Site-Specific Intestinal Delivery and/or Production of Biologically Active Substances

Field of the invention

The present invention relates to the site-specific production of biologically active substances by recombinant micro-organisms in the body of a human or animal subject. The invention provides transcription regulatory sequences, which are specifically induced in the gastrointestinal tract (GI-tract) of subjects and methods for their use for *in vivo* site-specific expression of nucleic acid sequences encoding biologically active substances.

Background of the invention

Harmless or even beneficial micro-organisms such as lactobacilli, enterococci, or other common bacteria, such as *E. coli* commonly populate human and animal body cavities. For example, many strains of *Lactobacillus*, a genus with a longstanding tradition in food industry, are present in large numbers on gastrointestinal (GI) mucosa together with other intestinal organisms (Mitsuoka, 1992 *in* The Lactic Acid Bacteria in Health and Disease, Volume I., E.B.J.B.Wood, Elsevier Science Publishing Ltd. pp 69-114). These beneficial micro-organisms are believed to play an important role in human and animal health by interfering with the invasion and colonisation of potential pathogens of the GI-tract. One of the mechanisms by which interference is believed to be achieved is competitive exclusion of pathogens by the beneficial micro-organisms through occupation of common receptors on the mucosal surfaces of the GI-tract and other body cavities such as mouth, nasopharynx, vagina and rectum.

The natural occurrence of these non-pathogenic micro-organisms in the various body cavities opens the possibility of using them as delivery vehicles for biologically active substances to those areas of the body. To this end the beneficial micro-organisms could be genetically engineered to produce a desired biologically active substance, such as a protein, an enzyme or reaction products thereof (Elmer *et al.*, 1996, JAMA 275: 870-876; Steidler *et al.* 2000, Science. 289:1352-5). The engineered micro-organism is then introduced into the desired body cavity where it will produce and deliver the desired biologically active substance(s).

Indeed, for some years now lactic acid bacteria are being developed as delivery vehicles for vaccines, as reviewed by Wells *et al.* (1996 Antonie Van Leeuwenhoek **70**: 317-330) and Pouwels *et al.* (1998 Int. J. Food Microbiol. 41: 155-167). The ubiquity of lactic acid bacteria in the mammalian GI-tract and their capability to adhere to squamous epithelial cells and mucus makes them useful as vehicles for vaccine delivery.

WO 99/47657 discloses the use of a transformed *Lactobacillus* species, in particular *Lactobacillus reuteri*, as vaccine delivery vehicles. A gene encoding a factor involved in bacterial aggregation, referred to as agg, is disclosed in WO 99/47657, as well as a gene encoding a factor involved in the adhesion of *L. reuteri* to mucous membranes, referred to as muc. WO 99/47657 further discloses fusions of agg- and/or muc-genes with genes encoding exogenous antigens to be used for vaccination of animals. Lactobacilli transformed with such fusion genes are said to be capable of provoking a desired immune response to the exogenous antigen and, therefore, useful as oral vaccines.

US 5,821,081 suggests using lactobacilli transformed with sequences encoding antiviral proteins referred to as cyanovirins, to deliver the antiviral proteins in the vagina for protection against viruses like HIV.

US 6,605,286 and US 2002019043 describe the use of *Lactobacillus lactis* for delivery of IL-10, IL-2, IL-6 and other biologically active peptides to the GI-tract.

However, none of the art teaches transcription regulatory sequences that are capable of driving efficient gene expression in lactobacilli or other non-pathogenic or enteric bacteria under conditions that occur while these bacteria are passing the vertebrate's GI-tract. In particular, transcription regulatory sequences which are specifically induced during passage of a micro-organism through the GI-tract, while not being active under non-inducing conditions such as outside the subject's body, are not available in the art. Such transcription regulatory sequences are extremely valuable tools for GI-tract specific expression of nucleic acid sequences and are one embodiment of the invention.

One species of lactobacilli, *Lactobacillus plantarum*, has a long tradition of utilization in industrial and artisanal fermentations and is 'generally regarded as safe' (GRAS; Adams and Marteau 1995, Int. J. Food Microbiol. 27:263-364). Next to the occurrence of *L. plantarum* in human and animal diets, this microorganism is frequently

encountered as a natural inhabitant of the GI-tract (Ahrne *et al.* 1998 J. Appl. Microbiol. 85, 88-94). The complete 3.3 Mbp genome sequence of *L. plantarum* WCFS1 has been determined (Kleerebezem, M. *et al.* 2003 Proc. Natl. Acad. Sci. USA 100, 1990-1995) and is available to the skilled person. Strain WCFS1 has been deposited under the Budapest Treaty at the CBS, Baarn, The Netherlands, on 27 June 2003 and was given accession number CBS 113118. The strain WCFS1 is a single colony isolate of strain NCIMB8826 (available at the National Collection of Industrial and Marine Bacteria, Aberdeen, UK) which reaches the ileum in high numbers and in an active form compared to other lactic acid bacteria strains and is detectable in the colon (Vesa, T. *et al.* 2000, Aliment. Pharmacol. Ther. 14, 823-8). Intriguingly, genome sequence comparison revealed that the closest relatives of *L. plantarum* include *Listeria inocua* and *monocytogenes* (Kleerebezem *et al.* supra), which also naturally inhabit and persist the human GI-tract (Glaser, P. *et al.* 2001, Science 294, 849-52). Despite the availability of the complete genome sequence of *L. plantarum* it is not known which genes are specifically induced during passage of the micro-organism through the GI-tract. Also, it is not known whether traditional methods for identifying genes that are highly expressed *in vivo* as compared to laboratory conditions can be used to identify *in vivo* expressed genes in food-grade bacteria such as *L. plantarum*, as these methods have so far mostly been applied to pathogenic bacteria, such as *S. typhimurium, Listeria monocytogenes, Pseudomonas aeruginosa* and *Histoplasma capsulatum*, in order to identify genes involved in pathogencitiy.

The three main strategies developed and used in pathogenic micro-organisms for the identification of genes that are highly expressed *in vivo* during infection as compared to laboratory conditions are selective capture of transcribed sequences (SCOTS, see Dozois, *et al.* 2003 PNAS 100, 247-52; Hou *et al.* 2002 Infect. Immun. 70, 3714-26; Daigle, *et al.* 2001 Mol. Microbiol. 41, 1211-22 and Liu, S. *et al.* 2002 Appl. Environ. Microbiol. 68, 1697-705), signature tagged mutagenesis (STM, for reviews see Mecsas, J. 2002 Curr. Opin. Microbiol. 5, 33-7 and West, N. P. *et al.* 2003 Trends Microbiol. 11, 338-44) and *in vivo* expression technology (IVET). The original IVET technology involves a tandem set of promoter-less reporter genes which were used to identify promoters that are specifically switched on in *S. typhimurium* during infection (Mahan, M. J. *et al.* 1993 Science 259, 686-8). This set-up has been modified in several ways and auxotrophic markers, antibiotic resistance genes and induction of site-specific

recombinases have been used to trap promoters specifically activated during infection of several pathogens (for reviews see Angelichio, M. J. & Camilli, A. 2002 Infect. Immun. 70, 6518-23 and Mahan, M. J. *et al.* 2000 Annu. Rev. Genet. 34, 139-164).

A recent study describes the use of IVET to identify *Lactobacillus reuteri* genes specifically induced in the GI-tract of mouse treated with the antibiotic lincomycin (Walter et al. 2003 Appl. Environ. Microbiol. 69, 2044-51). This study identified three *Lactobacillus reuteri* genes, which are induced during colonization of mice on an antibiotic-containing diet. However, continuous antibiotic administration is required during the animal experiment and will dramatically disturb the intestinal microflora natively present in these mice. Moreover, these experiments were performed in non-conventional, *Lactobacillus* free mice. These factors indicate that the GI-tract conditions encountered by *L. reuteri* in these experiments differ significantly from those in a conventional mouse. Thus, although in this study a food-grade bacterium was used, the artificial model applied and the method used were not suitable for identifying transcription regulatory sequences, which are induced in a conventional, undisturbed GI-tract.

A modification of IVET is the so-called resolvase-based IVET (R-IVET) strategy (Angelichio and Camilli 2002 supra), which makes use of a promoterless resolvase gene, such as *tnpR*. A major advantage of R-IVET over IVET is that no selection pressure is exerted on the micro-organism. R-IVET has also to date only been used in pathogenic micro-organism, and it is not known if it can be adapted to identify *in vivo* induced genes from food-grade bacteria.

The present inventors have for the first time used an R-IVET strategy in a non-pathogenic, food-grade lactic acid bacterium, *L. plantarum* strain WCFS1, to identify genes expressed *in vivo* during passage of the bacterium through the GI-tract. 71 transcription regulatory sequences have been identified that are specifically induced *in vivo* during passage of *L. plantarum* through the GI-tract of conventional mice.

Description of the invention

Definitions

The term "LPXTG-like anchoring motif" refers to a cell wall anchoring region commonly found in surface-anchored proteins of Gram-positive bacteria. These motifs can be used to attach proteins or hybrid proteins (e.g. covalently) to the cell surface of

Gram-positive bacteria, as described by Leenhouts *et al.* (Antonie van Leeuwenhoek 1999, 76(1-4):367-76). The definition as used herein encompasses the LPQTNE motif found in *L. plantarum* surface anchored proteins. The *L. plantarum srtA* gene encodes a sortase, which is involved in maturation and anchoring of proteins comprising an LPXTG-like anchoring motif to the cell wall (Navarre and Schneewind, 1999, Microbial Mol. Biol. Rev. 63:174-229).

The term "gene" means a DNA fragment comprising a region (transcribed region), which is transcribed into an RNA molecule (e.g. an mRNA) in a cell, operably linked to suitable regulatory regions (e.g. a promoter). A gene may thus comprise several operably linked fragments, such as a promoter, a 5' leader sequence, a coding region and a 3'nontranslated sequence comprising a polyadenylation site. "Expression of a gene" refers to the process wherein a DNA region which is operably linked to appropriate regulatory regions, particularly a promoter, is transcribed into an RNA, which is biologically active, i.e. which is capable of being translated into a biologically active protein or peptide. In one embodiment the coding sequence is preferably preceded by a nucleic acid sequence encoding a secretion signal, so that the encoded protein or peptide is secreted out of the cell. The coding sequence is preferably in sense-orientation and encodes a desired, biologically active protein or peptide.

A "chimeric" (or recombinant) gene refers to any gene, which is not normally found in nature in a species, in particular a gene in which different parts of the nucleic acid region are not associated in nature with each other. For example the promoter is not associated in nature with part or all of the transcribed region or with another regulatory region. The term "chimeric gene" is understood to include expression constructs in which a promoter or transcription regulatory sequence is operably linked to one or more coding sequences.

A "transcription regulatory sequence" is herein defined as a nucleic acid sequence that is capable of regulating the rate of transcription of a (coding) sequence operably linked to the transcription regulatory sequence. A transcription regulatory sequence as herein defined will thus comprise all of the sequence elements necessary for initiation of transcription (promoter elements), for maintaining and for regulating transcription, including e.g. attenuators or enhancers. Although mostly the upstream (5') transcription regulatory sequences of a coding sequence are referred to, regulatory sequences found downstream (3') of a coding sequence are also encompassed by this definition.

As used herein, the term "operably linked" refers to a linkage of polynucleotide elements in a functional relationship. A nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For instance, a promoter, or rather a transcription regulatory sequence, is operably linked to a coding sequence if it affects the transcription of the coding sequence. Operably linked means that the DNA sequences being linked are typically contiguous and, where necessary to join two protein encoding regions, contiguous and in reading frame.

The terms "signal sequence", "signal peptide" and "secretory leader" are used interchangeably and refer to a short (usually about 15-60 amino acids), continuous stretch of amino acids at the amino-terminus of secreted and membrane-bound polypeptides, which directs their delivery to various locations outside the cytosol. Signal sequences usually contain a hydrophobic core of about 4-15 amino acids, which is often immediately preceded by a basic amino acid. At the carboxyl-terminal end of the signal peptide there are a pair of small, uncharged amino acids separated by a single intervening amino acid that defines the signal peptide cleavage site (von Heijne, 1990). Despite their overall structural and functional similarities, native signal peptides do not have a consensus sequence.

The term "induced", "specifically induced" or "specifically expressed" as used herein refers to the induction of activity of a transcription regulatory sequence, specifically of a promoter. In particular, an initiation or increase in mRNA transcript production of a DNA sequence operably linked to the transcription regulatory sequence is referred to. Induction preferably refers to a change from no mRNA transcription to mRNA transcription, preferably to high mRNA transcription. In some instances "induction" may also refer to a change from a low basal transcription level (e.g. low basal expression levels present *ex vivo*) to a higher transcription level (e.g. induction *in vivo*). Induction can be triggered by physiological, developmental or environmental changes. In particular, an increase of at least 5%, 10%, 20%, 30%, 50%, or more, of mRNA transcription after induction is encompassed in this definition.

"Site-specific induction" or "site-specific expression" refers to the induction of transcription in a specific body cavity, such as in the GI-tract of a subject. Site specific induction may also comprise de-repression of the transcription regulatory activity. Preferably transcription remains long enough active after induction for a sufficient amount of a biologically active substance (encoded by a coding sequence operably

linked to the transcription regulatory sequence) to be produced. A sufficient amount in this respect is the minimum amount necessary for the substance to exert its biological effect. The time required will vary, depending on the nature of the biologically active substance. Transcription may, therefore, in one embodiment remain active for an extended period of time, e.g. until the host containing the transcription regulatory sequence has passed through the subject, while in another embodiment it may only be active for a short period following induction. In one particular embodiment, the transcription regulatory sequence is only active in the body cavity in which it is induced. The site-specific induction may occur in any cavity of the GI-tract, such as the stomach, the large intestine or a part thereof (e.g. the cecum, colon or rectum) or the small intestine or a part thereof (e.g. the duodenum, jejunum or ileum).

"GI-tract-induced transcription regulatory sequences" (or "GI-tract-inducible transcription regulatory sequences") refers to transcription regulatory sequences, especially promoters, which are induced *in vivo* in the GI-tract of a conventional subject or any part or parts thereof. GI-tract-induced transcription regulatory sequences may, for example, be only transiently active in the GI-tract.

"*In vivo*" refers to the presence or activity within a body cavity of a subject, in particular within the GI-tract of the subject. Vice versa, "*ex vivo*" refers to the presence or activity outside the body of a subject, e.g. in the laboratory.

The term "body cavity of a subject" is understood to mean a cavity in the body of a subject, in particular a vertebrate, that is externally accessible and that is usually lined with a mucosal surface. Examples of such cavities include the mouth, the nasopharynx, the upper and lower GI-tract, the rectum, the vagina and the non-mammalian counterparts of these mammalian cavities.

When referring herein to "conventional subjects", such as for example "conventional mice", subjects whose GI-tract has not been disturbed by the administration of e.g. drugs or other pharmaceutical substances, such as antibiotics. The GI-tract is thus in a natural, undisturbed condition.

A "nucleic acid construct" is herein understood to mean a man-made nucleic acid molecule resulting from the use of recombinant DNA technology. The term "nucleic acid construct" therefore does not include naturally occurring nucleic acid molecules although a nucleic acid construct may comprise (parts of) naturally occurring nucleic acid molecules. A chimeric gene as defined above is an example of a nucleic acid

construct, in particular it is an example of an expression construct. The term "vector" as used herein refers to a nucleic acid construct used for introducing exogenous DNA into host cells, such as for example bacterial cells. A plasmid vector thus at least comprises a replicon and a DNA segment that may be used for insertion of the exogenous DNA into the vector by recombinant techniques, preferably without interfering with the plasmids capability to replicate. Vectors usually comprise further genetic elements to facilitate their use in molecular cloning, such as e.g. selectable markers, multiple cloning sites and the like (see below). An "expression vector" as used herein refers to any plasmid-based cloning vector, in which a promoter and other regulatory elements are present to enable transcription of inserted exogenous DNA (e.g. one or more chimeric genes) when the expression vector is present in a suitable environment (e.g. a suitable host cell and/or body cavity). A "shuttle vector" refers to a plasmid vector that is capable of replication and stable maintenance in at least two different host organisms, e.g. two organisms of different species or different genera. For this capability the shuttle vector may rely on a single broad host-range replicon but usually a shuttle vector will comprise different replicons for different host-organisms or different groups of host-organisms.

A “recombinant micro-organism”, a "host" or a "recombinant host" are terms referring to a (micro)-organism comprising a (man made) nucleic acid construct within its cell(s), in particular one or more chimeric genes. The recombinant micro-organism preferably contains the nucleic acid construct as an episomally replicating molecule, or alternatively and more preferably, integrated into its genome. The latter has the advantage of greater genetic stability of the introduced DNA. It is immaterial by what method the nucleic acid construct is introduced into the micro-organism. Suitable transformation methods for introducing nucleic acid constructs into cells of micro-organisms, such as e.g. electroporation, are available to a skilled person. It is noted that sometimes the recombinant micro-organism itself is also referred to as “vector” in the art, as it can be used as a vehicle to deliver nucleic acid and protein molecules to desired locations.

“Food-grade” micro-organisms are in particular organisms, which are considered as not harmful, when ingested by a human or animal subject.

A “subject” refers herein to a human or non-human animal, in particular a vertebrate, such as but not limited to domestic animals.

The term "selectable marker" is a term familiar to one of ordinary skill in the art and is used herein to describe any genetic entity which, when expressed, can be used to select for a cell or cells containing the selectable marker. Selectable marker gene products confer for example antibiotic resistance. Genes conferring resistance to antibiotics such as kanamycin, rifampicin, erythromycin, actinomycin, chloramphenicol, tetracyclines, nisin and lactacin F are generally known in the art. The term "reporter" may be used interchangeably with marker, although it is mainly used to refer to visible markers, such as green fluorescent protein (GFP). Selectable markers may be dominant or recessive or bidirectional.

The term "homologous" when used to indicate the relation between a given (recombinant) nucleic acid or polypeptide molecule and a given host organism or host cell, is understood to mean that in nature the nucleic acid or polypeptide molecule is produced by a host cell or organisms of the same species, preferably of the same variety or strain. If homologous to a host cell, a nucleic acid sequence encoding a polypeptide will typically (but not necessarily) be operably linked to another (heterologous) promoter sequence or, if applicable, another (heterologous) secretory signal sequence and/or terminator sequence than in its natural environment. Such regulatory sequences may also be homologous to the host cell. In this context, the use of only "homologous" sequence elements allows the construction of "self-cloned" genetically modified organisms (GMO's) (self-cloning is defined herein as in European Directive 98/81/EC Annex II; see also below herein). When used to indicate the relatedness of two nucleic acid sequences the term "homologous" means that one single-stranded nucleic acid sequence may hybridize to a complementary single-stranded nucleic acid sequence. The degree of hybridization may depend on a number of factors including the amount of identity between the sequences and the hybridization conditions such as temperature and salt concentration as discussed later.

"Stringent hybridization conditions" can also be used to identify nucleotide sequences, which are substantially identical to a given nucleotide sequence. Stringent conditions are sequence dependent and will be different in different circumstances. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (Tm) for the specific sequences at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically stringent conditions will be chosen

in which the salt concentration is about 0.02 molar at pH 7 and the temperature is at least 60°C. Lowering the salt concentration and/or increasing the temperature increases stringency. Stringent conditions for RNA-DNA hybridizations (Northern blots using a probe of e.g. 100nt) are for example those which include at least one wash in 0.2X SSC at 63°C for 20min, or equivalent conditions. Stringent conditions for DNA-DNA hybridization (Southern blots using a probe of e.g. 100nt) are for example those which include at least one wash (usually 2) in 0.2X SSC at a temperature of at least 50°C, usually about 55°C, for 20 min, or equivalent conditions.

The term "substantially identical", "substantial identity" or "essentially similar" or "essential similarity" means that two peptide or two nucleotide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default parameters, share at least about 75 percent, preferably at least about 80 percent sequence identity, preferably at least about 90 percent sequence identity, more preferably at least 95 percent sequence identity or more (e.g., 99 percent sequence identity). GAP uses the Needleman and Wunsch global alignment algorithm to align two sequences over their entire length, maximizing the number of matches and minimizes the number of gaps. Generally, the GAP default parameters are used, with a gap creation penalty = 50 (nucleotides) / 8 (proteins) and gap extension penalty = 3 (nucleotides) / 2 (proteins). For nucleotides the default scoring matrix used is nwsgapdna and for proteins the default scoring matrix is Blosum62 (Henikoff & Henikoff, 1992). It is clear than when RNA sequences are said to be essentially similar or have a certain degree of sequence identity with DNA sequences, thymine (T) in the DNA sequence is considered equal to uracil (U) in the RNA sequence.

The term "comprising" is to be interpreted as specifying the presence of the stated parts, steps or components, but does not exclude the presence of one or more additional parts, steps or components. A nucleic acid sequence comprising region X, may thus comprise additional regions, i.e. region X may be embedded in a larger nucleic acid region.

The present invention is based on the finding that Cre recombinase (also referred to as resolvase) functions correctly and efficiently in *L. plantarum* and that the R-IVET method can successfully be used to identify *L. plantarum* transcription regulatory sequences, which are specifically induced *in vivo*, i.e. during the passage of *L. plantarum* through the GI-tract of a conventional subject. Abremski *et al.*, 1983 (Cell,

32: 1301-1311) discovered this site-specific recombination system of bacteriophage P1, consisting of a recombination site designated loxP and a recombinase designated Cre. Since then the Cre/loxP recombination system has been used for example in *Saccharomyces cerevisiae* (Sauer, Molecular and Cellular Biology, 7: 2087-2096, 1987), mouse (Sauer and Henderson, Nucl. Acids Res., 17: 147-161, 1989) and plants (US 5,658,772 and US 4,959,317).

Screening method of the invention

In one embodiment of the invention a method is provided for identifying transcription regulatory sequences of lactic acid bacteria (preferably food-grade), which are specifically induced *in vivo*, in the GI-tract of a subject, while preferably displaying very low or preferably no activity *ex vivo*. "Lactic acid bacteria" are bacteria, which produce lactic acid as an end product of fermentation, such as bacteria of the genus Lactobacillus, Streptococcus and Bifidobacterium. In one embodiment the method of the invention is used for the identification of *L. plantarum* genes and transcription regulatory sequences, which are specifically induced *in vivo* when *L. plantarum* passes through the GI-tract of a conventional subject. The method comprises the insertion of a selectable marker gene flanked by loxP sequences into the genome of a lactic acid bacterium from which *in vivo* induced regulatory sequences are to be isolated, constructing a chromosomal library of the lactic acid bacterium in a vector comprising a promoterless *cre* gene (wherein the chromosomal fragments are inserted upstream of the *cre* gene, introducing the vectors into the loxP-selectable marker-loxP comprising strain, counter-selecting against activity under *ex vivo* conditions by growing the cells for several generations *ex vivo* in the presence of the selectable marker, (orally) administering marker positive colonies to the subject, collecting an plating faecal samples and identifying marker negative strains. The transcription regulatory element trapped in the vector of these cells is then sequenced. The method is further described in the non-limiting Examples.

Transcription regulatory sequences of the invention

The present invention is further based on the surprising discovery that a large number of transcription regulatory sequence of *L. plantarum* are specifically induced in a body cavity of a subject, more preferably in the GI-tract of a subject and most preferably in

the GI-tract of a conventional subject *in vivo*. In one embodiment of the invention transcription regulatory sequences, which are specifically induced in a body cavity or in the GI-tract, or one or more parts of the GI-tract, of subjects are provided (transcription regulatory sequences of SEQ ID NO: 1 to SEQ ID NO: 62). The present invention relates to the practical application of such transcription regulatory sequences, in particular to the use of these regulatory sequences for the *in vivo*, site-specific production of biologically active substances in the GI-tract of a subject. As the whole genome of *L. plantarum* has been sequenced and is available in the art (EMBL database Accession number AL935263 and Kleerebezem *et al.* 2003 supra), SEQ ID NO's 1-62 are available to a skilled person. However, the present invention provides a previously unknown use for these 71 transcription regulatory sequences and variants thereof, such as active fragments thereof, or longer transcription regulatory sequences comprising any of the transcription regulatory sequences of SEQ ID NO: 1-62, or sequences, which are substantially identical to the transcription regulatory sequences of SEQ ID NO's 1-62 and have the capacity to be GI-tract-induced.

SEQ ID NO's 1-62 refers to the GI-tract-induced transcription regulatory sequences normally associated with the following *L. plantarum* genes, as listed in Table 1 below. It is noted that in eight instances one SEQ ID NO comprises more than one transcription regulatory sequence, which is why the 71 GI-tract induced transcription regulatory sequences identified are comprised within only 62 SEQ ID NO's.

SEQ ID NO:	GI-tract-induced genes (ORFs) and the translation start codon (bases within the SEQ ID NO)	product or function:
	Functional class of gene: nutrient acquisition and synthesis	
1	Lp_0185 (*pts1BCA*)[a] (553-555)	sucrose PTS EIIBCA
2	Lp_1164 (*pts14C*) (431-433)	cellobiose PTS, EIIC
3	Lp_2647 (*pts19A*) (509-511)	N-acetylglucosamine/galactosamine PTS, EIIA
4	Lp_3473 (*ram2*) (1010-1012)	alpha-L-rhamnosidase
5	Lp_3522 (*pts32BC*) (519-521)	sucrose PTS, EIIBC
6	Lp_3526 (*pbg10*)[a] (553-555)	6-phospho-beta-glucosidase
7	Lp_3618[2] (*pts37A*) (356-358)	sorbitol PTS EIIA

8	Lp_3659[3] (*rbsD*) (1276-1278)	ribose transport protein
8	Lp_3660[3] (*rbsK3*) (353-355)	ribokinase
9	Lp_0017 (*proA*) (68-70)	glutamate-5-semialdehyde dehydrogenase
10	Lp_0228 (*pepD1*) (916-918)	dipeptidase
11	Lp_0696 (417-419)	cytosine/adenosine deaminase
12	Lp_0775 (*argG*) (707-709)	argininosuccinate synthase
13	Lp_0854 (*birA2*)[a] (643-645)	biotin--[acetyl-CoA-carboxylase] ligase and biotin operon repressor
14	Lp_1058 (*adk*) (673-675)	adenylate kinase
15	Lp_1319 (*rsuA*) (391-393)	pseudouridylate synthase
16	Lp_1602[7] (116-118)	geranyltranstransferase / farnesyl di P transferase
17	Lp_2031 (*ribC2*) (128-130)	bifunctional protein: riboflavin kinase and FMN adenylyltransferase
	Funtional class of genes: stress	
18	Lp_1019 (*clpC*) (187-189)	ATP-dependent Clp protease
19	Lp_3055[4] (*copA*) (983-985)	copper transporting ATPase
20	Lp_3288 (484-486)	cation efflux protein
21	Lp_3303[5] (1022-1024)	putative multidrug transport protein
	Functional class of gene: extracellular	
22	Lp_0800 (797-799)	cell surface protein precursor, LPQTNE motif
23	Lp_2940 (44-46)	cell surface protein precursor, LPQTNE motif
24	Lp_0141 (441-443)	extracellular protein
25	Lp_1403[a] (506-508)	cell surface protein
	Functional class of gene: regulators	
26	Lp_3176 (*pkn2*) (770-772)	serine/threonine protein kinase
27	Lp_3514 (*bglG4*) (538-540)	transcription antitermination
28	Lp_3646[a] (740-742)	transcription regulator
	Functional class of gene: other	
29	Lp_0237 (575-577)	integral membrane protein
30	Lp_0299 (604-606)	ABC transporter, ATP-binding protein
31	Lp_0305 (*gcsH1*) (532-534)	glycine cleavage system, H protein
32	Lp_0393 (*thgA1*)(1150-1152)	galactoside O-acetyltransferase

33	Lp_0419 (*plnI*)[a] (266-268)	immunity protein
34	Lp_0489 (*tkt1-N*) (443-445)	transketolase
35	Lp_0698 (*dnaX*)[a] (115-117)	DNA-directed DNA polymerase III, gamma/tau subunit
36	Lp_0740[6] (*prfB-N*) (256-258)	peptide chain release factor 2, N-terminal fragment
36	Lp_0741[6] (*prfB-C*) (299-301)	peptide chain release factor 2, C-terminal fragment
16	Lp_1603[7] (*ispA*) (1021-1023)	hemolysin homologue
37	Lp_1847 (*codV*) (1149-1151)	integrase/recombinase
38	Lp_3051 (*dhaT*) (275-277)	1,3-propanediol dehydrogenase
39	Lp_3082 (432-434)	transport protein
40	Lp_3281 (1107-1109)	transport protein
41	Lp_3500 (137-139)	short-chain dehydrogenase/oxidoreductase
42	Lp_3505 (*est2*) (26-28)	acetylesterase
7	Lp_3617[2] (*tal3*) (756-758)	transaldolase
43	Lp_3662 (*adhE*) (218-220)	bifunctional protein: alcohol and acetaldehyde dehydrogenase
	hypothetical	
44, 45,	Lp_0026 (478-480), Lp_0139 (343-345),	
46, 47,	Lp_0190 (103-105), Lp_0291[a] (738-740),	
48, 48,	Lp_0476[8] (1115-1117), Lp_0477[8] (168-170),	
49, 50,	Lp_0907 (106-108), Lp_1969 (413-415),	
51, 52,	Lp_2337 (957-959), Lp_2507[9] (615-617),	
52, 53,	Lp_2508[9] (204-206), Lp_2713 (938-940),	
54, 19,	Lp_2718 (225-227), Lp_3057[4] (714-716),	
19, 55,	Lp_3058[4] (324-326), Lp_3312 (1062-1064)	
56, 57,	Lp_0118[a] (342-344), Lp_0292 (685-687),	
58, 59,	Lp_1788 (34-36), Lp_1872 (885-887),	
60, 61,	Lp_2112 (243-245), Lp_2584 (1208-1210),	
62, 21	Lp_3246 (1001-1003), Lp_3305[5] (394-396)	

[a] = identified twice during R-IVET screen; [2-9] genes with the same tag originate from one clone. The gene-annotation corresponds to that of Kleerebezem *et al.* 2003 (supra).

Although SEQ ID NO's: 1-62 provide the complete insert nucleic acid sequence of the clones identified in the R-IVET screen, it is understood that preferably only the transcription regulatory sequence is referred to, when referring to a specific SEQ ID NO. The transcription regulatory sequence is in particular the nucleic acid sequence (or an active part thereof) found upstream (5') of the prokaryotic translation start codon (ATG, GTG or TTG) of the nucleic acid sequence or its complement. One can also refer to the "lp_XXXX transcription regulatory sequence", which is understood to mean the transcription regulatory sequence normally driving the transcription of the nucleic acid comprising the lp_XXXX open reading frame (ORF) in *L. plantarum* (wherein X refers to the numbers as indicated in Table 1). This wording encompasses the whole or any active fragment of the naturally found transcription regulatory sequence, or active variants thereof.

The transcription regulatory sequences present in SEQ ID NO's 1-62 can be operably linked to nucleotide sequences encoding a variety of biologically active proteins or peptides, as described further herein below. Also, active fragments of SEQ ID NO's 1-62 may be used. Fragments which retain the ability to be specifically induced in the GI-tract of a conventional subject can be easily generated by a skilled person by, for example, deleting one or more nucleotides at either or both ends of the nucleic acid molecule and subsequently testing the activity and specificity of the modified transcription regulatory sequence. Similarly, single nucleotide insertions, replacements or deletions can be carried out and the specific activity of the modified nucleotide sequence can be tested.

As mentioned above, the skilled person will recognise that the sequences in SEQ ID NO's 1-62 at least comprise a functional part of the transcription regulatory sequences that are specifically induced in the GI-tract, but that these nucleic acid sequences may in addition comprise (parts of) coding sequences that are naturally linked to the transcription regulatory sequences as present in the *L. plantarum* genome. Such endogenous coding sequences may be removed from the transcription regulatory sequences for use in the present invention, preferably at or near the translation initiation codon (or 'start codon', as indicated in Tabel 1). Alternatively, a coding sequence, encoding a biologically active protein, may be linked in frame to the present coding sequence (or part thereof), so that a fusion protein is generated upon transcription and

translation (as long as the biological activity of the protein or peptide is not reduced significantly by the amino acids fused to it upstream). Similarly, the transcription regulatory sequences present within SEQ ID NO's 1-62 may comprise only a part of the transcription regulatory sequences that are specifically induced in the GI-tract.

Although the transcription regulatory fragment is active, it may function suboptimally. In such instances the skilled person will recognise that the transcription regulatory sequences comprised in SEQ ID NO's 1-62 may be extended with upstream sequence (relative to the direction of transcription) to complement the transcription regulatory sequences of the invention, using the genomic sequence information as available for *L. plantarum*. A skilled person can easily generate such sequences varying in length and identify whether the whole transcription regulatory sequence (as found endogenously in *L. plantarum*) or shorter fragments confer optimal GI-tract induction.

Transcription regulatory sequences of the invention are thus selected from:

(a) a sequence comprising any of the sequences of SEQ ID NO: 1-62;

(b) a sequence comprising a fragment of any of the sequences of SEQ ID NO: 1-62 whereby the fragment retains the ability to be specifically induced in a body cavity of a subject; or,

(c) a sequence which is substantially identical to any of the sequences of (a) or (b) and which retains the ability to be specifically induced in a body cavity of a subject.

The ability of a fragment or of a modified regulatory sequence to be GI-tract-induced can be tested by, for example, generating an expression vector comprising the (modified) regulatory sequence operably linked to a coding sequence, transforming *L. plantarum* with this vector and analysing spatial-temporal induction of the transcript during passage of the recombinant *L. plantarum* through the GI-tract of a subject. The assay used to analyse spatial-temporal expression depends on the nature of the operably linked coding sequence. This sequence may for example encode a light-emitting reporter (such as GFPuv, a Green Fluorescent Protein protein variant optimized for bacterial expression, see Geoffrey *et al.* 2000, Applied and Environmental Microbiology Vol. 66(1), 383-391) or an enzyme whose activity can be tested. Alternatively, quantitative reverse transcriptase PCR (RT-PCR) can be used to quantify the transcript levels produced. Relative transcript levels are determined by comparing steady state levels of the transcript(s) regulated by the transcription regulatory sequence *in vivo* and *ex vivo* or, if more convenient, indirectly by comparing the steady state

levels or enzymatic activities of the protein(s) or enzyme(s) respectively encoded by these transcripts.

Nucleotide sequences which are substantially identical to SEQ ID NO's 1-62, or parts thereof, and in particular to the sequences upstream of the translation start codon(s) comprised in SEQ ID NO's 1-62, and which retain the ability to be GI-tract-induced are also an embodiment of the invention. Such sequences can either be generated by modification of SEQ ID NO's 1-62 (or parts thereof) as described above (nucleotide deletions, insertions, replacements), by *de novo* synthesis of sequences which are substantially similar to SEQ ID NO's 1-62 (or parts thereof) or by isolation of substantially similar sequences from other strains of *L. plantarum* or other organisms, such as but not limited to other food-grade lactic acid bacteria. This can be done using, for example, hybridization methods known in the art, wherein all or part of SEQ ID NO's 1-62 is used as a probe. Preferably stringent hybridization conditions are used to isolate substantially identical sequences.

It is understood that substantially identical GI-tract-induced regulatory sequences can also be obtained by other known methods, such as e.g. those employing (degenerate) primers or probes. For example PCR primers can be synthesized, which correspond to parts of any one of nucleotide sequences SEQ ID NO: 1-62 or to parts of the coding sequence which is in nature associated with the particular regulatory sequence and a PCR reaction can be carried out using genomic DNA of another strain or organism as template DNA. Substantially identical sequences can also be identified *in silico*, by using publicly available algorithms and databases such as BLAST, FASTA and the like, using any one of SEQ ID No's 1-62 (or fragments or variants thereof) as input sequence, or any of the coding sequence listed in Table 1. Sequences identified in this way can then either be cloned using standard molecular biology techniques or synthesised chemically using for example an oligonucleotide synthesizer as supplied e.g. by Applied Biosystems Inc. (Fosters, CA, USA). In all cases, the ability of the resulting nucleotide sequence to be GI-tract-induced and to drive expression in the GI-tract of a subject need to be tested as described above.

Thus, the use of transcription regulatory sequence which have at least about 75%, more preferably at least about 80%, 85%, 90%, 95% or 99% substantial identity to any one of the transcription regulatory sequences of SEQ ID NO's 1-62 (or fragments thereof) and which are GI-tract induced in a conventional subject are one embodiment of the

invention. It is understood that identical sequences obtained from other organisms or other bacterial strains and which are GI-tract-induced are also part of the invention. The transcription regulatory sequences according to the invention are preferably specifically induced in at least one part of the GI-tract, although they may also be induced in several GI-tract parts or the whole of the GI-tract. They may be active transiently (e.g. only for a certain time) or only in a certain part(s) of the GI-tract of the subject. For delivery of a biologically active protein or peptide to a specific body cavity of the GI-tract, it is preferred that the transcription regulatory sequences are specifically induced in only this specific cavity and that the transcription regulatory sequences only remain active in this body cavity, so that the protein or peptide encoded by the operably linked nucleotide sequence is exclusively produced in this cavity. Preferably, the transcription regulatory sequences provide a high level of transcription of an operably linked coding sequence after induction, although this depends on the properties of the encoded biologically active substance. It may, for example, be preferable to express a coding sequence encoding a substance which is very active, but which has negative side effects, for only a short period and/or at a low level. A skilled person can analyse the relative transcription strength, the GI-tract part(s) where induction occurs, the time period during which transcription remains active or the induction cues (such as osmolarity, pH, temperature, activators and the like) using known methods. Using this information the transcription regulatory sequence most suitable for the *in vivo* production of a biologically active substance can be selected. For example, it may be desirable to specifically induce protein production in e.g. the duodenum, the stomach or the colon. For specific delivery of a biologically active substance to the duodenum the transcription regulatory sequences SEQ ID NO: 29 or SEQ ID NO: 12 (or active fragments or variants thereof) may be particularly suitable, while for specific delivery to the stomach for example SEQ ID NO: 18 may be suited and for specific delivery to the colon SEQ ID NO: 4 or SEQ ID NO: 9 may be suited.

Vectors

In another embodiment of the invention vectors comprising one or more of the transcription regulatory sequences of the invention (as described above) are provided. Usually the nucleic acid vectors of the invention will consist of DNA. These nucleic acid vectors may be provided using techniques known per se, for which reference is

made to standard handbooks such as e.g. Sambrook *et al.* (1989) and Sambrook and Russell (2001). In particular expression vectors comprising one or more of the GI-tract-induced regulatory sequences of the invention are provided. An expression vector comprises at least one transcription regulatory sequence according to the invention and one or more coding sequence(s), whereby the transcription regulatory sequence is operably linked to the coding sequence. Suitable coding sequences are those encoding biologically active substances as described elsewhere herein.

The 'empty' vector, only comprising one or more of the transcription regulatory sequences of the invention, but lacking a suitable coding sequence is also an embodiment of the invention. Ideally the empty vector comprises one or more restriction sites downstream of the regulatory sequence, which can be used to easily insert a coding sequence downstream of the regulatory sequence, so that the coding sequence is operably linked to the regulatory sequence.

A vector according to the invention may comprise more than one chimeric gene, each of which may comprise a GI-tract-induced regulatory sequence. The GI-tract-induced regulatory sequences may be different or identical. In this way the timing, location and expression levels of the coding sequences *in vivo* can be regulated. For example, nucleotide sequences encoding two biologically active peptides may be expressed concomitantly or consecutively in the GI-tract of a subject. Similarly, one coding sequence may be expressed at a high level, while the other may be expressed at a low level in a particular part of the GI-tract by using two different transcription regulatory sequences of the invention.

Further, a single regulatory sequence may drive the expression of two or more coding sequences. For example, nucleotide sequences encoding different proteins (e.g. proteins having a synergistic effect) may be linked by IRES (internal ribosome entry sites) elements, providing bicistronic or multicistronic transcripts under control of a single regulatory sequence. Suitable IRES elements are described in e.g. Hsieh *et al.* (1995, Biochemical Biophys. Res. Commun. 214:910-917). The arrangement of several coding sequences in a single transcription unit allows for the co-ordinate expression of several different polypeptides. This may be advantageous for the expression of multisubunit enzymes or in metabolic pathway engineering when several enzymes of a biosynthetic pathway need to be co-ordinately expressed.

The one or more coding sequences operably linked to the regulatory sequence are preferably each provided with the appropriate signals for initiation of translation such as e.g. a Shine-Dalgarno sequence and a translation initiation codon. Optionally a coding sequence in the expression construct comprises a signal sequence encoding a signal peptide causing translocation of the polypeptide encoded by the coding sequence across the cell membrane, and preferably secretion of the polypeptide from the host cell. Signal sequences are well known to the skilled person and a variety of suitable signal sequences is available in the art. Usually the skilled person will select a signal sequence known to effect secretion of polypeptides from a given host cell of choice. Preferably, the signal sequence to be used in the invention will be homologous to the host organism or will originate from an organism closely related to the host organism. Signal sequences that are preferably used in the context of the present invention include signal sequences from lactic acid bacteria of which preferably *Lactobacillus*. Examples of such signal sequences include e.g. signal sequences from α-amylase, *slpA*, and *prtP*. Further suitable signal sequence s may be identified using dedicated genetic screening methods as described by Poquet *et al.* (J. Bacteriol. 1998, 180: 1904-1912.) and Hols *et al.* (Gene, 1992, 118: 21-30). In the expression construct, the transcription unit comprising the one or more coding sequences linked to the transcription regulatory sequence is preferably terminated by a suitable 3'nontranslated region (comprising a transcription terminator), such e.g. *tldH* or *tslP*. Suitable bacterial 3' nontranslated sequences may be those naturally associated with the coding sequence or may be identified as described by Ermolaeva *et al.* (J. Mol. Biol. 2000, 301: 27-33.).

The nucleic acid constructs of the invention may further comprise additional sequence elements for various functions known in the art *per se*. Such sequence element include e.g.: autonomously replicating sequences or origins of replication for episomal multiplication and maintenance of the vector; sequences for promoting integration of the vector or chimeric gene into the host's genome; sequences for homologous recombination; sequences for site-specific integration and/or recombination; selection marker genes; reporter or indicator genes; sequences for promoting conjugation or other means of transfer of genetic material etc.

Hosts

In a further aspect of the invention a host organism, preferably a microbial and/or recombinant host organism, comprising one or more transcription regulatory sequences of the invention, and in particular comprising one or more chimeric genes which comprise at least one regulatory sequence of the invention, is provided. The chimeric gene may be integrated into the genome of the host organism or may be present as an episomal entity, for example on an expression vector, as described above. Integration has the advantage of greater genetic stability. The "host" refers to the organism comprising the chimeric gene(s). The host or host cells may be any micro-organism (such as a bacteria, fungi such as yeasts, viruses) or any plant or animal. A host or host cells may also be a mutant, generated by classical (X-ray or chemical) mutagenesis or the host may already have been transformed previously. However, the preferred host is a bacterium, preferably one of the following: A food-grade bacterium, a commensal bacterium or a (attenuated) pathogenic bacterium. In one embodiment the host is a food-grade bacterium, particularly a gram-positive bacterium, and more preferably a lactic acid bacterium. Other preferred hosts are food-grade fungi and yeasts, such as bakers or brewers yeast, such as species of the genus Saccharomyces, Pichia, or Hansenula. The host may also be a probiotic bacterium, which in itself has a beneficial effect when ingested by a subject.

A preferred host is a bacterium that belongs to a genus selected from the group consisting of *Lactobacillus*, *Lactococcus, Leuconostoc, Carnobacterium, Streptococcus, Bifidobacterium, Bacteroides, Eubacterium, Clostridium, Fusobacterium, Propionibacterium, Enterococcus, Staphylococcus, Peptostreptococcus,* and *Escherichia*. A further preferred host is a bacterium that is a *Lactobacillus* or *Bifidobacterium* species selected from the group consisting of *L. reuteri, L. fermentum, L. acidophilus, L. crispatus, L. gasseri, L. johnsonii, L. plantarum, L. paracasei, L. murinus, L. jensenii, L. salivarius, L. minutis, L. brevis, L. gallinarum, L. amylovorus, B. bifidum, B. longum, B. infantis, B. breve, B. adolescente, B. animalis, B. gallinarum, B. magnum,* and *B. thermophilum*.

Preferred hosts or host cells are (1) capable of initiating transcription, and preferably also regulating transcription, from the transcription regulatory sequence of the invention; (2) capable of surviving passage through a subject's GI-tract and preferably capable of colonizing a mucosal surface lining a subjects GI-tract; and (3) preferably,

not pathogenic to the subject in which the host or host cells are to be employed. With respect to the third aspect it is to be understood that the invention comprises the use of hosts that may normally be pathogenic to the subject (e.g. *Listeria* spp., *Salmonella* spp. or *Campylobacter* spp.) but that have been modified such that they are no longer pathogenic or virulent. Thus, a preferred microbial host will usually be a non-pathogenic bacterium capable of colonizing mucosal surfaces of vertebrates, which bacterium is transformed with a nucleic acid construct or vector described herein above. In a further embodiment of the invention, the host or host cell may be modified as to its capability to adhere to mucosal surfaces. A range of bacterial proteins are known to be involved in adherence to mucosal surfaces. The genes encoding these may be modified. A number of *L. plantarum* genes also display significant sequence homology to such genes identified in other bacterial species and may thus be modified with a reasonable expectation of successfully modifying the capability of *L. plantarum* to adhere to mucosal surfaces. These include genes involved in mucus-binding, fibronectin binding, aggregation promotion, intercellular adhesion, cell clumping, etc., such as for example: lp_0304: aggregation promoting protein [*Lactobacillus gasseri*], lp_0520: autoaggregation-mediating protein [*Lactobacillus reuteri*], lp_1229: mucus binding protein precursor; Mub [*Lactobacillus reuteri*], lp_1643: mucus binding protein precursor; Mub [*Lactobacillus reuteri*], lp_1793: fibronectin-binding protein-like protein A [*Streptococcus gordonii*], lp_1873: aggregation-relevant adhesin [*Streptococcus gordonii*], lp_2278: autoaggregation-mediating *protein [Lactobacillus reuteri*], lp_2845: aggregation promoting protein [*Lactobacillus gasseri*], lp_2847: aggregation promoting protein [*Lactobacillus gasseri*], lp_3050: aggregation promoting protein [*Lactobacillus gasseri*], lp_3059: mucus binding protein precursor; Mub [*Lactobacillus reuteri*], lp_3114: mucus binding protein precursor; Mub [*Lactobacillus reuteri*], lp_3209: mucus adhesion promoting protein [*Lactobacillus reuteri*], lp_3214: collagen binding protein [*Lactobacillus reuteri*] / mucus adhesion promoting protein [*Lactobacillus reuteri*]. Similarly, the expressing of mucosal adhesion proteins MapA may be altered in the host. Such alterations may include changing the expression of endogenous mucosal adhesion proteins, adding the capability to express an exogenous, i.e. heterologous mucosal adhesion proteins, or changing the mucus binding properties of a mucosal adhesion protein. The alterations may be effected in various ways: e.g. endogenous genes encoding mucosal adhesion proteins may be added or deleted,

exogenous of genes encoding mucosal adhesion protein may be introduced, mucus binding properties of mucosal adhesion proteins may be modified, the regulation of expression of the mucosal adhesion proteins may be altered, e.g. by linking a non-native or modified transcription regulatory region to a sequence coding for a mucosal adhesion protein. Some non-limiting specific example include e.g. the inactivation of the endogenous *mapA* gene of *L. reuteri*, or the expression of the *L. reuteri mapA* gene from a promoter other than its native promoter, such as a constitutive or differently regulated promoter.

Methods for transforming bacterial cells are known in the art and are for example described in "Genetics and Biotechnology of Lactic Acid Bacteria", Gasson and de Vos, eds., Chapman and Hall, 1994. Similarly, methods for introducing recombinant DNA into fungi, plants and animals are well known, such as particle gun transformation, protoplast transformation, electroporation, *Agrobacterium* mediated transformation and the like.

In case a host of the invention is constructed through genetic engineering such that the resulting recombinant host comprises only sequences derived from the same species as the host is, albeit in recombined form, the host is said to be obtained through "self-cloning". Hosts obtained through self-cloning have the advantage that there application in food (or pharmaceuticals) is more readily accepted by the public and regulatory authorities as compared hosts comprising foreign (i.e. heterologous) nucleic acid sequences. The present invention thus allows the construction of self-cloned *L. plantarum* and other lactobacillus hosts for food, pharmaceutical or nutraceutical applications (see also de Vos, 1999, Int. Dairy J. 9: 3-10) and such self-cloned hosts are one further aspect of the invention.

Method for *in vivo* site-specific delivery to the GI-tract

Another aspect of the invention relates to a method for producing a polypeptide or a biologically active substance in a body cavity or in several body cavities of a subject, wherein the method comprises the step of administering to the subject a recombinant host or recombinant host cells (preferably recombinant bacterial cells) capable of producing the biologically active substance. In one preferred embodiment the body cavity is the GI-tract, or one or more parts thereof. For this purpose the host comprises at least one chimeric gene, which comprises a transcription regulatory sequence of the invention, so that transcription of the coding sequence is specifically induced when the

host is present *in vivo* in the body cavity (e.g. in the GI-tract, or a part thereof) of a conventional subject. The coding sequence that is operably linked to the transcription regulatory sequence preferably encodes a polypeptide that is (1) the biologically active substance, (2) a precursor thereto, or (3) a polypeptide involved in the biosynthesis of the biologically active substance.

A "biologically active substance" as used herein refers to a substance, which has biological activity when present in at least part of the GI-tract of a subject. Suitable nucleotide sequences encoding biologically active proteins or polypeptides are available to a skilled person. The type of coding sequence that is linked to the transcription regulatory sequence depends on the biologically active substance to be produced and delivered site-specifically by the recombinant host to the GI-tract (or parts thereof) of the subject. Usually, the coding sequence is not natively associated with the transcription regulatory sequence. The coding sequence may be homologous but preferably is heterologous to *Lactobacillus*. The coding sequence may encodes a polypeptide selected from the group consisting of an antigen from a human or animal pathogen, such as e.g. tetanus toxin fragment C (TTFC) or a protective antigen of *Bacillus anthracis*; an enzyme, such e.g. a β-galactosidase, an amylase, a protease, a pancreatic (phospho-)lipase or a superoxidedismutase; a biologically active peptide, such as e.g. oxyntomodulin, oryzatensin, prosomatostatin-derived peptides, exorphin A5, bioactive antinutritional peptides derived from cereal prolamins and others as reviewed by Dziuba *et al.* (Nahrung, 1999, 43: 190-195), gastrointestinal (peptide) hormones, such as somatostatin and others as reviewed by Rehfeld (Physiol. Rev., 1998, 78: 1087-1108), therapeutic proteins such as e.g. interleukin-10, TGF-β, lactoferrin. The coding sequence may also encode an antibody (raised against any desired antigen), a hybrid antibody, a altered antibody or an Fab fragment or any part which exhibits the immunological binding properties of the parent antibody molecule. In one embodiment of the invention the coding sequence, which is operably linked to the regulatory sequence(s) according to the invention, encodes a peptide or protein which is able to cause lysis of the bacterial host cell. Such a host is herein referred to as a "GI-tract lytic strain", as the cell auto-lyses *in vivo* upon induction of the transcription regulatory sequence. The coding sequence may for example comprise a bacteriophage-derived lytic coding sequence or cassette, an autolysin encoding sequence (e.g. *acmA*), a sequence encoding any other bacteriolytic enzyme or toxin, or a sequence encoding

an activator of cell autolysis. The GI-tract lytic strain is preferably "pre-loaded" with a biologically active compound, which is released into the GI-tract lumen upon cell lysis. Pre-loading can for example be achieved by recombinant expression of a biologically active substance under laboratory conditions (e.g. using a constitutive promoter), or at least prior to arrival of the GI-tract lytic strain in the cavity in which lysis is induced. The biologically active substance is preferably present in significant amounts in the cytoplasm of the lytic strain prior to cell lysis. Controlled expression of lytic coding sequences has been described by de Ruyter *et al.* 1997(Nat. Biotechnol. 15:976-9) and Buist G *et al.* 1997(Appl Environ Microbiol. 63(7):2722-8), incorporated herein by reference.

Preferably the bioactive polypeptides originate from the subject species in who's GI-tract the polypeptides are to be produced, e.g. human sequences coding for interleukin-10, TGF-β, or lactoferrin are preferably used for production in man. Alternatively, the coding sequence(s) may encode one or more enzyme(s) involved in the biosynthesis, transport or release of a biologically active substance. Examples of such biologically active substances are vitamins like folic acid and other B-vitamins; organic acids or alcohols such as lactate, acetate, ethanol or butanediol; antioxidants like thiols; amino acids; non-digestible oligo- and polysaccharides such as those synthesized by Gram-positive bacteria; or antimicrobial agents such as e.g. bacteriocins or lantibiotics. Generally secretion of the biologically active substance into the GI-tract is desired and can be achieved by operably linking a nucleic acid sequence encoding a suitable signal peptide to the nucleic acid sequence encoding the biologically active substance (as described above). Likewise, it may be sufficient that only part of the biologically active protein or peptide is presented externally on the host's cell(s).

A further aspect of the invention relates compositions comprising host cells as described herein above, as well as to methods for the production of such compositions. Host cells of the invention, preferably bacteria, are cultured under appropriate conditions, optionally recovered from the culture medium and optionally formulated into a composition suitable for the intended use. Methods for the preparation of such compositions are known *per se*. Compositions for enteral or oral administration may be either food compositions or pharmaceutical compositions whereas compositions for nasal, vaginal or rectal administration will usually be pharmaceutical compositions.

Pharmaceutical compositions will usually comprise a pharmaceutical carrier in addition to the host cells of the invention. The preferred form depends on the intended mode of administration and (therapeutic) application. The pharmaceutical carrier can be any compatible, nontoxic substance suitable to deliver the host cells to the body cavity, e.g. the GI-tract of a subject. E.g. sterile water, or inert solids may be used as the carrier usually complemented with pharmaceutically acceptable adjuvants, buffering agents, dispersing agents, and the like. Compositions will either be in liquid, e.g. a stabilized suspension of the host cells, or in solid forms, e.g. a powder of lyophilized host cells. E.g. for oral administration, the host cells can be administered in solid dosage forms, such as capsules, tablets, and powders, or in liquid dosage forms, such as elixirs, syrups, and suspensions. The host cells can be encapsulated in gelatin capsules together with inactive ingredients and powdered carriers, such as e.g. glucose, lactose, sucrose, mannitol, starch, cellulose or cellulose derivatives, magnesium stearate, stearic acid, sodium saccharin, talcum, magnesium carbonate and the like.

A preferred composition according to the invention is suitable for consumption by a subject, preferably a human or an animal. Such compositions may be in the form of a food supplement or a food or food composition, which besides the host cells of the invention also contains a suitable food base. A food or food composition is herein understood to include liquids for human or animal consumption, i.e. a drink or beverage. The food or food composition may be a solid, semi-solid and/or liquid food or food composition, and in particular may be a dairy product, such as a fermented dairy product, including but not limited to a yoghurt, a yoghurt-based drink or buttermilk. Such foods or food compositions may be prepared in a manner known per se, e.g. by adding the recombinant host cells of the invention to a suitable food or food base, in a suitable amount. In a further preferred embodiment, the host cells are micro-organisms that are used in or for the preparation of a food or food composition, e.g. by fermentation. Examples of such micro-organisms include baker's or brewer's yeast and lactic acid bacteria, such as probiotic lactic acid strains. In doing so, the host cells of the invention may be used in a manner known per se for the preparation of such fermented foods or food compositions, e.g. in a manner known per se for the preparation of fermented dairy products using lactic acid bacteria. In such methods, the host cells of the invention may be used in addition to the micro-organism usually used, and/or may replace one or more or part of the micro-organism usually used. For

example, in the preparation of fermented dairy products such as yoghurt or yoghurt-based drinks, a food grade lactic acid bacterium of the invention may be added to or used as part of a starter culture or may be suitably added during such a fermentation. Preferably, the above compositions will contain the host cells in amounts that allow for convenient (oral) administration of the host cells, e.g. as or in one or more doses per day or per week. In particular, the preparations may contain a unit dose of the host cells.

In another aspect the invention relates to a method for (site-specific) production of a polypeptide or a biologically active substance at a mucosal surface of a subject. The method comprises the step of administering to the subject a composition comprising host cells, preferably bacteria as described above. In a preferred embodiment, the polypeptide is an antigen from a human or animal pathogen. The *in situ* production of the pathogen's antigen(s) allows for the vaccination of the human or animal to which the composition is administered. In another preferred embodiment the polypeptide or the biologically active substance is a health-promoting factor.

Another aspect of the invention relates to the use of a host cell according to the invention in the preparation of a medicament for preventing or treating an infection by pathogenic micro-organisms of a mucosal surface in a subject.

The following non-limiting Examples describe the use of R-IVET in *L. plantarum* and the identification of GI-tract-inducible transcription regulatory sequences. Unless stated otherwise, the practice of the invention will employ standard conventional methods of molecular biology, virology, microbiology or biochemistry. Such techniques are described in Sambrook *et al.* (1989) Molecular Cloning, A Laboratory Manual (2nd edition), Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press; in Sambrook and Russell (2001) *Molecular Cloning: A Laboratory Manual,* Third Edition, Cold Spring Harbor Laboratory Press, NY; in Volumes 1 and 2 of Ausubel *et al.* (1994) *Current Protocols in Molecular Biology, Current Protocols*, USA; and in Volumes I and II of Brown (1998) *Molecular Biology LabFax*, Second Edition, Academic Press (UK); *Oligonucleotide Synthesis* (N. Gait editor); *Nucleic Acid Hybridization* (Hames and Higgins, eds.), all incorporated herein by reference.

Description of the figures

Figure 1: Basic principles of R-IVET. *L. plantarum* strain NZ7109 harbors a chromosomally located *loxP-ery-loxP* cassette and pNZ7125. Chromosomal loci of the *L. plantarum* genome (black triangle) are cloned upstream of *cre* in pNZ7125 and promoter activities in the resulting library can be trapped by monitoring the erythromycin phenotype, since *cre* expression will lead to excision of the erythromycin marker from the chromosome by a homologous recombination event between the two *loxP* sites. The colonies in the library appearing as erythromycin resistant in the laboratory (no active promoters under laboratory conditions) are intragastrically incubated in mice and changes in the erythromycin phenotype (promoter activations) can be monitored in faecal samples.

Examples

Example 1 – Materials and methods

1.1 Bacterial strains and plasmids

Escherichia coli strains MC1061 (Casadaban and Cohen 1980, J. Mol. Biol. 138, 179-207) was used as cloning host during construction of pNZ7109 (see below) and was grown aerobically in TY medium (Sambrook *et al.* 1989, supra). *Lactococcus lactis* MG1363 (Gasson, 1983, J. Bacteriol. 154, 1-9) was used as a cloning host during construction of pNZ7125, pNZ7126 and the *Lactobacillus plantarum* promoter library (see below). *Lactococcus lactis* was grown without aeration at 30 °C in M17 medium (Merck, Darmstadt, Germany), supplemented with 0.5% (wt/vol) glucose (GM17). *L. plantarum* WCFS1 (Kleerebezem *et al.* 2003, supra) and its *loxP-ery-loxP* derivative NZ7109 (see below) were grown at 37 °C in MRS medium (Difco, Surrey, U.K.) without aeration. When appropriate, antibiotics were added to the media; for *E. coli* ampicillin (50 μg/ml); for *Lactococcus lactis* chloramphenicol (5 μg/ml); for *L. plantarum* chloramphenicol (5 μg/ml), erythromycin (5 or 30 μg/ml, for selection after transformation or replica-plating, respectively), lincomycin (10 μg/ml) and rifampicin (50 μg/ml).

1.2 DNA techniques and sequence analysis

Plasmid DNA was isolated from *E. coli* on a small scale using the alkaline lysis method (Sambrook et al. 1989 and Birnboim and Doly 1979, Nucleic Acid Res. 7, 1513-1523). Large scale plasmid DNA isolations were performed using Jetstar columns according to the manufacturer's instructions (Genomed GmbH, Bad Oberhausen, Germany). DNA isolation and transformation in *Lactococcus lactis* and *L. plantarum* were performed as described previously (de Vos *et al.* 1989, Gene 85, 169-176; Ferain et al. 1994, J. Bacteriol. 176, 596-601; Josson et al. 1989, Plasmid 21, 9-20). Standard procedures were applied for DNA manipulations in *E. coli* (Sambrook *et al.* 1989). Restriction endonucleases, *Taq* and *Pwo* polymerase, T4 DNA ligase and calf intestine alakaline phosphatase (CIAP) were used following the recommendations of the manufacturer (Promega, Leiden, the Netherlands and Boehringer, Mannheim, Germany). Primers were purchased from Pharmacia Biotech (Roosendaal, the Netherlands). The sequences of the inserts present in the pNZ7125 derivatives (see below), were amplified by PCR using the vector-based primers cre-R2 (5'-GTCCATCAGGTTCTTGCG-3') and BglII-*cre* (5'-ATAGTTTACCCCGTCAGC-3'), followed by amplicon purification using Sephadex-G50 and multiscreen HV 96-well plates (Millipore, Amsterdam, The Netherlands). Partial insert sequences were determined using either primer *cre*-R2 or BglII-*cre*, approximately 100 ng of the purified amplicons and the ABI Prism® BigDye (tm) Terminator Cycle Sequencing Ready Reaction Kit protocol provided by the manufacturer (Applied Biosystems, Nieuwekerk a/d IJsel, The Netherlands). Sequence reaction products were analyzed using an ABI-prism 3700 DNA analyzer. The determined insert sequences were assigned to *L. plantarum* WCFS1 chromosomal loci using Blast-N (Kleerebeezem *et al.* 2003).

1.3 Construction of *L. plantarum* NZ7109

The intergenic locus flanked by the convergent genes Lp_3503 and Lp_3504 (Fig. 1) was chosen for the integration of a *loxP-ery-loxP* cassette (the *ery* gene confers erythromycin resistance) into the chromosome of *L. plantarum* WCFS1 using a two step double cross-over strategy (Ferain *et al.* 1994, supra). For this purpose pNZ7109, a pUC18 (Yanisch-Perron et al. 1985, Gene 33, 103-119) derivative which harbours a *loxP-ery-loxP* cassette flanked by PCR amplified 5'- truncated fragments of Lp_3503 (using primers 5'-TGCTTTCCAAGAGCAAGCTG–3' and 5'-ACGTCTGCAGTCAGGTGTGAAGTTGGCACT-3') and Lp_3504 (using primers 5'-ACGTCTGCAGCCACTCACCGACTAACACTC-3'and 5'-

CGTAAGCTTGACCCACGAGTTACCAACACG-5') (Kleerebezem *et al.* 2003) was constructed and integrated into the chromosome of *L. plantarum* WCFS1 (Fig 1). One integrant possessing the anticipated genotype, as analyzed by PCR and Southern blotting, was cultured with increasing concentrations of rifampicin (maximum concentration 50 μg/ml) and the resulting rifampicin resistant strain was designated NZ7109. The rifampicin and the plasmid-derived chloramphenicol resistant phenotype, were used for selective recovery of NZ7109 from faecal samples (see below).

1.4 pNZ7125, pNZ7126 and R-IVET library construction

To implement R-IVET in *L. plantarum* the low-copy vector pNZ7125, a pJIM2246 (Renault *et al.* 1996, Gene 183, 175-182) derivative, was constructed. This vector harbors the *prtP* ribosome binding site originating from pUC-*Nco*I (Fernandez et al. 2002, J. Bacteriolo. 184, 82-90) and a promoterless copy of the *cre* gene followed by the *pepN* terminator, both originating from pNZ8048-Cre (Campo et al. 2002, Appl. Env. Microbiol. 68, 2359-67). To confirm correct chromosomal *loxP-ery-loxP* resolution (Ambreski *et al.* 1983, Cell 32, 1301-11) in strain NZ7109 upon *cre* expression, a pNZ7125 derivative containing the *L. plantarum ldhL1* promoter (Ferain *et al.* 1994, supra) upstream of the *cre* gene was constructed (pNZ7126). Moreover, a *L. plantarum* WCFS1 chromosomal R-IVET library was constructed in pNZ7125. Chromosomal DNA was partially digested with *Sau*3AI and size-fractionated on 1% agarose gels. Fragments ranging from 1-2 kb were purified using Sephaglas™ Bandprep (Pharmacia Biotech, Roosendaal, The Netherlands). These purified fragments were cloned into *Bgl*II digested and CIAP dephosphorylated pNZ7125 (Fig 1). Ligation mixtures were transformed to *Lactococcus lactis* MG1363 (Gasson *et al.* 1983, supra) and approximately 50,000 of the obtained colonies were collectively resuspended in GM17. Plasmid DNA was isolated from these pooled cells and was introduced into NZ7109 cells. Approximately 37,000 colonies were obtained on MRS plates containing 5 μg/ml chloramphenicol, collectively resuspended in MRS containing 15% glycerol, and stored in aliquots at –80 °C.

1.5 R-IVET animal experiments

To counter select against clones in the R-IVET library that harbour pNZ7125 derivatives containing a promoter element that is active under the laboratory conditions

applied, the collective library was grown for approximately 20 generations in MRS containing 5 μg/ml chloramphenicol, 30 μg/ml erythromycin and 50 μg/ml rifampicin. Subsequent animal experiments were performed in an accredited establishment (N° A59107) according to guidelines N°86/609/CEE of the French government. Seven weeks-old female Balb/c mice were purchased from Iffa Credo (St Germain sur l'Arbresle, France) and had free access to tap water and standard mice chow during the experiments. Four mice received an oral dose of freshly prepared bacterial suspensions (1×10^9 CFU in 100 μl) by intragastric administration at 2 consecutive days, and 24h after the last administration faecal samples were collected and resuspended in MRS medium at a final concentration of 100 mg/ml (wet weight). After extensive homogenization, the complete R-IVET library was recovered from the faecal samples by plating appropriate dilutions of the suspensions on MRS plates containing 5 μg/ml chloramphenicol and 50 μg/ml rifampicin. After 72h full grown colonies were replica-plated to plates containing 5 μg/ml chloramphenicol and 50 μg/ml rifampicin, with or without 30 μg/ml erythromycin. Another 24h later the plates were compared, leading to the identification of cells displaying an erythromycin sensitive phenotype. The inserts present in the corresponding pNZ7125 derivatives harboured by these resolved clones were amplified by PCR and from the resulting amplicons the DNA sequences were analyzed.

To verify the primary R-IVET results 132 of the selected clones were divided in four groups of 33 clones which were used for collective plasmid DNA isolation. These mixtures of pNZ7125 derivatives were reintroduced into *L. plantarum* NZ7109 and from all four groups several hundreds (minimum of 200) colonies were collectively stored at –80 °C. These four groups were separately subjected to a second round passage using 2 mice per group, following the same procedure as applied in the first round passage, including recovery from faecal samples and analysis of the erythromycin phenotype of the clones.

Example 2 - Results

2.1 Implementation and functionality of R-IVET in *L. plantarum*

Previously R-IVET has been used exclusively for the identification of genes in pathogens which are induced during infection of host tissues (for reviews see Angelichio and Camilli 2002; Mahan *et al.* 2000, supra). To evaluate the functionality

of the *cre-loxP* system in *L. plantarum*, a *L. plantarum* WCFS1 derivative strain that harbors a chromosomally located *loxP-ery-loxP* cassette, was constructed and designated NZ7109 (Fig. 1). The growth rate of NZ7109 did not differ from that observed for wild-type *L. plantarum* WCFS1, and after 50 generations of growth without antibiotic selection pressure, replica-plating revealed that all NZ7109 cells were erythromycin resistant. Moreover, PCR analysis revealed that the chromosomal *loxP-ery-loxP* insertion could be amplified from all colonies, indicating that the chromosomal insertion is stable. Both pNZ7125, a low-copy vector encoding a promoterless copy of *cre*, and a derivative containing the *cre* gene under control of the *L. plantarum* WCFS1 *ldhL1* promoter (pNZ7126) were introduced in NZ7109 cells and plated on MRS medium with chloramphenicol. Replica-plating of the transformants revealed that all NZ7109 colonies harbouring pNZ7125 were erythromycin resistant, while all NZ7109 colonies harbouring pNZ7126 were erythromycin sensitive. In addition, a PCR analysis confirmed that the *loxP-ery-loxP* cassette could only be amplified from erythromycin resistant colonies. These results establish the functional implementation of the *cre-loxP* resolution system in *L. plantarum* WCFS1 and confirm the suitability of pNZ7125 as an R-IVET vector.

2.2 Construction of a *L. plantarum* R-IVET library in pNZ7125

A genomic library of *L. plantarum* WCFS1 was constructed in pNZ7125, using *Lactococcus lactis* MG1363 as an intermediate cloning host. Approximately 37,000 colonies were obtained in *L. plantarum* NZ7109 and the quality of the library was assessed in several ways. First, 100 colonies were randomly picked from MRS plates and their plasmids were used as template for insert amplification by PCR, demonstrating that over 95% of the investigated clones contains an insert, of which the average size was estimated to be 1.3 kb. To assess insert redundancy, all amplicons were digested with *Sau*3AI and the resulting fragments were separated by 2% agarose gel electrophoresis. No common restriction profiles were detected among these 100 amplicons, indicating that redundancy in the R-IVET library is low. Moreover, 28 of these amplicons were used for partial sequence analysis, which revealed no apparent over- or underrepresentation of a specific region of the *L. plantarum* genome. These results support the randomness of the library and genome coverage was calculated to be approximately 98%. The R-IVET library was replica-plated to plates with and without

erythromycin which indicated that 10% of the R-IVET clones contain a pNZ7125 derivative harbouring a properly oriented promoter element that is active under the laboratory conditions applied during the experiments. Moreover, the resulting *cre* expression level in these clones is sufficient to result in excision of the *loxP-ery-loxP* cassette from the chromosome of NZ7109.

2.3 R-IVET screen in mice

To counter select against clones displaying *cre* expression under laboratory conditions, the R-IVET library was cultured for approximately 20 generations in the presence of erythromycin. Subsequently, full-grown cultures were used for gastric administration to Balb/c mice. After recovery from faecal samples 6,000 of the R-IVET clones were analyzed for their erythromycin resistance or sensitivity phenotype by replica-plating, revealing 198 (3.3%) clones that displayed an erythromycin sensitive phenotype. The partial sequence of 132 of the chromosomal inserts present in the pNZ7125 derivatives originating from the selected clones was determined and corresponded to 119 unique loci of the *L. plantarum* genome, since one locus was found three times, while 11 loci were found twice. According to the current genome annotation database of *L. plantarum* WCFS1 (Kleerebezem *et al.* 2003) these loci harbour 71 unique genes and their upstream transcription regulatory sequences (SEQ ID NO: 1-62 or fragments or variants thereof) in the proper orientation to explain the observed induction of *cre* expression (Table 1, supra). As noted above, 8 loci contained more than one putative 5'-end of an annotated ORF and their regulatory sequence. Remarkably, two independent *Sau*3AI clones corresponding to Lp_0291 and its upstream sequence (SEQ ID NO: 47) were found that differ in size (about 1.0 and 1.5 kb). One clone contained the whole of SEQ ID 47 as insert (1551 bp, i.e. about 1.5 kb), while the other contained a smaller fragment of about 1 kb (965 bp, starting at nucleotide 148 of SEQ ID NO: 47). Hence, the *in vivo* induction of this gene was independently confirmed twice during the R-IVET procedure. The identified GI-tract-induced genes appeared to be randomly located within the *L. plantarum* genome. Moreover, the genes appeared to be randomly distributed among the functional classes of the *L. plantarum* genome. The genes identified in this R-IVET screen were functionally categorized in groups involved in nutrient acquisition and synthesis (18 genes), stress response and adaptation (4 genes), extracellular proteins (4 genes), regulation (3 genes) and others (18 genes).

The remaining 24 genes encoded (conserved) hypothetical proteins of unknown function (Table 1, supra).

To verify the primary results of the R-IVET screen, the 132 sequenced pNZ7125 derivatives originating from the resolved clones after the first animal experiment were divided in four groups of 33 pNZ7125 derivatives which were collectively reintroduced into NZ7109. Replica-plating of the transformants revealed that all tested colonies displayed an erythromycin resistant phenotype, confirming the absence of *cre* expression in these clones under laboratory conditions. A second round passage in mice revealed a dramatic increase in the percentage of erythromycin sensitive colonies that was recovered from the faecal samples of all individual mice used during this animal experiment as compared to the experimentally identical first round passage (Table 2). The fact that the second round passage resulted in an average of 38.1% erythromycin sensitive clones in the 8 mice as compared to 3.3% in the first round passage confirmed the R-IVET based enrichment of *L. plantarum* chromosomal fragments harbouring *in vivo* induced transcription regulatory sequences.

Table 1: see above

Table 2: A second round passage of R-IVET positives through mice results in an increased percentage of erythromycin sensitive (ery^S) NZ7109 colonies as compared to the first round passage (3.3%).

Mouse:	Promoter groups	ery^S (%):
1	1	58.0
2	1	71.7
3	2	52.9
4	2	28.8
5	3	34.2
6	3	33.3
7	4	31.5
8	4	37.5
1-8	1-4	38.1

2.4 Analysis of *in vivo* GI-tract-induced genes

This is the first R-IVET approach in a food-grade bacterium, which resulted in the identification of 71 transcription regulatory sequences (SEQ ID NO: 1-62) that are induced *in situ* in *L. plantarum* WCFS1 during passage of the mouse GI-tract. The identified genes appeared to be randomly distributed over the chromosome. Moreover, genes from many functional classes were identified and grouped into 6 functional classes (Table 1, supra).

Nine of the genes identified using R-IVET are involved in sugar transport and utilization (Table 1; transcription regulatory sequences SEQ ID NO: 1-8), including 5 (components of) PTS (phosphotransferase) systems, specific for N-acetylglucosamine, sorbitol, sucrose (2x) and cellobiose, a ribose permease and a ribose kinase, and two di- and polysaccharide hydrolyzing enzymes. A diverse carbohydrate potential has been associated with several gram-positive microbes inhabiting the GI-tract, including *L. plantarum* (Kleerebezem *et al.* 2003), *Listeria inocua* and *Listeria monocytogenes* (Glaser *et al.* 2001, Science 294, 849-52) and *Bifidobacterium longum* (Schell *et al.*, 2002, PNAS 99, 14422-7). The finding that several of these functions are induced *in situ* in the GI-tract, confirms their importance for survival and persistence under GI-tract conditions. Moreover, genes involved in the metabolism of the same sugars have been reported to be induced *in vivo* in a host system in several pathogens, including *Klebsiella pneumoniae* (ribose, Lai *et al.* 2001, Infect. Immun. 69, 7140-5), *Salmonella typhi* (ribose, Daigle *et al.* 2001, Mol. Microbiol. 41, 1211-22, *Pasteurella multocida* (sorbitol, Hunt *et al.* 2001, Infect. Immun. 69, 3004-12), *Listeria monocytogenes* (cellobiose, Gahan and Hill 2000, Mol. Microbiol. 36, 498-507) and *S. gordonii* (sucrose and cellobiose, Kilic *et al.* 1999, Plasmid 42, 67-72). Remarkably, the IIC transport component of the cellobiose PTS system found (Lp_1164; transcription regulatory sequence SEQ ID NO: 2) is not located in an operon-like structure. In *Listeria monocytogenes* similar "orphan" cellobiose-PTS-IIC components have been shown to play a role in host-specific signalling, leading to modulation of virulence gene expression (Kreft and Vazquez-Boland, 2001, Int. J. Med. Microbiol. 291, 145-57), suggesting host-factor mediated gene regulation in bacteria, possibly including *L. plantarum*.

Nine genes were identified that are involved in the acquisition of non-sugar compounds, including factors involved in amino acid, nucleotide and cofactor metabolism (Table 1; transcription regulatory sequences SEQ ID NO: 9-17). These results suggest that limiting amounts of these compounds are available under the starvation conditions in the GI-tract, leading to activation or derepression of these *L. plantarum* genes. Accordingly, *in vitro* studies in different bacteria confirm the induction of several of these metabolic pathways under end-product limiting conditions (Muse *et al.* 2003, J. Bacteriol. 185, 2920-6; Khan & Isaacson 2002, Infect. Immun. 70, 3404-12; Cronan 1988, J. Biol. Chem. 263, 10332-6; Mack et al. 1998, J. Bacteriol. 180, 950-5). Moreover, *in vivo* approaches have demonstrated that genes in arginine and proline metabolism in *V. cholerae* (Camilli & Mekalanos 1995, Mol. Microbiol. 18, 671-83) and *Helicobacter pylori* (Kavermann et al. 2003, J. Exp. Med. 197, 813-22), respectively, are induced during mouse infection. In one of the closest relatives of *L. plantarum* (*Listeria monocytogenes*) proline metabolism is induced under high osmolarity conditions (Sleator *et al.* 2001, Appl. Environm. Microbiol. 67, 2571-7). Such conditions could potentially be found in the colon and might suggest differential colonic expression of Lp_0017 (transcription regulatory sequence SEQ ID NO: 9). Similarly, other experiments in the inventors lab have identified Lp_3473 (transcription regulatory sequence SEQ ID NO: 4), encoding an α-rhamnosidase, as induced by high osmolarity, supporting its colonic induction.

The copper transporting ATPase identified here (Lp_3055, Table 1, transcription regulatory sequence SEQ ID NO: 19) could be involved in copper acquisition. Alternatively, this transporter could act as an exporter, thereby preventing accumulation of copper in the cytoplasm. Arbitrarily, this gene was categorized as a stress-related protein, involved in copper detoxification. Three other genes were categorized as stress-related, namely *clpC* (transcription regulatory sequence SEQ ID NO: 18), a multidrug transporter (transcription regulatory sequence SEQ ID NO: 21) and a cation efflux protein (Table 1, transcription regulatory sequence SEQ ID NO: 20), suggesting that the efficient transport of toxic compounds is important for GI-tract persistence of *L. plantarum*. Genes important in the transport of metals have been identified in many IVET screens (Mahan *et al.* 2000, supra), including a heavy metal transporter in *S. typhimurium* possibly involved in Cu^{2+} homeostasis during infection (Heithoff et al. 1997, PNAS 94, 934-9). Moreover, in *Helicobacter pylori* an STM

strategy revealed two genes encoding multidrug transporters to be essential for gastric colonization of mice (Kavermann *et al.* 2003, supra). In several gram-positive microorganisms *clp* genes were demonstrated to be involved in stress response (Varmanen *et al.* 2000, Microbiol. 146, 1447-55; Nair *et al.* 2000, Mol. Microbiol. 35, 800-11; Derre *et al.* 1999, Mol. Microbiol. 31, 117-31). In *S. mutants* the *clpC* operon is induced at low pH (Lemos & Burne 2002, J. Bacteriol. 184, 6357-66), which could suggest that SEQ ID NO: 18 might be induced in *L. plantarum* in the mouse stomach. In *L. plantarum* the *clpC* gene is chromosomally located directly downstream of the regulator encoding gene *ctsR*. In *Listeria monocytogenes* a similar operon structure was found and mutants in the *ctsR* gene display a lower ability to survive the initial stages of murine infection in mice (Karatsas et al. 2003, Mol. Microbiol. 49, 1227-38).

Four genes encoding extracellular proteins were identified in the R-IVET screen, including two proteins (Lp_0800 and Lp_2940, Table 1 transcription regulatory sequences SEQ ID NO: 22 and 23) that contain an LPXTG-like motif involved in anchoring them to the bacterial cell wall (Mazmanian et al. 1999, Science 285, 760-3). Lp_0141 (transcription regulatory sequence SEQ ID NO: 25) contains a high number of positive charges, which could be involved in the interaction of the encoded protein with the cell wall. No putative binding domains were found in the protein encoded by Lp_1403 (transcription regulatory sequence SEQ ID NO: 25), suggesting secretion of this protein. The bacterial surface is the primary site of interaction with the host and numerous surface exposed adhesion factors have been described (Klemm & Schembri, 2000, Int. J. Med. Microbiol. 290, 27-35). Therefore, the two proteins containing an LPXTG-like motif identified here might represent factors that directly interact with host cells in the GI-tract or with components excreted in the GI-tract lumen of the host. Lp_0800 encodes a protein that is extremely rich in serine and threonine. For a serine-rich surface protein encoded by *Streptococcus pneumoniae* it has been suggested that the serine residues might be glycosylated by glycosyltransferases that are encoded by genes flanking the surface protein encoding gene (Tettelin *et al.* 2001, Science 293, 498-506). These glycosylated serine residues could resemble mucin like structures that coat the bacterial surface or interact with host cell mucins (Tettelin *et al.* 2001). Although no glycosyltransferase encoding genes appear to be genetically linked to Lp_0800, a similar role might be fulfilled in *L. plantarum* by the Lp_0800 encoded protein.

Three regulators were found to be induced *in vivo*, including a regulator belonging to the BglB, AraC and serine/threonine protein kinase family (corresponding transcription regulatory sequence SEQ ID NO: 26-28). The *bgl* operon in *L. plantarum* was previously shown to be down-regulated in the presence of glucose (Marasco et al. 2000, FEMS Microbiol. Letter 186, 269-73). Therefore, the BglG transcription antiterminator might be involved in the regulation of the response to the different sugars *L. plantarum* ferments during passage of the GI-tract. Remarkably, among the best homologues of Lp_3514 in the *Listeria monocytogenes* genome is the *bvrA* gene (33% identity), which encodes a BglG-family antiterminator involved in regulation of virulence gene expression (Brehm et al. 1999, J. Bacteriol. 181, 5024-32).
Eighteen genes are involved in diverse pathways, including DNA and energy metabolism, protein fate and synthesis, and fermentation (Table 1, corresponding transcription regulatory sequences SEQ ID NO: 29-43, 7, 16). Several genes in these pathways have previously been described as being important for pathogenesis in various bacteria, including the 1,3-propanediol regulator and a short chain dehydrogenase of *K. pneumonia* (Maroncle *et al.* 2002, Infect. Immun. 70, 4729-34), a bifunctional protein possessing alcohol and acetaldehyde dehydrogenase activity in *E. coli* K1 (Martindale *et al.* 2000, Mol. Microbiol. 37, 1293-305) and a DNA polymerase in *S. gordonii* (Kilic *et al.* 1999, supra). Another interesting observation is the apparent induction of PlnI, an immunity protein against plantaricin (Diep *et al.* 1996, J. Bacteriol. 178, 4472-83), suggesting that the production of this bacteriocin is important for *L. plantarum* in the highly competitive environment that exists in the GI-tract. Finally, 24 hypothetical proteins (16 conserved and 8 unique) apparently play a role during passage of *L. plantarum* through the GI-tract (Table 1, transcription regulatory sequence SEQ ID NO: 44-62, 19, 21). Although, these putative genes were used for extensive analysis, using the available (R)-IVET literature and blast searches, no significant homologies could be found between these *L. plantarum* hypothetical proteins and other *in vivo* induced genes found in other species. The role of these genes in *L. plantarum* GI-tract persistence remains to be determined.
Overall, a striking amount of parallels can be drawn between the pathogenic and non-pathogenic *in vivo* response, strongly suggesting that survival rather than virulence is the explanation for the importance of these genes during host residence. This suggestion is further enhanced by the fact that the gene encoding peptide methionine

sulfoxide reductase has previously been identified using IVET in the food-associated microbe *L. reuteri* (Walter *et al.* 2003, supra) and the pathogen *S. gordonii* (Kilic *et al.*, 1999, supra). The similarities found for the *L. plantarum* R-IVET screen presented here are most prominent with *in vivo* screens that are performed in the same host (mice) using pathogens that infect the same host-organ (GI-tract), suggesting that organ- and host-specific factors play a key role in the determination of the microbial response. In addition, a number of functions that are specifically induced in *L. plantarum* during passage of the mouse GI-tract, which (so far) have not been identified in similar *in vivo* studies in other bacteria, might contribute to specific interactions between this bacterium and its host. Moreover, parallels between the *in vivo* study presented here and *in vitro* studies performed in the same bacterium (osmolarity induction, see above) or closely related species (osmolarity induction or low pH, see above) might hint towards spatial differentiation of *L. plantarum* gene expression during passage of the mouse GI-tract, i.e. specific induction in the stomach, small intestine or colon (see above). Interesting in this respect is the recent observation that expression of Lp_0237 (transcription regulatory sequence SEQ ID NO: 29) and Lp_0775 (transcription regulatory sequence SEQ ID NO: 12) is induced by bile, suggesting *in vivo* induction in the duodenum, as this is the site of bile release by the host. In conclusion, R-IVET performed in *L. plantarum* is a very efficient method for identifying *in vivo* induced genes and their transcription regulatory sequences.

2.5 Activity of GI-tract induced transcription regulatory sequences in human subjects

Preliminary experiments using human GI-tract biopsies have shown that several of the GI-tract specific genes identified above are also found to be expressed by *L. plantarum* in the human GI-tract. The identified transcription regulatory sequences of the invention are thus also specifically induced in the human GI-tract and in the vertebrate GI-tract. This is due to the fact that the physicochemical conditions encountered by bacteria in the GI-tract of mice, humans and other vertebrates are comparable (for example low pH in the stomach, stress conditions induced by bile in the duodenum, etc.). Moreover, and without limiting the invention, it may be that the interactions between the bacteria and the mucosal surfaces found in vertebrate GI-tracts require expression of the same (or very similar) genes.

Claims

1. A method for producing a biologically active substance *in vivo* in the body cavity of a subject, wherein the method comprises the step of administering to the subject a bacterial host capable of producing the biologically active substance, whereby the bacterial host contains a chimeric gene comprising a transcription regulatory sequence that is induced in at least a body cavity of the subject, whereby the transcription regulatory sequence is operably linked to a sequence encoding the biologically active substance or to a sequence encoding a polypeptide involved in the biosynthesis of the biologically active substance.

2. The method according to claim 1, whereby the transcription regulatory sequence is induced *in vivo* in at least one of the following body cavities of the subject: the stomach, the duodenum, the jejunum or the ileum, the cecum, the colon, the rectum.

3. The method according to claim 1 or 2, whereby the bacterial host is a lactic acid bacterium.

4. The method according to any of the preceding claims, wherein the bacterial host is selected from the group consisting of a food grade bacterium, a commensal bacterium or an attenuated pathogenic bacterium.

5. The method according to claim 4, wherein the bacterium is one of the genus *Lactobacillus*, *Lactococcus*, *Leuconostoc*, *Streptococcus*, *Carnobacterium*, *Bifidobacterium*, *Bacteroides*, *Eubacterium*, *Clostridium*, *Fusobacterium*, *Propionibacterium*, *Enterococcus*, *Staphylococcus*, *Peptostreptococcus*, or *Escherichia*.

6. The method according to any one of the preceding claims, whereby the transcription regulatory sequence is selected from:

(a) a sequence comprising any of the sequences of SEQ ID NO: 1-62;

(b) a sequence comprising a fragment of any of the sequences of SEQ ID NO: 1-62 whereby the fragment retains the ability to be specifically induced in a body cavity of a subject; or,

(c) a sequence which is substantially identical to any of the sequences of (a) or (b) and which retains the ability to be specifically induced in a body cavity of a subject.

7. A chimeric gene comprising a GI-tract-inducible transcription regulatory sequence and further comprising a coding sequence operably linked to the transcription regulatory sequence, whereby the coding sequence is not natively associated with the transcription regulatory sequence.

8. A chimeric gene according to claim 7, whereby the GI-tract-inducible transcription regulatory sequence is selected from:

(a) a sequence comprising any of the sequences of SEQ ID NO: 1-62;

(b) a sequence comprising a fragment of any of the sequences of SEQ ID NO: 1-62 whereby the fragment retains the ability to be specifically induced in the GI-tract of a subject; or,

(c) a sequence which is substantially identical to any of the sequences of (a) or (b) and which retains the ability to be specifically induced in the GI-tract of a subject.

9. A nucleic acid vector, characterized in that the vector comprises the chimeric gene according to claim 7 or 8.

10. A host cell characterized in that the host cell comprises a chimeric gene according to claim 7 or 8.

11. A host cell according to claim 10, whereby the host cell is a bacterium that belongs to a genus selected from the group consisting of *Lactobacillus, Lactococcus, Leuconostoc, Carnobacterium, Streptococcus, Bifidobacterium, Bacteroides, Eubacterium, Clostridium, Fusobacterium, Propionibacterium, Enterococcus, Staphylococcus, Peptostreptococcus,* and *Escherichia.*

12. A composition comprising a host cell according claim 11.

13. A method for the production of a composition as defined in claim 12, wherein the method comprises culturing a host cell as defined in claim 11, and optionally recovery of the recombinant host cell.

14. Use of GI-tract-inducible transcription regulatory sequence in the production of a biologically active substance *in vivo* in a body cavity of a subject.

15. Use according to claim 14, characterized in that the GI-tract-inducible transcription regulatory sequence is selected from:

(a) a sequence comprising any of the sequences of SEQ ID NO: 1-62;

(b) a sequence comprising a fragment of any of the sequences of SEQ ID NO: 1-62 whereby the fragment retains the ability to be specifically induced in the GI-tract of a subject; or,

(c) a sequence which is substantially identical to any of the sequences of (a) or (b) and which retains the ability to be specifically induced in the GI-tract of a subject.

16. Use of a chimeric gene as defined in claim 7 or 8, a nucleic acid vector as defined in claim 9, a recombinant host cell as defined in claim 10 or 11, in the production of a polypeptide or a biologically active substance in the body cavity of a subject.

17. Use of a recombinant host cell as defined in claim 10 or 11 in the preparation of a composition for producing a biologically active substance *in vivo* in a body cavity of a subject.

18. Use of a recombinant host cell as defined in claim 17, wherein the composition is a medicament for preventing or treating an infection by a pathogenic micro-organism of a body cavity of a subject.

19. Use according to any one of claims 16 to 18, wherein the body cavity is the gastrointestinal tract or part thereof.

20. A GI-tract lytic strain comprising a transcription regulatory sequence selected from:

(a) a sequence comprising any of the sequences of SEQ ID NO: 1-62;

(b) a sequence comprising a fragment of any of the sequences of SEQ ID NO: 1-62 whereby the fragment retains the ability to be specifically induced in a body cavity of a subject; or,

(c) a sequence which is substantially identical to any of the sequences of (a) or (b) and which retains the ability to be specifically induced in a body cavity of a subject, wherein said GI-tract lytic strain releases its cytoplasmic content following induction of said transcription regulatory sequence and wherein said cytoplasmic content comprises a biologically active substance.