-galactosyltransferase	LNT, LNFP-I, LNFP-II, LNFP-V, LNDFH-I, LNDFH-II, pLNH, F-pLNH I
<i>cpsIBJ</i>	AB050723.1	β -1,3-galactosyltransferase	LNT, LNFP-I, LNFP-II, LNFP-V, LNDFH-I, LNDFH-II, pLNH, F-pLNH I
<i>jhp0563</i>	AEZ55696.1	β -1,3-galactosyltransferase	LNT, LNFP-I, LNFP-II, LNFP-V, LNDFH-I, LNDFH-II, pLNH, F-pLNH I
<i>galTK</i>	homologous to BD182026	β -1,3-galactosyltransferase	LNT, LNFP-I, LNFP-II, LNFP-V, LNDFH-I, LNDFH-II, pLNH, F-pLNH I
<i>futC</i>	WP_080473865.1	α -1,2-fucosyl-transferase	2'-FL, DFL, LNFP-I, LNDFH-I
<i>FucT2_HpUA802</i>	AAC99764.1	α -1,2-fucosyl-transferase	2'-FL, DFL, LNFP-I, LNDFH-I
<i>FucT2_EcO126t</i>	ABE98421.1	α -1,2-fucosyl-transferase	2'-FL, DFL, LNFP-I, LNDFH-I
<i>FucT2_Hm12198</i>	CBG40460.1	α -1,2-fucosyl-transferase	2'-FL, DFL, LNFP-I, LNDFH-I
<i>FucT2_Pm9515</i>	ABM71599.1	α -1,2-fucosyl-transferase	2'-FL, DFL, LNFP-I, LNDFH-I

Gene	Protein Sequence ID (GenBank)	Description	HMO example
<i>FucT2_HpF57</i>	<i>BAJ59215.1</i>	α -1,2-fucosyl-transferase	2'-FL, DFL, LNFP-I, LNDFH-I
<i>FucT6_3_Bf</i>	<i>CAH09151.1</i>	α -1,3-fucosyl-transferase	2'-FL, 3-FL, DFL, LNFP-I, LNFP-III, LNFP-V, LNFP-VI, LNDFH-II, F-pLNH I
<i>FucT7_3_Bf</i>	<i>CAH09495.1</i>	α -1,3-fucosyl-transferase	2'-FL, 3-FL, DFL, LNFP-I, LNFP-III, LNFP-V, LNFP-VI, LNDFH-II, F-pLNH I
<i>FucT_3_Am</i>	<i>ACD04596.1</i>	α -1,3-fucosyl-transferase	2'-FL, 3-FL, DFL, LNFP-I, LNFP-III, LNFP-V, LNFP-VI, LNDFH-II, F-pLNH I
<i>MAMA_R764</i>	<i>AGC02224.1</i>	α -1,3-fucosyl-transferase	2'-FL, 3-FL, DFL, LNFP-I, LNFP-III, LNFP-V, LNFP-VI, LNDFH-II, F-pLNH I
<i>Mg791</i>	<i>AEQ33441.1</i>	α -1,3-fucosyl-transferase	2'-FL, 3-FL, DFL, LNFP-I, LNFP-III, LNFP-V, LNFP-VI, LNDFH-II, F-pLNH I
<i>Moumou_00703</i>	<i>YP_007354660</i>	α -1,3-fucosyl-transferase	2'-FL, 3-FL, DFL, LNFP-I, LNFP-III, LNFP-V, LNFP-VI, LNDFH-II, F-pLNH I
<i>futA</i>	<i>NP_207177.1</i>	α -1,3-fucosyl-transferase	2'-FL, 3-FL, DFL, LNFP-I, LNFP-III, LNFP-V, LNFP-VI, LNDFH-II, F-pLNH I
<i>fucT</i>	<i>AAB81031.1</i>	α -1,3-fucosyl-transferase	2'-FL, 3-FL, DFL, LNFP-I, LNFP-III, LNFP-V, LNFP-VI, LNDFH-II, F-pLNH I
<i>fucTIII</i>	<i>AY450598.1</i>	α -1,4-fucosyl-transferase	LNDFH-I, LNDFH-II
<i>fucTa</i>	<i>AF194963.1</i>	α -1,3/4-fucosyl-transferase	LNFP-II, LNDFH-I, LNDFH-II
<i>Pd2,6ST</i>	<i>BAA25316.1</i>	α -2,6-sialyltransferase	6'-SL
<i>PspST6</i>	<i>BAF92026.1</i>	α -2,6-sialyltransferase	6'-SL
<i>PIST6_145</i>	<i>BAF91416.1</i>	α -2,6-sialyltransferase	6'-SL
<i>PIST6_119</i>	<i>BAI49484.1</i>	α -2,6-sialyltransferase	6'-SL
<i>NST</i>	<i>AAC44541.1</i>	α -2,3-sialyltransferase	3'-SL

Heterologous nucleic acid sequence

An aspect of the present invention is the provision of a nucleic acid construct comprising a heterologous nucleic acid sequence(s) encoding a polypeptide capable of sugar transportation
5 which is a major facilitator superfamily (MFS) polypeptide as shown in SEQ ID NO: 1, or a functional homologue thereof which amino acid sequence is more than 95.4 % identical to SEQ ID NO: 1, wherein the nucleic acid sequence encoding the sugar transportation polypeptide has at least 70 % sequence identity to SEQ ID NO: 2.

By the term "heterologous nucleic acid sequence", "recombinant gene/nucleic acid/DNA encoding"
10 or "coding nucleic acid sequence" is meant an artificial nucleic acid sequence (i.e. produced in vitro using standard laboratory methods for making nucleic acid sequences) that comprises a set of consecutive, non-overlapping triplets (codons) which is transcribed into mRNA and translated into a polypeptide when under the control of the appropriate control sequences, i.e. a promoter. The boundaries of the coding sequence are generally determined by a ribosome binding site located
15 just upstream of the open reading frame at the 5' end of the mRNA, a transcriptional start codon (AUG, GUG or UUG), and a translational stop codon (UAA, UGA or UAG). A coding sequence can include, but is not limited to, genomic DNA, cDNA, synthetic, and recombinant nucleic acid sequences.

The term "nucleic acid" includes RNA, DNA and cDNA molecules. It is understood that, as a result
20 of the degeneracy of the genetic code, a multitude of nucleotide sequences encoding a given protein may be produced. The term nucleic acid is used interchangeably with the term "polynucleotide".

An "oligonucleotide" is a short chain nucleic acid molecule.

"Primer" is an oligonucleotide, whether occurring naturally as in a purified restriction digest or
25 produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, (i.e., in the presence of nucleotides and an inducing agent such as DNA polymerase and at a suitable temperature and pH). The primer is preferably single stranded for maximum efficiency in amplification but may alternatively be double stranded. If double
30 stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is a deoxyribonucleotide sequence. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer and the use of the method.

35 The recombinant nucleic sequence of the invention may be a coding DNA sequence, e.g. a gene, or non-coding DNA sequence, e.g. a regulatory DNA, such as a promoter sequence. One aspect of

the invention relates to providing a recombinant cell comprising recombinant DNA sequences encoding enzymes necessary for the production of one or more HMOs and a DNA sequence encoding Fred transporter. Accordingly, in one embodiment the invention relates to a nucleic acid construct comprising a coding nucleic sequence, i.e. recombinant DNA sequence of a gene of interest, e.g. a glycosyltransferase gene or the *fred* gene, and a non-coding DNA sequence, e.g. a promoter DNA sequence, e.g. a recombinant promoter sequence derived from the promoter of *lac* operon or an *gfp* operon, or a promoter sequence derived from another genomic promoter DNA sequence, or a synthetic promoter sequence, wherein the coding and promoter sequences are operably linked. The term "operably linked" refers to a functional relationship between two or more nucleic acid (e.g., DNA) segments. Typically, it refers to the functional relationship of a transcriptional regulatory sequence to a transcribed sequence. For example, a promoter sequence is operably linked to a coding sequence if it stimulates or modulates the transcription of the coding sequence in an appropriate host cell or other expression system. Generally, promoter transcriptional regulatory sequences that are operably linked to a transcribed sequence are physically contiguous to the transcribed sequence, i.e., they are cis-acting.

In one embodiment, the nucleic acid construct of the invention may be a part of the vector DNA, in another embodiment the construct it is an expression cassette/cartridge that is integrated in the genome of a host cell. Accordingly, the term "nucleic acid construct" means an artificially constructed segment of nucleic acid, in particular a DNA segment, which is intended to be 'transplanted' into a target cell, e.g. a bacterial cell, to modify expression of a gene of the genome or express a gene/coding DNA sequence which may be included in the construct. In the context of the invention, the nucleic acid construct contains a recombinant DNA sequence comprising two or more recombinant DNA sequences: essentially, a non-coding DNA sequence comprising a promoter DNA sequence and a coding DNA sequence encoding a gene of interest, e.g. Fred protein, a glycosyltransferase, of another gene useful for production of an HMO in a host cell. Preferably, the construct comprises further non-coding DNA sequences that either regulate transcription or translation of the coding DNA of the construct, e.g. a DNA sequence facilitating ribosome binding to the transcript, a leading DNA sequence that stabilize the transcript.

Integration of the recombinant nucleic acid of interest comprised in the construct (expression cassette) into the bacterial genome can be achieved by conventional methods, e.g. by using linear cartridges that contain flanking sequences homologous to a specific site on the chromosome, as described for the attTn7-site (Waddell C.S. and Craig N.L., Genes Dev. (1988) Feb;2(2):137-49.); methods for genomic integration of nucleic acid sequences in which recombination is mediated by the Red recombinase function of the phage λ or the RecE/RecT recombinase function of the Rac prophage (Murphy, J Bacteriol. (1998);180(8):2063-7; Zhang et al., Nature Genetics (1998) 20: 123-128 Myrers et al., EMBO Rep. (2000) 1(3): 239-243); methods based on Red/ET recombination (Wenzel et al., Chem Biol. (2005), 12(3):349-56.; Vetcher et al., Appl Environ

Microbiol. (2005);71(4):1829-35); or positive clones, i.e. clones that carry the expression cassette, can be selected e.g. by means of a marker gene, or loss or gain of gene function.

A single copy of the expression cassette comprising a gene of interest may be sufficient to secure production of a desired HMO and achieve the desired effects according to the invention.

- 5 Accordingly, in some preferred embodiments, the invention relates to a recombinant HMO producing cell that comprises one, two or three copies of a gene of interest integrated in the genomic DNA of the cell. In some embodiments the single copy of the gene is preferred.

- In one preferred embodiment, recombinant coding nucleic acid sequence of the nucleic acid construct of the invention is heterologous with respect to the promoter, which means that in the
10 equivalent native coding sequence in the genome of species of origin is transcribed under control of another promoter sequence (i.e. not the promoter sequence of the construct). Still, with respect to the host cell, the coding DNA may be either heterologous (i.e. derived from another biological species or genus), such as e.g. the DNA sequence encoding Fred protein expressed in *Escherichia coli* host cells, or homologous (i.e. derived from the host cell), such as e.g. genes of
15 the colonic acid operon, the *wca* genes.

- Preferably, the construct of the invention comprising a gene related to biosynthetic production of an HMO, a promoter DNA sequence, and other regulatory sequences, such as a ribosomal binding site sequence (e.g. Shine-Dalgarno sequence), expressed in the genetically modified cell enables production of the HMO at the level of at least 0,03 g/OD (optical density) of 1 liter of the
20 fermentation media comprising a suspension of genetically modified cells, e.g., at the level of around 0.05 g/l/OD to around 0,5 g/l/OD. For the purposes of the invention, the later level of HMO production is regarded as "sufficient" and the genetically modified cell capable of producing this level of a desired HMO is regarded as "suitable genetically modified cell", i.e. the cell can be further modified to express the HMO transporter protein, e.g. Fred, to achieve at least one effect
25 described herein that is advantageous for the HMO production.

The genetically modified cell or the nucleic acid construct of the present invention comprises a nucleic acid sequence such as a heterologous gene encoding a putative MFS (major facilitator superfamily) transporter protein.

- An MFS transporter of particular interest in the present invention is Fred protein. A nucleic acid
30 construct of the present invention therefore contains a nucleic acid sequence having at least 70% sequence identity to the gene, *fred*, SEQ ID NO: 2.

The nucleic acid sequence contained in the genetically modified cell or in nucleic acid construct encodes for a protein of SEQ ID NO: 1, or a functional homologue thereof which amino acid sequence is more than 95.4 % identical to SEQ ID NO: 1.

- 35 A functional homologue of the protein of SEQ ID NO: 1, may be obtained by mutagenesis. The functional homologue should have a remaining functionality of at least 50%, such as 60%, 70%, 80

%, 90% or 100% compared to the functionality of amino acid sequence of SEQ ID NO: 1. The functional homologue can have a higher functionality compared to the functionality of amino acid sequence of SEQ ID NO: 1. The functional homologue of SEQ ID NO: 1, should be able to enhance HMO production of the genetically modified cell according to the invention.

5 *A genetically modified cell*

A "genetically modified cell" as used herein is understood as a cell which has been transformed or transfected, by a heterologous polynucleotide sequence. In the present context, the terms a "genetically modified cell" and "a host cell" are used interchangeably.

10 Accordingly, a "genetically modified cell" or "genetically modified cell" is in the present context understood as a host cell which has been transformed or transfected by an exogenous polynucleotide sequence.

The genetically modified cell is preferably a prokaryotic cell. Appropriate microbial cells that may function as a host cell include yeast cells, bacterial cells, archaeobacterial cells, algae cells, and fungal cells.

15 *Host cell*

The genetically modified cell (host cell or recombinant cell) may be e.g. a bacterial or yeast cell. In one preferred embodiment, the genetically modified cell is a bacterial cell.

Regarding the bacterial host cells, there are, in principle, no limitations; they may be eubacteria (gram-positive or gram-negative) or archaeobacteria, as long as they allow genetic manipulation for
 20 insertion of a gene of interest and can be cultivated on a manufacturing scale. Preferably, the host cell has the property to allow cultivation to high cell densities. Non-limiting examples of bacterial host cells that are suitable for recombinant industrial production of an HMO(s) according to the invention could be *Erwinia herbicola* (*Pantoea agglomerans*), *Citrobacter freundii*, *Pantoea citrea*, *Pectobacterium carotovorum*, or *Xanthomonas campestris*. Bacteria of the genus *Bacillus* may also
 25 be used, including *Bacillus subtilis*, *Bacillus licheniformis*, *Bacillus coagulans*, *Bacillus thermophilus*, *Bacillus laterosporus*, *Bacillus megaterium*, *Bacillus mycoides*, *Bacillus pumilus*, *Bacillus lentus*, *Bacillus cereus*, and *Bacillus circulans*. Similarly, bacteria of the genera *Lactobacillus* and *Lactococcus* may be modified using the methods of this invention, including but not limited to *Lactobacillus acidophilus*, *Lactobacillus salivarius*, *Lactobacillus plantarum*,
 30 *Lactobacillus helveticus*, *Lactobacillus delbrueckii*, *Lactobacillus rhamnosus*, *Lactobacillus bulgaricus*, *Lactobacillus crispatus*, *Lactobacillus gasseri*, *Lactobacillus casei*, *Lactobacillus reuteri*, *Lactobacillus jensenii*, and *Lactococcus lactis*. *Streptococcus thermophiles* and *Propionibacterium freudenreichii* are also suitable bacterial species for the invention described herein. Also included as part of this invention are strains, modified as described here, from the genera *Enterococcus*
 35 (e.g., *Enterococcus faecium* and *Enterococcus thermophiles*), *Bifidobacterium* (e.g., *Bifidobacterium longum*, *Bifidobacterium infantis*, and *Bifidobacterium bifidum*), *Sporolactobacillus*

spp., *Micromonospora spp.*, *Micrococcus spp.*, *Rhodococcus spp.*, and *Pseudomonas* (e.g., *Pseudomonas fluorescens* and *Pseudomonas aeruginosa*).

Bacteria comprising the characteristics described herein are cultured in the presence of lactose, and an oligosaccharide, such as an HMO, produced by the cell is retrieved, either from the
5 bacterium itself or from a culture supernatant of the bacterium. In one preferred embodiment, the genetically modified cell of the invention is an *Escherichia coli* cell.

In another preferred embodiment the host cell is a yeast cell e.g. *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Kluyveromyces lactis*, *Kluyveromyces marxianus*, etc.

Genetically modified cells of the invention can be provided using standard methods of the art e.g.
10 those described in the manuals by Sambrook et al., Wilson & Walker, "Maniatis et al.", and Ausubel et al.

A host suitable for the HMO production, e.g. *E. coli*, may comprise an endogenous β -galactosidase gene or an exogenous β -galactosidase gene, e.g. *E. coli* comprises an endogenous *lacZ* gene (e.g., GenBank Accession Number V00296 (GI:41901)). For the purposes of the invention, an
15 HMO-producing host cell is genetically manipulated to either comprise any β -galactosidase gene or to comprise the gene that is inactivated. The gene may be inactivated by a complete or partial deletion of the corresponding nucleic acid sequence from the bacterial genome, or the gene sequence is mutated in the way that it is transcribed, or, if transcribed, the transcript is not translated or if translated to a protein (i.e. β -galactosidase), the protein does not have the
20 corresponding enzymatic activity. In this way the HMO-producing bacterium accumulates an increased intracellular lactose pool which is beneficial for the production of HMOs.

In some embodiments, the engineered cell, e.g. bacterium, contains a deficient sialic acid catabolic pathway. By "sialic acid catabolic pathway" is meant a sequence of reactions, usually controlled and catalyzed by enzymes, which results in the degradation of sialic acid. An exemplary sialic acid
25 catabolic pathway described herein is the *E. coli* pathway. In this pathway, sialic acid (Neu5Ac; N-acetylneuraminic acid) is degraded by the enzymes NanA (N-acetylneuraminic acid lyase) and NanK (N-acetylmannosamine kinase) and NanE (N-acetylmannosamine-6-phosphate epimerase), all encoded from the *nanATEK-yhcH* operon, and repressed by NanR (<http://ecocyc.org/ECOLI>). A deficient sialic acid catabolic pathway is rendered in the *E. coli* host by introducing a mutation in
30 the endogenous *nanA* (N-acetylneuraminic acid lyase) (e.g., GenBank Accession Number D00067.1 (GL216588)) and/or *nanK* (N-acetylmannosamine kinase) genes (e.g., GenBank Accession Number (amino acid) BAE77265.1 (GL85676015)), and/or *nanE* (N-acetylmannosamine-6-phosphate epimerase, GI: 947745, incorporated herein by reference). Optionally, the *nanT* (N-acetylneuraminic acid transporter) gene is also inactivated or mutated. Other
35 intermediates of sialic acid metabolism include: (ManNAc-6-P) N-acetylmannosamine-6-phosphate; (GlcNAc-6-P) N-acetylglucosamine-6-phosphate; (GlcN-6-P) Glucosamine-6-phosphate, and (Fru-6-P) Fructose-6-phosphate. In some preferred embodiments, *nanA* is

mutated. In other preferred embodiments, *nanA* and *nanK* are mutated, while *nanE* remains functional. In another preferred embodiment, *nanA* and *nanE* are mutated, while *nanK* has not been mutated, inactivated or deleted. A mutation is one or more changes in the nucleic acid sequence coding the gene product of *nanA*, *nanK*, *nanE*, and/or *nanT*. For example, the mutation may be 1, 2, up to 5, up to 10, up to 25, up to 50 or up to 100 changes in the nucleic acid sequence. For example, the *nanA*, *nanK*, *nanE*, and/or *nanT* genes are mutated by a null mutation. Null mutations as described herein encompass amino acid substitutions, additions, deletions, or insertions, which either cause a loss of function of the enzyme (i.e. reduced or no activity) or loss of the enzyme (i.e. no gene product). By "deleted" is meant that the coding region is removed completely or in part such that no (functional) gene product is produced. By inactivated is meant that the coding sequence has been altered such that the resulting gene product is functionally inactive or encodes for a gene product with less than 100 %, e.g. 90 %, 80 %, 70 %, 60 %, 50 %, 40 %, 30 % or 20 % of the activity of the native, naturally occurring, endogenous gene product. A "not mutated" gene or protein does not differ from a native, naturally-occurring, or endogenous coding sequence by 1, 2, up to 5, up to 10, up to 20, up to 50, up to 100, up to 200 or up to 500 or more codons, or to the corresponding encoded amino acid sequence.

Furthermore, the bacterium (e.g., *E. coli*) may also comprise a sialic acid synthetic capability. For example, the bacterium comprises a sialic acid synthetic capability through provision of an exogenous UDP-GlcNAc 2-epimerase (e.g., *neuC* of *Campylobacter jejuni* (GenBank AAK91727.1) or equivalent (e.g. (GenBank CAR04561.1), a Neu5Ac synthase (e.g., *neuB* of *C. jejuni* (GenBank AAK91726.1) or equivalent, (e.g. *Flavobacterium limnosediminis* sialic acid synthase, GenBank WP_023580510.1), and/or a CMP-Neu5Ac synthetase (e.g., *neuA* of *C. jejuni* (GenBank AAK91728.1) or equivalent, (e.g. *Vibrio brasiliensis* CMP-sialic acid synthase, GenBank WP_006881452.1).

Production of neutral N-acetylglucosamine-containing HMOs in modified bacteria is also known in the art (see e.g. Gebus C et al. (2012) Carbohydrate Research 363 83-90).

For the production of N-acetylglucosamine-containing HMOs, such as Lacto-N-triose 2 (LNT-2), Lacto-N-tetraose (LNT), Lacto-N-neotetraose (LNnT), Lacto-N-fucopentaose I (LNFP-I), Lacto-N-fucopentaose II (LNFP-II), Lacto-N-fucopentaose III (LNFP-III), Lacto-N-fucopentaose V (LNFP-V), Lacto-N-difucohexaose I (LDFH-I), Lacto-N-difucohexaose II (LDFH-II), and Lacto-N-neodifucohexaose II (LNDFH-II), as described above, and it is modified to comprise an exogenous UDP-GlcNAc:Gal α / β -R β -3-N-acetylglucosaminyltransferase gene, or a functional variant or fragment thereof. This exogenous UDP-GlcNAc:Gal α / β -R β -3-N-acetylglucosaminyltransferase gene may be obtained from any one of a number of sources, e.g., the *lgtA* gene described from *N. meningitidis* (Genbank protein Accession AAF42258.1) or *N. gonorrhoeae* (Genbank protein Accession ACF31229.1). Optionally, an additional exogenous glycosyltransferase gene may be co-expressed in the bacterium comprising an exogenous UDP-GlcNAc:Gal α / β -R β -3-N-acetylglucosaminyltransferase. For example, a β -1,4-galactosyltransferase gene is co-expressed

with the UDP-GlcNAc:Gal α / β -R β -3-N-acetylglucosaminyltransferase gene. This exogenous β -1,4-galactosyltransferase gene can be obtained from any one of a number of sources, e.g., the one described from *N. meningitidis*, the *lgtB* gene (Genbank protein Accession AAF42257.1), or from *H. pylori*, the HP0826/*galT* gene (Genbank protein Accession NP_207619.1). Optionally, the additional exogenous glycosyltransferase gene co-expressed in the bacterium comprising an exogenous UDP-GlcNAc:Gal α / β -R β -3-N-acetylglucosaminyltransferase gene is a P-1,3-galactosyltransferase gene, e.g., that described from *E. coli* 055:H7, the *wbgO* gene (Genbank protein Accession WP_000582563.1), or from *H. pylori*, the *jhp0563* gene (Genbank protein Accession AEZ55696.1), or from *Streptococcus agalactiae* type Ib O12 the *cpsIBJ* gene (Genbank protein Accession AB050723). Functional variants and fragments of any of the enzymes described above are also encompassed by the disclosed invention.

A N-acetylglucosaminyltransferase gene and/or a galactosyltransferase gene, can also be operably linked to a *Pglp* and be expressed from the corresponding genome-integrated cassette. In one embodiment, the gene that is genome integrated is a gene encoding for a galactosyltransferase, e.g. HP0826 gene encoding for the GalT enzyme from *H. pylori* (Genbank protein Accession NP_207619.1); in another embodiment, the gene that is genome integrated is a gene encoding a β -1,3-N-acetylglucosaminyltransferase, e.g. *lgtA* gene from *N. meningitidis* (Genbank protein Accession AAF42258.1). In these embodiments, the second gene, i.e. a gene encoding a β -1,3-N-acetylglucosaminyltransferase or galactosyltransferase, correspondingly, may either be expressed from a genome-integrated or plasmid borne cassette. The second gene may optionally be expressed either under the control of a *glp* promoter or under the control of any other promoter suitable for the expression system, e.g. *Plac*.

HMO producing host cells typically comprise a functional *lacY* and a dysfunctional *lacZ* gene.

The HMOs produced by recombinant cells of the invention may be purified using a suitable procedure available in the art (e.g. such as described in WO2015188834, WO2017182965 or WO2017152918).

Encoded polypeptide capable of sugar transportation

Sugar transportation relates to the transport of a sugar, such as, but not limited to, an oligosaccharide, and in relation to the invention, influx and/or efflux transport of one/or more HMOs, from the cytoplasm or periplasm of a genetically modified cell to the production media and/or from the production media to the cytoplasm or periplasm of a genetically modified cell. Thus, a polypeptide expressed in the genetically modified cell, capable of transporting HMOs from the cytoplasm or periplasm to the production medium and/or from the production media to the cytoplasm or periplasm of a genetically modified cell is a polypeptide capable of sugar transportation. Thus, in the present invention, sugar transportation can mean efflux and/or influx transportation of sugar, such as, but not limited to, an oligosaccharide.

In that regard, the polypeptide capable of sugar transportation is a polypeptide belonging to the Major Facilitator Superfamily (MFS). In particular, the polypeptide has more than 95.4%, such as at least 95.5 %, such as at least 96%, 97%, 98% or 99% sequence identity with SEQ ID NO:1 or it is a functional variant thereof as described herein. SEQ ID NO:1 is the amino acid sequence of the Fred protein.

The genetically modified cell or the nucleic acid construct of the present invention comprises a nucleic acid sequence such as a heterologous gene encoding a Fred protein. Said nucleic acid sequence has at least 70% sequence identity to the *fred* gene as shown in SEQ ID No: 2, such as at least 75%, 80%, 85%, 90%, 95% or 99%.

The nucleic acid sequence construct encodes a protein of SEQ ID NO: 1, or a functional homologue thereof which amino acid sequence is more than 95.4%, such as at least 95.5 %, 96%, 97%, 98, or 99% identical to SEQ ID NO: 1.

A functional homologue of the protein of SEQ ID NO: 1, may be obtained by mutagenesis. The functional homologue should have a remaining functionality of at least 50%, such as 60%, 70%, 80 %, 90% or 100% compared to the functionality of amino acid sequence of SEQ ID NO: 1. The functional homologue can have a higher functionality compared to the functionality of amino acid sequence of SEQ ID NO: 1. The functional homologue of SEQ ID NO: 1, should be able to enhance HMO production of the genetically modified cell according to the invention.

The genetically modified cell or the nucleic acid construct may contain one or more nucleic acid sequences encoding the polypeptide capable of sugar transportation. More often, the genetically modified cell or the nucleic acid construct of the invention, encodes a single copy of a polypeptide capable of sugar transportation.

A single copy of the expression cassette comprising a gene of interest may be sufficient to secure production of a desired HMO and achieve the desired effects according to the invention.

Accordingly, in some preferred embodiments, the invention relates to a recombinant HMO producing cell that comprises one, two or three copies of a gene or genes of interest integrated in the genomic DNA of the cell. In some embodiments the single copy of the gene or genes is/are preferred.

The genetically modified cell or the nucleic acid construct, may also contain one or more regulatory elements for the regulation of the expression of a nucleic acid sequence encoding the sugar transportation polypeptide and wherein said nucleic acid sequence has at least 70 % sequence identity with SEQ ID NO: 2.

Regulatory element

As mentioned above, the genetically modified cell or the nucleic acid construct may further comprise a nucleic acid sequence comprising a regulatory element for the regulation of the

expression of the nucleic acid sequence having at least 70 % sequence identity to SEQ ID NO: 2. The nucleic acid sequence of the regulatory region may be heterologous or homologous.

The term, a "regulatory element" or "promoter" or "promoter region" or "promoter element" is a nucleic acid sequence that is recognized and bound by a DNA dependent RNA polymerase during initiation of transcription. The promoter, together with other transcriptional and translational regulatory nucleic acid sequences (also termed "control sequences") is necessary to express a given gene or group of genes (an operon). In general, the transcriptional and translational regulatory sequences include, but are not limited to, promoter sequences, ribosomal binding sites, transcriptional start and stop sequences, translational start and stop sequences, and enhancer or activator sequences. The "transcription start site" means the first nucleotide to be transcribed and is designated +1. Nucleotides downstream of the start site are numbered +2, +3, +4 etc., and nucleotides in the 5' opposite (upstream) direction are numbered -1, -2, -3 etc. The promoter of the construct can derive from a promoter region of any gene encoded in the genome of a species. Preferably, a promoter region of the genomic DNA of *E. coli*. Accordingly, any promoter that is able to bind to an RNA polymerase and initiate transcription is suitable for practicing the invention. In principle, any promoter can be used to control transcription of the recombinant gene, such as the MFS transporter or the glycosyltransferases of the invention. In carrying out the invention, different or identical promoter sequences may be used to drive transcription of different genes of interest integrated in the genome of the host cell or in expression vector DNA. In one example promoter sequence A is promoting the expression of the MFS transporter, and another promoter sequence B or identical promoter sequence A is promoting the expression of the glycosyltransferase.

To have an optimal expression of the recombinant genes included in the construct, the construct may comprise further regulatory sequences, e.g. a leading DNA sequence, such as a DNA sequence derived from 5'-untranslated region (5'UTR) of a *glp* gene of *E. coli*, a sequence for ribosomal binding. Examples of the later sequences are described in WO2019123324 (incorporated herein by reference) and are illustrated in non-limiting working examples herein.

In one aspect of the invention, one or more regulatory elements may be inserted into to the DNA construct encoding the *fred* gene according to SEQ ID NO: 2 and/or into to the DNA construct encoding one or more glycosyltransferases, according to the invention. In another aspect of the invention, one or more copies of the *fred* gene according to SEQ ID NO: 2 may be inserted into the DNA construct, according to the invention. In yet another aspect of the invention, the *fred* gene according to SEQ ID NO: 2 may be inserted into one or more identical or non-identical DNA constructs, according to the invention.

In one aspect of the invention, the regulatory element for the regulation of the expression of a recombinant gene included in the construct of the invention is *glpFKX* operon promoter, *PglpF*. In another aspect of the invention, the promoter is *lac* operon promoter, *Plac*. And in yet another aspect of the invention, the regulatory element is *PglpF_SD4* and/or *PglpF_SD7*, which are

modified version of the *PglpF* sequence comprising a modified ribosomal binding site sequence downstream of the promoter sequence. However, any promoter enabling transcription and/or regulation of the level of transcription, of one or more recombinant nucleic acids that encode one or more polypeptides according to the invention are suitable for practicing the invention.

- 5 Typically, promoters for the expression of a heterologous gene according to the present invention are selected from table 2.

Table 2.

Promoter name	Description	Reference	SEQ ID NO:
PgatY_70UTR	<i>E. coli</i> promoter for gatYZABCD; tagatose-1,6-bisP aldolase	NA	3
PglpF	<i>E. coli</i> promoter for glpFKX operon; glycerol uptake	WO2019123324	4
PglpF_SD1	<i>E. coli</i> promoter for glpFKX operon; glycerol uptake	WO2019123324	
PglpF_SD10	<i>E. coli</i> promoter for glpFKX operon; glycerol uptake	WO2019123324	
PglpF_SD2	<i>E. coli</i> promoter for glpFKX operon; glycerol uptake	WO2019123324	
PglpF_SD3	<i>E. coli</i> promoter for glpFKX operon; glycerol uptake	WO2019123324	
PglpF_SD4	<i>E. coli</i> promoter for glpFKX operon; glycerol uptake	WO2019123324	5
PglpF_SD5	<i>E. coli</i> promoter for glpFKX operon; glycerol uptake	WO2019123324	
PglpF_SD6	<i>E. coli</i> promoter for glpFKX operon; glycerol uptake	WO2019123324	
PglpF_SD7	<i>E. coli</i> promoter for glpFKX operon; glycerol uptake	WO2019123324	6
PglpF_SD8	<i>E. coli</i> promoter for glpFKX operon; glycerol uptake	WO2019123324	
PglpF_SD9	<i>E. coli</i> promoter for glpFKX operon; glycerol uptake	WO2019123324	
Plac_16UTR	<i>E. coli</i> promoter for lacZYA: lac operon	WO2019123324	
Plac	<i>E. coli</i> promoter for lacZYA: lac operon	WO2019123324	19
PmglB_70UTR	<i>E. coli</i> promoter for mglBAC; galactose/methyl-galactosidase transporter	NA	7
PmglB_70UTR_SD4	<i>E. coli</i> promoter for <i>mglBAC</i> ; galactose/methyl-galactosidase transporter	NA	8

The preferred regulatory elements present in a genetically modified cell or in a nucleic acid construct of the present invention, is selected from the group consisting of, *PgatY_70UTR*, *PglpF*, *PglpF_SD1*, *PglpF_SD10*, *PglpF_SD2*, *PglpF_SD3*, *PglpF_SD4*, *PglpF_SD5*, *PglpF_SD6*, *PglpF_SD7*, *PglpF_SD8*, *PglpF_SD9*, *Plac_16UTR*, *Plac*, *PmglB_70UTR* and

5 *PmglB_70UTR_SD4*.

Especially preferred regulatory elements present in a genetically modified cell or in a nucleic acid construct of the present invention, is selected from the group consisting of *PglpF*, *PglpF_SD4* and *PglpF_SD7*.

Genetically modified cells for biosynthetic production

10 In the present invention, promoters may be either necessary or beneficial for achieving an optimal level of biosynthetic production of one or more HMOs in the genetically modified cell and allowing to achieve the desired effects according to the invention. Thus, a promoter sequence, of this invention, enables transcription and/or regulates the expression of a polypeptide capable of sugar transportation and/or the glycosyltransferases of the invention, resulting in optimized biosynthesis

15 and transport of HMOs or HMO precursors and/or degradation of by-products of the HMO production.

In the genetically modified cell of the invention, a nucleic acid construct is comprised in the genetically modified cell, that encodes at least one gene related to biosynthetic production of one or more HMOs, a promoter DNA sequence, and other regulatory sequences, such as a ribosomal

20 binding site sequence (e.g. Shine-Dalgarno sequence). The expression of the gene or genes related to the biosynthetic production of one or more HMOs in a genetically modified cell, enables production of the HMOs, making the host cell a "suitable host cell" for carrying out the invention as described. In the genetically modified cell, the expression, of the gene or genes related to the biosynthetic production of HMOs, as mentioned above, enables the production of one or more

25 HMOs at level of 0,03 g//OD (optical density) from 1 liter of fermentation media comprising a suspension of the genetically modified cell s. Thus, the HMO level could be approx. 0.05 g//OD to approx. 0,5 g//OD, such as at least 0.4g/L/OD. For the purposes of the invention, the level of HMO production is regarded as "sufficient" and the genetically modified cell capable of producing this level of a desired HMO or mixture of HMOs is regarded as the suitable genetically modified cell for

30 carrying out the invention.

Thus, in. the light of the invention, the "suitable genetically modified cell " can be further modified, as described above, to express the sugar polypeptide capable of sugar transportation of the MFS family, e.g. fred, to achieve an, in one way or another, advantageous HMO production, such as but not limited to, a higher HMO level pr biosynthetic production, a higher level of purity in the

35 biosynthetic production, a faster production time and/or a more efficient biosynthetic production of HMOs.

Method of production

A second aspect of the invention related to a method for the production of one or more HMOs, the method comprising the steps of:

- 5 (i) providing a genetically modified cell capable of producing an HMO, wherein said cell comprises a recombinant nucleic acid encoding a protein of SEQ ID NO: 1, or a functional homologue thereof which amino acid sequence is more than 95.4%, such as at least 95.5 % identical, preferably at least 96 % identical, more preferably at least 99.9 % identical to SEQ ID NO: 1;
- 10 (ii) culturing the cell of (i) in a suitable cell culture medium to allow the HMO production and expression of the DNA sequence to produce the protein having the amino acid sequence of SEQ ID NO: 1, or a functional thereof which amino acid sequence is more than 95.4%, such as at least 95.5 % identical, preferably at least 96 % identical, more preferably at least 99.9 % identical to SEQ ID NO: 1;
- 15 (iii) harvesting the HMOs produced in step (ii).

According to the invention, the term "culturing" (or "cultivating" or "cultivation", also termed "fermentation") relates to the propagation of bacterial expression cells in a controlled bioreactor according to methods known in the industry.

To produce one or more HMOs, the HMO-producing bacteria as described herein are cultivated according to the procedures known in the art in the presence of a suitable carbon source, e.g. glucose, glycerol, lactose, etc., and the produced HMO is harvested from the cultivation media and the microbial biomass formed during the cultivation process. Thereafter, the HMOs are purified according to the procedures known in the art, e.g. such as described in WO2015188834, WO2017182965 or WO2017152918, and the purified HMOs are used as nutraceuticals, pharmaceuticals, or for any other purpose, e.g. for research.

Manufacturing of HMOs is typically accomplished by performing cultivation in larger volumes. The term "manufacturing" and "manufacturing scale" in the meaning of the invention defines a fermentation with a minimum volume of 5 L culture broth. Usually, a "manufacturing scale" process is defined by being capable of processing large volumes of a preparation containing the product of interest and yielding amounts of the protein of interest that meet, e.g. in the case of a therapeutic compound or composition, the demands for clinical trials as well as for market supply. In addition to the large volume, a manufacturing scale method, as opposed to simple lab scale methods like shake flask cultivation, is characterized by the use of the technical system of a bioreactor (fermenter) which is equipped with devices for agitation, aeration, nutrient feeding, monitoring and control of process parameters (pH, temperature, dissolved oxygen tension, back pressure, etc.). To a large extent, the behavior of an expression system in a lab scale method, such as shake flasks,

benchtop bioreactors or the deep well format described in the examples of the disclosure, does allow to predict the behavior of that system in the complex environment of a bioreactor.

With regard to the suitable cell medium used in the fermentation process, there are no limitations. The culture medium may be semi-defined, i.e. containing complex media compounds (e.g. yeast
5 extract, soy peptone, casamino acids, etc.), or it may be chemically defined, without any complex compounds.

By the term "one or more HMOs" is meant that an HMO production cell may be able to produce a single HMO structure (a first HMO) or multiple HMO structures (a second, a third, etc. HMO). In some embodiments, it may be preferred a genetically modified cell that produces a single HMO, in
10 other preferred embodiments, a genetically modified cell producing multiple HMO structures may be preferred. Non-limiting examples for genetically modified cells producing single HMO structures are 2'-FL, 3-FL, 3'-SL, 6'-SL or LNT-2 producing cells. Non-limiting examples of genetically modified cells capable of producing multiple HMO structures can be DFL, FSL, LNT LNnT, LNFP I, LNFP II, LNFP III, LNFP IV, LNFP V, pLNnH, pLNH2 producing cells.

15 In particular, the present invention relates to genetically modified cells producing one or more of the single HMO structures which are selected from the group consisting of 2'-FL, 3-FL, DFL and LNT producing cells.

The term "harvesting" in the context in the invention relates to collecting the produced HMO(s) following the termination of fermentation. In different embodiments it may include collecting the
20 HMO(s) included in both the biomass (i.e. the genetically modified cells) and cultivation media, i.e. before/without separation of the fermentation broth from the biomass. In other embodiments the produced HMOs may be collected separately from the biomass and fermentation broth, i.e. after/following the separation of biomass from cultivation media (i.e. fermentation broth). The separation of cells from the medium can be carried out with any of the methods well known to the
25 skilled person in the art, such as any suitable type of centrifugation or filtration. The separation of cells from the medium can follow immediately after harvesting the fermentation broth or be carried out at a later stage after storing the fermentation broth at appropriate conditions. Recovery of the produced HMO(s) from the remaining biomass (or total fermentation) include extraction thereof from the biomass (the production cells). It can be done by any suitable methods of the art, e.g. by
30 sonication, boiling, homogenization, enzymatic lysis using lysozyme, or freezing and grinding.

After recovery from fermentation, HMO(s) are available for further processing and purification.

Purification of HMOs produced by fermentation can be done using a suitable procedure described in WO2016095924, WO2015188834, WO2017152918, WO2017182965, US20190119314 (all incorporated by reference).

35 In some embodiments of the invention, a genetically modified cell may produce several HMOs, wherein one HMO is the "product" HMO and some/all the other HMOs are "by-product" HMOs.

Typically, by-product HMOs are either the major HMO precursors or products of further modification of the major HMO. In some embodiments, it may be desired to produce the product HMO in abundant amounts and by-product HMOs in minor amounts. Cells and methods for HMO production described herein allow for controlled production of an HMO product with a defined HMO profile, e.g. in one embodiment, the produced HMO mixture wherein the product HMO is a dominating HMO compared to the other HMOs (i.e. by-product HMOs) of the mixture, i.e. the product HMO is produced in higher amounts than other by-product HMOs; in other embodiments, the cell producing the same HMO mixture may be tuned to produce one or more by-product HMOs in higher amount than product HMO. For example, during the production of 2'-FL and/or 3-FL, the product HMO, often a significant amount of DFL, the by-product HMO, is produced. With the genetically modified cells of the present invention the level of DFL in the 2'-FL and/or 3-FL product can be significantly reduced.

Advantageously, the invention provides both a decreased ratio of by-product to product and an increased overall yield of the product (and/or HMOs in total). This, less by-product formation in relation to product formation facilitates an elevated product production and increases efficiency of both the production and product recovery process, providing superior manufacturing procedure of HMOs.

In different preferred embodiments, different genetically modified cells producing 2'-FL, 3-FL, DFL and/or LNT, as the product or by-product HMO, may be selected. In one preferred embodiment, the product is 3-FL. In another preferred embodiment, the product is 2'-FL and by-product is DFL. In another preferred embodiment, the product is LNT-2, and by-products are LNT and LNFP I.

Preferred HMO products of the invention as described are 2'-FL, LNT and 3-FL.

Use

The present invention also related to the use of a host cell as described herein for use in the production of one or more Human Milk Oligosaccharides (HMOs). In particular, the invention relates to the use of a host cell as described herein for the production of a specific HMO, wherein the host cell is selected with the aim of generating a majority of one specific HMO, preferably selected from 2'-FL, 3-FL and LNT, presently most preferably 3-FL.

General

It should be understood that any feature and/or aspect discussed above in connections with the described invention apply by analogy to the methods described herein.

The following figures and examples are provided below to illustrate the present invention. They are intended to be illustrative and are not to be construed as limiting in any way.

EXAMPLES

Materials and methods

- Unless otherwise noted, standard techniques, vectors, control sequence elements, and other expression system elements known in the field of molecular biology are used for nucleic acid manipulation, transformation, and expression. Such standard techniques, vectors, and elements can be found, for example, in: Ausubel *et al.* (eds.), *Current Protocols in Molecular Biology* (1995) (John Wiley & Sons); Sambrook, Fritsch, & Maniatis (eds.), *Molecular Cloning* (1989) (Cold Spring Harbor Laboratory Press, NY); Berger & Kimmel, *Methods in Enzymology* 152: Guide to Molecular Cloning Techniques (1987) (Academic Press); Bukhari *et al.* (eds.), *DNA Insertion Elements, Plasmids and Episomes* (1977) (Cold Spring Harbor Laboratory Press, NY); Miller, J.H. *Experiments in molecular genetics* (1972.) (Cold spring Harbor Laboratory Press, NY)

The embodiments described below are selected to illustrate the invention and are not limiting the invention in any way.

Strains

- Strains utilized in the present Examples are described in the following table 3.

Table 3.

Strain IDs	Product	Relevant Genotype
DH1	-	<i>F⁻ λ⁻ endA1 recA1 relA1 gyrA96 thi-1 glnV44 hsdR17(r^k m^k)</i>
MDO	-	<i>E coli DH1 ΔlacZ ΔlacA, ΔnanKETA, ΔmelA, ΔwcaJ, ΔmdoH wcaF::Plac</i>
Strain 1	2'-FL	<i>MDO PglpF-CA PglpF-futC ΔlacI</i>
Strain 2		<i>MDO PglpF-CA PglpF-futC ΔlacI PglpF-fred</i>
Strain 3	3-FL	<i>MDO PglpF-CA PglpF-futA ΔlacI</i>
Strain 4		<i>MDO PglpF-CA PglpF-futA ΔlacI PglpF-fred</i>
Strain 5	LNT	<i>MDO PglpF-IgtA PglpF-galTK</i>
Strain 6		<i>MDO PglpF-IgtA PglpF-galTK PglpF_SD4-fred</i>
Strain 7		<i>MDO PglpF-IgtA PglpF-galTK PglpF_SD7-fred</i>

MP3672	LNnT	<i>MDO PglpF-GlcNAcT PglpF-Gal4T (version 1)*</i>
MP3708		<i>MDO PglpF-GlcNAcT PglpF-Gal4T Plac-yberC0001_9420</i>
MP3972		<i>MDO PglpF-GlcNAcT PglpF-Gal4T Plac-bad</i>
MP3979		<i>MDO PglpF-GlcNAcT PglpF-Gal4T Plac-fred</i>
MP4011		<i>MDO PglpF-GlcNAcT PglpF-Gal4T Plac-marc</i>
MP3994		<i>MDO PglpF-GlcNAcT PglpF-Gal4T Plac-vag (version 1)**</i>
MP3984		<i>MDO PglpF-GlcNAcT PglpF-Gal4T Plac-nec</i>
MP4362		<i>MDO PglpF-GlcNAcT PglpF-Gal4T Plac-yabM</i>
MP3020		<i>MDO PglpF-GlcNAcT PglpF-Gal4T(version 2)*</i>
MP4064		<i>MDO PglpF-GlcNAcT PglpF-Gal4T PglpF-vag</i>
MP4065		<i>MDO PglpF-GlcNAcT PglpF-Gal4T Plac-vag (version 2)**</i>
MP4473	LNT	<i>MDO PglpF-GlcNAcT PglpF-Gal3T</i>
MP4546		<i>MDO PglpF-GlcNAcT PglpF-Gal3T PglpF-vag</i>

CA = *gmd-wcaG-wcaH-wcaI-manC-manB*

* Strains MP3672 and MP3020 bear the same heterologous GlcNAcT and Gal4T, but differ in the copy number of the corresponding GlcNAcT-encoding gene

5 ** Strains MP3994 and MP4065 bear the same heterologous GlcNAcT and Gal4T, but differ in the copy number of the corresponding GlcNAcT-encoding gene

Cultivation

Unless otherwise noted, *E. coli* strains were propagated in Basal Minimal medium containing 0.2 % glucose at 37°C with agitation. Agar plates were incubated at 37°C overnight.

Basal Minimal medium had the following composition: NaOH (1 g/L), KOH (2.5 g/L), KH₂PO₄ (7 g/L), NH₄H₂PO₄ (7 g/L), Citric acid (0.5 g/l), Trace mineral solution (5 mL/L). The trace mineral stock solution contained: ZnSO₄*7H₂O 0.82 g/L, Citric acid 20 g/L, MnSO₄*H₂O 0.98 g/L, FeSO₄*7H₂O 3.925 g/L, CuSO₄*5H₂O 0.2 g/L. The pH of the Basal Minimal Medium was adjusted to 7.0 with 5 N NaOH and autoclaved. Before inoculation, the Basal Minimal medium was supplied with 1 mM MgSO₄, 4 µg/mL thiamine, 0.5 % of a given carbon source (glycerol (Carbosynth)).

15 Thiamine, and antibiotics, were sterilized by filtration. All percentage concentrations for glycerol are expressed as v/v and for glucose as w/v.

Chemical competent cells and transformations

E. coli was inoculated from LB plates in 5 mL LB containing 0.2 % glucose at 37°C with shaking until OD₆₀₀ ~0.4. 2 mL culture was harvested by centrifugation for 25 seconds at 13.000 g. The supernatant was removed, and the cell pellet resuspended in 600 µL cold TB solutions (10 mM PIPES, 15 mM CaCl₂, 250 mM KCl). The cells were incubated on ice for 20 minutes followed by pelleting for 15 seconds at 13.000 g. The supernatant was removed, and the cell pellet resuspended in 100 µL cold TB solution. Transformation of plasmids were done using 100 µL

- competent cells and 1 to 10 ng plasmid DNA. Cells and DNA were incubated on ice for 20 minutes before heat shocking at 42°C for 45 seconds. After 2 min incubation on ice 400 µL SOC (20 g/L tryptone, 5 g/L Yeast extract, 0.5 g/L NaCl, 0.186 g/L KCl, 10 mM MgCl₂, 10 mM MgSO₄ and 20 mM glucose) was added and the cell culture was incubated at 37°C with shaking for 1 hour before plating on selective plates.

Plasmid were transformed into TOP10 chemical competent cells at conditions recommended by the supplier (ThermoFisher Scientific).

DNA techniques

- Plasmid DNA from *E. coli* was isolated using the QIAprep Spin Miniprep kit (Qiagen). Chromosomal DNA from *E. coli* was isolated using the QIAmp DNA Mini Kit (Qiagen). PCR products were purified using the QIAquick PCR Purification Kit (Qiagen). DreamTaq PCR Master Mix (ThermoFisher), Phusion U hot start PCR master mix (ThermoFisher), USER Enzym (New England Biolab) were used as recommended by the supplier. Primers were supplied by Eurofins Genomics, Germany. PCR fragments and plasmids were sequenced by Eurofins Genomics. Colony PCR was done using DreamTaq PCR Master Mix in a T100™ Thermal Cycler (Bio-Rad).

- The heterologous proteins expressed in the genetically modified HMO producing cells of this invention is described in table 4, the promoter elements which was used in the below exemplifications of the invention is described in table 5, and the oligos used for amplification of plasmid backbones, promoter, elements, and *fred* is described in table 6.

Table 4: The heterologous proteins expressed in the genetically modified HMO producing cells

Gene	Origin of Genes	Protein Accession Number	Protein Function
<i>futC</i>	<i>Helicobacter pylori</i> 26695	WP_080473865.1	α -1,2-fucosyl-transferase
<i>futA</i>	<i>Helicobacter pylori</i> 26695	WP_000487428.1	α -1,3-fucosyl-transferase
<i>lgtA</i>	<i>Neisseria meningitidis</i> 053442	WP_002248149.1	β -1,3-acetylglucosaminyltransferase
<i>galTK</i>	<i>Helicobacter pylori</i> 43504	BD182026	β -1,3-galactosyltransferase
<i>fred</i>	<i>Yersinia frederiksenii</i>	WP_087817556.1	MFS transporter

<i>yberC0001_9420</i>	<i>Yersinia bercovieri</i>	EEQ08298.1	Major facilitator superfamily MFS_1
<i>nec*</i>	<i>Rosenbergiella nectarea</i>	WP_092672081.1	MFS transporter
<i>marc*</i>	<i>Serratia marcescens</i>	WP_060448169.1	MFS transporter
<i>bad*</i>	<i>Rouxiiella badensis</i>	WP_017489914.1	MFS transporter
<i>vag*</i>	<i>Pantoea vagans</i>	WP_048785139.1	MFS transporter
<i>yabM</i>	<i>Erwinia pyrifoliae</i>	CAY73138.1	Putative MFS sugar efflux transporter

* the gene name is given to identify the nucleic acid encoding the protein having amino acid sequence of the corresponding GenBank Accession Number for the purposes of the present invention.

Table 5: The synthetic DNA elements utilized for expression of *fred*.

Sequence name	SEQ ID NO	Sequence (5'-3')	Description
<i>PglpF</i>	4	GCGGCACGCCTTGCAGATTACGGTTTGC CACACTTTTCATCCTTCTCCTGGTGACAT AATCCACATCAATCGAAATGTTAATAAAT TTGTTGCGCGAATGATCTAACAAACATGC ATCATGTACAATCAGATGGAATAAATGGC GCGATAACGCTCATTTTATGACGAGGCAC ACACATTTTAAGTTCGATATTTCTCGTTTT TGCTCGTTAACGATAAGTTTACAGCATGC CTACAAGCATCGTGGAGGTCCGTGACTTT CACGCATACAACAAACATTAACCAAGGAG GAAACAGCT	300-nucleotide DNA expression element
<i>PglpF_SD 4</i>	5	GCGGCACGCCTTGCAGATTACGGTTTGC CACACTTTTCATCCTTCTCCTGGTGACAT AATCCACATCAATCGAAATGTTAATAAAT TTGTTGCGCGAATGATCTAACAAACATGC ATCATGTACAATCAGATGGAATAAATGGC GCGATAACGCTCATTTTATGACGAGGCAC ACACATTTTAAGTTCGATATTTCTCGTTTT TGCTCGTTAACGATAAGTTTACAGCATGC CTACAAGCATCGTGGAGGTCCGTGACTTT	300-nucleotide DNA expression element

		CACGCATACAACAAACATTAACCAACTAG GAAACAGCT	
<i>PglpF_SD</i> 7	6	GCGGCACGCCTTGCAGATTACGGTTTGC CACACTTTTCATCCTTCTCCTGGTGACAT AATCCACATCAATCGAAAATGTTAATAAAT TTGTTGCGCGAATGATCTAACAAACATGC ATCATGTACAATCAGATGGAATAAATGGC GCGATAACGCTCATTTTATGACGAGGCAC ACACATTTTAAGTTCGATATTTCTCGTTTT TGCTCGTTAACGATAAGTTTACAGCATGC CTACAAGCATCGTGGAGGTCCGTGACTTT CACGCATACAACAAACATTAACCAAGAGC AAACAGCT	300-nucleotide DNA expression element
<i>fred</i>	2	ATGAAGAGCGCGCTGACCTTCAGCCGTC GTATTAACCCGGTTTTTCTGGCGTTCTTT GTGGTTGCGTTCCTGAGCGGTATTGCGG GTGCGCTGCAGGCGCCGACCCTGAGCCT GTTCTGAGCACCGAGGTGAAAGTTCGT CCGCTGTGGGTGGGCCTGTTCTACACCG TTAACGCGATTGCGGGTATCACCGTGAG CTTTGTTCTGGCGAAGCGTAGCGACCTG CGTGGCGATCGTCGTAAACTGATCCTGG TGTGCTACCTGATGGCGGTTGGTAACTG CCTGCTGTTGCGTTTAACCGTGACTATC TGACCCTGATTACCGCGGGCGTGCTGCT GGCGGCGGTTGCGAACACCGCGATGCC GCAGATTTTCGCGCTGGCGCGTGAGTAC GCGGATAACAGCGCGCGTGAAGTGGTTA TGTTTAGCAGCATTATGCGTGCGCAACTG AGCCTGGCGTGGGTTATCGGTCCGCCGC TGAGCTTCATGCTGGCGCTGAACTATGG CTTACCCTGATGTTTTGCATTGCGGCGG GTATCTTCGTGCTGAGCGCGCTGGTTGT GTGGTTTATTCTGCCGAGCGTGACGCGT GCGGAACCGGTTATGGATGCGCCGACCG TGGCGCAAGGCAGCCTGTTTCGCGGACAA GGATGTTCTGCTGCTGTTTATTGCGAGCA TGCTGATGTGGACCTGCAACACCATGTAC	DNA encoding MFS transporter

		ATCATTGATATGCCGCTGTATATCACCGC GAGCCTGGGTCTGCCGGAGCGTCTGGC GGGTCTGCTGATGGGCACCGCGGCGGG TCTGGAAATCCCGATTATGCTGCTGGCG GGCTACAGCGTGCCTCGTTTTGGCAAGC GTAAAATCATGCTGTTTCGCGGTGCTGGC GGGCGTTCTGTTTTATACCGGTCTGGTTC TGTTCAAGTTTAAAAGCGCGCTGATGCTG CTGCAGATTTCAACGCGATCTTTATTGG TATCGTGGCGGGTATCGGCATGCTGTAC TTCCAAGACCTGATGCCGGGTCGTGCGG GTGCGGCGACCAACCCTGTTTACCAACAG CATTAGCACCGGCGTTATCCTGGCGGGC GTGCTGCAAGGTGTTCTGACCGAAACCT GGGGTCACAACAGCGTGTATGTTATGGC GATGATTCTGGCGATCCTGAGCCTGATCA TTTGCGCGCGTGTGCGTGAAGCGTAA	
<i>Plac</i>	19	ATGCGCAAATTGTGAGTTAGCTCACTCATTAG GCACCCAGGCTTTACACTTTATGCTTCCGGC TCGTATGTTGTGTGGAATTGTGAGCGGATAAC AATTTACACAGGAAACAGCTATGACCATGAT TACGCCAAGCGCGCAATTAACCCTCACTAAAG GGAACAAAAGCTGGGTACCTAAGGAGGAAAC AGCT	195-nucleotide DNA expression element

Table 6: Oligos used for amplification of plasmid backbones, promoter, elements, and genes of interest, including *fred*.

Name	SEQ ID NO	Oligonucleotide Sequence 5'-3'	Description
O40	9	ATTAACCCUCCAGGCATCAAATAAACGAAAGGC	Backbone.for
O79	10	ATTTGCGCAUCACCAATCAAATTCACGCGGC	Backbone.rev
O261	11	ATGCGCAAUAGCGGCACGCCTTGCAGATTACG	<i>PglpF</i> .for
O262	12	AGCTGTTUCCTCCTTGGTTAATGTTTGTGTATGCG	<i>PglpF</i> .rev
O459	13	AGCTGTTUCCTAGTTGGTTAATGTTTGTGTATGCG	<i>PglpF_SD4</i>

O462	14	AGCTGTTUTGCTCTTGGTTAATGTTTGTGTA TGCG	<i>PglpF_SD7</i>
KABY733	15	AAACAGCUATGAAGAGCGCGCTGACCTTCA G	<i>fred.for</i>
KABY734	16	AGGGTTAAUTTACGCTTCACGCACACGCG	<i>fred.rev</i>
O48	17	CCCAGCGAGACCTGACCGCAGAAC	<i>galK.for</i>
O49	18	CCCCAGTCCATCAGCGTGACTACC	<i>galK.rev</i>
O68	20	ATGCGCAAAUTGTGAGTTAGCTCACTCATTAG	<i>Plac.for</i>
O113	21	AGCTGTTUCCTCCTTAGGTACCCAGCTTTTGTTC C	<i>Plac.rev</i>
KABY745	22	AAACAGCUATGAAGAGCCTGCTGACCCGTAAAC	<i>vag.for</i>
KABY746	23	AGGGTTAAUTTAACGTTTTTCACACGCGCG	<i>vag.rev</i>
KABY721	24	AAACAGCUATGAAGAGCGCGCTGACCTTTAGC	<i>yberC0001_942 0.for</i>
KABY722	25	AGGGTTAAUTTACGCCTCACGCACACGCG	<i>yberC0001_942 0.rev</i>
KABY729	26	AAACAGCUATGAGCAGCCGTCGTCTGAGC	<i>bad.for</i>
KABY730	27	AGGGTTAAUTTACACGTTTTTAACACGGGTCATCA G	<i>bad.rev</i>
KABY741	28	AAACAGCUATGCAGAGCTTCACCCCGCC	<i>nec.for</i>
KABY742	29	AGGGTTAAUTTACGCCTGCTCTTTAACACGCAGC	<i>nec.rev</i>
KABY737	30	AAACAGCUATGCAGCGTCTGAGCCGTCTGAG	<i>marc.for</i>
KABY738	31	AGGGTTAAUTTAACCTTCACGCACTTTTCGCGC	<i>marc.rev</i>
MP1217	32	AAACAGCUATGAAGGCGCTGTGGAGCCGTCG	<i>yabM.for</i>
MP1218	33	AGGGTTAAUCGCCAGCGGAACGCTCTTCACG	<i>yabM.rev</i>

Construction of Plasmids

Plasmid backbones containing two I-SceI endonuclease sites, separated by two DNA fragments appropriated for homologous recombination into the *E. coli* genome and a T1 transcriptional terminator sequence were synthesized. For example, in one plasmid backbone (pUC57::gal) the *gal* operon (required for homologous recombination in *galK*), and a T1 transcriptional terminator sequence was synthesized (GeneScript). The DNA sequences used for homologous recombination in the *gal* operon covered base pairs 3.628.621-3.628.720 and 3.627.572-3.627.671 in sequence *Escherichia coli* K-12 MG155 complete genome GenBank: ID: CP014225.1. Insertion by homologous recombination would result in a deletion of 949 base pairs of *galK* and a *galK*-phenotype. In similar ways backbones based on pUC57 (GeneScript) or an any other appropriated vector containing two I-SceI endonuclease sites, separated by two DNA fragments appropriated for homologous recombination into the *E. coli* genome and a T1 transcriptional terminator sequence could be synthesized. Standard techniques well-known in the field of molecular biology were used for designing of primers and amplification of specific DNA

- sequences of the *Escherichia coli* K-12 DH1 chromosomal DNA. Such standard techniques, vectors, and elements can be found, for example, in: Ausubel et al. (eds.), Current Protocols in Molecular Biology (1995) (John Wiley & Sons); Sambrook, Fritsch, & Maniatis (eds.), Molecular Cloning (1989) (Cold Spring Harbor Laboratory Press, NY); Berger & Kimmel, Methods in
- 5 Enzymology 152: Guide to Molecular Cloning Techniques (1987) (Academic Press); Bukhari et al. (eds.).

Chromosomal DNA obtained from *E. coli* DH1 was used to amplify a 300 bp DNA fragment containing the promoter *PglpF* using oligos O261 and O262, *PglpF_SD4* using oligos O261 and O459, and *PglpF_SD7* using oligos O261 and O462 (Table 6).

- 10 A 1.182 bp DNA fragment containing a codon optimized version of the *fred* gene, SEQ ID NO: 2, originating from *Yersinia frederiksenii* was synthesized by Genescript (Table 4). The *fred* gene was amplified using oligos KABY733 and KABY734.

- All PCR fragments (plasmid backbones, promoter containing elements and the *fred* gene) were purified, and plasmid backbones, promoter elements (*Plac*, *PglpF*, *PglpF_SD4* or *PglpF_SD7*), and
- 15 a *fred* (or other gene of interest, see table 4) containing DNA fragment were assembled. The plasmids were cloned by standard USER cloning. Cloning in any appropriated plasmid could be done using any standard DNA cloning techniques. The plasmids were transformed into TOP10 cells and selected on LB plates containing 100 µg/mL ampicillin (or any appropriated antibiotic) and 0.2 % glucose. The constructed plasmids were purified and the promoter sequence and the 5'end
- 20 of the *fred* gene was verified by DNA sequencing (MWG Eurofins Genomics). In this way, a genetic cassette containing any promoter of interest linked to the *fred* (or other gene of interest, see table 4) gene was constructed.

Construction of strains

- The bacterial strain used, MDO, was constructed from *Escherichia coli* K-12 DH1. The *E. coli* K-12 DH1 genotype is: *F⁻*, *λ⁻*, *gyrA96*, *recA1*, *relA1*, *endA1*, *thi-1*, *hsdR17*, *supE44*. In addition to the *E. coli* K-12 DH1 genotype MDO has the following modifications: *lacZ*: deletion of 1.5 kbp, *lacA*: deletion of 0.5 kbp, *nanKETA*: deletion of 3.3 kbp, *melA*: deletion of 0.9 kbp, *wcaJ*: deletion of 0.5 kbp, *mdoH*: deletion of 0.5 kbp, and insertion of *Plac* promoter upstream of the *gmd* gene.
- 25

- Insertion of an expression cassette containing a promoter linked to the *fred* gene and to a T1 transcriptional terminator sequence in the chromosomal DNA of *E. coli* K-12 DH1 MDO was performed by Gene Gorging essentially as described by Herring et al. (Herring, C.D., Glasner, J.D. and Blattner, F.R. (2003). *Gene* (311). 153-163). Briefly, the donor plasmid and the helper plasmid were co-transformed into MDO and selected on LB plates containing 0.2 % glucose, ampicillin (100 µg/mL) or kanamycin (50 mg/mL) and chloramphenicol (20 µg/mL). A single colony was inoculated
- 30
- 35 in 1 mL LB containing chloramphenicol (20 µg/mL) and 10 µL of 20 % L-arabinose and incubated at 37 °C with shaking for 7 to 8 hours. For integration in the *galK* loci of *E. coli* cells were then plated on M9-DOG plates and incubated at 37 °C for 48 hours. Single colonies formed on MM-

- DOG plates were re-streaked on LB plates containing 0.2 % glucose and incubated for 24 hours at 37 °C. Colonies that appeared white on MacConkey-galactose agar plates and were sensitive for both ampicillin and chloramphenicol were expected to have lost the donor and the helper plasmid and contain an insertion in the *galK* loci. Insertions in the *galK* site was identified by colony PCR
- 5 using primers O48 (SEQ ID NO: 17) and O49 (SEQ ID NO: 18) and the inserted DNA was verified by sequencing (Eurofins Genomics, Germany).

Insertion of genetic cassettes at other loci in the *E. coli* chromosomal DNA was done in a similar way using different selection marker genes and different screening methods.

Deep well assay

- 10 The deep well assay was performed as originally described to Lv et al (Bioprocess Biosyst Eng (2016) 39:1737–1747) and optimized for the purposes of the current invention.
- More specifically, the strains disclosed in the examples were screened in 24 deep well plates using a 4-day protocol. During the first 24 hours, shaker culture cells were grown at 34 °C with 700 rpm shaking, to high densities while in the next 48 hours for 2'-FL and 3-FL producing cells, and 72
- 15 hours for LNT producing cells, cells were transferred to a medium that allowed induction of gene expression and product formation. Specifically, during day 1 fresh inoculums were prepared using a basal minimal medium supplemented with magnesium sulphate, thiamine and glucose. After 24 hours of incubation, cells were transferred to a new basal minimal medium (2 ml) supplemented with magnesium sulphate and thiamine with addition of an initial bolus consisting of 20 % glucose
- 20 solution (1 µl) and 10 % lactose solution (0.1 ml) were added, then 50 % sucrose solution (0.04 ml) as carbon source was provided to the cells accompanied by the addition of sucrose hydrolase (invertase, 5 µl of a 0.1 g/L solution) so that glucose was provided at a slow rate for growth by cleavage of sucrose by the invertase. After inoculation of the new medium, cells were shaken at 700 rpm at 28 °C for 48 hours. After denaturation and subsequent centrifugation, the supernatants
- 25 were analysed by HPLC. For the analysis of total samples, the cell lysate prepared by boiling was pelleted by centrifugation for 10 minutes at 4.700 rpm. The HMO concentration in the supernatant was determined by HPLC or HPAC methods.

Example 1 – Production of 2'-FL

- 30 *Engineering of Escherichia coli for 2'-FL production expressing the fred gene*
- The *Escherichia coli* K-12 (DH1) MDO strain can be manipulated to express heterologous genes of interest. For instance, Strain 1 is a 2'-FL producing strain overexpressing the α1,2-fucosyltransferase gene, *futC*, and the colonic acid genes (*gmd-wcaG-wcaH-wcaI-manC-manB*). Insertion of an expression cassette containing a promoter element (*PglpF*) linked to *fred* in a single
- 35 copy into the Strain 1 background resulted in Strain 2. Results from deep-well assays showed that expression of the *fred* gene using *PglpF* i) increase the 2'-FL production by 1.15-fold (15%

increase in the total 2'-FL production and ii) improves the 2'-FL product distribution by lowering the amount of 2'-FL in the cell fraction and increasing the amount of 2'-FL in the media.

Example 2 – Production of 3-FL

5 Engineering of *Escherichia coli* for 3-FL production expressing the *fred* gene

The *Escherichia coli* K-12 (DH1) MDO strain can be manipulated to express heterologous genes of interest. For instance, Strain 3 is a 3-FL producing strain overexpressing the α -1,3-fucosyltransferase gene, *futA*, and the colonic acid genes (*gmd-wcaG-wcaH-wcaI-manC-manB*). Insertion of an expression cassette containing a promoter element (*PglpF*) linked to *fred* in a single
 10 copy into the Strain 3 background resulted in Strain 4. Results from deep-well assays showed that expression of the *fred* gene using *PglpF* i) increase the 3-FL production almost 2-fold (90% increase in the total 3-FL production) and ii) improves the 3-FL product distribution by lowering the amount of 3-FL in the cell fraction and increasing the amount of 3-FL in the media.

15 Example 3 – Production of LNT

Engineering of *Escherichia coli* for LNT production expressing the *fred* gene

The *Escherichia coli* K-12 (DH1) MDO strain can be manipulated to express heterologous genes of interest. For instance, Strain 5 is an LNT producing strain overexpressing the β -1,3-N-acetylglucosaminyl-transferase, *lgtA*, and the β -1,3-galactosyltransferase, *galTK*. Insertion of an
 20 expression cassette containing a promoter element (*PglpF_SD4*) linked to *fred* in a single copy into the Strain 5 background resulted in Strain 6. Insertion of an expression cassette containing a promoter element (*PglpF_SD7*) linked to *fred* in a single copy into the Strain 5 background resulted in Strain 7. Results from deep-well assays showed that expression of the *fred* gene using
 25 *PglpF_SD4* or *PglpF_SD7* i) increase the LNT production by 1.2-1.3-fold (20-30% increase in the total LNT production and ii) slightly improves the LNT product distribution by lowering the amount of LNT in the cell fraction and increasing the amount of LNT in the media.

Summary of examples 1-3

Chromosomal expression of *fred* in a 2'-FL, LNT or 3-FL production strain increased the HMO
 30 production, and in the case of 2'-FL and 3-FL, the expression of *fred* decreased the amount of HMO inside the cells (pellet fraction).

Hence, export of HMOs by the Major facilitator superfamily proteins *fred* increase the HMO production capacity of the production strains.

Table 7.

Background Strain*	ID of new strain expressing fred	HMO	HMO production	Reduction of HMO in pellet
Strain 1	Strain 2	2'-FL	115%	yes
Strain 3	Strain 4	3-FL	190 %	yes
Strain 5	Strain 6	LNT	130%	no
	Strain 7		120%	no

*Strain background for HMO production

Example 4 -The Vag transporter is a novel LNT-2-core transporter with high selectivity for LNnT

Table 6: Examples of helper and donor plasmids used for strain construction

Plasmid	Relevant genotype	Marker gene
pACBSR	<i>Para-I-SceI-λ Red, p15A ori, cam*</i>	cam
pUC57	<i>pMB1, bla</i>	bla
pUC57::gal	<i>pUC57::galTK' / T1-galKM'</i>	bla

5

DNA sequences of heterologous genes coding for transporters or glycosyltransferases of interest were codon optimized and synthesized by GenScript. The genes of interest encoding for the transporter proteins as shown in Table 4 were amplified by PCR using appropriated primers covering the start codon, ATG, and the stop codon, TAA, of the gene (Table 3). To construct donor plasmids with any heterologous gene of interest, standard USER cloning was employed to combine the purified PCR fragments of the relevant plasmid backbone, promoter element and gene of interest. Cloning in an appropriated plasmid could be done using any standard DNA cloning technique. Following cloning the DNA was transformed into TOP10 cells and selected on LB plates containing 100 µg/mL ampicillin (or 50 mg/mL kanamycin depending on the backbone applied) and 0.2 % glucose. The constructed plasmids were purified, and the promoter sequence and the 5'end of the gene of interest were verified by DNA sequencing (MWG Eurofins Genomics).

10

15

Construction of strains

The bacterial strain used, MDO, was constructed from *Escherichia coli* K-12 DH1. The *E. coli* K-12 DH1 genotype is: *F⁻, λ⁻, gyrA96, recA1, relA1, endA1, thi-1, hsdR17, supE44*. In addition to the *E. coli* K-12 DH1 genotype MDO has the following modifications: *lacZ*: deletion of 1.5 kbp, *lacA*:

20

deletion of 0.5 kbp, *nanKETA*: deletion of 3.3 kbp, *melA*: deletion of 0.9 kbp, *wcaJ*: deletion of 0.5 kbp, *mdoH*: deletion of 0.5 kbp, and insertion of *Plac* promoter upstream of the *gmd* gene.

Insertion of an expression cassette containing a promoter linked to the *vag* gene and to a T1 transcriptional terminator sequence was performed by Gene Gorging essentially as described by Herring et al (Herring, C.D., Glasner, J.D. and Blattner, F.R. (2003). *Gene* (311). 153-163), and example of helper and donor plasmids used to construct strains presented in the present application are provided in Table 6. Briefly, the donor plasmid and the helper plasmid were co-transformed into MDO and selected on LB plates containing 0.2 % glucose, ampicillin (100 µg/mL) or kanamycin (50 mg/mL) and chloramphenicol (20 µg/mL). A single colony was inoculated in 1 mL LB containing chloramphenicol (20 µg/mL) and 10 µL of 20 % L-arabinose and incubated at 37°C with shaking for 7 to 8 hours. For integration in the *galK* loci of *E. coli* cells were then plated on M9-DOG plates and incubated at 37°C for 48 hours. Single colonies formed on MM-DOG plates were re-streaked on LB plates containing 0.2 % glucose and incubated for 24 hours at 37°C. Colonies that appeared white on MacConkey-galactose agar plates and were sensitive for both ampicillin and chloramphenicol were expected to have lost the donor and the helper plasmid and contain an insertion in the *galK* loci. Insertions in the *galK* site was identified by colony PCR using primers O48 (SEQ ID NO: 17) and O49 (SEQ ID NO: 18) and the inserted DNA was verified by sequencing (Eurofins Genomics, Germany).

Insertion of genetic cassettes at other loci in the *E. coli* chromosomal DNA was done in a similar way using different selection marker genes.

Sequence alignments

Heuristic pairwise sequence alignments as implemented in BLAST 2.1.2 (Basic Local Alignment Search Tool) (Altschul et al. 1990) on the NCBI database (<http://www.ncbi.nlm.nih.gov>) were performed to test the homology among the transporter protein sequences mentioned in the present invention. All parameters were kept at their default values for every BLAST alignment.

In the present invention, we tested the ability of selected heterologous MFS transporters to export LNT out of the host cell using a reference strain, namely MP3672. The strain has one *PglpF*-driven copy of each heterologous glycosyltransferase that is required for product formation.

One copy of each of the selected heterologous transporter genes, i.e. *vag*, *yberC0001_9420*, *marc*, *bad*, *nec*, *yabM* and *fred* (Table 4), was individually integrated in the genome of the strain MP3672 under the control of the *Plac* promoter, which is known to provide moderate transcript levels. The ribosomal binding site of the *Plac* promoter (i.e., the 16 bp upstream of the translational start site) has been modified (see Table 5) to strengthen ribosomal binding to the synthesized transcripts and in this manner positively regulate *Plac*-driven expression at the translational level.

Following the above strain engineering approach and testing strain performance in DWA, we report here that only *Vag*-expressing LNNt producing cells show a relative increase both in the efflux of the produced LNNt and LNNt total production, compared to the reference strain.

As shown in Figure 7, *Plac*-driven expression of most transporter genes (*marc*, *fred*, *bad*, *nec*,
5 *yabM*, *yberC0001_9420*) either has no significant effect or diminish LNNt production (down to 45 %). On the contrary introduction of the *vag* gene in the genetic background of the strain MP3672 results in a marked increase in LNNt titer (Figure 7).

The alignment of amino acid sequences of the seven tested putative MSF transporters revealed that the *Vag* transporter has a very high sequence coverage (99 to 100 %) to the rest six
10 transporters tested in the present invention, with sequence identity ranging from 65% to 75%. The highest sequence identity (75%) was scored for the sequence of *Vag* and sequence of the MFS transporter encoded by the *yabM* gene (Table 8). Interestingly, the MFS transporter *YabM*, described in WO2017042382 as a putative effective exporter of LNT, did not show any significant impact on neither the final total LNNt titer nor LNNt efflux (Figure 7, 8). Thus, contrary to its
15 relatively high protein sequence similarity to *Vag*, the *YabM* transporter seems to be ineffective in relation to LNNt transport, which is clearly reflected by the product concentrations detected in the extracellular fraction of the corresponding host cell cultures (Figure 8). Specifically, as revealed by the analysis of the supernatant fractions of bacterial cultures, extracellular LNNt concentrations were much higher for the strain MP3994 (*vag*-expressing cells) than for the strains MP4362 (*yabM*-
20 expressing cells) and MP3672 (reference cells expressing no heterologous transporter) (Figure 8).

Taken together, this data suggest that the *Vag* transporter is an efficient transporter for LNNt. This conclusion derives from both the results of a screening procedure of the seven transporter genes genomically expressed in the same reference strain, under the same culture conditions and control of the same promoter, *Plac*, and analysis of the transporter protein sequences, that revealed that
25 transporter proteins with relatively high homology to *Vag*, such as *YabM* and *Marc* (75% and 71%), do not export LNNt at all, or, if export, the efficiency is very low.

Table 8: Homology between different heterologous transporters of the present invention:

Protein name	Identification	From organism	Identical to Vag WP_048785139.1 (coverage)
YberC0001_9420	EEQ08298.1 Major facilitator superfamily MFS_1	<i>Yersinia bercovieri</i>	68% (99%)
Fred	WP_087817556.1 MFS transporter	<i>Yersinia frederiksenii</i>	69% (99%)
Bad	WP_017489914.1 MFS transporter	<i>Rouxiella badensis</i>	65% (99%)
Nec	WP_092672081.1 MFS transporter	<i>Rosenbergiella nectarea</i>	66% (99%)
Marc	WP_060448169.1 MFS transporter	<i>Serratia marcescens</i>	71% (99%)
YabM	CAY73138.1 Putative MFS sugar efflux transporter	<i>Erwinia pyrifoliae</i>	75% (99%)