Direct selection and phage display of the *Lactobacillus rhamnosus* HN001 secretome

A thesis presented to Massey University in partial
fulfillment of the requirements
for the degree of Doctor of Philosophy

by
Dragana Jankovic
2008

Acknowledgments

I would like to thank the following people for the time, help, and support they have given me during my PhD work.

Firstly and primarily I would like to thank my supervisor Dr Jasna Rakonjac for her encouragement, help, support, and expedience which were most appreciated. Her great scientific enthusiasm and amazing energy have been and will always be the inspiration for me. Also many thanks go to my cosupervisors Dr Mark Lubbers and Dr Michael Collett who were always a well of new ideas and discussion topics off all kinds. Their support was invaluable.

I thank my cosupervisor Dr John Tweedie for helpful discussions and critical comments on my work.

Special thanks to my office mates from Helipad lab, especially David Sheerin and Nicholas Bennett whose tolerance, patience and entertaining discussions were sometimes the only thing that kept me sane.

Thanks to everyone at the Institute of Molecular BioSciences for their support. Thanks to them I looked forward to every coffee break to discuss new and interesting current issues. It was a pleasure and privilege working with such a great group of people.

A special thanks goes to my sisters, Tatijana and Vesna, and my nephews Fedja and Filip for all their support and encouragement, without which I would never have made it this far.

Finally all this would not have been possible without amazing support from my husband Milan, my mum and my daughter Mina whose presence and encouragement was always with me.

Contents

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

BLAST	Basic Local Alignment Search Tools
BRB	Bromphenol blue
BSA	Bovine Serum Albumin
cfu	Colony forming units
Da	Dalton
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside triphosphate
ECM	Extracellular matrix
EDTA	Ethylenediaminetetra Acetic Acid Disodium salt
ELISA	Enzyme-Linked Immunosorbent Assay
EtBr	Ethidium bromide
Fn	Fibronectin
HRP	Horseradish Peroxidase
IAA	Iso Amyl Alcohol
IPTG	Isopropylthio-β-D-galactoside
Kbp	Kilobase pair
LAB	Lactic acid bacteria
LTA	Lipoteichoic acid
min	minutes
Mb	megabases
OD	Optical density
ORF	Open reading frame
ori	Origin of replication
PBS	Phosphate buffer saline
PCR	Polymerase chain reaction
pfu	Plaque forming units
RBS	Ribosome binding site
s	seconds
SDS	Sodium dodecyl sulfate

SDS-PAGE	SDS-polyacrylamide gel electrophoresis
Tm	Melting temperature
TMB	3,3’,5,5’ tetramethylbenzidine
UV	Ultraviolet
v/v	volume/volume
wt	wild type
w/v	weight/volume

ABSTRACT

Bacteria communicate with their hosts in part via surface, secreted and transmembrane proteins (collectively the secretome) resulting in probiotic (beneficial) or pathogenic (harmful) outcomes to the host. Therapeutic benefits of probiotic bacteria have been shown previously, but the molecular mechanisms and the health-promoting effector components involved are still being elucidated. Some evidence suggests that probiotic bacteria can competitively adhere to intestinal mucus and displace pathogens. The adherence of probiotic bacteria to human intestinal mucus and cells appears to be mediated, at least in part, by secretome proteins.

Secretome proteins-encoding open reading frames can be identified in bacterial genome sequences using bioinformatics. However, functional analysis of the translated secretome is possible only if many secretome proteins are expressed and purified individually. Phage display technology offers a very efficient way to purify and functionally characterise proteins by displaying them on the surface of the bacteriophage. While a phage display system for cloning secretome proteins has been previously reported it is not efficient for enrichment and display of Gram-positive secretome proteins.

In this study a new phage display system has been developed and applied in direct selection, identification, expression and purification of Gram-positive *Lactobacillus rhamnosus* strain HN001 secretome proteins. The new phage display system is based on the requirement of a signal sequence for assembly of sarcosyl-resistant filamentous phage virions. Using this system 89 secretome open reading frames were identified from a library of only 10^6 clones, performing at least 20-fold more efficiently than the previously reported enrichment method. Seven of the identified secretome proteins are unique for *L. rhamnosus* HN001.

A *L. rhamnosus* HN001 shot-gun phage display library was also constructed to capture proteins that mediate adhesion or aggregation, initial steps in establishing host-microbe contact or forming multicellular aggregates, both of which may lead to beneficial effects – colonisation of the gastro-intestinal tract and exclusion of pathogens. In search for proteins involved in adhesion, a *L. rhamnosus* HN001 shot-gun phage display library was screened against the human extracellular matrix component fibronectin commonly used as binding

target by bacteria that colonise diverse tissues. This screen selected, instead of a fibronectin-binding protein, a protein that binds to avidin, used to immobilise biotinylated fibronectin.

Affinity screening of the shot-gun library for binding to *L. rhamnosus* HN001 cells identified a secretome protein, Lrh33, as an HN001-cell surface binding protein. This protein contains two bacterial immunoglobulin-like domains type 3. Analysis of phage-displayed nested deletions of Lrh33 determined that the proximal (N-terminal) immunoglobulin-like domain is not sufficient for binding; only the constructs displaying both domains demonstrated binding to HN001. Lrh33 does not have any similarity to previously identified *Lactobacillus*-binding proteins and no match in the NCBI database (at a cutoff value of $> e^{-13}$), hence it represents potentially a new type of bacterial auto-aggregation protein.

Chapter IA

Probiotic effects of lactic acid bacteria with special reference to *Lactobacillus rhamnosus* HN001

1.1 Probiotic bacteria

The human gastrointestinal tract (GIT) contains a complex, dynamic, and diverse society of between 300 and 1000 microbial species and some 10^{14} microorganisms in total [1, 2]. This indigenous microbial ecosystem is essential to the overall health of the host by performing important physiological functions, including protection against pathogens, nutrient processing, stimulation of angiogenesis, development of the immune system during neonatal life and regulation of host fat storage [3-6]. The microbial consortia are acquired at birth and the process of re-colonisation occurs throughout life. The climax population of commensal bacteria is dependent on environmental and genetic factors, whereas the influence of factors including mode of infant delivery and diet remain controversial [7-9]. Thus this make-up of the GIT microbial community is unique to each individual, in terms of the actual species present and their relative proportions.

In the normal physiological state, the mucosal surface of the intestine must remain immunologically hyporesponsive to commensal bacteria while retaining its capacity to respond to a pathogenic challenge [10]. This homeostatic balance between tolerance and immunity is exemplified by evidence that dysregulation of the balance can contribute to the pathogenesis of numerous inflammatory conditions, including food allergies, inflammatory bowel diseases (IBD) and intestinal cancer [11-15]. Given that over the past 50 years in societies with improved medical care and hygiene, the steady increase of these diseases was reported, the search for biotherapeutic agents that can ameliorate these pathological conditions has been initiated. During the last two decades several studies have shown that supplementation with preparations of “mono- or mixed cultures of live microorganisms which, when applied to animal or man, beneficially affect host by improving the properties of the indigenous microflora” [16]. Such microorganisms, collectively named probiotics (from Greek “*pro*” and “*biota*” –“for life”) may not necessarily be constant inhabitants of the GIT but should be viable and have beneficial effects on the host. Although in recent years, it was demonstrated that some cell wall components or DNA derived from a few probiotic strains duplicated beneficial effects to the host by live bacteria [17, 18].

The health benefits attributed to probiotics are either nutritional or therapeutic in nature. Enhanced bio-availability of minerals [19] and vitamins and increased digestibility of proteins

in yoghurt are just some of the nutritional benefits that probiotic bacteria contribute to the well-being of the host [20]. Numerous therapeutic benefits have been proposed, ranging from treatment of conditions including gastrointestinal disorders, hyper-cholesterolaemia, lactose intolerance, to suppression of pro-carcinogenic enzymes, immunomodulation and treatment of food-related allergies [21, 22]. However, despite a vast literature on efficacy of probiotics (e.g. >200 articles on probiotic efficacy of *Lactobacillus rhamnosus* GG [23, 24]), the precise mechanistic basis of probiotic action remains a major research goal [25].

1.1.2 Selection of probiotic strains

Bacterial strains selected as probiotic candidates are often of human origin or are those utilized in the food industry. The potential probiotic organism should not be contaminated with pathogenic organisms and it should have GRAS status (Generally Recognised As Safe) [26]. Some additional requirements have been proposed that successful probiotic strain needs to have in order to be able to exert its beneficial effects. These includes the ability to (i) adhere to host intestinal epithelial cells; (ii) exclude or reduce pathogenic adherence; (iii) persist and multiply; (iv) coaggregate and form a normal, balanced flora [4]. However the probiotics used today have not been selected on the basis of all these criteria but just several of them. Though the efficacy of probiotics ultimately needs to be demonstrated in controlled clinical trials to justify their use in specific disease prophylaxis, their primary selection is based on a series of well-defined *in vitro* and *in vivo* tests aimed to demonstrate some of the requirements mentioned above. [27].

Most probiotic bacteria fall into the group of microorganisms known as lactic-acid producing bacteria, with majority belonging to two bacterial genera: *Lactobacillus* and *Bifidobacterium*. A detailed description of probiotic activity of *Bifidobacterium* is beyond the scope of this introduction and is covered by recent reviews [28, 29].

1.2 Lactic Acid Bacteria

1.2.1 General characteristics of lactic acid bacteria

Lactic acid bacteria (LAB) consist of a number of bacterial genera within the phylum *Firmicutes*. The genera *Aerococcus*, *Alloiococcus*, *Carnobacterium*, *Dolosigranulum*,

Enterococcus, *Globicatella*, *Lactobacillus*, *Lactococcus*, *Lactosphaera*, *Leuconostoc*, *Oenococcus*, *Pediococcus*, *Streptococcus*, *Tetragenococcus*, *Vagococcus* and *Weissella* are recognised as LAB (Figure 1.1). Lactic acid-producing Gram-positive bacteria which are often considered in the same context as LAB even though they are phylogenetically distinct from the phylum *Firmicutes* are genera *Microbacterium*, *Propionibacterium* and *Bifidobacterium* of the phylum *Actinobacteria* [30]. LAB are Gram-positive, non-sporulating bacteria that ferment carbohydrates to lactic acid. Depending on the organism, metabolic pathways differ when glucose is the main carbon source: homofermentative genera such as *Lactococcus* and *Streptococcus* yield two moles of lactic acid from one glucose molecule, whereas the heterofermentative genera (e.g. *Leuconostoc*) catabolise glucose into 1 mole of lactic acid and additional products such as ethanol, carbon dioxide, acetate, formate and succinate.

LAB are widely used in food fermentation for two reasons: to improve the aroma and texture of food and to act as a preservative, inhibiting the growth of spoilage bacteria. However, LAB are also found naturally associated with foods of animal origin such as dairy and meat, fresh vegetables and natural plant environments, where fermentation can occur. Another important type of niche occupied by LAB are animal and human cavities, including gastrointestinal tract, the oral cavity and the vaginal cavity. LAB are present throughout the GIT in variable amounts. For instance, species of *Lactobacillus* genus are among the dominant bacteria in the small intestine (duodenum and jejunum) while in the colon they represent about 1% of the microorganisms per gram of luminal content [31].

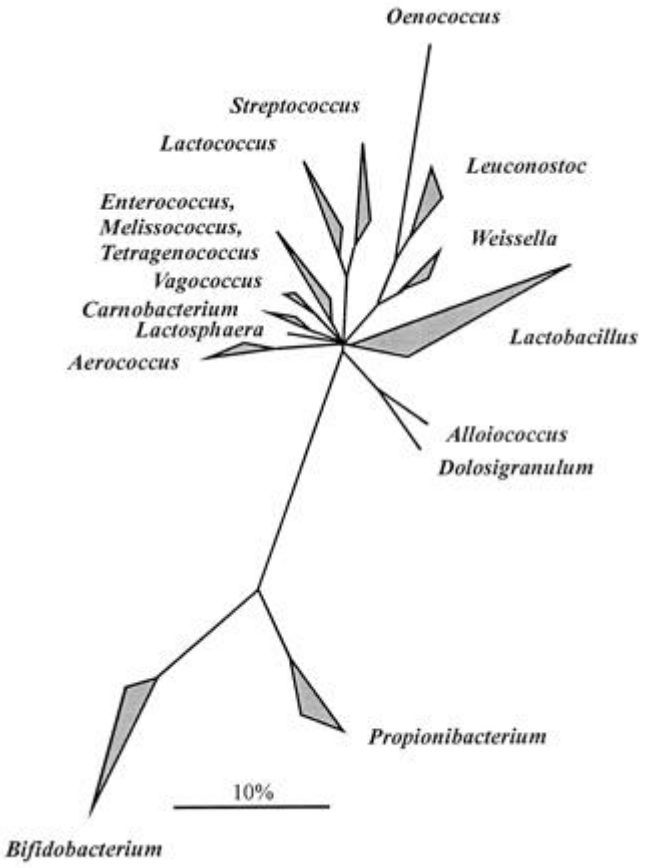


Figure 1.1. Consensus tree of major phylogenetic groups of lactic acid bacteria
The tree is based on comparative sequence analysis of 16S rRNA, showing major phylogenetic groups of lactic acid bacteria with low mol% G+C in DNA and the nonrelated Gram-positive genera *Bifidobacterium* and *Propionibacterium*. The bar indicates 10% expected sequence divergence; the size of the triangle indicates a number of species in the genus. Adopted from [30] with permission.

1.2.2 Lactobacilli

1.2.2.1 General characteristics

Lactobacillus is the most numerous genus of LAB. It belongs to the class *Bacilli*, order *Lactobacillales* and family *Lactobacillaceae*. About one hundred species of genus *Lactobacillus* have been identified thus far [32], subdivided by 16S rRNA analysis, DNA-DNA hybridisation and other phylogenetic methods, into three major subgroups: *Lactobacillus delbrueckii* group, *Lactobacillus casei-Pediococcus* group and *Leuconostoc* group [33]. Recently, complete genomes of all major branches of the order *Lactobacillales*

have become available, enabling a more definitive analysis of evolutionary relationships within this order [34]. Phylogenetic analysis of multiple ribosomal protein sequences showed that the *Pediococcus* group is a sister to the *Leuconostoc* group, while *Lactobacillus casei* is placed at the base of the *L. delbruckii* group, which contradicts the earlier classification, in which *L. casei* and *Pediococcus* belong to the same phylogeny (Figure 1.2).

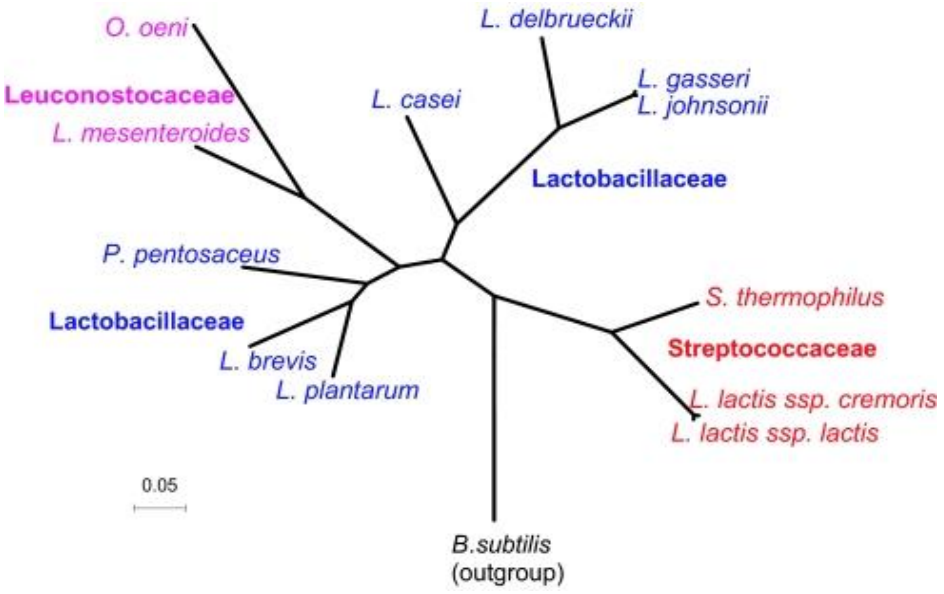


Figure 1.2. Phylogenetic tree of *Lactobacillales* based on concatenated alignments of ribosomal proteins.

All branches are supported at >75% bootstrap values. Species are colored according to the current taxonomy: *Lactobacillaceae*, blue; *Leuconostocaceae*, magenta; *Streptococcaceae*, red. Adopted from [34] with permission.

The cells of lactobacilli are mostly rod-shaped, varying in dimensions from 0.5x1.0 to 1.2x10.0 μm. Lactobacilli are chain-forming and usually non-motile, although some are motile with peritrichious flagella. A few strictly anaerobic species of lactobacilli exist, but the majority are facultative anaerobes and anaerobic growth is generally enhanced by 5% CO_2. The optimum growth temperature for lactobacilli is between 30 and 40°C. Each species within the *Lactobacillus* genus has its own characteristic nutritional requirements. Using glucose as a carbon source, lactobacilli may be either homofermentative or hemofermentative.

Generally, at least half of the carbon-based end product is lactic acid [35]. Lactobacilli are aciduric or acidophilic, with an optimal pH for growth at less than 5.5. Lactobacilli are thus particularly well-adapted to habitats where fermentation occurs which are characterised by low pH.

Lactobacilli are important in the production of foods that require lactic acid fermentation, notably dairy products, fermented meats and breads. Their use in the food industry has a long history, and the functions of these bacteria in the industrial setting have been well studied. However, *Lactobacillus* species that inhabit the mucosal surfaces of animals and humans and mechanisms of their action are much less known, despite an almost 100-year long research interest.

1.2.2.2 General features of the *Lactobacillus* genomes

In recent years the rapidly growing wealth of *Lactobacillus* (and other LAB) genome sequence data has shone not only a new light on their evolution but also on their conserved/varying biosynthetic and metabolic capabilities. *Lactobacillus* genomes sequenced to date range between 1.8 and 3.3 Mbp in size, with a G+C content ranging from 33% to 52%. Plasmids were found in most *Lactobacillus* isolates. Some of these episomes are essential for growth in specific environments and these carry genes encoding proteins involved in metabolic pathways, membrane transport, and bacteriocin production [36].

The comparison of the number of predicted protein-coding genes between *Lactobacillus* species and the most recent common ancestor revealed substantial gene loss and metabolic simplification but also acquisitions of unique genes and a considerable number of lineage-specific duplications [34]. Substantial gene loss manifests in smaller genomes with a proportionally large number of pseudogenes and reduction of enzymes for biosynthesis of amino acids. Compensating for this latter group of deficiencies, the *Lactobacillus* genomes generally encode a large number of peptidases, permeases and transporters that allow use of available extracellular protein as a source of amino acids. This trait occurred in species which are adapted to life in nutritionally rich mediums such as milk and human GIT (e.g. *L. delbruckii* group). For example, 1.86 Mbp genome of *L. delbruckii* spp. *bulgaricus* (hereafter referred to as a *L. bulgaricus*) contains a very high number of pseudogenes (12% of the total number of genes) and has extremely narrow biosynthetic repertoire for amino acids, indicating an active and ongoing process of genome degeneration due to specialisation for life

in the protein-rich milk environment [37]. In the species with larger genomes such as *L. plantarum* (3.3 Mbp) and *L. casei* (2.9 Mbp), it is proposed that the loss of the ancestral genes was compensated by emergence of many new genes via duplication and horizontal gene transfer. Interestingly, the number of pseudogenes in these species is significantly lower (1.3-3% of total number of genes) than in *L. delbruckii* group [34, 38]. Consistently, these two species have broader adaptation that includes less nutritionally rich habitats such as plants and vegetables, in addition to protein-rich environments such as meat, dairy products and the human oral cavity and GIT.

Transposable elements and, particularly, IS elements as potential means of gene transfer are present in all *Lactobacillus* genomes. They occupy from 0.2% of the genome in *L. gasseri* to 3.25% in *L. casei* genome [34]. With exception of *L. bulgaricus*, the similar trend was observed for prophage-like elements and prophage remnants which contribute significantly to bacterial interstrain variability [37, 39].

Promoter mapping by primer-extension analysis showed that lactobacilli -35 and -10 regions closely resembled those of *Escherichia coli* and *Bacillus subtilis*. The region upstream of the -35 motif, the AT-rich UP element present in *E. coli* and *B. subtilis*, is also conserved in *Lactobacillus* species [40, 41].

1.2.2.3 Cell surface of Lactobacilli

The surface properties of lactobacilli are of major importance in fermentation technology. For example, in dairy product manufacturing, adhesion of lactobacilli to a material is a first step leading to biofilm formation, which allows continuous inoculation in yogurt or cheese. The cell surface of lactobacilli also has an important role in host-microbe and microbe-microbe interaction in the GIT. The Gram-positive cell wall of *Lactobacillus* consists mainly of peptidoglycan, (lipo)teichoic acids, proteins and polysaccharides. The peptidoglycan layer is composed of linear chains of alternating *N*-acetylglucosamine and *N*-acetyl-muramic acid units, which are themselves connected by a β–(1,4)-glycosidic bond, and short peptide chains of three to five amino acids attached to the N-acetylmuramic acid [42]. A variety of molecules cover or are embedded in the peptidoglycan layer. Lipoteichoic acids (LTA) are anchored in the cytoplasmic membrane by their lipidic tail whereas teichoic acids are covalently attached to the peptidoglycan. Both polymers are potent immunogens and can be regarded as the Gram-positive equivalent of Gram-negative lipopolysaccharides. LTA was

also found to act as an adhesion factor in some lactobacilli [43]. Neutral polysaccharides can be covalently or loosely bound to the cell wall, or as in case of extracellular polysaccharide (EPS) they can be released into the medium. In some Gram-positive bacteria polysaccharides/EPS form a thick capsular shell while in others they only decorate the cell envelope. EPS and polysaccharides in general differ in various Gram-positive bacteria not only in the nature of sugar monomers, but also in their mode of linkage, branching and substitutions (noncarbohydrate substances). These polysaccharides are extensively studied in pathogens as they contribute to their virulence. A few examples of immunomodulatory activity of the surface polysaccharide from commensal and probiotic bacteria have also been reported (e.g. *Bacterioides fragilis*, *L. bulgaricus* and *Lactobacillus rhamnosus* RW-9595M) [21, 44, 45].

The most abundant surface proteins (10-15% of the proteome) in several *Lactobacillus* species are the S-layer proteins [46]. They are noncovalently bound to the cell wall and assemble into surface layers often completely covering the cell wall. S-layers of some lactobacilli, such as *L. brevis* ATCC 8287, *L. crispatus* and *L. helveticus*, were shown to be involved in adhesion to different host surfaces and competitive exclusion of enteropathogens [18, 47, 48]. However, for many other species of the *Lactobacillus* genus specific functions of the S-layer are still unknown.

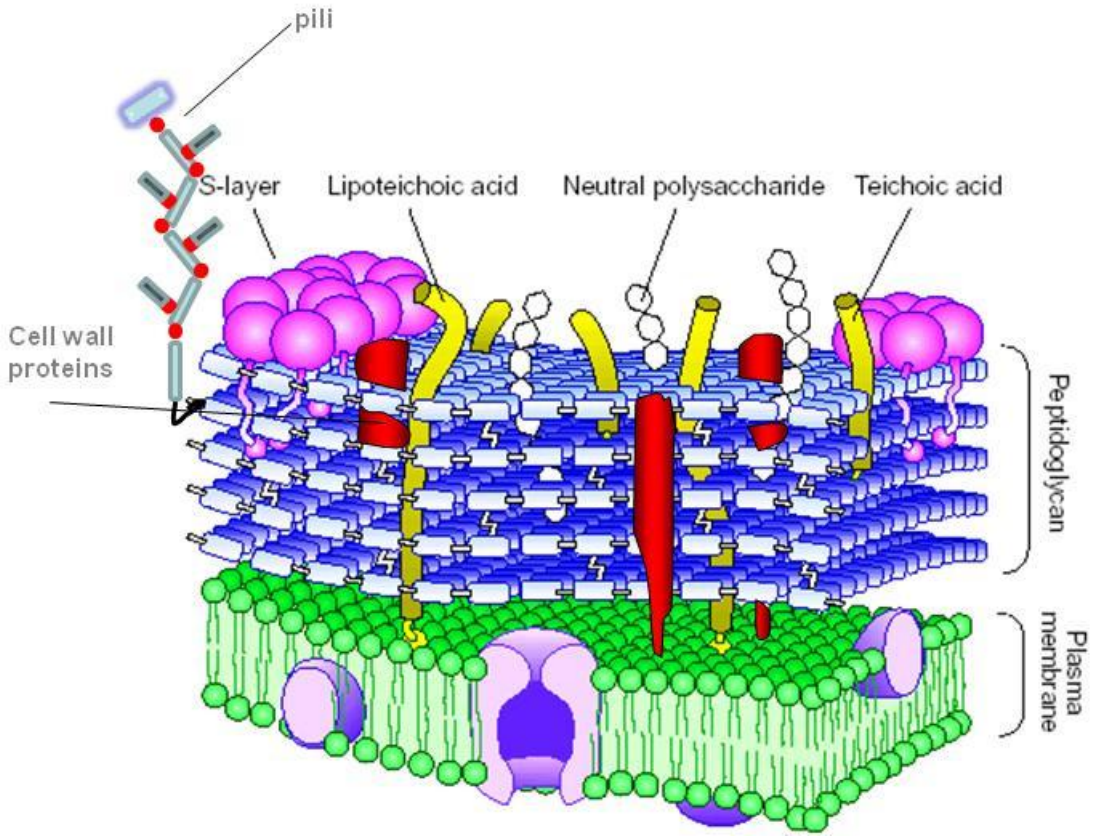


Figure 1.3. The cell wall of a Gram-positive bacterium.
The bilipidic plasma membrane with embedded proteins is covered by a multilayered peptidoglycan shell decorated with neutral polysaccharides, cell wall-associated proteins, lipoteichoic acids and teichoic acids, surrounded by an outer envelope of S-layer proteins. Non-flagellar appendages, pili, are built of covalently linked pilin subunits which are after assembly attached to the cell wall by a transpeptidase called housekeeping sortase. Adapted from [42] with permission (not to scale).

Another complex structure allegedly present on *Lactobacillus* cell surface are pili [49]. In pathogenic bacteria pili play key roles in the adhesion and invasion processes and in pathogenesis in general. Emerging evidence suggests that pili of commensal bacteria have a role in enhancing adhesion and subsequent colonisation of the GIT [50]. Components and mechanisms of pilus assembly differ between Gram-positive and Gram-negative bacteria. For example, subunits of Gram-negative type I or P pili are assembled by donor strand exchange mechanism, in which incomplete immunoglobulin-like fold of each subunit is complemented by a β strand of the neighbouring subunit [51]. In Gram-positive bacteria, major (or shaft)

pilin subunits and tip are covalently cross-linked one to another as well as to the cell-wall by the action of the house-keeping sortase (pilus attachment to the peptidoglycan) and pilin polymerase (subunit polymerisation) [52, 53]. Individual proteins displayed on the cell surface of lactobacilli are attached either to the cytoplasmic membrane or cell wall components of the cell wall envelope. Chapter IB, section 1.4, describes in detail cell surface anchoring and roles of these proteins.

1.3 *Lactobacillus rhamnosus* HN001

1.3.1 History

In the mid-1990's a large culture collection (of over 2000 strains) of Lactic Acid Bacteria (LAB) held at Fonterra Research Centre was subjected to a series of assays to identify strains with functional characteristics typical of probiotic bacteria. Following an initial screening of strains for withstanding GIT environment and colonisation, four strains were identified as putative probiotic strains [54]. Three of the selected strains were of dairy origin and one was of human origin. Further on, these strains were subjected to a "health" screen in animal models to establish some of their probiotic properties and test their safety. Through these tests, a strain of dairy origin, isolated from cheddar cheese, *Lactobacillus rhamnosus*, was designated HN001 and shown to have an excellent set of basic probiotic characteristics: it was safe for human consumption, it colonised the human gut and stabilised the microflora within, and enhanced parameters of both innate and acquired immunity [55-58]. Being a cheese-fermentation isolate, *L. rhamnosus* HN001 was also suitable for use as an adjunct during cheese manufacture, to reduce adventitious microflora, accelerate cheese ripening, and improve cheese flavour. As a result of flavour-enhancing and probiotic attributes, HN001 has been marketed by Fonterra under the DR20™ trademark and is used today in a variety of dairy products, predominantly on international markets.

1.3.2 HN001 genome

To fully unlock the potential value of HN001, a draft genome sequence of this strain has been obtained (unpublished at the time the thesis was written but available since September 2008 in GeneBank; accession # ABWJ00000000). Initially the sequence information was generated by shot-gun sequencing of small and large insert libraries. By the time this research work

commenced about nine-fold coverage was achieved. The complete genome is estimated to be ~ 3 Mb in size by pulse-field agarose gel electrophoresis (B.Kelly, AgResearch Ltd – New Zealand; unpublished observations), with an average G+C content of 46.4%, within the range reported for other *L. rhamnosus* and species of *L. casei* phylogeny group (*L. casei, L. paracasei, L. zeae* and *L. rhamnosus*). Strain HN001 harbors two plasmids, 8.75 Kbp and 31.55 Kbp in size. The draft genome sequence was compared against publicly available databases and to date over 300 genes with potential involvement in flavour, survival and probiotic activity have been identified (Patent number EP1711609). Subsequently, the HN001 genome was sequenced with 454 technology, giving a further 36-fold coverage and the two plasmids were completely sequenced (GeneBank accession # NC_011223 and NC_011225).

1.3.3 Probiotic characteristics of *L. rhamnosus* HN001

In initial screening for the health-promoting bacterium, *L. rhamnosus* HN001 showed ability to survive and propagate at conditions likely to be encountered in human GIT, such as low pH (1.0-3.0) and high bile acids concentration (0.4-1%) [54]. This newly isolated strain did not have previous history of use in food products therefore it was necessary to establish its safety. General safety of HN001 was assessed in both mice and human feeding trials. Results suggested that HN001 is non-toxic, non-invasive at the mucosal surfaces and free of antibiotic resistance-linked plasmids, hence likely to be safe for human use [57, 58].

Presumably, adhesion to intestinal cells and exclusion or reduction of pathogen attachment are important attributes of probiotics. HN001 adheres to various intestinal cell lines (Caco-2, HT-29 and HT29-MTX) and is able to colonise the human GIT and “stabilise” microflora in the intestine [56]. With regard to exclusion or reduction of pathogen adhesion, there is evidence that HN001 reduces the severity of *E. coli* O157:H7 infection in mice [59]. Furthermore, HN001 protects mice against infection by *Salmonella* [60]. This protection effect correlates with enhanced immunity manifested as increased anti-pathogen antibody titre and phagocytosis responses. The mechanisms by which HN001 inhibits pathogen infection are not fully understood at present. It has been reported that other probiotic strains hamper the virulence of pathogens through mechanisms other than competitive adhesion to target cells, for example through production of antibacterial substances and bacterial aggregation [61-63]. HN001 aggregates in laboratory culture, but it is still unknown whether the same is true *in*

vivo (in the GIT). If so, aggregation could extend HN001 retention in GIT and biofilm formation.

The ability to modulate immune function has been well-established for only a few LAB strains, however a large variation exists in the level and mechanisms of immunomodulation [64]. Research involving *ex vivo* cell and animal models, as well as human clinical studies, has shown that several indices of natural and acquired immunity were enhanced after consumption or application of HN001. Phagocytic (monocytes, macrophages, polymorphonuclear leukocytes) and natural killer (NK) cells are the major effectors of natural immunity and their activity was found to be significantly increased by HN001 [55, 65, 66].

T cells are the main effectors and regulators of cell-mediated immunity. On activation by antigen or pathogen via antigenic-presenting cells, T cells synthesize and secrete a variety of cytokines that serve as growth, differentiation and activation factors for other immunocompetent cells. T helper cells can be subdivided into at least three functional types, Th1, Th2 and Th17, based on their cytokine profile. Th1 cells produce cytokines IL-2, IFN-γ and TNF and they are vital for cell-mediated immunity. The Th2-type cytokines include IL-4, IL-5 and IL-13,which are associated with humoral immunity and allergic responses [65]. In excess, Th2 responses counteract the Th1 mediated microbicidal action. Optimally (in healthy subjects), Th1 and Th2 response are well balanced and suited to the immune challenge. In animal studies, mice fed with HN001 produced significantly greater amounts of IFN-γ [67]. In contrast to reported effects of *L. casei* Shirota, HN001 feeding did not appear to down-regulate Th2 cytokines in antigen-primed mice, suggesting a different mode of immune regulation from pro-Th1/anti-Th2 [67, 68]. In summary, the results indicate that *L. rhamnosus* HN001 is a general immunostimulatory agent.

1.4 Colonisation of the human GIT by lactobacilli

1.4.1 Intestinal epithelium

Intestinal epithelium forms a barrier that plays a fundamental role in cross-talk between the host and the luminal content. It consists of a single layer of columnar epithelium supported by connective tissue known as *lamina propria* (Figure 1.4). The *lamina propria* is traversed by

blood and lymphoid vessels and contains innate and adaptive immune cells. The intestinal epithelial cells (IECs) are joined firmly together by the intercellular tight junctions to prevent paracellular traffic. The apical surface of IECs contains microvillar extensions that create a brush border, while the basolateral side is covered with a sheet of specialised extracellular matrix, basement membrane [10]. Extracellular matrix (ECM) is also located between adjacent IECs and is sealed off from the lumen by a tight junction. The ECM is a complex mixture of matrix proteins, such as fibronectin, collagens, laminins, various proteoglycans, and non-matrix proteins, including growth factors. Collagen I and fibronectin can be also shed into the mucus. The ECM is exposed to bacteria in the lumen during the sloughing off of epithelial cells and in case of tissue trauma. Adherence of pathogens to ECM has been investigated thoroughly, demonstrating that ECM-binding is often the mechanism by which they colonise and infect the host [69].

The physical barrier made of IECs and *lamina propria* is reinforced by the presence of a continuous layer lining, at the apical side of IECs, composed of the mucus and the glycocalyx. The mucus is widely distributed throughout the GIT, and is produced by Goblet cells present in the epithelium throughout the intestine. This polysaccharide network forms a relatively rigid and sticky structure, in which bacteria are trapped. Intestinal mucins, the major protein component of the mucus, provide attachment sites for commensal and pathogenic microbes by interacting with bacterial cell-surface polysaccharides and protein appendages like pili [70]. Lined in mucus, bacteria are eliminated by the intestinal peristalsis and the massive and permanent hydrous flow from the digestive tract. In addition to the physical barrier, the mucosal lining is characterised by the production of a broad spectrum of antimicrobial peptides including defensins, calthelidins and calprotectins [71]. These antimicrobial peptides represent a significant biological barrier at the mucosal surface against bacterial pathogens.

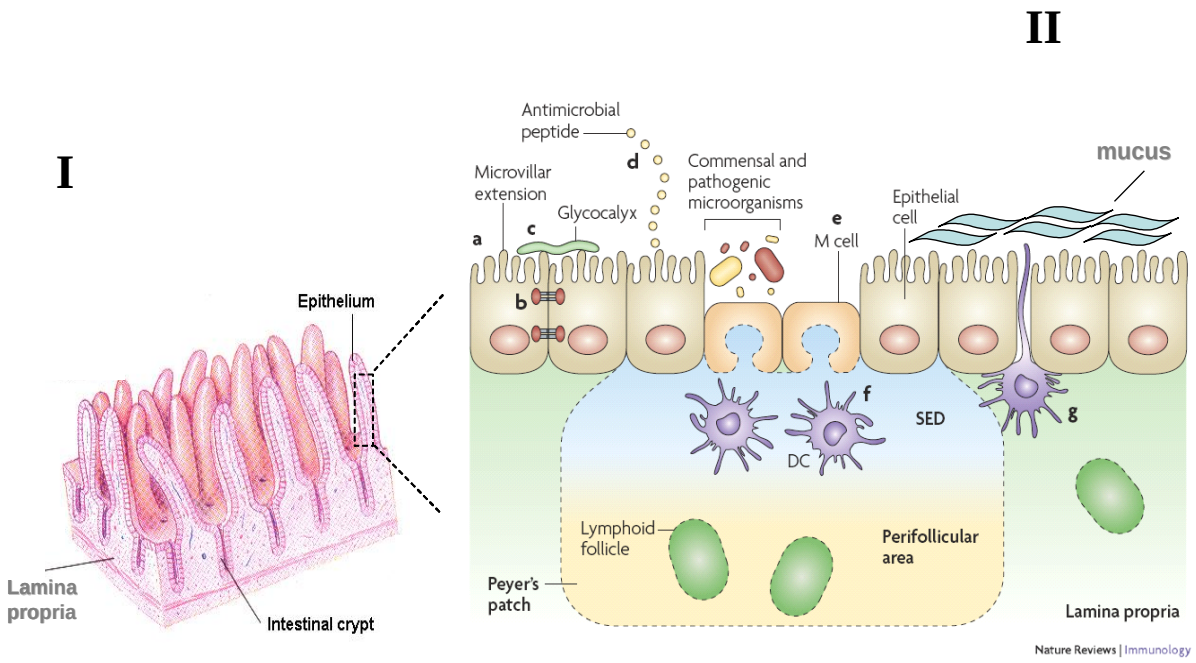


Figure 1.4. Surface of small intestine of humans.

I) The left diagram (adapted from [72] with permission) shows detail of the surface of the small intestine with villi, intestinal crypts and supporting *lamina propria*. **II)** The right diagram (adapted from [10] with permission) shows simple columnar epithelial cells that exhibit physical and biochemical adaptations to maintain barrier integrity including actin-rich microvillar extensions (a), epithelial-cell tight junctions (b), apically attached and secreted mucus and a glycocalyx layer (c) and the production of various antimicrobial peptides (d). Specialised intestinal epithelial cells known as M cells overlie lymphoid organs, Peyer's patches and lymphoid follicles, to facilitate luminal sampling. M cells exhibit reduced mucin secretion and have modified apical and basolateral surfaces (e) to promote uptake of luminal contents and its transport to professional antigen-presenting cells that inhabit the subepithelial dome (SED) of the Peyer's patches and lymphoid follicles (f). Specialized dendritic cell (DC) subsets can also extend dendrites between the tight junctions of intestinal epithelial cells to sample the luminal contents (g).

Sampling the luminal content by the immune system is facilitated by specialised lymphoid structures that constitute the mucosal immune system, including mesenteric lymph nodes, Peyer’s patches in the small intestine, colonic patches and isolated lymphoid follicles. These structures are overlaid by follicle-associated epithelium containing specialised cells, called M

(microfold) cells. Although M cells are derived from IECs they lack microvilli on their apical side and mucus and glycocalyx coating. M cells are adapted to uptake and transport luminal antigens, including antigens from enteropathogenic and commensal bacteria as well as food-derived antigens, to the underlying lymphoid tissue. Probably due to the lack of protective structures at their apical surface, M cells are often utilised by pathogenic bacteria for invasion. In addition to M cells, specialised intestinal dendritic cells located in the *lamina propria* of the small intestine can directly sample the luminal environment [73].

1.4.2 Aggregation

The ability to form multicellular aggregates has been shown to play an important role in colonisation of the oral cavity [74] and the urogenital tract [75]. It is a rather complex phenomenon due to many physical and chemical forces that can drive the interaction between cell surfaces of bacterium and host tissues. The aggregation ability comprises autoaggregation, characterised by clumping of cells of the same strain, and coaggregation in which phylogenetically distinct bacteria become attached to one another via specific surface molecules. Several aggregation-promoting factors (Apf), located on the cell surface or secreted, showing great diversity, have been described previously in the lactobacilli, including the DEAD-box helicase AggH of *L. reuteri* 1063 [63], the S-layer protein CbsA of *L. crispatus* JCM 5810 [76], and a 2 kDa hydrophilic peptide of *L. gasseri* 2459 (Table 1.1) [77]. *In vivo* studies indicate that bacteria with aggregating phenotypes persist longer in the GIT compared with non-aggregating phenotypes [78, 79] and that adhesion, auto-aggregation and co-aggregation are correlated [80, 81]. As in the case of autoaggregating strains, strains with co-aggregating ability have a great advantage by forming large co-aggregates to already established colonizing biofilm of other species [82]. Conversely, co-aggregation of lactobacilli with pathogens may exclude or reduce adhesion of pathogens to the host tissue, possibly by promoting removal of co-aggregates or competition for binding to the mucosal tissue. In summary, co-aggregating strains have competitive advantage over non-aggregating planktonic bacteria and present one of the GIT defence mechanisms. Co-aggregation-promoting factors (Cpf) have been analysed widely in oral and dental environments where biofilm formation is crucial for survival of bacteria, but only a few studies described Cpf(s) of the *Lactobacillus* species. One example is cell-wall associated protein, Cpf, of *L. coryniformis* DSM 20001^T which was reported to mediate co-aggregation with *E. coli* K88, *Campylobacter coli*, and *Campylobacter jejuni* [83].

Table 1.1. Auto-aggregation and co-aggregation promoting factors identified in lactobacilli

Species	Aggregation promoting factors	Subcellular localisation	Reference
L. reuteri 1063	AggH	Cell wall protein	[63]
L. crispatus JCM 5910	CbsA	Cell wall	[76]
L. coryniformis DSM 20001^T	Cpf	Cell wall protein	[83]
L. gasseri 2459	2 kDa peptide	Secreted	[77]
L. paracasei BGSJ2-8	High molecular mass Apf	Cell wall protein	[84]
L. crispatus M274	Metabolic enzymes, EF-Tu	Cytoplasm, Cell wall	[85, 86]
L. reuteri TMW1.106	Extracellular polysaccharides	Secreted	[87]

1.4.3 Adhesion to human GIT

Adhesion, even temporary, of probiotics to the GIT epithelium prolongs their persistence in the GIT and as a consequence may influence the host health by affecting the local microbial composition or by stimulating the mucosal immune system. Adherence to the surface of a healthy GIT usually begins with adhesion to the mucus layer, followed by adhesion to the intestinal epithelial cells (IECs). In the case of invasive pathogens, attachment to the IECs is a requirement for subsequent internalisation and invasion. Initial attachment may be mediated by non-specific forces, particularly hydrophobicity and surface charge, but persistence is achieved by interactions between specific bacterial surface components, known as adhesins, and host receptor molecules on the colonised surface. Bacterial adhesins include carbohydrates, lipids, phosphodiester-containing polymers ((lipo)techoic acids) and cell surface proteins or protein structures, such as pili. Apart from surface-associated components,

bacteria may also secrete soluble components that associate with other bacterial cells to promote adhesion [88] (see Figure 1.3 for a diagram of the Gram-positive cell wall surface).

Pathogenic bacteria have evolved an incredibly large and diverse array of adhesion and invasion molecules that enable them to exploit a variety of IECs surface components. Besides complex protein structures such as pili [89-91] , a plethora of simple adhesins exist which recognise many different elements of IECs surfaces, including components of the extracellular matrix (ECM) such as collagens, laminins and proteoglycans. Host glycoproteins such as vitronectin, laminin, fibrinogen, and especially fibronectin are also recognised by many different species of bacterial pathogens [92].

One of the largest group of microbial adhesins belongs to sugar-binding proteins, lectins. Bacterial lectins display exquisite specificity for the target tissue. They interact with carbohydrates such as D-mannose, D-fucose, D-galactose, sialic acid and N-acetyl-galactosamine exposed on the host cell surfaces in the form of glycoproteins, glycolipids and polysaccharides [93]. Another class of microbial surface adhesins are MSCRAMMs (microbial surface components recognising adhesive matrix molecules), which specifically interact with the ECM ligands. To be classified as a MSCRAMM, the interaction with the ECM ligand must have high affinity and must be specific. Many examples of MSCRAMMs exist and a number of these are described in several reviews [94-96]. A large number of MSCRAMMs are fibronectin-binding proteins, originally identified in Gram-positive pathogens Streptococci and Staphylococci. One of the fibronectin-binding proteins identified in *Streptococcus pyogenes* is Serum Opacity Factor (Sof) [97]. This multifunctional virulence determinant is a cell-wall bound protein consisting of two distinct domains; an N-terminal apolipoprotein A1-binding domain and a C-terminal repeat domain that binds fibronectin and fibrinogen.

There are many other characterised adhesins that bind to components on intestinal surfaces other than those in the ECM. Examples of well-studied adhesins that also bind ligands other than ECM components are the M proteins from *Streptococcus pyogenes* which can bind a range of ligands, including fibrinogen, CD46, galactose, fibronectin, laminin, fucose/fucosylated glycoproteins and sialic acid-containing ligands, depending on the sero-type of M protein [98].

Another functionally diverse group of proteins often involved in attachment and invasion of host IECs and ECM by pathogenic bacteria are the cell surface-associated proteins with immunoglobulin-like (Big) domains [99]. Ubiquitous Big domains can be found in all kingdoms of life, including dsDNA phages. Although Big domains have a common basic topology (seven β-strands arranged into two distinct sheets packed in a parallel manner), their amino acid sequences are highly divergent [100]. Immunoglobulin-like domains can bind small molecules, hormones, or large protein complexes. The well-characterised bacterial representatives of this superfamily are type 1 pilus-associated adhesin (e.g. F17-G) of *E. coli* [101], intimin of enteropathogenic *E. coli* [102] and invasin of *Yersinia pseudotuberculosis* [103]. A Big domain plays a direct binding role in the type I pilus adhesin F17-G [101], whereas Big domains in intimin and invasin, arrayed as repeats, act as linkers that project the receptor-binding C-terminal domain away from the bacterial surface therefore providing higher accessibility for interaction.

1.4.4 Characterised adhesins of *Lactobacillus* species

To study the complex process of adhesion to the IECs, many *in vitro* model systems have been developed, because *in vivo* experiments are difficult to carry out. The most routinely used are human intestinal cell lines, Caco-2, HT-29 and mucus producing cell line HT-29-MTX, as well as immobilised matrix proteins and protein mixtures, including mucus and ECM proteins. However, to date, there has been no standardisation of the conditions used in above *in vitro* assays, leading to great variability of results obtained for identical bacterial strains in different laboratories. In spite of these difficulties, some of the *Lactobacillus* key adhesion factors have been characterised in detail.

1.4.4.1 Mucus binding proteins

Mucus-adhering bacteria usually have short residence in the mucus layer due to the constant change of composition and expression of mucin, main glycoprotein of mucus (Figure 1.4). This dynamics of mucous layer assures protection of the host against pathogens. Nevertheless, mucus also provides a habitat for commensal bacteria, including *Lactobacillus* species, and this association has been confirmed by several *in vitro* and *in vivo* studies. In most cases, adhesion of lactobacilli to the mucus layer is mediated by proteins located on bacterial cell surface, although examples of mucus binding mediated by secreted proteins [104], LTA and carbohydrates have been reported [105]. Mucus-binding proteins (MucBPs) have been

identified and fully characterized from four different *Lactobacillus* species: Mub from *L. reuteri* 1063 [106], the mannose-specific adhesin, Msa, from *L. plantarum* WCFS1 [107], Mub from *L. acidophilus* NCFM [108] and LspA of *L. salivarius* UCC118 [109, 110]. These proteins are covalently attached to the cell wall by a cell wall-anchoring transpeptidase called sortase [52]. MucBPs share conserved mucus-binding domains, which consist of repeats. Although these repeats are typically 50 amino acid residues in length, *in silico* search of several publicly available databases identified MucBP-like domains (or MUB) ranging in size up to 200 residues [111]. It appears that these longer MUB domains are unique for LAB. According to the same study, MUB-containing proteins are more abundant in GIT-associated lactobacilli than in cheese or environmental isolates, which supports the hypothesis that this domain is involved in the adherence to intestinal mucus.

Table 1.2. Functionally characterised adherence factors in lactobacilli.

Adapted from [112].

Lactobacillus strain	Adhesin name	Target for binding	Number and type of adhesive domains	Reference
L. acidophilus NCFM	Mub	Human epithelial cell lines and mucus	17 MucBP (PF06458)	[108]
	SlpA	Human epithelial cell lines	2 MucBP (PF06458)	[108]
	FbpA	Human epithelial cell lines and fibronectin	1 fibronectin-binding domain (PF05833) 1 Domain of unknown function (PF05670)	[108]
L. brevis ATCC 8287	SlpA	Human epithelial cell lines, fibronectin, collagen and laminin	No hits in Pfam	
L. crispatus JCM 5810	CbsA	Collagen I, IV and laminin containing regions in human colon and ileum and bacterial LTA	1 SLAP Bacterial surface layer domain (PF03217)	
L. helveticus R0052	Slp	Human epithelial cell lines	1 SLAP Bacterial surface layer domain (PF03217)	[18]
L. johnsonii NCC 533	EF-Tu	Human epithelial cell lines and mucus	1 EF-Tu, GTP binding domain (PF00009) 1 EF-Tu domain 2 (PF03144) 1 EF-Tu C-terminal (PF03143)	[113]
	GroEL	Human epithelial cell lines and mucus	1 TCP-1/cpn60 chaperonin family (PF00118)	[114]
L. plantarum WCSF1	Msa	Mucus via Mannose binding	1 legume lectin domain (PF00139) 4 MucBP (PF06458)	[38] [107]
L. reuteri 1063	Mub	Mucus components	14 MucBP (PF06458)	[106]
L. reuteri NCIB 11951	CnBP	Collagen	1 bacterial extracellular solute-binding domain (PF00497)	[67, 115]
L. reuteri 104R	MapA	Caco-2 cells and mucus	1 bacterial extracellular solute-binding domain (PF00497)	[116]
L. salivarius UCC118	LspA	Human epithelial cell lines	8 MucBP (PF06458)	[109, 110]

Not all mucus-binding proteins contain MucBP-domain. For example, Pfam domain analysis [117] indicated that MapA, a mucus adhesion-promoting protein of *L. reuteri* 104R, did not contain MucBP domains. Instead it was predicted that MapA belongs to the family of ABC transporters because two domains characteristic for extracellular substrate-binding domains of ABC transporters were detected [116, 118].

1.4.4.2 Proteins mediating adhesion to the intestinal epithelium and extracellular matrix

The adherence capacity of some lactobacilli to intestinal epithelial cells and surrounding extracellular matrix has been associated mostly with surface proteins or proteinaceous arrays such as S-layer proteins. Binding of lactobacilli to ECM components, which may be exposed if the epithelial layer is injured, could prevent the adhesion to and colonisation of damaged tissue by invading pathogens [119]. A possible ECM adhesin is the collagen-binding protein (CnBP) of *L. reuteri* NCIB 11951 (Table 1.2) [120]. The translated amino acid sequence of CnBP shows two motifs typical of extracellular solute-binding domain of bacterial ABC transporters [115]. Another example of an extracellular component of an ABC transporter binding not only to the mucus but also to intestinal epithelial cells is the above mentioned MapA protein from *L. reuteri* 104R [116].

In spite of the identification of many S-layer protein-encoding genes, the biological functions of most of these proteins are still to be defined. Adhesion to IECs via S-layer proteins has been described for several *Lactobacillus* species. Four *Lactobacillus* S-layer proteins have been fully characterized: CbsA of *L. crispatus* JCM 5810 [67, 76], SlpA of *L. acidophilus* NCFM [108], Slp of *L. helveticus* R0052 [18] and SlpA of *L. brevis* ATCC 8287 [48]. These proteins mediate adhesion to intestinal epithelial cells. In addition to IECs-binding capacity, SlpA of *L. brevis* ATCC 8287 [48] also binds fibronectin and CbsA of *L. crispatus* JCM 5810 [67] binds laminin and collagen.

Another type of *Lactobacillus* ECM adhesins are fibronectin binding proteins (Fbp). Surprisingly, only a few lactobacilli Fbp(s) have been fully characterised, in spite of the strong evidence that a plethora of Gram-positive pathogens use fibronectin-binding as a major colonisation route. In addition to S-layer (SlpA) protein of *L. brevis* ATCC 8287, *L. acidophilus* NCFM cell surface-associated protein FbpA was found to bind fibronectin [108].

FbpA of NCFM is the only protein found in lactobacilli that contains fibronectin-binding domain homolog to those in fibronectin-binding proteins in *Streptococcus mutans* and *Streptococcus gordonii*. However, this domain was not detected in *L. brevis* Fn-binding protein SlpA. By inactivating the FbpA, reduction in adherence to Caco-2 cells by 76% was observed, indicating that FbpA is an important factor contributing to the ability of *L. acidophilus* NCFM to adhere to epithelial cells. *L. rhamnosus* HN001 binds to fibronectin, either soluble or immobilised, and to Caco-2 cells; however the mediator of the adhesion has not been identified as yet [121].

1.4.4.3 Unconventional *Lactobacillus* surface proteins and non protein molecules mediating adhesion

Several atypical *Lactobacillus* adhesins have been reported. The elongation factor Tu (EF-Tu), a cytosolic guanosine nucleotide-binding protein involved in protein biosynthesis, was found located at the cell surface of *L. johnsonii* NCC 533 [113]. None of presently known domains or motifs needed for transport and attachment to the cell wall were detected in EF-Tu. Interestingly, cell-surface localisation of EF-Tu was observed in other lactobacilli and *Mycoplasma pneumoniae* [113, 122]. The *L. johnsonii* NCC 533 surface-associated EF-Tu was shown to mediate adhesion to human intestinal epithelium as well as mucus. The second unexpected *L. johnsonii* NCC 553 protein found to bind to human IECs and mucus is the heat shock protein GroEL [114]. This protein is normally located in the cytoplasm and is a mediator of protein folding. Similar to EF-Tu, no membrane-targeting or anchoring motifs were detected.

Recently, a cell surface-associated GAPDH of *L. plantarum* LA 318 was found to mediate adherence of this strain to human colonic mucin [123]. The authors suggest that GAPDH may act as lectin-like protein, recognising the sugar chains on the mucus. This is the only example of *Lactobacillus* GAPDH being involved in mucus adhesion although the surface-associated GAPDH (and enolase) from *L. crispatus* ST1 was previously shown to bind plasminogen [124]. A few reports also showed that some pathogens possess a surface-localised GAPDH which can bind to various host proteins such as plasmin(ogen), fibronectin, laminin and actin [125, 126]. As in case of EF-Tu and GroEL, GAPDH is localised on the cell surface in spite of the lack of a conventional N-terminal signal sequence or a membrane-anchoring motif.

Two non-protein cell surface polymers, LTA and EPS, have been demonstrated for some *Lactobacillus* strains to act as adhesins. LTA has been reported to affect *Lactobacillus* adhesion to GIT and more specifically to mediate the adhesion of *L. johnsonii* NCC 533 to Caco-2 cells [127]. EPS, which can be loosely attached to the cells or secreted, plays a significant role in biofilm development and in immunomodulation of some Lactobacilli. However, the adhesion capacity of EPS is still to be defined. A study of *L. rhamnosus* GG EPS-mediated adhesion to human intestinal mucus suggests that EPS has no significant effect on adhesion to this fraction of intestinal epithelium [128]. Recently, *L. rhamnosus* HN001 EPS adhesion and immunomodulation properties have been assessed (M. Collett, Fonterra – New Zealand; unpublished observations), and this data suggests that HN001 EPS has a significant inhibitory effect on adhesion to fibronectin and the mucus-producing HT29-MTX cell line, as well as on auto-aggregation, and has no direct role in the immuno-modulation.

*Chapter I**B***

Secretome proteins

1.1 Definition of the secretome

The term secretome will be used in this thesis for the collection of secreted, surface and integral membrane proteins excluding secretion pathways components. The secretome comprises a wide range of proteins that mediate interactions with the environment, such as receptors, adhesins, transporters, proteins that form complex cell surface structures such as pili and secreted proteins such as enzymes, toxins and virulence factors. In bacteria that colonise the human organism, secreted proteins mediate attachment to the host, destruction of the host tissue or interference with the immune response [129-131]. In pathogenic bacteria, variation of a surface protein between strains of a species can indicate its role in evading the immune response [132-135]; conversely, conserved surface proteins that are capable of inducing a protective immune response are sought as vaccine candidates [136].

1.2 Secretion pathways

Secretome proteins have been reported to constitute 10-30% of the total number of encoded proteins in bacteria [137, 138]. To be released from the cell membrane or get attached to it, secretome proteins must pass through one or more membranes. A remarkable array of mechanisms that catalyse protein secretion have been invented in the course of evolution. Fifteen such systems, which handle protein secretion, sorting and membrane integration, are present in the *Bacteria* alone [139]. The Sec (secretion) pathway for translocation of proteins across the cytoplasmic membrane is ubiquitous and essential in every organism. In *E. coli* it consists of two protein-targeting pathways, S(ignal)R(ecognition)P(article)/Fth and SecB/A, and a multisubunit protein-translocase complex, SecYEG. Three other systems can transport proteins across the cytoplasmic membrane independently of Sec translocon. One consists of YidC, for transport of transmembrane proteins with small extracellular domains, present in bacteria, some chloroplasts and mitochondria of eukaryotes [140, 141]. This protein can work by itself or in complex with the SecYEG translocase. The other system, Tat (twin arginine transport), is exclusively independent of Sec machinery [142]. The Tat system is the only translocation pathway characterised thus far able to transport prefolded and often multimeric proteins across the cytoplasmic membrane. Finally, some proteins can self-integrate into the cytoplasmic membrane of bacteria and export a specific substrate protein. Such proteins are

holins, small membrane proteins of phage origin, mainly involved in energy-independent secretion of autolysins in processes relevant to apoptosis [143, 144].

The other secretory pathways currently recognised in bacteria are classified as Sec(Tat)-dependent or Sec(Tat)-independent based on their dependence or otherwise from Sec or Tat machinery for targeting of secretome proteins to the cytoplasmic membrane.

1.2.1 Sec-dependent and Sec-independent secretion systems of Gram-negative and Gram-positive bacteria

While the transport across the cytoplasmic membrane is sufficient for Gram-positive secretome proteins to be released into the extracellular milieu, in Gram-negative bacteria secretome proteins need to cross an additional membrane, the outer membrane (OM). Several dedicated secretion pathways (T1SS-T6SS) fulfil this role in Gram-negative bacteria. In addition, systems exist in Gram-negative bacteria for assembly of complex surface structures such as pili, flagella or fimbriae, as well as systems for assembly/secretion of filamentous phages. For some of these secretion systems, equivalents have been identified in Gram-positive bacteria.

Secretion systems T2SS, T5SS and the Chaperon/Usher (CU) pathway depend on the Sec machinery for transport of proteins across the cytoplasmic membrane [137]. Once released in the periplasm by Sec (or in few cases Tat) machinery many enzymes and toxins of Gram-negative bacteria use the T2SS pathway for secretion across the OM [145]. Translocation through the OM is mediated by multi-subunit “channels” formed by proteins termed secretins (e.g. outer membrane secretin PulD of *Klebsiella* spp. [146]). Conversely, substrates of T5SS, autotransporters, do not need any special channel or pore for the transport across OM. These modular, extracellular or OM-associated, proteins secrete their N-terminal auto-catalytic protease/effector domain through the OM via their channel-forming C-module. Many Gram-negative enzymes (e.g. IgA-protease of *Neisseria* spp.[147]) and adhesins (e.g. invasin YadC of *Yersinia* spp. [148]) are autotransporters (T5SS). Another Sec-dependent secretion pathway, Chaperone-Usher pathway (CU), is linked to the assembly of adhesive surface structures such as type I pili and P pili [145]. No equivalents of T2SS, T5SS and CU pathways are identified in Gram-positive bacteria.

Trans-envelope secretion systems encompassing cytoplasmic and outer membrane of Gram-negative bacteria, independent from Sec pathway, include T1SS, T3SS, T4SS T6SS and holins. T1SS or ABC transporter-dependent systems are found in archaea, eubacteria and eukaryotes. In general, ABC transporter-dependent systems are heterotrimeric complexes consisting of ATP-binding subunits, integral membrane subunits (permeases) and periplasmic (or lipoproteins in Gram-positive bacteria) substrate-binding subunits. In Gram-negative bacteria, T1SS in addition contains an outer membrane channel that interacts with the ABC transporter. A wide range of substrates (proteinaceous and non-proteinaceous), including various virulence factors, are secreted in a single step through ABC transporters to the extracellular milieu.

Type III secretion systems (T3SS) are exclusively found in Gram-negative bacteria including some important pathogens (*Salmonella*, *Yersinia*, *Shigella*, *Escherichia*, *Pseudomonas*). Evolutionarily it is related to the flagellum assembly system. The function of T3SS secretion is transport of virulence factors from bacterial cytoplasm directly into the cytoplasm of the host cell. This requires formation of a complex molecular structure, the "molecular needle" that spans three biological membranes. Similar to T3SS, type IV secretion system (T4SS) forms multi-subunit protein complex to transport proteins and nucleic acids across bacterial and host cell membranes into the host cell cytoplasm (e.g. CagA virulence factor from *Helicobacter pylori* [149] and *A. tumefaciens* T plasmid [150]). Some substrates secreted to the extracellular environment by T4SS depend on the Sec system (e.g. pertussis toxin of *Bordetella pertussis* [151]). T4SS is identified in both Gram-negative and Gram-positive bacteria [152]. Recently, investigation of the virulence-associated secretion cluster from *Vibrio cholerae* led to discovery of a novel class of Gram-negative secretion systems, termed type VI (T6SS) [153]. How exported proteins are targeted and secreted across the cell envelope in this system is not yet clear, but it is likely to be Sec-independent, considering that T6SS substrates do not contain corresponding membrane-targeting sequences.

Other specialised secretion systems present in both Gram-negative and Gram-positive bacteria are type IV pili assembly system (T4PS) and flagellum assembly system (termed FEA in Gram-positive bacteria) [144, 145]. T4PS is one of the most common pilus biogenesis pathways and is structurally related to T2SS, whereas flagellum assembly system is related to T3SS. Type IV pili are formed at the cytoplasmic membrane by polymerisation of pilin

subunits. The hallmark of T4PS biology is that they are able to retract through the bacterial cell wall while the pilus tip remains attached to a receptor. This enables bacteria to move across the surface in the specialised way termed "twitching motility". The system for assembly/secretion of filamentous phage is related distantly to T2SS and will be described in Chapter IC, section 1.2.

Several secretion systems are unique for Gram-positive bacteria such as Sec-dependent sortase-mediated assembly system and the recently recognised Sec-independent Esx secretion system [154, 155]. Sortases are transpeptidases responsible for covalent attachment to the cell wall of surface proteins containing an LPXTG motif which serves as an anchor to the cell wall. The mechanism of sortase-dependent cell wall attachment of surface proteins will be discussed further in this chapter, section 1.4.3. Esx (or WXG100) system has been described as a secretion system for exporting small, around 100 residues long proteins, containing a WXG motif [156]. No apparent membrane-targeting signal was detected in Esx-dependent proteins. Esx has been studied extensively in *Mycobacterium tuberculosis*; however, the mechanism of Esx-depandent translocation of proteins across the cytoplasmic membrane is still not understood [157].

1.3 Transport across the cytoplasmic membrane

Secretome protein export is a multi-stage process which can be divided into three distinct stages: sorting and targeting of proteins to the cytoplasmic membrane, transmembrane crossing and maturation/ release of translocated protein.

1.3.1 Secretome protein sorting

In the first stage of cytoplasmic membrane transport the polypeptides (pre-proteins) destined for export must be discriminated from the cytoplasmic-resident proteins. This sorting of secretome proteins is in general based on presence of membrane-targeting sequences, e.g. signal sequences and transmembrane α-helices, which act as a "zip code" and are recognised by particular secretory pathway-associated molecular chaperone [158-161].

A transmembrane α-helix (TMH) of any membrane protein may be used as a membrane-targeting signal. TMH is a hydrophobic segment usually composed of 15-30 amino acids

which are recognised by SRP. In addition to their membrane-targeting role, TMH of bitopic membrane proteins serves as a membrane anchor [140, 162].

There are several types of signal sequences: the "classic" or type I signal sequence, the Tat signal sequence, the lipoprotein or type II signal sequence, pseudopilin-like or type IV signal sequence and the bacteriocin/pheromone signal sequences (Figure 1.5). They vary considerably in sequence but show three distinct zones based on hydrophobicity and charge: an amino-terminal (N-) region with positively charged amino acids, a central hydrophobic (H-) region of at least 15 residues and a polar carboxy-terminal (C-) region containing the signal peptidase (SPase) cleavage-site [161, 163]. High positive charge of the N-region determines the topology of the pre-protein in the membrane, obeying "positive-inside rule" by which positively charged regions remain located at the *cis* (cytoplasmic) side of the cytoplasmic membrane. The exceptions from the classical signal sequence three-partite structure are signal sequences of pheromones and antimicrobial peptides. These cleavable signal sequences consist of only N-and C-domains and completely lack a hydrophobic H-domain. Dedicated ABC transporters (T1SS) are responsible for the removal of this type of signal sequence and translocation of the mature proteins across the cytoplasmic membrane [164]. Upon signal sequence cleavage, the mature protein is released from the cytoplasmic membrane, provided that it does not contain membrane anchors or additional TMHs.

Signal sequences of Gram-positive bacteria differ from those in Gram-negative bacteria by being usually longer, more hydrophobic and by having an increased number of charged residues on their N-terminal ends. Although signal sequences from different organisms are often interchangeable, there are also examples of signal sequences that do not function efficiently when expressed in a foreign host [155, 165, 166].

Non-classical secretion

Some bacterial proteins are secreted without any apparent signal sequence. The first examples of this signal sequence-independent or so-called non-classical secretion were eukaryotic interleukin 1β and thioredoxin. In bacteria, proteins secreted via the recently discovered Gram-negative T6SS and Gram-positive Esx as well as holin-mediated secretion systems are examples of non-classical secretion. T3SS and T4SS exported proteins also lack classical signal sequences. Furthermore, in several bacterial species (*Mycobacterium*, *Listeria*,

Streptococcus) a specific subset of proteins contributing to bacterial virulence, which lack signal sequence are exported by Sec-pathway via an alternative accessory protein, SecA2. For example, fibronectin-binding protein A (FbpA) of *L. monocytogenes* and SodA (superoxide dismutase A) from *M. tuberculosis* do not contain signal sequence and are secreted by SecA2-pathway [167]. Some cytoplasmic proteins with well-defined intracellular functions are also found in the extracellular environment where they play entirely different roles from those in the cytoplasm and are therefore named "moonlighting" proteins. Two types of the moonlighting proteins, GroEL [114], a protein folding chaperone and EF-Tu [113, 122], a translation elongation factor, have been identified in several bacterial species (e.g. *M. pneumoniae*, *H. pylori*, *L. johnsonii*).

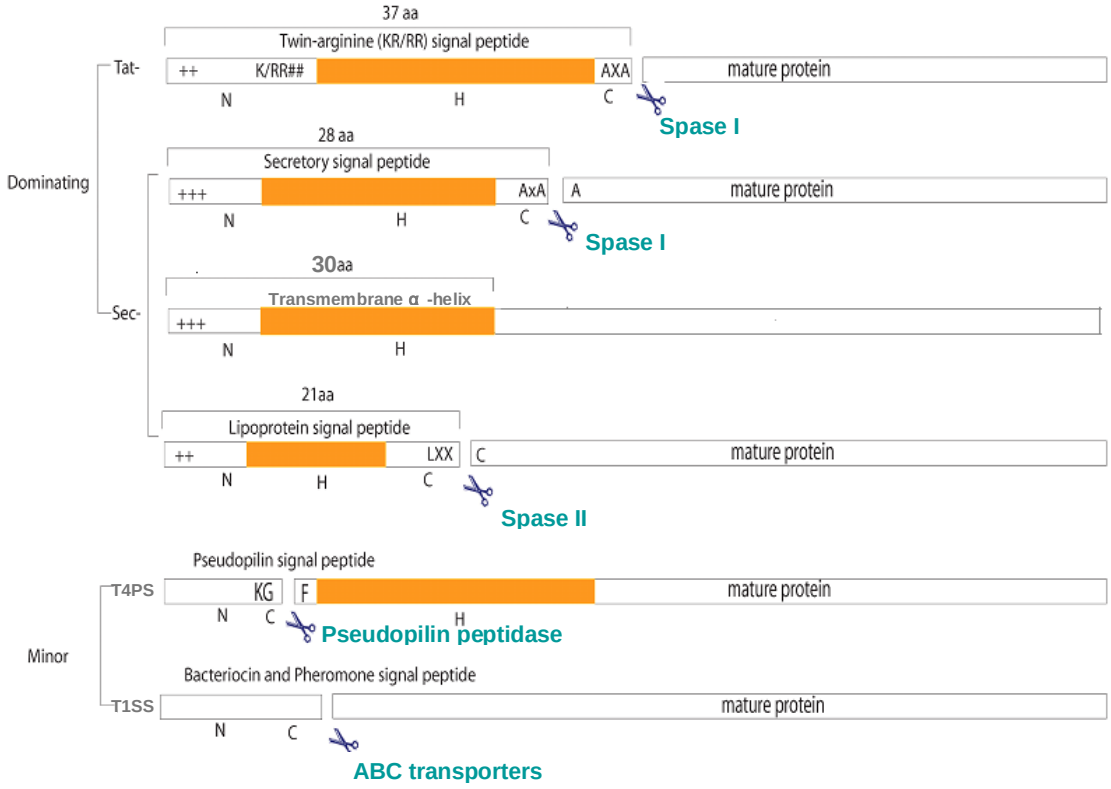


Figure 1.5. The structure of known cytoplasmic membrane-targeting signals.
On the basis of signal peptidase (SPase) cleavage sites (or absence thereof) and the export pathways by which the pre-proteins are exported, signal sequences are divided into distinct classes: Tat (RR/KR) signal sequence, secretory (Sec-type) signal sequences, lipoprotein signal sequences, pseudopilin-like signal sequences and signal sequences of bacteriocins and pheromones. Transmembrane α-helices are recognised by the Sec machinery but lack SPase cleavage sites. The export pathways and the SPases responsible for pre-protein cleavage are indicated. Most signal sequences have a tripartite structure: a positively charged N-domain (N), containing lysine and/or arginine residues (indicated by +), a hydrophobic H-domain (H, indicated by an orange box), and a C-domain (C) that specifies the cleavage site for their specific SPases. Where appropriate, the most frequently occurring first amino acid of the mature protein is indicated. The signal sequences of proteins targeted for minor secretion pathways, such as T1SS for pheromones/bacteriocin translocation in Gram-positive bacteria and T4PS for type IV pili assembly do not follow the N-H-C structure. Adapted from [168] with permission.

1.3.2 Membrane-targeting, translocation and release of secretome proteins

In the Sec-pathway, the nascent pre-protein coming from the ribosome is intercepted by piloting factors, SRP or SecB. Transmembrane proteins and secreted proteins with long, highly hydrophobic signal sequences are targeted to the membrane via SRP as soon as they emerge from the ribosome (co-translational translocation). Other Sec-dependent secretome proteins are mainly chaperoned to the cytoplasmic membrane by SecB in co-translational and post-translational fashion, meaning that translation of the pre-protein is almost completed by the time SecB binds. After binding of the pre-protein to SRP or SecB, the resulting protein complexes are targeted to the translocase at the membrane. The SRP-pre-protein complex docks to its membrane receptor FtsY while the SecB-pre-protein complex binds to the SecA subunit of the translocase. SecB homologues are not found in Gram-positive bacteria, although protein CsaA from *Bacillus subtilis* probably has an analogous function. In *Listeria* spp., *Mycobacterium* spp. and *Streptococcus* spp. a paralogue of SecA, SecA2, has been identified. SecA2 is required for membrane-targeting of some specific proteins, with or without signal sequences, usually involved in bacterial virulence.

In Sec-dependent export, the majority of pre-proteins, irrespective of the targeting route, cross the membrane via a membrane-embedded SecYEG protein-conducting channel that is built of three subunits, SecY, SecE and SecG. This translocase involves also several accessory transmembrane proteins that stimulate translocation (SecDFYajC) and membrane protein insertion (YidC) [140]. Translocation of the SecB-chaperoned proteins is driven by the SecA ATPase, at the expense of metabolic energy in the form of ATP and the proton-motive force (PMF). In 2000, a new translocator protein, YidC, was identified and shown to act not only as an auxiliary translocation protein, but also as a SecYEG-independent translocase, enabling translocation of several transmembrane proteins, usually with small extracellular domains [141].

At the last stage of Sec-dependent export, pre-proteins are converted into mature proteins. After signal peptidase cleaves the signal sequence, correct folding of the polypeptide chain is initiated at the *trans* side of the membrane. In Gram-positive bacteria mature proteins can be (i) secreted into the extracellular medium, (ii) integrated into the plasma membrane, (iii) anchored to the membrane or cell-wall. In Gram-negative bacteria Sec-dependent secretion

systems, proteins can be (i) released into the periplasm, (ii) integrated into the cytoplasmic membrane; (iii) transported across the outer membrane by specific secretion systems (T2SS, T5SS and CU).

The Tat pathway transports a small group of cofactor-containing and therefore folded proteins across the cytoplasmic membrane [160]. The translocation machinery is composed of only four proteins: TatA, TatB, TatC and TatE. TatBC proteins act as a receptor complex that recognises twin arginine motif (RR)-containing signal sequence of the substrate protein and transport it post-translationally to the membrane. At the membrane, the TatBC-substrate complex associates with TatA which probably forms the transport channel. This association is proposed to persist until completion of the substrate protein transport across the membrane. Subsequently the Tat signal sequence is cleaved by signal peptidase and the protein is released into the periplasmic space or extracellular milieu.

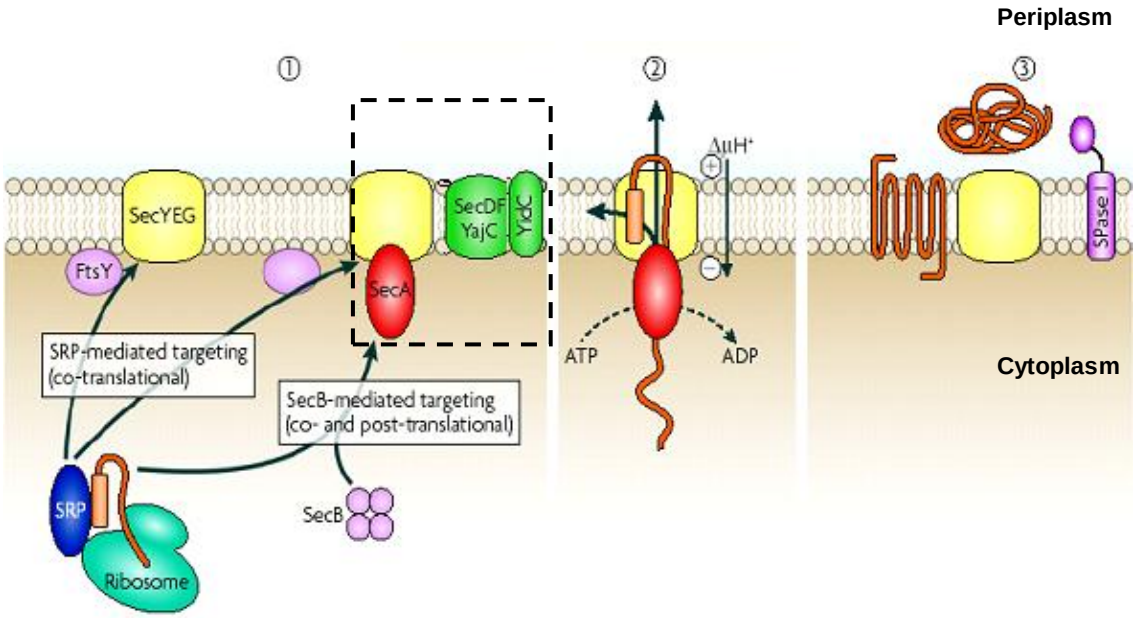


Figure 1.6. Schematic representation of Sec/SRP mediated pre-protein export through the translocase channel.

The translocase consists of the SecYEG pre-protein-conducting channel (yellow) and the ATPase motor SecA (red). SecYEG can associate with the auxiliary translocator proteins SecDFYajC and YidC (green) [black dotted rectangle]. **1)** Secretome pre-proteins (thick orange line) are synthesized with N-terminal signal sequences (orange rectangle) and are targeted to the translocase either by the SRP as soon as they emerge from the ribosome exit tunnel (this is co-translational translocation), or by the tetrameric SecB chaperone after translation has largely been completed (this is co- and post-translational translocation). For simplicity, auxiliary proteins are not presented beside all diagrams of translocase. Both targeting routes merge at the membrane at SecYEG (SRP route) or SecYEG complexed with SecA (Sec route). FtsY (pink) and SecA act as receptors for SRP and SecB, respectively. **2)** Pre-proteins are translocated through SecYEG and **3)** they fold at the *trans* side of the membrane or integrate/anchor into the lipid bilayer after signal-peptide cleavage by the signal peptidase (SPase I). Energy is provided by ATP binding and hydrolysis by the SecA ATPase and the proton-motive force ($\Delta\mu H^+$). Adapted from [167] with permission.

1.4 Anchoring of proteins to the plasma membrane and cell wall

In the Gram-positive model organism *Bacillus subtilis*, approximately 21% of total cellular protein content is attached to the cell envelope, either to the cytoplasmic membrane or the cell wall [164, 169]. The envelope-attached proteins include DNases, RNases, proteases, cell wall hydrolases, enzymes involved in synthesis of peptidoglycan, substrate-binding proteins and adhesins. Four major types of envelope proteins are currently recognised based on the mode of anchoring: (i) integral membrane proteins anchored in the membrane via hydrophobic transmembrane helices, (ii) lipoproteins, covalently attached to the membrane lipids, (iii) cell wall proteins containing a C-terminal LPXTG-like motif, covalently linked to the cell wall and (iv) proteins with various cell wall-binding domains (Figure 1.7). Macromolecular structures such as S-layer, pili, flagellum and cellulosome are anchored to the cell surface by one of these mechanisms [170].

1.4.1 Transmembrane anchors

Transmembrane α-helices serve as membrane-targeting signals but due to the absence of an SPase cleavage site they also act as membrane anchors. Bitopic membrane proteins contain one hydrophobic TM α-helix traversing the phospholipid bilayer that is flanked by two polar domains. Polytopic membrane proteins contain two or more TM α-helices that are connected by loops exposed to aqueous environment. These proteins can also contain a cleavable N-terminal signal sequence or N-terminal amphiphilic region (as recently described for Kdp sensor protein of *E. coli*) [171], in which case additional TMs at their C-terminus can function as membrane anchors. Two topologies with respect to the membrane are possible for bitopic membrane proteins: the N-terminus is exposed on the outside face of the membrane (type I) or the cytoplasm (type II). The latter one is regarded as an uncleaved N-terminal anchor. The topology of polytopic membrane proteins is determined by the "positive inside" rule, in which positively charged residues flanking the TMH must be localised on the *cis* (cytoplasmic) side of the membrane [164].

Most integral membrane proteins are targeted to the membrane by the SRP pathway where the folding and insertion are mediated by either by the main SecYEG translocase and its auxiliary proteins or by the membrane insertase/translocase YidC [140, 141].

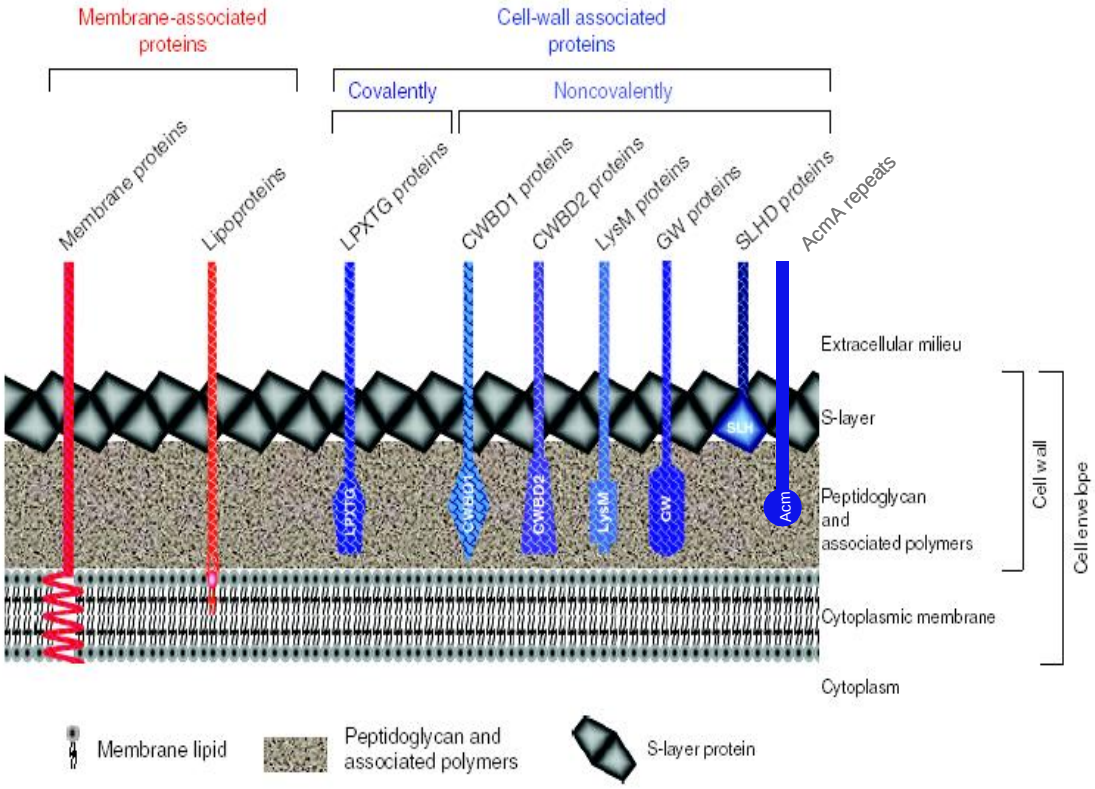


Figure 1.7. Schematic representation of the different types of surface proteins found in Gram-positive bacteria.

Membrane-associated proteins can either (i) possess N- or C- terminal transmembrane α-helix or (ii) be lipoproteins, and thus be covalently attached to long chain fatty acids of the cytoplasmic membrane. Cell wall-associated proteins can either possess (i) C-terminal LPXTG motif (or related motifs) and be covalently anchored to peptidoglycan via sortase activity, or (ii) cell wall binding domains (CWBDs) which act as non-covalent anchors to the cell wall. Six CWBDs are currently characterised: CWBD1, CWBD2, LysM, GW modules, SLHD and AcmA repeats. Adapted from [170] with permission.

1.4.2 Lipoprotein anchors

Lipoproteins are targeted to the membrane by specific signal sequence type II which exhibits a conserved lipobox motif. The lipobox contains a cysteine residue at the SPase II cleavage site [172]. Before processing by SPase II, prolipoprotein diacylglyceryl transferase attaches a diacyl-glycerol group to the cysteine residue of the lipoprotein precursor. The diacyl-glyceryl group inserts into the lipid bilayer, preventing release of mature lipoprotein into the

environment after cleavage by SPase II. Besides lipoprotein signal sequence, some lipoproteins can also contain transmembrane domains. In this case, the lipoprotein signal sequence is required primarily for protein function rather than for anchoring in the membrane. Lipoproteins are involved in many physiological functions, including adhesion, transport, enzymatic reactions and virulence [173].

1.4.3 LPXTG anchors

A range of important surface proteins including adhesins, antigens, receptors and enzymes are anchored to the cell wall in Gram-positive bacteria by sortases. In addition to type I signal sequence, these proteins contain at the C-terminus a cell wall sorting domain which is comprised of an LPXTG motif followed by a sequence of hydrophobic amino acids (15-19 residues) and then a positively charged tail [174]. A specific transpeptidase called sortase is responsible for cleavage of the cell wall sorting signal and covalent attachment to the cell wall. There are several types of sortases: the house-keeping sortase (SrtA) and pilus-specific sortases. SrtA is anchored in the membrane, typically in the vicinity of the SecYEG translocase. It processes the majority of LPXTG-containing proteins. After cleavage of the signal sequence by SPase, the translocated protein is retained in the membrane by the C-terminal hydrophobic α-helix and charged tail. SrtA recognises the LPXTG pentapeptide motif followed by an acidic residue cleaves it between T and G residues and attaches threonine by amide linkage to the peptide crossbridge in the peptidoglycan of the cell wall.

In addition to SrtA-type sortase, three additional classes of sortases are currently recognised, SrtB, SrtC and SrtD [175]. They anchor far fewer proteins than does the SrtA and these proteins carry different sorting motifs: SrtB and SrtD-dependent proteins contain NPQTN and LPXTA/LAXTG motifs, respectively, whereas SrtC-dependent proteins contain the LPXTG motif, followed by a glycine instead of an acidic residue.

Biogenesis of Gram-positive pili also involves sortases [52, 176]. The pilin subunits are proposed to be sequentially cross-linked by transpeptidase sortase which belongs to the class C but it is not excluded that other sortases could participate in this process.

1.4.4 Noncovalent binding of surface proteins to the cell wall

A plethora of noncovalent cell wall-binding domains have been reported in Gram-positive bacteria. Enzymes are the largest group of proteins anchored via these domains, although some well-characterised virulence factors and adhesins also bind to the cell wall in this way. Furthermore, complex proteinaceous surface structures such as S-layer and cellulosome are also anchored noncovalently to the cell wall.

Choline-binding domain or type I cell wall binding domain (CWBD1) recognises phosphorylcholine on the cell wall-embedded teichoic and lipoteichoic acids (LTA). CWDB1 was found in many enzymes (sucrases, lactamases, and proteases) and bacteriocins. The major autolysin (LytA) and surface adhesin (CbpA) of *Streptococcus pneumonie* are two examples of such CWDB1-anchored proteins [177, 178]. Several autolysins, adhesins and S-layer proteins contain CWDB2 in multiple copies [170, 179]. Cell wall components involved in interaction with CWDB2 are still to be defined.

The LysM domain is one of the most common modules in bacterial surface proteins and is also present in a number of eukaryotic proteins [180]. It is ~ 40 amino acids in length and binds to the peptidoglycan. The LysM module is most often found in cell wall degrading enzymes, but can also be found in proteins involved in bacterial pathogenesis, including intimin from enteropathogenic *E. coli* and Ig-binding proteins from staphylococci [102, 181].

Other examples of noncovalently attached cell wall-binding domains are GW, SLHD and AcmA-repeats [46, 182, 183]. GW proteins are mostly autolysins which bind the cell wall via 3-8 repeats of the Gly-Trp (GW) dipeptide. SLDHs (S-layer homology domains) are commonly found in S-layer proteins, enzymes involved in polysaccharide degradation and in scaffoldins of the macromolecular structure cellulosome. They can interact with cell wall acidic and neutral polysaccharide complexes and/or peptidoglycan itself. The AcmA-repeats anchor is another noncovalent cell wall-binding domain discovered as an anchoring mechanism of lactococcal cell wall hydrolase AcmA. Repeats homologous to those found in AcmA are present in many cell wall or membrane-associated proteins of Gram-positive and Gram-negative bacteria [184].

1.5 Secretomes of *Lactobacillus* species

1.5.1 Membrane-targeting and anchoring to the cell envelope

In *Lactobacillus* species prediction and characterisation of secretome proteins provides insight into possible mechanisms of interactions of these species with their variable environments. To date, genome sequences of 14 *Lactobacillus* species have been published and for some of those comparative and functional genomic analyses have been performed. Four reports describing specifically lactobacilli secretomes either using bioinformatic approach and/or functional analysis have been published thus far [110, 111, 185, 186].

In silico secretome analyses of the completely sequenced *Lactobacillus* genomes revealed that 20-25% of total number of proteins are targeted to the cytoplasmic membrane. Of those ~ 7% were secreted/cell envelope bound and ~ 15% were transmembrane proteins. This prediction is in agreement with number of secretome proteins reported for other Gram-positive bacteria such as *Bacillus subtilis* (21%) [164]. Approximately 30-40% of secretome proteins contain a cleavable signal sequence recognised by either signal peptidases I (SPase I) or SPase II, with or without additional transmembrane α-helices. The remaining 60-70% are bitopic membrane proteins that contain a single hydrophobic N- or C-terminal transmembrane α -helix lacking an SPase cleavage site. This suggests that the majority of secretome proteins remain associated with the cell membrane [110].

All of the Gram-positive cell wall and membrane anchors described above except choline-binding domain have been found in *Lactobacillus* secretome proteins. A proportionally high number (~ 10%) of secretome proteins contains an LPXTG motif in probiotic *Lactobacillus* species. Other Gram-positive cell wall binding domains such as LysM (~ 6% of secretome) and GW (~ 4%) were also identified in lactobacilli secretomes. Recently, a novel C-terminal cell wall binding domain WXL [111] was identified in *Lactobacillus* species *L. plantarum*, *L. salivarius* and *L. sakei* [110].

When bioinformatics is used to identify cell envelope-anchoring domains, for some of the anchors the reported numbers vary significantly depending on algorithm used for their prediction. For example, in one report the number of predicted lipoproteins of *L. plantarum*

WCFS1 was 3 (0.7% of total secretome) [110] while in the separate study, 48 (21% of total secretome) lipoproteins were identified [111].

1.5.2 Functional groups of *Lactobacillus* secretome proteins

For around 30 % of secretome genes identified in the genomes of lactobacilli thus far the function can not be predicted. Often, secretome proteins can contain several domains indicating that they can have more than one function [115]. Furthermore, a number of secretome proteins, especially adhesins, contain domain repeats (multiple copies of similar domains) and regions of low complexity [111]. The low complexity repeats presumably act as spacers, separating and positioning the functional domains of the protein away from the cell surface [187]. Secretome proteins can be divided into groups reflecting predicted protein function, based on functional annotation and domain composition. Six major functional categories are currently recognised: enzymes, transport proteins, host/microbial interaction proteins, sensor-kinase proteins and unknown conserved and unconserved proteins.

The largest (30-35%) group of secretome proteins for which function could be assigned are enzymes. This diverse group includes enzymes with known biological role in processes such as cell wall biosynthesis and turnover and enzymes for which only the general type of catalysed reaction can be predicted, e.g. serine proteases, metalloproteinases, hydrolases, transglycosidases. The majority of enzymes are anchored to the cell envelope by an N-terminal transmembrane anchor or LysM domain, although a significant number is also secreted into the extracellular medium. Another large class of secretome proteins comprises proteins involved in transport (10-15%). Bitopic, polytopic proteins and lipoproteins are predominant among transport proteins. In all lactobacilli around 15% of transport proteins are substrate-binding subunits of ABC transporters, anchored in the membrane via lipoprotein anchor.

Adhesion and aggregation factors are generally considered to play an important role in host-microbial interactions. It is predicted that approximately 2-5% of lactobacillar secretome proteins are involved in these interactions and some of them have been discussed previously (Chapter IA, sections 1.4.2 and 1.4.4). Retention of this group of proteins in the cell wall is mediated mainly but not exclusively via LPXTG-like peptidoglycan anchor. In addition to cell wall binding domain these proteins usually have substrate binding domains which are present

in multiple copies. This is very prominent in, for example, mucus-binding proteins which can contain up to 15 MubBP (or MUB) domains. Several other domains were identified in proteins of this group (as described in Pfam [117]) such as collagen-binding, mannose-binding, fibronectin-binding, Ig-like and phospholipid-binding domains. Recently ten new putative domains have been identified in *L. plantarum* WCSF1: five new repeat domains, a new mucus binding domain, two cell wall binding domains (WXL1 and 2), cell surface hydrolase (CSH) domain and Trp-Tyr (WY) domain present in putative transglycosylases [111]. Moreover, Boekhorst *et al.*, [111] identified 16 secretome proteins, some of which have mucin and collagen binding domains, and low complexity repeat regions usually preceding LPXTG anchor. It is suggested that these repeats act as a spacers, positioning functional domains above the cell wall layer and therefore exposing these domains to the extracellular milieu.

In bacteria effective adaptation to changing environmental conditions is commonly mediated by two-component regulatory systems (TCSs) which consist of transmembrane histidine kinase (HK) acting as a sensor and a cytoplasmic response regulator. Various changes in environmental conditions such as osmolarity, nutrient availability and temperature are monitored by sensor HKs. In addition HKs also respond to specific secreted signalling peptides (AIPs) or non-peptide small molecules involved in inter/intraspecies quorum sensing. HKs are polytopic membrane proteins and therefore part of the secretome. In lactobacilli whose habitat is restricted to specific environments such as the intestine, one or two HKs could be identified. In contrast, lactobacilli adapted to multiple environments contain multiple HKs, reflecting the ecological flexibility of this species [188]. An example of latter group *L. plantarum* WCFS1, which has six HKs involved in five quorum sensing TCSs.

Chapter IC

Phage display

1.1 Phage display - overview

The physical link between phenotype and genotype of a protein displayed on the surface of the virion, the high replication capacity of bacteriophage and subsequent affinity selection are the elements that underpin the phage display technology. The origins of phage display date to 1985 when George Smith first expressed a foreign segment of a protein on the surface of bacteriophage M13 virus particles [166]. Since then *E. coli* filamentous phages of the Ff class including strains M13, fd, and f1 (Figure 1.8) have been extensively used in phage display technology. These phages infect *E. coli* by attachment to the F-pilus [189], and progeny phages are subsequently secreted from infected bacteria without cell lysis (Chapter IC, section 1.2) [190, 191].

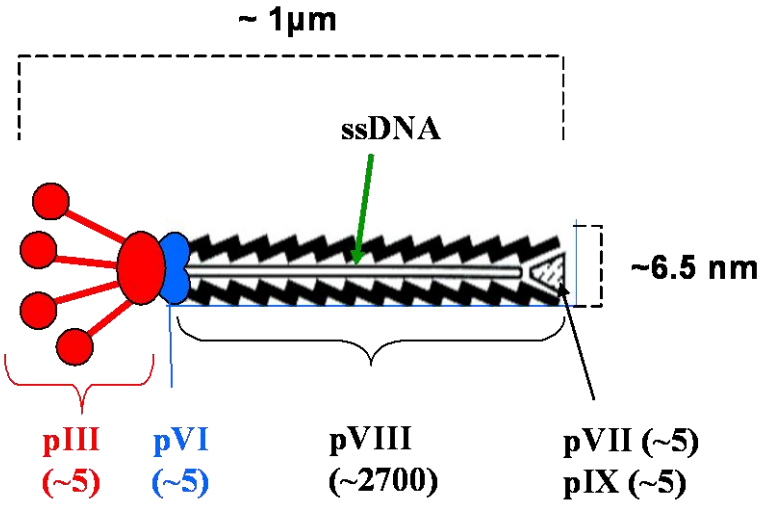


Figure 1.8. Schematic representation of M13, f1 and fd. (not to scale).
pIII, pVI, pVII, pVIII, pIX - virion proteins. The approximate number of copies of each coat protein is indicated.

Display is achieved by translational fusion of a protein or a library of proteins of interest to any of the five virion proteins, although the pIII and pVIII proteins are used most frequently [192, 193]. Filamentous phage virion proteins are themselves secretome proteins, translocated from cytoplasm via the Sec-dependent pathway and anchored in the cytoplasmic membrane prior to assembly into the virion [194, 195]. Therefore, if the secretome proteins are to be displayed, they would be targeted to, and folded in, the cellular compartment in which they normally reside. When a library of fusion proteins is constructed and displayed on the phage surface, a recombinant phage clone displaying a certain binding affinity can be isolated from

the majority of other (non-binding) recombinant phages present in a library, by an affinity selection procedure known as bio-panning (Figure 1.9)[196]. Through successive rounds of binding, washing, elution and amplification, the originally very diverse phage display library (up to 10^{11} variants) is increasingly enriched for the phage library clones with a propensity to bind to the target molecule in question. Ultimately, monoclonal phage populations with desired specificities can be identified and analysed.

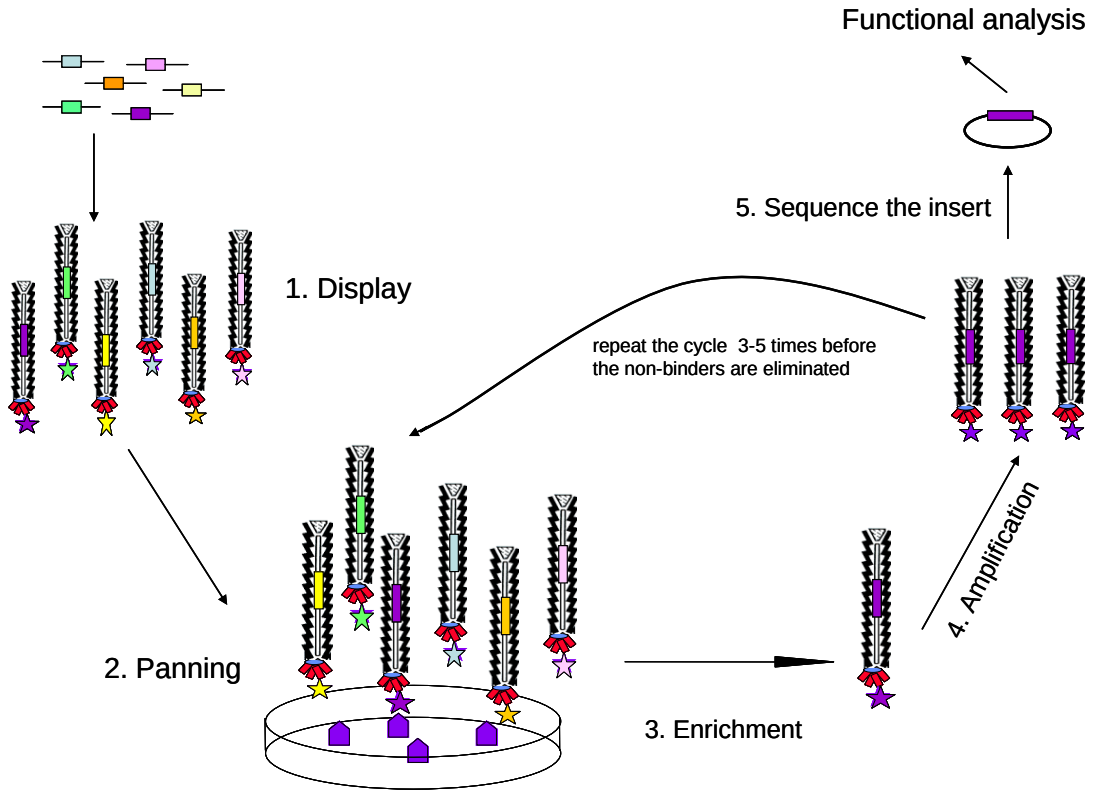


Figure 1.9. Panning of a phage display library against immobilised target.

1) A library of genes encoding variants of proteins/peptides/antibodies is created, cloned into phage or phagemid vectors as fusions to a coat protein pIII gene and displayed on the surface of the virions. 2) The phage library displaying variant peptides or proteins (different coloured stars) is exposed to immobilised ligand (purple pentagon) and phage with appropriate binding specificity is captured (purple star). 3) Non-binding phages are washed off and bound phage(s) is (are) eluted by conditions that disrupt the peptide-ligand interactions, leading to enrichment for a specific binder. 4) Eluted phage is then amplified by infection of a suitable *E. coli* strain. This amplified phage population is greatly enriched in recombinant phage clones displaying peptides that bind to the target. The biopanning steps (2 to 4) are repeated for several (three to five) rounds, ultimately resulting in a clonal population of recombinant phages that bind to the target used for affinity panning of the library. Captured putative binder can then be identified by sequencing (5) and functionally analysed.

Phage protein pIII is the most frequently used display platform. The structure of pIII includes two N terminal domains (N1 and N2) separated by a glycine linker and a C-terminal domain which is separated from the N-terminal domains by another glycine linker. At the N-terminus, pIII contains a type I signal sequence. A signal sequence is necessary for correct targeting of pIII to the inner membrane and incorporation into the virion [197]. Moreover, assembly of pIII into the virion is required to complete the phage assembly. When pIII is absent, virions either stay associated with the host cells as long filaments composed of multiple sequentially packaged genomes, or are broken off by mechanical shearing. pIII is required for formation of the stabilising cap structure at the terminus of the virion; hence, the broken-off pIII-deficient virions are structurally unstable and are easily disassembled by sarcosyl, to which the pIII-containing virions are resistant [198, 199].

1.2 The Ff phage lifecycle

Ff filamentous bacteriophage begins the process of infecting a host *E. coli* by binding to the tip of the F pilus (Figure 1.10) [195]. This is the primary receptor for Ff phage and interacts with N2 terminal domain of pIII. Binding causes a structural rearrangement between the N domains releasing the N1 domain and exposing the binding site that interacts with cytoplasmic membrane complex of TolA, R and A. The N2 binding to the F pilus also causes the F pilus to retract, bringing the phage into closer contact with the cell, allowing the interaction between the N1 domain and TolA. After interaction of N1 with TolA, virion coat becomes integrated into the host cell membrane and virion ssDNA enters the cytoplasm. Once the Ff phage ssDNA genome enters the cytoplasm of the host cell a negative strand is synthesized by the host DNA replication machinery. The dsDNA form of Ff phage genome is referred to as the replicative form (RF). The RF of the genome is a template for replication of the phage positive strand (+) and for transcription of phage genes. During early infection newly synthesized (+) strands are recycled to create more dsDNA replicative forms. Later in the infection, (+) strands are coated with the ssDNA-binding protein pV in the cytoplasm and this complex serves as a packaging substrate for the phage assembly. Newly synthesized phage proteins are targeted to the predetermined locations. Proteins pII, pV and pX remain in the cytoplasm and pI, pIII, pIV, pVI, pVII, pVIII, pIX and pXI are targeted to the membranes.

Ff assembly and export from the cells resembles T2SS secretion in Gram-negative bacteria. Proteins pI and pXI form an assembly complex in the cytoplasmic membrane while 14 pIV

monomers form a large gated pore in the outer membrane [197, 200]. The ssDNA genomes coated with pV dimers are brought to the phage export complex in the membrane where the packaging signal interacts with pVII, pIX and pVIII. As the ssDNA genome passes through the cytoplasmic membrane complex, pV is removed and replaced by pVIII, forming a tube or sheath around the ssDNA genome. When the phage ssDNA is fully packaged, pVI and pIII are added to the end of phage causing the termination of elongation and the release of the phage.

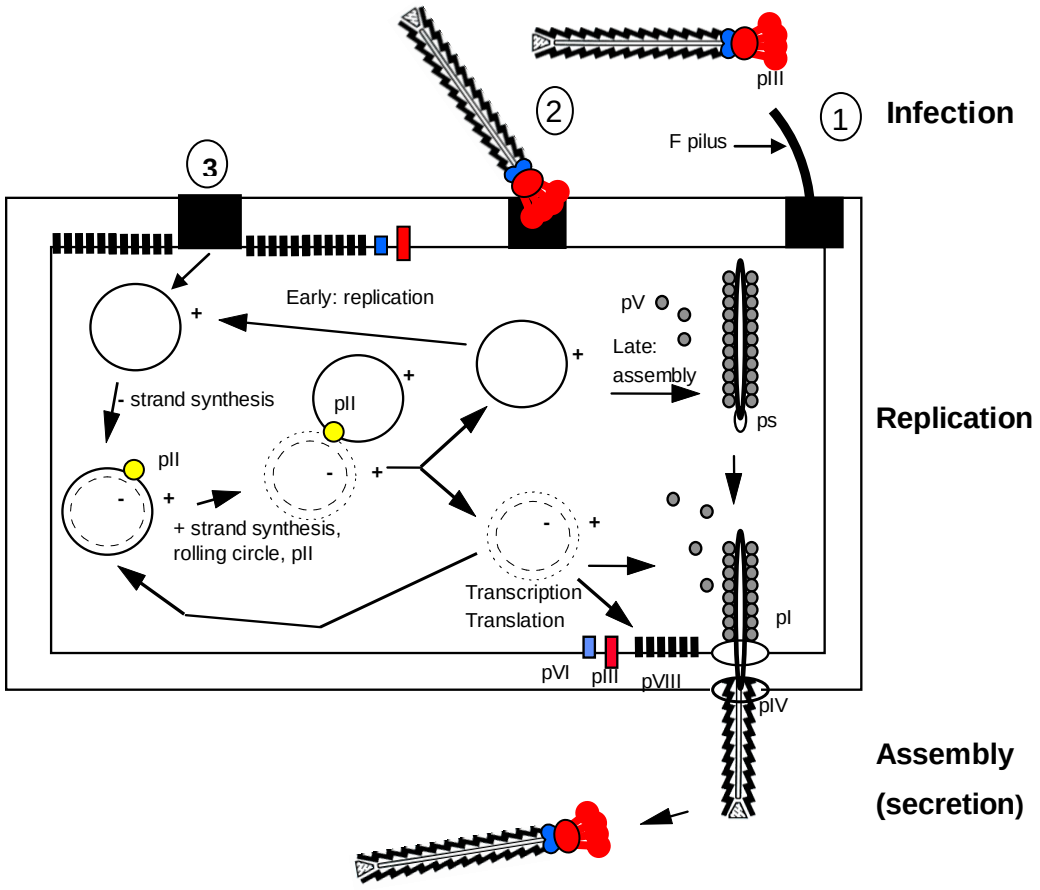


Figure 1.10. Life cycle of filamentous bacteriophage in *Escherichia coli*.

The drawing is schematic and not to scale. **Infection** by binding of the tip of the phage to the tip of the F pilus **(1)**. Secondary binding to protein complex TolQRA embedded in the cytoplasmic membrane and spanning the periplasm **(2)**. Entry of ssDNA (+ strand) into the cell; coat protein integrates into the inner membrane as DNA enters the cytoplasm **(3)**. **Replication**: In the first stage minus strand is synthesised by host proteins, starting from minus strand origin, primed by host RNA polymerase and carried out by the host DNA polymerase. At completion of negative (–) strand synthesis, the replicative form (RF) ds DNA is created. In the second stage phage protein II (pII) nicks the supercoiled dsDNA at the + strand origin, creating primer for the rolling circle synthesis of + strand. After one round, the product is cleaved off, again by pII, and the ends ligated. Early in the infection, + strand is recycled for replication. Later, + ssDNA genome is coated by pV. This complex is the substrate for packaging into the viral particles. **Assembly/secretion**: pI/pXI phage proteins form multimers in the inner membrane while pIV multimer forms a large gated channel in the outer membrane; together they form the assembly machinery. An exposed hairpin loop at the tip of pV/ssDNA packaging substrate interacts with pI, and then distal phage tip proteins pVII and pIX, to initiate assembly. As ssDNA is extruded through the *E. coli* envelope, pVIII packs on around it and in the end pIII/pVI release phage from membranes. If pIII or pVI are non-functional, phage is not released from the cell membrane, forming long filaments which are broken off by mechanical shearing.

1.3 The phagemid-based phage display system

There are two general types of phage display systems based on whether the library is constructed in modified Ff phage vectors or in specially modified plasmids called phagemids. The advantage of a phage vector system is in its simplicity. However, many investigators have observed that some sequences are not well displayed on the surface of the phage if the phage vector is used, as a result of interference with the viral particle assembly, stability and infectivity [201]. Some of these problems are alleviated by using phagemid vector systems. Phagemids are plasmids that contain the origin of replication of Ff phage, a plasmid origin of replication, a gene for display of protein encoded by the insert through fusion formation to an Ff virion protein, a multiple cloning site and an antibiotic marker. The Ff origin, packaging signal and the virion protein-encoding gene allow packaging of the phagemid genome and display of protein fusion. A helper phage is used as a source of proteins necessary for replication from Ff origin and assembly of a complete virion. Helper phages typically have defective phage origin of replication and/or packaging signal. Therefore, phagemid single stranded DNA is preferentially replicated and packaged over the helper phage genome (Figure 1.11).

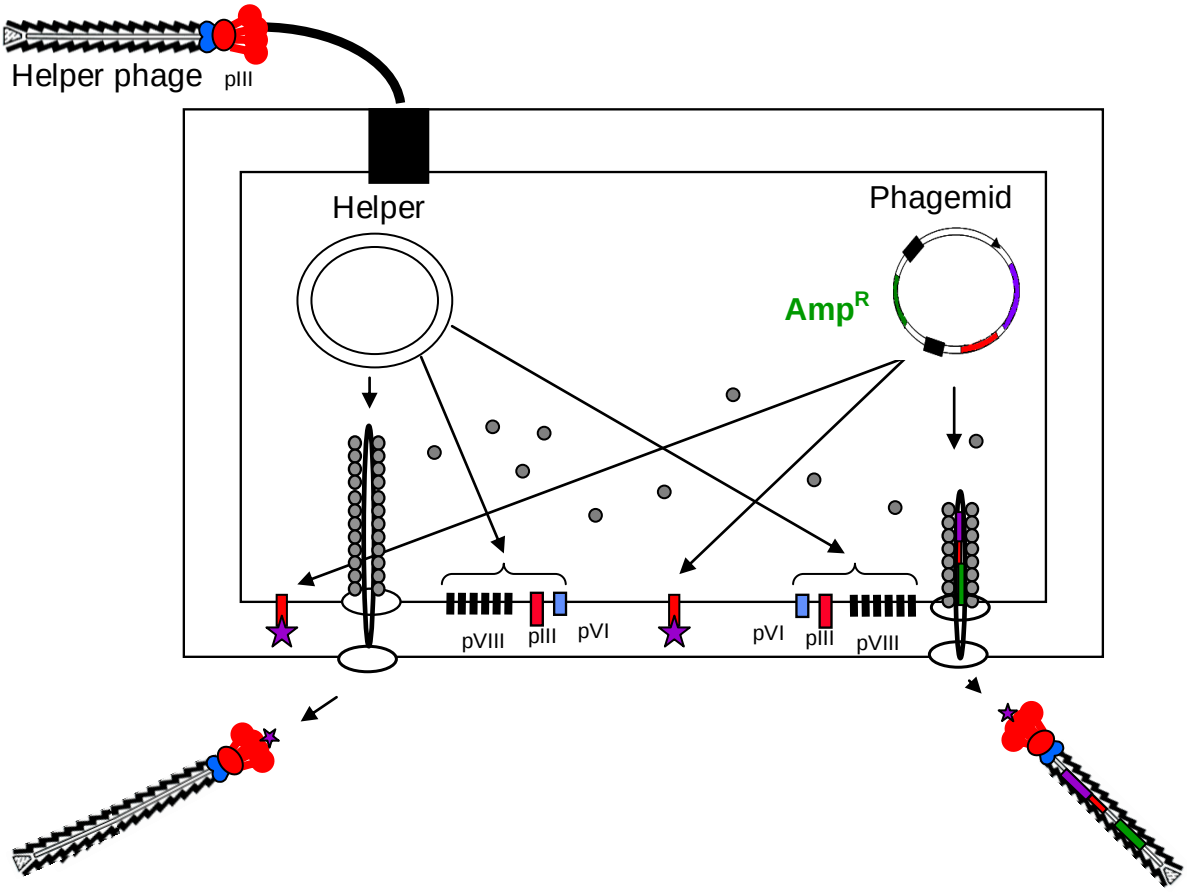


Figure 1.11. Schematic drawing of phagemid/helper phage replication and assembly in *E. coli*

Phagemids can be replicated as plasmids or alternatively, in the presence of a helper phage, packaged as recombinant phagemid particles. This figure shows a phagemid vector that uses phage protein pIII as a platform for display. The resulting phagemid particles may incorporate either pIII derived from the helper phage (red lollipop-like structure) or the polypeptide-pIII fusion protein (red lollipop-like structure decorated with a purple star), encoded by the phagemid. Phagemid particles are produced at a 10-100-fold excess over the helper phage.

Helper phage can contribute all virion proteins to the phagemid encapsidating virions, including wild-type pIII. This type of helper phage is herein referred to as a wild type helper. The phagemid/wild-type helper system results in monovalent display of recombinant protein; i.e. virions or phagemid particles are mosaic for recombinant and wild-type capsid proteins produced by phagemid encoding the fusion protein and by a helper, respectively (Figure 1.12). To increase the copy number of displayed fusion proteins, helper phage lacking the virion protein provided by the phagemid can be used. In the case of the most frequently used

phagemid vectors that encode for virion protein pIII, helper phages that carry deletion of corresponding gene III (*gIII*) are used. The obtained virions exhibit polyvalent display [202].

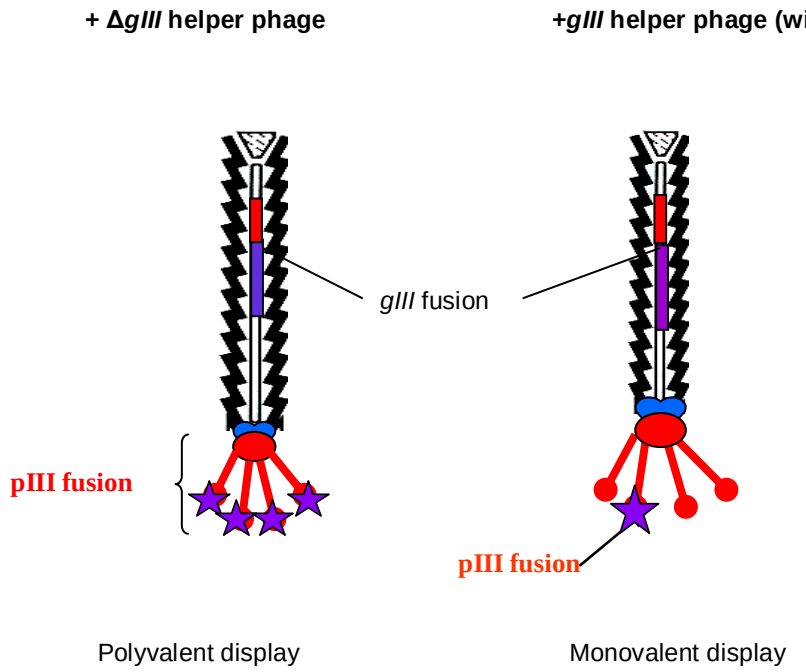


Figure 1.12. Valence of the recombinant protein in phagemid-based display systems
Depending on the helper phage used, recombinant protein can be displayed on the phagemid particle in multi-copy (polyvalent display using gene III deleted helper) or in one copy (monovalent display using wild-type helper phage ($gIII^+$).

Additional refinements of some phage display phagemid vectors are peptide tags (c-myc or E-tag), followed by an *amber* stop codon between the insert and downstream pIII (Figure 1.13). These additions allow a soluble version of foreign protein to be produced if phagemids are transformed into an appropriate non-suppressing *E. coli* strain.

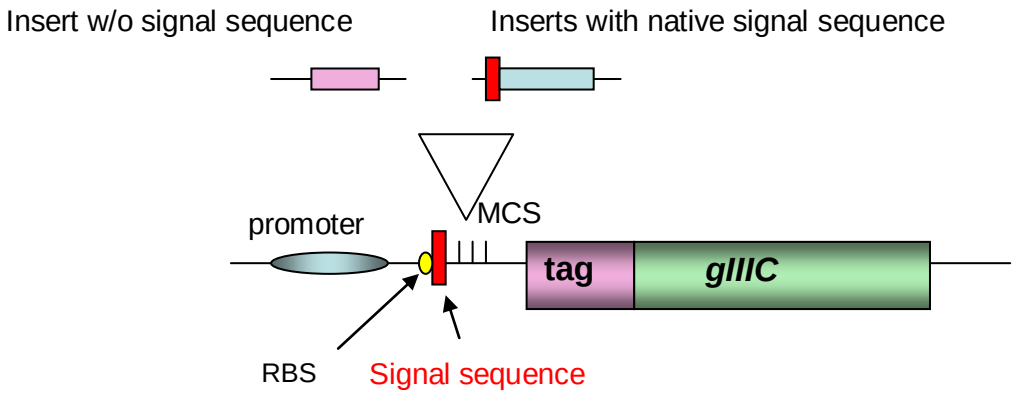


Figure 1.13. Schematic drawing of display module of phagemid used for "classical" phage display

"Classical" phage display phagemid features: promoter, a ribosomal binding site (RBS), signal sequence (commonly used PelB signal sequence from *Erwinia carotovora* [203]), multiple cloning site (MCS), affinity tag (tag) and C-terminal part of pIII (*gIIIC*; required for assembly of the virion). All inserts, irrespective of presence of the signal sequence, form functional pIII fusions provided they are inserted in frame with the upstream signal sequence and downstream pIII-encoding sequence.

1.4 Methods for identification of secretome proteins

A secretome can be deduced from a completely sequenced genome by using a range of available algorithms that can identify signal sequences and transmembrane α-helices described in Chapter IB. Most commonly used algorithms are SignalP 3.0, TMHMM 2.0, LipoPred and pSORT [204-209]. In addition, prediction algorithm SecretomeP can be used for detection of exported proteins with atypical signal sequences. However, obtaining complete genome sequences of multiple bacterial strains in order to identify their secretomes is inefficient because the secretome is a minor portion of the genome, typically comprising only 10-30% of the total number of the open reading frames (ORFs) [137].

Purely bioinformatic analysis is not only inefficient for secretome protein identification, but also does not provide the means for direct functional characterisation of identified proteins. In the post-bioinformatics phase of genome research, candidate ORFs are usually chosen based on a sequence motif or homology to a protein of known function, and then are either mutated or the protein products are expressed, purified and directly characterised. Both of these approaches are very demanding. The former requires that genetic tools exist for the organism of interest; the latter is complicated by the fact that some secretome proteins are notoriously hard to express and purify [210]. An approach in which the secretome sequences were specifically selected prior to sequence analysis would dramatically increase the efficiency of identifying secretome proteins, compared to the conventional shot-gun sequencing approach [1, 211].

Display technologies, bacterial cell-surface display and phage display, have been utilised for identification and functional characterisation of the secretome. The principle underlying these display methods is the ability to physically link phenotypes of polypeptides displayed on a certain platform, such as cell surface or bacteriophage, to their corresponding genotype. In cell-surface display a foreign peptide is fused to an export-specific reporter protein. In Gram-negative bacteria, most often, foreign gene products are fused to outer-membrane proteins or lipoproteins (e.g. OmpA, maltoporin LamB, phosphate-inducible porin PhoE and Lpp and TraT lipoproteins) although autotransporters and surface organelles such as pili and flagella have also been used for display [212-215]. Several Gram-positive bacteria (*Staphylococcus* spp., *Streptococcus* spp., *Bacillus* spp., *and L. lactis*) have also been used for bacterial surface-display. In these hosts, most commonly, a foreign peptide is fused to the cell-wall anchoring region (LPXTG motif) of the host surface protein, as in examples of protein A (SpA) and M proteins cell-surface displays [213, 216-218].

The use of fusion to Gram-negative reporter genes as a genetic tool to screen Gram-positive secretome proteins has had limited success due to improper folding, decreased expression and the creation of bias during preselection in surrogate hosts. Fusion to Gram-positive reporters generated more stable export-specific activity but considerable amount of false-positive clones was detected. Phage display is advantageous over bacterial surface display because of lower toxicity, more efficient folding, and stability of Ff phage to pH extremes and detergents [219].

Another advantage of phage display over cell surface display is that phage display technology offers a very efficient way to purify and characterise proteins by displaying them on the surface of the bacteriophage virion [220, 221], since filamentous phage are simply purified from culture supernatants by precipitation with polyethylene glycol (PEG). Phage display combinatorial libraries are widely used to identify rare protein variants that bind to complex ligands of interest; the most complex example reported being an *in vivo* screen for peptides that bind endothelial surfaces of the capillaries in an organ-specific fashion [222]. Furthermore, phage display screening methods for selection and *in vitro* evolution of enzymes have been developed and used successfully [223].

1.4.1 Examples of bacterial adhesins identified using phage display technology

In the last decade, the phage display technique has proven to be a powerful tool for identification of genes encoding bacterial adhesins and characterisation of the respective binding domains that mediate interactions with host cell [221]. The shot-gun phage display (display of random fragments of bacterial genomic DNA) was first exploited by Jacobsson and Frykberg to identify *Staphylococcus aureus* proteins that interact with extracellular matrix components and host serum or plasma [224, 225]. Following their initial description of *gIII*-phagemid-based phage display many genes encoding for proteins involved in host-microbial interactions from different bacterial species have been identified using the phage display approach [221, 222, 226-228].

In general, a shot-gun phage display library contains several recombinant phagemids carrying portions of a gene that binds to a ligand of interest. These inserts usually overlap across the binding domain. Thus many studies used random fragmentation of bacterial genomic DNA to identify the gene of interest in the genome or random fragmentation of DNA of the gene of interest to generate a single-gene phage display library and map a domain of interest within a particular gene. For example, this approach was used to identify and map the cell-surface-associated agglutinin (RapA) from *Rhizobium leguminosarum* genome [226]. Recently, a shot-gun phage display library of *Borrelia burgdorferi* genome was used for the *in vivo* screening for potential adhesins, resulting in discovery of at least five new adhesion proteins [229]. Single-gene phage display libraries were utilised to map a fibronectin-binding motif of

the fibronectin-binding cell surface protein, Fnz, of *Streptococcus equi*, the IgG- and β2-glycoprotein-I-binding domains of *S. aureus* receptin Sbi and many others [230, 231].

1.5 Aims of the Project

Overall, our knowledge of cellular components that mediate probiotic effects and availability of technologies to identify them in a simple, cheap and rapid way is relatively poor. The understanding of molecular mechanisms of probioosis is very limited. Only a handful of bacterial molecular effectors involved in probiotic effects are known, and none of them have been identified in *L. rhamnosus*.

As surface and secreted (secretome) proteins are essential for interaction of bacteria with the colonised host, the main aim of this thesis was to improve technology for identification and analysis of this group of proteins using phage display and to apply it to the probiotic bacterium *L. rhamnosus* HN001.

Therefore the two specific aims of the work presented in this thesis were to design a new phage display system for direct selection, expression and display of the secretome of the probiotic bacterium *Lactobacillus rhamnosus* strain HN001 and to "mine" the secretome library for proteins that mediate probiotic effects. The first set of experiments in Chapter III describes the construction of a new system for selective display of the bacterial secretome and its testing on *L. rhamnosus* HN001. Using this approach, the secretome library inserts were sequenced and 89 putative secretome ORFs were annotated. To functionally identify proteins involved in establishing the host-microbe interaction (adherence) or microbe-microbe contact (aggregation) a shot-gun phage display library of *L. rhamnosus* HN001 was constructed. Virions displaying entire HN001 proteome were screened for proteins that interact with ligands known to be involved in adhesion and autoaggregation processes, which in turn are essential for colonisation and biofilm formation (Chapter IV).

Chapter II

Materials and Methods

2.1 Materials

2.1.1 Laboratory Chemicals and enzymes

Restriction endonucleases and DNA modifying enzymes were obtained from the following companies: Invitrogen Life technologies Inc. (USA); Roche Molecular Biochemicals, (Germany); New England Biolabs Inc. (USA); USB Corporation (USA). Thermophilic DNA polymerases were purchased from Roche Molecular Biochemicals (Germany).

Oligonucleotides were manufactured by Invitrogen Life Technologies Inc. (USA). Bacteriological media, Difco™ 2xYT (Yeast Extract Tryptone) and MRS were purchased from Becton-Dickinson and Company (USA), and Scharlau Chemie S.A (Spain). Antibiotics were purchased from Sigma-Aldrich (Australia). General laboratory chemicals were purchased from Sigma-Aldrich (Australia), Merck Ltd. (Australia), BDH (USA). X-ray film, photographic developer and fixer were obtained from Eastman Kodak (USA).

2.1.2 Buffers, solutions, and media

Standard buffers, solutions, and media were prepared as described in Sambrook and Russell [232].

2.1.3 Bacterial strains, plasmids and phage

Bacterial strains, plasmids and phage used in this study are listed in Tables 2.1 and 2.2.

Table 2.1. Bacterial strains used in this study

Strains	***Genotype***	***Reference***
Escherichia coli TG1	*supE thi-1* Δ(*lac-proAB*) Δ(*mcrBhsdSM*)*5* (rK⁻ mK⁻) [F´ *traD36 proABlacI*^{q}*ZΔM15*]	[233]
E. coli K1976	TG1// pJARA112	This study
E. coli K1931	TG1//pDJ01	This study
E. coli K1937	TG1//pSOF22	This study
E. coli K1939	TG1//pAK-SOF22	This study
E. coli K1978	TG1//pDJ08	This study
E. coli K1981	TG1//pYW01	This study
E. coli K2087	TG1//pFn01	This study
E. coli K2088	TG1//pFn02	This study
E. coli K2089	TG1//pLrh33S	This study
E. coli K2090	TG1//pLrh33A	This study
E. coli K3000-K3490	TG1// L001-490 (TG1 carrying secretome library clones)	This study
Lactobacillus rhamnosus HN001		[54]

Table 2.2. Plasmids and phage used in this study

Plasmids	***Reference***	***Resistance***
pAK100	[234]	Chloramphenicol
pJARA144	J.Rakonjac unpublished	Ampicillin
pJARA112	[202]	Ampicillin
pDJ01	This study	Chloramphenicol
pDJ08	This study	Chloramphenicol
pSOF22	This study	Chloramphenicol
pAK-SOF22	This study	Chloramphenicol
pGZ119EH	[235]	Chloramphenicol
pYW01	This study	Chloramphenicol
pFn01	This study	Chloramphenicol
pFn02	This study	Chloramphenicol
pLrh33S	This study	Chloramphenicol
pLrh33A	This study	Chloramphenicol
L001-L490	This study	Chloramphenicol
Phage		
VCSM13	Stratagene (USA)	Kanamycin
VCSM13d3	[202]	Kanamycin
λEMBL4::SOF	[97]	

2.1.4 Oligonucleotides

Oligonucleotide primers used for cloning, sequencing, PCR reactions and dot blot hybridisation are listed in Table 2.3.

Table 2.3. Oligonucleotide primers used for cloning, sequencing, PCR reactions and hybridisations

Name	***Sequence***	***Restriction site***	***Application***
pspF01	CCGCCGAATCCCGGAAGAG CTGCAGCATGATGAAATTC	EarI	cloning
pspR01	CCGCCGCCG<u>GAATTC</u>*TCTAG A***CCCGGG***G<u>C</u>ATGC<u>AT</u>*TGTCC TCTTG	*Nsi*I, SphI, **SmaI**, *XbaI*, EcoRI	cloning
pDJ01F03	ATGTTGCTGTTGATTCTTCA		sequencing
pDJ01R02	CCGGAAACGTCACCAATGA A		sequencing
pJARA144R02	GCCGCC<u>ATGCAT</u>GGGCCATG GCCAT	NsiI	cloning
Sof22F01	CCGCCG<u>ATGCAT</u>TGACAAAT TGTAAG	NsiI	cloning
Sof22R01	CCGCCG<u>GAATTC</u>CTCGTTAT CAAAGTG	EcoRI	cloning
Sof22F02	CGC<u>CCGCGG</u>ACTGAGACGA GTGCT	SacII	cloning
L001F01	ACAACCCGCAGTTCCCGAGC		PCR for hybridisation probe
L001R01	TTGACATCGGCGAAAACAC A		PCR for hybridisation probe
L014F01	TCGTAATTCTCACTGGCGGA		PCR for hybridisation probe
L014R01	TCCGAGCGGCAGTTTCATTG		PCR for hybridisation probe
L050F01	CTTAACCCAGCAAGGACAAT CG		PCR for hybridisation probe
L050R01	ACTCACTTGTTCCGTTGGCT CGC		PCR for hybridisation probe
L062F01	GACCTTGTCCGATGCACCTA		PCR for hybridisation probe
L062R01	GGCTGATTGTAAAACCTGCG C		PCR for hybridisation probe
L064F01	TCCATGCCGCGACAACAACT		PCR for hybridisation probe
L064R01	ACTGGCAGACTCACGATCAT ATC		PCR for hybridisation probe
L073F01	CCAGATGAAGATTGAGAAC AACTCC		PCR for hybridisation probe
L073R01	ATCACTTCAGTGTTAGCTGC TTCGA		PCR for hybridisation probe
L108F01	CCACTGACAAGTTCGGCATC T		PCR for hybridisation probe
L108R01	AGCTGGTAGCGACTGGCAAT C		PCR for hybridisation probe

L157F01	TTGCAGGCTCGCTTTGTCGT	PCR for hybridisation probe
L157R01	GAAGTTGATCCAAGCACGC AT	PCR for hybridisation probe
L182F01	CCTAGTGCCGTCATCGTATC CT	PCR for hybridisation probe
L182R01	GCGCCCAACTGAGCTTTACA	PCR for hybridisation probe

2.2 Methods

2.2.1 General molecular biology methods

General molecular biology methods were carried out as described in Sambrook and Russell [232].

2.2.2 Bacterial growth conditions

Escherichia coli strains were propagated at 37ºC in 2xYT broth with aeration or Yeast Extract Tryptone (2xYT) agar plates. The media were supplemented with suitable antibiotics. Chloramphenicol (Cm) was used at 20 µg/ml, Ampicillin (Amp) at 60 µg/ml and Kanamycin (Km) at 30 µg/ml, unless indicated otherwise.

Lactobacillus rhamnosus HN001 was routinely cultured in Man-Rogosa-Sharpe (MRS) broth in a stationary 37°C waterbath or on MRS agar plates. The plates were placed in a sealed plastic bag together with an anaerobic generator (Pouch-Anaero, BioMerieux) and incubated overnight in a 37°C incubator. Anaerobic conditions were confirmed by indicator strips (BBL™ Dry Anaerobic indicator strips, Becton Dickinson).

For long-term storage, all cultures were stored at -80ºC in 7% (v/v) dimethyl sulfoxide (DMSO). Short-term storage (up to two weeks) of bacterial strains was achieved by storing solid plates at 4°C.

2.2.3 Nucleic acid techniques

2.2.3.1 Isolation of total genomic DNA from *L. rhamnosus* HN001

For library construction, chromosomal DNA was isolated from an overnight culture of *L. rhamnosus* HN001 using a modification of the method described previously [54]. Briefly, an overnight culture was diluted 1:100 into 80 ml MRS broth and incubated overnight at 37ºC. Cells were harvested by centrifugation at 5,500×g for 10 minutes, resuspended in 80 ml of MRS broth and incubated for a further period of 2 h at 37ºC. Cells were washed twice in 16 ml of lysis buffer [30 mM Tris-HCl (pH 8.0), 50 mM NaCl and 5 mM EDTA] and resuspended in 2 ml of the same buffer containing 25% (w/v) sucrose, 20 mg ml^{-1} lysozyme (Sigma-Aldrich) and 20 μg ml^{-1} mutanolysin (Sigma-Aldrich). The suspension was incubated for 1 h at 37ºC. Further lysis of the cells was accomplished by adding 2 ml of EDTA/SDS solution [0.25 M EDTA, 800 μl 20% (w/v) SDS]. After addition of SDS the suspension was carefully mixed and incubated at 65ºC for 30 minutes. Next, Proteinase K (Roche) was added to a final concentration of 200 μg ml^{-1} and the suspension was incubated at 65ºC for 15 minutes. The DNA was then purified by adding equal volume of phenol:chloroform:isoamyl alcohol mix (in ratio 25:24:1), vortexing and centrifuging at 5000xg for 5 min. The aqueous phase was collected and phenol:chloroform:isoamyl alcohol extraction was repeated. To remove traces of phenol, equal amount of chloroform was added to the aqueous phase containing DNA, followed by vortexing and centrifugation steps as described above. Residual RNA contamination was removed by adding DNase-free RNase A (Roche) to a final concentration of 100 μg ml^{-1} and the incubation at 37ºC for 30 min. The DNA was extracted again by phenol:chloroform:isoamyl alcohol as described above. After chloroform extraction, the DNA was precipitated by addition of 1/10 volume 3 M sodium acetate (pH 5.2) and 2.5 volumes 95% (v/v) ethanol. The DNA was pelleted by centrifugation, washed with 70% (v/v) ethanol, air dried and resuspended in an appropriate volume of 10 mM Tris-HCl (pH 8.0).

2.2.3.2 Plasmid DNA isolation

Small scale plasmid DNA from *E. coli* (mini prep) was prepared using a modified alkaline lysis method and the High Pure Plasmid Isolation kit (Roche, Germany), which uses a modified alkaline lysis method [236]. If larger quantities of plasmid were required the QIAGEN plasmid Midi kit (QIAGEN, Germany) was used. *E. coli* TG1 cultures containing various plasmids and phagemids were propagated with appropriate antibiotic selection for

maintenance of the particular episome. Two ml of stationary phase culture was harvested for DNA extraction when using the miniprep kit, whereas 50 ml of stationary phase culture was harvested for DNA extraction when using the midiprep kit. DNA extractions were performed as outlined in the manufacturers' instructions (QIAGEN and Roche). Briefly, cells were disrupted by alkaline lysis, the bacterial lysate was cleared by centrifugation and then applied to a silica-gel membrane (in a column). After washing the column to remove impurities, plasmid DNA was eluted from the column. In the midiprep protocol, DNA in the eluate was precipitated by isopropanol, pelleted by centrifugation and resuspended in 10 mM Tris, pH 8.0. In miniprep protocol, DNA was eluted in a small volume of low-salt buffer, eliminating the need for precipitation.

2.2.3.3 Polymerase chain reaction (PCR) amplification

Bacterial colony PCR

Bacterial colony PCR was used for identifying bacterial transformants directly from bacterial cells, using as a template DNA released from the cells during the PCR protocol. Individual colonies were picked from plate, resuspended in 100 µl 1X TBS and streaked on a replica plate. Replica plates were incubated at 37ºC overnight and used a source for the plasmid extraction of desirable clones. A volume of 5 µl from above suspension of colony in 1X TBS was used for the PCR reaction. Primers for the reaction were complementary to the vector sequence flanking the insert. PCR amplifications were carried out with *Taq* polymerase (Roche) in the Whatman BioMetra Thermal Cycler (UK). The components of PCR mixes and reaction conditions are described in Tables 2.4 and 2.5, respectively.

Table 2.4. Component mixtures used in colony PCR reactions

Components	*Volume (20 µl)*	*Final Concentration*
10X PCR buffer, (including 15 mM $MgCl_2$)	2 µL	1X
25 mM dNTP mixture	0.16 µL	0.2 mM each
10 µM sense primer	0.4 µL	0.2 µM
10 µM anti-sense primer	0.4 µL	0.2 µM
Colony pre-mix	5 µL	n/a
Taq DNA Polymerase	0.1 µL	1 unit
H_2O	11.94 µL	n/a

Table 2.5. Thermal profile of the colony PCR reactions

Thermal profile		
1	94°C	5 min
2	94°C	20 sec
3	Annealing*	30 sec
4	72°C	Ext*
5	Repeat steps 2-4	25
6	72°C	7 min
7	4°C	∞

* Annealing temperature was set according to the primers used. Extension was adjusted according to the length of the amplicon.

General PCR

In this study PCR amplifications of the DNA used for cloning were carried out with *Pwo* DNA polymerase (Roche) in the Whatman BioMetra Thermal Cycler (UK). The composition of PCR mixes and the reaction conditions are listed in Tables 2.6 and 2.7, respectively.

Table 2.6. Component mixtures used in general PCR reactions

Components	*Volume (50 µl)*	*Final Concentration*
Master Mix 1:	25 µl	
25 mM dNTP mixture	0.16 µL	0.2 mM each
10 µM sense primer	0.4 µL	0.2 µM
10 µM anti-sense primer	0.4 µL	0.2 µM
Template DNA	2 µl	0.05 µg
H_2O	22.04 µl	n/a
Master Mix 2:	25 µl	
10X PCR buffer, (including 20 mM $MgSO_4$)	5 µL	1X
Pwo DNA polymerase	0.5 µl	2.5 units
H_2O	19.5 µl	n/a

Master mixes were prepared on ice to avoid the degradation of primers or template in the absence of dNTPs through the 3'-5' exonuclease activity of *Pwo* polymerase.

Table 2.7. Thermal profile of the general PCR reactions

Thermal profile		
1	94°C	2 min
2	94°C	15 sec
3	Annealing*	30 sec
4	72°C	Ext*
	Repeat steps 2-4	10
6	94°C	15 sec
7	Annealing*	30 sec
8	72°C	Ext + cycle elongation of 2 sec for each cycle*
	Repeat steps 6-8	16
9	72°C	7 min
10	4°C	∞

* Annealing temperature was set according to the primers used. Extension was adjusted according to the length of the amplicon.

2.2.3.4 Dot blot hybridisations

After the sequence analysis of the first 192 clones, the clones whose sequences were detected more than five times were excluded from further sequencing using dot blot hybridisation. Probes of approximately 0.4 Kbp in length for each of the nine promiscuous clones were obtained by PCR amplification using primers that did not overlap with the vector. PCR-amplified probes were gel-purified and individually labeled using ECL Direct Nucleic Acid Labelling and Detection kits (GE Healthcare BioSciences). The purified phagemid DNA of individual library clones (~ 0.1 µg) were spotted directly on a positively charged nylon membrane (Roche) and cross-linked. Labelling, hybrisation and detection were performed according to the procedure recommended by the ECL Direct Nucleic Acid Labelling and Detection kit manual.

2.2.4 Phage protocols

2.2.4.1 Phage infection, growth and purification

In the phagemid system used in this work, virions containing helper phage genomes are referred as phage, whereas virions containing phagemid genomes are referred to as phagemid particles (PPs). In all experiments in which phage or PPs were produced, an exponential phase culture (around 10^8 cells/ml) of the appropriate strain was infected with the appropriate helper phage at a multiplicity of infection (m.o.i) of 50 (50 phage to 1 bacterium), for 30 min at 37ºC. Infected cells were separated from unabsorbed helper phage by centrifugation at 5000xg for 10min at RT and resuspended in fresh medium. The cultures were then incubated for 4 h at 37ºC with aeration unless otherwise stated. At the end of incubation, cells were chilled, titrated and then separated from the phage or PPs by centrifugation at 13200xg for 20 min at 4ºC. Bacterial pellets were resuspended in the equal volume of the chilled medium supplemented with 7% DMSO. Generally 1 ml aliquots were taken, frozen on dry ice, stored at -80ºC and later used for the analysis of the proteins by SDS-PAGE and Western blotting.

The supernatant containing phage/PPs was centrifuged again and then passed through a 0.45 μm filter to eliminate any remaining bacterial cells. Phagemid particles were further purified and concentrated by precipitation in PEG/NaCl buffer [5% (w/v) PEG, 500 mM NaCl] and resuspended in TN buffer (10 mM Tris-HCl pH 7.6; 50 mM NaCl) unless otherwise stated. Phage and PPs were enumerated as described (section 2.2.4.2) and stored at 4ºC for short term storage or at -80ºC in 7% DMSO for long term storage.

2.2.4.2 Phage enumeration

Titration of infectious phage and phagemid particles

The number of infectious phage was determined by titration. "Wild-type" (*gIII*$^+$) helper phage VCSM13 was propagated and titrated on TG1 strain. Stocks of Δ*gIII* helper phage VCSM13d3 were generated on *E. coli* strain K1976, which contains complementring plasmid pJARA112 with *gIII* under the control of phage-inducible promoter *psp*. This strain allows plaque formation by Δ*gIII* phage (VCSM13d3). Approximate number of phage was first estimated by placing 10 μl drops of phage dilutions onto soft agar containing 0.2 ml of overnight culture of appropriate *E. coli* strain. Plaques in the area of absorbed drops were

counted to determine approximate numbers of phage. To accurately quantify phage, 10 µl of optimal dilution (determined by previous method) was mixed with 100 µl of appropriate *E. coli* strain and 2.5 ml of 2xYT soft agar (0.6% w/v). The mixture was poured over the 2xYT agar plate. At least three hundred plaques from the total of three plates were counted to determine the titre accurately. Helper phage titre was expressed as plaque-forming units per ml (pfu/ml).

The infectious phagemid particles carrying a Cm^R marker were titrated by the same method, except that specially prepared double-layer plates were used, to allow in-agar infection of the indicator strain TG1 prior to exposure to Cm [200]. Briefly, 2xYT agar plates were prepared with 25 µg ml^{-1} chloramphenicol. These plates were overlaid with 9 ml chloramphenicol-free 2xYT agar shortly before use. PP stock was then mixed with *E. coli* strain TG1 in the soft agar layer as described in previous paragraph. PP titre was expressed as colony-forming units per ml (cfu/ml).

Agarose electrophoresis of the phage and phagemid particles and quantification

Phage and PPs were quantified by densitometry based on the amount of encapsulated DNA. Due to different lengths of particles the amount of phage or PPs were expressed in monophage equivalents. A monophage equivalent is a measure of particle mass, and is defined as a particle containing one encapsulated genome (Appendix I, Figure A.1). Thus, a particle containing 10 genomes represent ten genome equivalents, as do 10 particles containing one genome each [199]. The number of genome equivalents was determined from agarose electrophoresis of phage ssDNA, released from the SDS-disassembled virions [237]. Prior to electrophoresis, virions were disassembled by incubation in SDS-containing buffer (1% SDS, 1X TAE, 5% glycerol, 0.25% BPB) at 70ºC for 20 min, and then subjected to agarose electrophoresis in 1X TAE buffer. After electrophoresis, phage ssDNA was stained with EtBr and quantified densitometrically. Since the amount of ssDNA in a band is not linearly proportional to the intensity of the fluorescence, every gel contained a set of twofold dilutions, typically 10 to 360 ng per lane of purified VCSM13d3 ssDNA, used for calibration. The gel was photographed with a CCD camera (BioRad, USA), and quantitative analysis performed using software packages Image Gauge V4.0 (Fuji Film, Japan) and Excel (Microsoft). The second order polynomial function was used to fit the standard curve over the

range. Conversion of the calculated amount of ssDNA in the samples into the amount of genome equivalents or phage was carried out based on the molecular weight of ssDNA genome, which was again calculated from its base composition and length of a phage or phagemid. Native virion agarose electrophoresis was used to separate virions of various lengths and to detect free phage DNA when stability of phage was analysed [237]. Samples were loaded on very low density agarose gels (0.4%) in DNA loading buffer (1X TAE, 5% glycerol, 0.25% BPB). Electrophoresis was performed at 2 V/cm for 17 h. After electrophoresis, free phage ssDNA was first detected by staining the gel in EtBr. To detect the position of the native phage in the gel, virions were disassembled by soaking the gel in 0.2 M NaOH followed by neutralisation with 0.45 M Tris pH 7.1 and visualisation with EtBr.

2.2.5 Protein electrophoresis and western blots

Proteins from the cell lysates and phage/PPs samples were separated by SDS-PAGE using a tricine gel system [238], transferred to nitrocellulose filters, and then detected using a monoclonal anti-pIII antibody [239] and anti-mouse anti-serum conjugated to HRP. The ECL Plus detection system (GE Healthcare Biosciences) was used for detection of antibodies bound to the filter. The basic buffer was TBS (30 mM Tris, 150 mM NaCl, pH 7.6) supplemented with 0.05% Tween 20. Blocking and antibody binding buffers also contained, respectively, 0.5% and 0.1% casein based I-block (Tropix, USA).

2.2.6 Phagemid constructs

2.2.6.1 Construction of the phagemid vector pDJ01

Primers pDJ01F01 containing an *Ear*I site and pDJ01R01containing *Eco*RI, *Xba*I, *Sma*I and *Sph*I sites at the 5' end (Table 2.3) and template pJARA144 (Table 2.2) were used to generate a PCR product containing the *psp* promoter (phage infection induced promoter) followed by a ribosomal binding site and a multiple cloning site. The product was cleaved with *Ear*I and *Eco*RI and ligated into *Ear*I-*Eco*RI digested phagemid pAK100 (Table 2.2) [234]. The ligation placed the *psp* promoter, ribosomal binding site and the multiple cloning cassette directly upstream of a sequence encoding the peptide tag c-myc, followed by suppressible a *amber* (TAG) stop codon and a coding sequence for the C-domain of pIII. The phagemid was named pDJ01 and used in the construction of *L. rhamnosus* HN001 secretome-selective phage display library.

2.2.6.2 Construction of the phagemid displaying the SOF of *S. pyogenes*

Primers pSOF22F01 containing an *Nsi*I site and pSOF22R01, containing an *Eco*RI site (Table 2.3) and the template, purified DNA of a λEMBL4::SOF clone of the *sof22* from *S. pyogenes* M type 22 strain D734 (Table 2.2), were used to generate a PCR product encoding the SOF of the M22 strain (SOF22), including the signal sequence but excluding the cell wall and membrane anchor sequences (963 residues). The PCR product was cleaved with *Nsi*I and *Eco*RI and ligated to the *Nsi*I-*Eco*RI-cleaved vector pDJ01. This phagemid was named pSOF22.

Primers pSOF22F02 containing a *Sac*II site and pSOF22R01, containing an *Eco*RI site and the template, purified DNA of a λEMBL4::SOF clone of the *sof22* from the *S. pyogenes* M type 22 strain D734 (Table 2.2), were used to generate a PCR product encoding the SOF, without its native signal sequence, the cell wall and membrane anchor sequences. The PCR product was cleaved with *Sac*II and *Eco*RI and ligated to the *Sac*II-*Eco*RI-cleaved vector pAK100 generating pAK-SOF22 phagemid which contained the *sof22* fused to the *pelB* signal sequence.

2.2.6.3 Construction of the phagemid pYW01 carrying PelB signal sequence and colD origin of replication

The phagemid pYW01, was constructed in two steps: first, the *psp* (phage infection induced) promoter, ribosomal binding site and *pelB* signal sequence (from *Erwinia carotovora* pectate lyase – PelB) were amplified by PCR using pJARA144 as a template and primers pDJ01F01, containing an *Ear*I site, and pJARA144R02, containing an *Nsi*I restriction site (Table 2.3). The PCR product was cleaved and ligated into the *Ear*I-*Nsi*I digested phagemid pDJ01 (Table 2.2) [234]. This step introduced *pelB* signal sequence between the *psp* promoter and multiple cloning site cassette of pDJ01. The resulting phagemid was designated pDJ08. Next, the origin of replication in pDJ08 (*colEI*) was replaced with the *colD* origin of replication from the vector pGZ119EH (Table 2.2).This was achieved by cleavage of the *psp-pelB*-c-myc-*gIIIC* cassette and origin of replication from pDJ08 using *Pst*I-*Sca*I followed by removal of 3' protruding end from *Pst*I by T4 DNA polymerase (Roche) and ligation of the resulting fragment to pGZ119EH *Eco*RV-*Sca*I fragment carrying *colD* origin of replication. The resulting phagemid pYW01 was used in the construction of *L. rhamnosus* HN001 shot-gun phage display library.

2.2.7 Production and functional assays of the SOF-displaying phagemid particles

The phagemid particles were generated by infection of 100 ml of exponentially growing cultures of K1931 (TG1//pDJ01) and K1937 (TG1//pSOF22) with helper phage stocks at a multiplicity of infection of 50 phage per bacterium. Helper phages VCSM13 and VCSM13d3 were used for production of phagemid particles of the pSOF22 (named pSOF22 PP/wt and pSOF22 PP/d3, respectively) and VCSM13 only for production of pDJ01 (negative phagemid particle control; named pDJ01 PP/wt). VCSM13d3 helper phage could not be used for the production of pDJ01 phagemid particles because of the lack of functional pIII. Infected cells were incubated for 4 h at 37ºC with aeration. The host cells were pelleted by centrifugation and phagemid particles collected in the supernatant. The phagemid particles were purified and quantified as described previously (section 2.2.4.2) [199].

2.2.7.1 Serum opacity assay

The serum opacity assay was carried out by mixing 1 ml of heat-inactivated horse serum with 10^{11} phagemid particles displaying the SOF (pSOF22 PP/wt; pSOF22 PP/d3) or negative control (pDJ01 PP/wt). Sodium azide (0.02%) was included to prevent multiplication of residual bacterial cells. The reactions were incubated at 37ºC and the time course of increase of optical density over time was monitored at a wavelength of 405 nm.

2.2.7.2 Fibronectin-binding assay

The fibronectin-binding assay was carried out by phage enzyme-linked immunosorbent assay (Phage ELISA [240]). The microtiter wells (Nunc-Immuno MaxySorp™, Denmark) were coated with human plasma fibronectin (Sigma, Australia) at a final concentration of 20 µg ml^{-1} in PBS (pH 7.2), 100 µl per well, for 1 h at 37ºC. The wells were washed once with 300 µl PBS, 0.05 % Tween 20 buffer (PBST) and then blocked with 1% (w/v) bovine serum albumin (BSA) in PBS for 2 h at room temperature. The wells were then washed (three times) with 300 µl of PBST buffer. Phagemid particles (2×10^8) in 100 µl of PBS were added to the wells. Negative buffer controls were TE (10 mM Tris, 1 mM, EDTA, pH 8.0), PBS, and 0.05% (w/v) BSA in PBS, and the negative phagemid particle control was pDJ01 PP/wt, generated as described above. The plates were incubated for 2 h at room temperature. The unbound phagemid particles were removed by washing with PBST (seven times). To detect bound phagemid particles, 100 µl mouse anti-pVIII (monoclonal antibody to M13, fd and f1,

Progen Biotechnik (Germany)) was added at a concentration of 0.1 µg ml^{-1} in 1X PBS containing 0.1% (w/v) BSA and incubated for 1 h at room temperature. The wells were then washed with 300 µl PBST buffer (five times) and 100 µl of secondary HRP-conjugated anti-mouse antibody was added at a dilution of 1:2000. The reaction was incubated for 1 h at room temperature. The plate was then washed seven times with PBST buffer and developed using the ImmunoPure TMB substrate kit (Pierce, USA). The absorbance was read at 450 nm. The phagemid particles were quantified as described in section 2.2.4.2.

2.2.8 Construction of the genome library of HN001 in vector pDJ01

The library was constructed from mechanically (nebulization) sheared *L. rhamnosus* HN001 DNA and cloned into the phagemid vector pDJ01. A disposable medical nebulizer containing 1.5 ml chromosomal DNA (approximately 20 µg) in TBS buffer and 25% (v/v) glycerol was subjected to nitrogen gas at a pressure of 10 psi for 90 s. The fragments obtained varied in size between 0.3 and 4 Kbp, with the majority between 0.5 and 1.6 Kbp. Blunt ends were achieved by treatment with T4 DNA polymerase (Roche), Klenow fragment of DNA polymerase I (Roche) and OptiKinase™ (USB Corporation, USA). To eliminate fragments below 0.3 Kbp, Sepharose CL-4B 200 (Sigma) size exclusion resin was used. The phagemid vector pDJ01 was digested with the restriction enzyme *Sma*I (Roche) and dephosphorylated with shrimp alkaline phosphatase (Roche). The DNA manipulations were performed according to standard methods [232].

Approximately 10 µg of the genomic fragments were ligated to 3 µg of the vector pDJ01 using T4 ligase (Roche). After phenol and chloroform extraction, the ligated DNA was ethanol-precipitated, washed with 70% (v/v) ethanol and dissolved in 25 µl H_2O. The ligation mix was transformed into *E. coli* TG1 by electroporation (2.5 kV, 25 µF, 400 Ω) in 2-mm-gap cuvettes. The transformed cells were transferred to 50 ml of 2xYT and incubated for 1 h at 37ºC with rotatory agitation. After the incubation a 2 ml aliquot was taken to determine the number of transformants by plating Cm selective plates. The remaining bacteria were amplified overnight at 37ºC with aeration.

2.2.9 Direct selection of the secretome phage display library

A 1 ml aliquot of the overnight culture containing the whole genome library, constructed as described in section 2.2.8), was used to inoculate 25 ml of 2xYT broth containing 20 µg ml^{-1}

chloramphenicol. The exponentially growing culture (OD_{600} approximately 0.2) was infected with helper phage VCSM13d3 (m.o.i = 50) for 1 h. Cells were then harvested by centrifugation at 3200×g for 10 minutes; the pellet was resuspended in 1 ml of 2xYT, mixed with 10 ml of soft agar [(2xYT broth containing 0.5% (w/v) agarose)] and poured over four Cm double-layer selective plates described in section 2.2.4.2. Both the soft agar and the plates contained molecular biology grade agarose instead of bacteriological agar. The plates were incubated overnight at 37ºC, then the phagemid particles were extracted from plates by adding 5 ml of 2xYT broth onto each plate followed by slow rotatory agitation at room temperature for 4 h. Extracted phagemid particles were precipitated by PEG/NaCl solution [5% (w/v) PEG, 0.5 M NaCl] and resuspended in TN buffer (10 mM Tris, 150 mM NaCl, pH 7.6). To eliminate unstable (defective) phagemid particles, the precipitate was treated with sarcosyl at a final concentration of 0.1% (w/v). The ssDNA released from defective phagemid particles was removed by DNase I (100 μg ml^{-1}) in the presence of 5 mM $MgCl_2$. DNase I was then inactivated by EDTA (20 mM). The resulting sarcosyl-resistant phagemid particles corresponded to the secretome phage display library clones. The ssDNA was first released from these sarcosyl-resistant phagemid particles by incubation at 70ºC for 10 minutes in the presence of 1.2% (w/v) SDS and then purified using a plasmid mini prep kit (Roche). To amplify the secretome library from the plasmid origin of replication, *E. coli* strain TG1 was transformed with purified ssDNA.

2.2.10 Sequence analysis of selected *L. rhamnosus* HN001 secretome library inserts

After transformation of ssDNA from sarcosyl-resistant phagemid particles (section 2.2.9), 491 clones were randomly selected for analysis. The phagemid DNA from these clones was purified using the 96-easy Mini-prep Kit (V-Gene Biotechnology, China). The inserts were sequenced using primer pDJ01R02 (Table 2.3) and BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA). Fragments were separated on the ABI3730 Genetic Analyzer (Applied Biosystems) at Allan Wilson Genome Services (Massey University). All inserts were sequenced using a primer complementary to *gIII* sequence of the vector at the 3′ end relative to the insert, since our interest was to investigate ORFs fused with *gIII*. Sequencing and sequence analysis was carried out on the batches of 96 clones. After the sequence analysis of the first 192 recombinant phagemids, those whose insert sequences were detected more than five times were excluded from further sequencing using dot blot

hybridisation as described in section 2.2.3.4. The sequences obtained were analyzed with the Vector NTI software Suite 10 (Invitrogen) and GLIMMER version 3.02, the latter analysis using a training set generated against the *L. rhamnosus* HN001 draft genome sequence, a position weight matrix representing ribosome binding sites for HN001 genes and an iterative approach, as described in the software documentation, to predict an ORFs [241]. If the 5′ end of the ORF was not reached, a sequencing reaction using the forward vector-complementary primer pDJ01F03 (Table 2.3) was carried out followed by primer walking if necessary.

SignalP 3.0 [242] and TMHMM 2.0 [233] were used for prediction of the signal sequence and transmembrane helices, respectively, using the default settings (for Gram-positive bacteria) and cut-off values [159, 204]. Amino-terminally located transmembrane helices that in SignalP 3.0 analysis showed a score for the signal peptidase cleavage site (C-score) below 0.52 were considered to be amino-terminal membrane anchors. The presence of a transmembrane helix in those cases was confirmed by using the TMHMM prediction program [243]. Lipoprotein signal sequences were predicted by the LipPred server [208, 244] using the default settings and cut-off values [245]. TATFIND 1.4 was used for prediction of Tat signal sequences [60]. Subcellular localization of the proteins in cells was predicted using PSORTb v.2.0 [207].

All translated insert sequences were examined with BlastP [246] at the NCBI website [81] with default settings to identify similarities with other bacterial proteins. An e-value lower then e^{-10} was used as a cut-off for notable similarity. Furthermore, conserved domains were identified in our query sequences by the Conserved Domain Architecture Retrieval Tool (CDART) engine in the course of the search; known domains being derived from either clusters of orthologous groups of proteins (COG) [247], Pfam [117], SMART (The Simple Modular Architecture Research Tool) and InterProScan databases [248, 249].

2.2.11 Construction of the *L. rhamnosus* HN001 shot-gun phage display library

The library was constructed from mechanically (nebulization) sheared *L. rhamnosus* HN001 DNA and cloned into the phagemid vector pYW01 (section 2.2.6.3). The fragments obtained varied in size between 0.3 and 4 Kbp, with majority between 0.8 and 2 Kbp. To eliminate

fragments below 0.5 Kbp, a protein concentrator with 100 kDa cut-off was used (Vivascience).

Approximately 50 μg of the genomic fragments were ligated to 14 μg of the vector pYW01 using T4 ligase (Roche). After phenol and chloroform extraction, the ligated DNA was ethanol-precipitated, washed with 70% (v/v) ethanol and dissolved in 70 μl H_2O. *E. coli* TG1electrocompetent cells [225] were transformed with 3 μl of ligation mix (2.5 kV, 25 μF, 400 Ω) in 2-mm-gap cuvettes. A total of 23 electro-transformations were carried out. The transformed cells were pooled, transferred to 230 ml of 2xYT and incubated for 1 h at 37ºC with rotatory agitation. After the incubation a 1 ml aliquot was taken to determine the number of transformants by plating on 2xYT agar with 20 μg ml^{-1} chloramphenicol. An aliquot (10 ml) of the remaining bacteria was transferred to 990 ml of 2xYT containing 20 μg ml^{-1} chloramphenicol and incubated at 37ºC with aeration until exponential phase (OD_{600} approximately 0.2). The exponentially growing culture was infected with wild-type helper phage VCSM13 (multiplicity of infection = 50) for 30 min. Cells were then harvested by centrifugation at 3200×g for 10 minutes and the pellet was resuspended in 1 L of 2xYT containing 20 μg ml^{-1} chloramphenicol. Infected cells were incubated for 4 h at 37ºC with aeration. The host cells were pelleted by centrifugation and phagemid particles collected in the supernatant. Phagemid particles were precipitated and purified as described in section 2.2.4.1.

2.2.12 Affinity screening of the shot-gun phage display library for *L. rhamnosus* HN001 cells binders

L. rhamnosus HN001-binding phagemid particles were selected from the *L. rhamnosus* HN001 shot-gun library by affinity-enrichment protocol similar to the one described for the *Rhizobium leguminosarum* cell-surface-associated agglutinin [226] with slight modification. Briefly, strain HN001 to be used as bait was cultivated in MRS medium supplemented with 1 mM $CaCl_2$ at 37ºC. Cells from 1.5 ml of the late-exponential-phase (~1 x 10^9 cells) were collected by centrifugation and the supernatant was carefully removed. The cell pellet was resuspended in 0.6 ml 1X TBS containing 2 mM $CaCl_2$ and mixed with the *L. rhamnosus* HN001 phage display library (~ 3 x 10^9 PPs). Binding was allowed to proceed for 3 h at room temperature on a roller. Unbound PPs were removed by washing the cells five times with 1.5 ml 1X TBS. After every wash the bacteria were gently spun down at 1100 x g for 3 min in a

microfuge. After the last wash, the cell pellet was resuspended in 0.05 ml 2xYT, transferred to a new tube and mixed with 0.5 ml of the exponentially growing *E. coli* TG1 culture to recover bound phage. This mixture was incubated for 30 min at 37ºC to allow phage to infect the *E. coli* cells. Subsequently, infected *E. coli* TG1 cells were plated on selective double layer chloramphenicol plates (section 2.2.4.2). Cm^R colonies were counted and collected by scraping the plates with 2 ml 2xYT. An aliquot (0.5 ml) was used to inoculate 200 ml of 2xYT containing Cm. When the culture reached exponential phase it was infected with wild type helper phage VCSM13 in order to produce PPs of the enriched library. After the helper phage infection the culture was processed as described (section 2.2.4.1) to extract, concentrate and purify PPs. These enriched PPs were then used in the second round of panning according to the same protocol. Affinity enrichment procedure was repeated one more time, a total of three rounds. After each round the total number of eluted PPs was determined and profiles of plasmid dsDNA isolated from amplified library pools were monitored to estimate enrichment frequency. Distinct plasmid bands after the third cycle of panning were purified from preparative agarose gel and used to transform the *E. coli* TG1 cells. Ten recombinants containing inserts of different sizes were further analysed by DNA sequencing.

2.2.12.1 Purification of phagemid particles displaying Lrh33 and HN001-binding assay

E. coli clones containing the recombinant phagemids pLrh33A, selected from the HN001 shot-gun library after panning on *L. rhamnosus* HN001 cells (K2089), and recombinant phagemid pLrh33 identified as a secretome clone in the HN001 secretome library (K2090) were used to generate phagemid particles displaying corresponding fragments of protein Lrh33. A volume of 100 ml of exponentially growing cultures of corresponding strains K2089 and K2090 were infected with helper phage VCSM13 at a multiplicity of infection of 50 phage per bacterium. The negative phagemid particle control was pYW01 PP/wt, generated by infection of strain K1978 (TG1//pYW01) as described above. Infected cells were incubated for 4 h at 37ºC with aeration. The *E. coli* cells were pelleted by centrifugation and phagemid particles were collected in the supernatant. The phagemid particles were precipitated with PEG/NaCl resuspended in 200 ul 1XTBS and quantified as described previously (section 2.2.4.2).

To determine specificity of binding, pLrh33A PP, pLrh33 PP and pYW01 PP were assayed for binding to HN001 cells as described in section 2.2.12.

2.2.13 Biopanning of *L. rhamnosus* HN001 shot-gun phage display library on immobilised fibronectin

For the first two rounds of biopanning, the microtiter wells (Ninclon multiwell #150628, Nunc) were coated with human plasma fibronectin at a final concentration of 100 µg ml^{-1}, 1 ml per well in 50 mM sodium carbonate (pH 9.5) overnight at 4ºC. The wells were washed once with 1 ml 1X PBS buffer containing 0.05% Tween 20 (PBST) and then blocked with 1% (w/v) BSA in 1X PBS for 2 h at room temperature. The wells were then washed (three times) with 1 ml of PBST buffer. Phagemid particles (3×10^{11}) in 100 µl of PBS supplemented with 0.1% BSA were added to the wells (phage library and BSA were mixed 1 h prior to addition to the wells). Negative phagemid particle control was pYW01 PPs generated using VCSM13 helper phage. The plates were incubated for 4 h at room temperature. The unbound phagemid particles were removed by washing with PBST (ten times). The bound phagemid particles were eluted in 1 ml elution buffer (100 mM glycine-HCl, 1 mg/ml BSA, pH 2.2) for 30 min followed by neutralization with 60 µl 1 M Tris (unbuffered). An aliquot (20 µl) was taken to determine the total number of bound phage as described before (section 2.2.4.2). The rest of elution was used to infect 1 ml of the exponentially growing *E. coli* TG1 cells. The infected cells were transferred to 8 ml Cm-containing medium and incubated overnight at 37ºC with aeration to amplify the eluted library pool. A 1 ml aliquot of the overnight culture containing the whole genome library was used to inoculate 100 ml of Cm-containing medium. The exponentially growing culture was infected with helper phage VCSM13 (m.o.i = 50) for 30 min. The PPs were amplified and purified as described before (section 2.2.4.2). They were used as input into the next round of panning.

The third round of biopanning was carried out on biotinylated human plasma fibronectin (Bio-Fn) immobilized on streptavidin magnetic beads (Roche). Fibronectin was biotinylated using EZ-Link Sulfo-NHS-LC Biotin (Pierce, USA) according to manufacturer-recommended procedure. To remove unspecific binders or "background" from the second round-selected phagemid particles (polypeptides displayed on the PPs binding to streptavidin magnetic beads independently of fibronectin), 0.5 mg of streptavidin magnetic beads were resuspended in 1ml blocking agent (2 % skim-milk powder in 1X PBS) and combined with ~ 10^{11} phagemid

particles. The mixture was incubated with end-over-end rotation for 1 h at room temperature. A tube containing streptavidin magnetic beads/ PPs mixture was placed in the magnetic field to separate bead-bound PPs from the unbound fraction of the library. A supernatant containing the unbound PPs was then transferred to a new tube, mixed with 175 pmol of Bio-Fn and incubated for 1h at room temperature on the roller. In the next step the interacting complex, Bio-Fn/PPs, was combined with 0.5 mg pre-washed and blocked streptavidin magnetic beads, (three washes with 1X PBS and blocking with 2 % skim-milk powder in PBS). The mixture was incubated 2 h at room temperature with end-over-end rotation. The tube was then placed in the magnet for 45 s to immobilise magnetic bead-associated phagemid particles and a series of seven 1 ml washes were performed with PBST followed by two washes 1X PBS supplemented with 2% skim-milk powder and a final wash with 1X PBS. Following the last wash, 0.5 ml elution buffer (0.1 M glycine-HCl, 1 mg/ml BSA, pH 2.2) was added to the interacting PPs/Bio-Fn/beads complex. After a 30 min elution, the beads were collected by magnet, the supernatant was transferred to a new tube and neutralized with 30 µl 1 M Tris (unbuffered). Output PPs were titrated and amplified as described in section 2.2.4.2.

After each round of panning the total number of eluted phage was determined by titration and plasmid profiles were monitored on an agarose gel. Distinct plasmid bands identified after the third round of panning were purified and used to transform the *E. coli* strain TG1. Ten recombinant phagemids containing inserts of different sizes were further analysed by DNA sequencing.

2.2.13.1 Preparation of phagemid particles displaying fibronectin-binding protein candidates and binding assay for fibronectin

E. coli clones containing phagemids Fn01 (K2087) and Fn02 (K2088), selected from HN001 shot-gun library after panning on fibronectin, were used to generate phagemid particles displaying corresponding proteins. A volume of 100 ml of exponentially growing cultures K2087 and K2088 were infected with helper phage VCSM13 at m.o.i of 50 phage per bacterium. Additional PP preparations for use as positive and negative controls, pSOF22 PP and pYW01 PP, were generated by infection of strains K1937 (TG1//pSOF22) and K1981 (TG1//pYW01) as described above. Infected cells were incubated for 4 h at 37ºC with aeration. The *E. coli* cells were pelleted by centrifugation and phagemid particles were

collected in the supernatant. The phagemid particles were purified and quantified as described in 2.2.4.2.

Binding of PPs derived from pure recombinant clones was examined by panning on Fn-coated wells of a microtiter plate. The microtiter wells (Nunc-Immuno MaxySorp™, Nunc) were coated with human plasma fibronectin at a final concentration of 100 µg ml^{-1}, 200 µl per well in 50 mM sodium carbonate (pH 9.5) overnight at 4ºC. Additional wells were coated with unspecific ligands, BSA (final concentration 1 mg ml^{-1} in 1X PBS) and streptavidin (final concentration 1 mg ml^{-1} in 1X PBS). The wells were washed once with 300 µl PBST and then blocked with 1% (w/v) BSA in 1X PBS for 2 h at room temperature. The wells were then washed (three times) with 300 µl of PBST buffer. Phagemid particles (3×10^{11}) in 200 µl of 1X PBS were added to the wells. The plates were incubated for 2 h at room temperature. The unbound phagemid particles were removed by washing with PBST (seven times). Bound phagemid particles were eluted with 200 µl of elution buffer (0.1 M glycine-HCl, 1 mg/ml BSA, pH 2.2) for 30 min followed by neutralisation with 12 µl of 1 M Tris (unbuffered). Output PPs were titrated as described previously (section 2.2.4.2).

Chapter III

Direct selection, expression and display of the secretome of probiotic bacterium *Lactobacillus rhamnosus* HN001

3.1 Construction of the secretome-selective phage display vector

A typical phage display system consists of two components: phagemid vector and a helper phage [26]. The phagemid vectors most commonly encode the carboxy-terminal domain of pIII, preceded by a signal sequence (Chapter IC). Inserts are placed between the signal sequence and mature portion of pIII. If an insert is translationally in-frame with both the signal sequence and the mature portion of pIII, then the encoded protein will be displayed on the surface of the phage. The first step in development of the secretome selection and display system was construction of a new phagemid vector, pDJ01, containing a pIII C-domain cloning cassette from which the signal sequence was deleted (Figure 3.1). Phagemid pDJ01 contains a phage infection-inducible promoter *psp* [250]. The *psp* promoter has very low level of background expression hence it decreases the chance of negative effect of potential toxicity of foreign protein fusions to pIII. Downstream of the *psp* promoter are Shine-Dalgarno site, multiple cloning cassette and coding sequence for the C-domain of pIII. To add the possibility for production and purification of soluble protein encoded by the insert, a purification tag (c-myc) and suppressible *amber* stop codon were engineered downstream of the cloning site. Thus soluble protein could be produced and affinity-purified in suppressor-positive strains. The stop codon is read as glutamic acid in the host strain TG1 (*supE*) used for the library construction and screening, allowing read-through into the in-frame *gIII*-coding sequence and display on the phagemid particles.

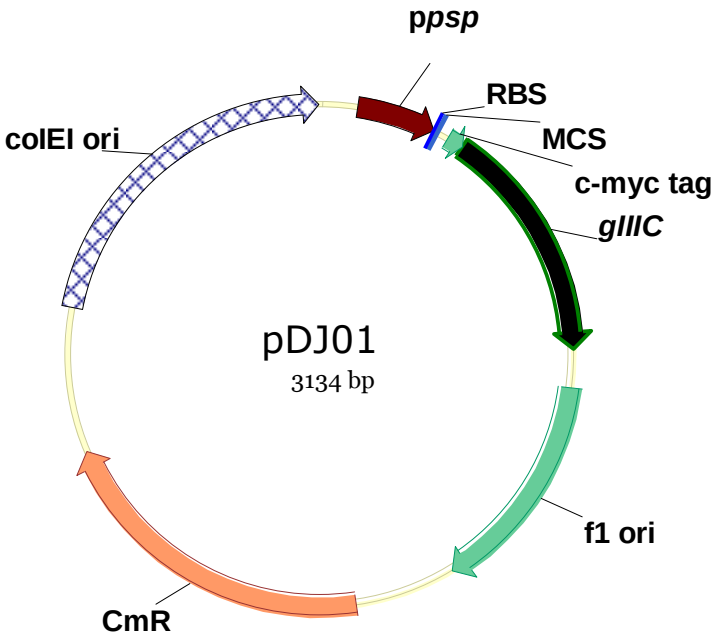


Figure 3.1. Phage display vector for selective secretome display.
gIIIC, carboxyterminal domain of *gIII*; **CmR**, chloramphenicol resistance cassette; **colEI ori**, the *colEI* plasmid origin of replication; **p*psp***, phage shock protein promoter; **MCS**, Multiple Cloning Site; **RBS**, ribosomal binding site; **c-myc**, a common peptide tag followed by a single *amber* stop codon; **f1 ori**, the f1 phage origin of replication for generation of ssDNA and for packaging into the phagemid particles.

The helper phage component of a phage display system is normally used to provide the f1 replication protein pII that mediates the rolling circle replication of the phagemid vector from the f1 origin, resulting in a single-stranded DNA (ssDNA) genome that is packaged into the virion [244]. The helper phage also provides other phage-encoded proteins essential for packaging of the phagemid ssDNA into the virion, to form phagemid particles. However, the helper phage that was used had the entire coding sequence for pIII (*gIII*) removed (Figure 3.2) [202]. Hence, the only pIII protein produced in the system was the phagemid vector-encoded pIII that lacked a signal sequence.

.

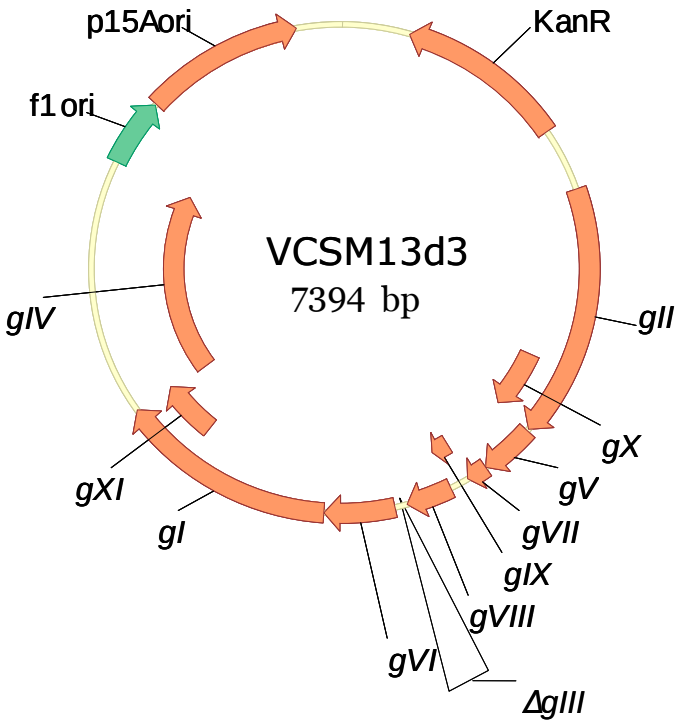


Figure 3.2. Helper phage VCSM13d3.
KanR, kanamycin resistance cassette; **p15A ori,** plasmid origin of replication; **f1 ori,** phage origin of replication; ***gI-XI,*** phage genes encoding: **pII** and **pX,** DNA replication; **pV,** ssDNA binding protein; **pVIII,** major coat protein; **pVI, pVII** and **pIX,** minor coat proteins; **pI** and **pXI,** assembly proteins in the cytoplasmic membrane*;* **pIV,** assembly protein/channel in the outer membrane; ***ΔgIII,*** entire *gIII* deleted.

3.2 Testing of the new phagemid pDJ01

3.2.1 Testing the pDJ01 vector for replication and packaging into phagemid particles

Packaging of a phagemid to obtain phagemid-encapsidating virions or phagemid particles is essential for phage display. Typically, a helper phage has a defective origin of replication or packaging signal, which allows the preferential packaging of the phagemid genome over the helper phage genome, resulting in domination of phagemid particles (PPs) over helper phage. The number of the phagemid particles usually exceeds the number of helper phage by a factor 10 to 100. Packaging of the phagemid vectors and helper phage can be impaired by

arrangement of the ORFs on the phagemid leading to low packaging efficiency of phagemid relative to the helper phage. Therefore, wild-type helper phage-assisted replication and packaging of the ssDNA of the phagemid pDJ01 into phagemid particles was tested. The amount of produced phagemid particles after infection with VCSM13 wild-type helper was determined by titration of infectious phagemid particles and the helper phage as described in section 2.2.4.1 (Table 3.1).

Table 3.1. Relative production of pDJ01 phagemid particles and helper phage.

[a]*PPs* *(cfu/ml)*	[b]*Helper phage* *(pfu/ml)*	*PPs : helper phage ratio*
[c]1.2×10^{11}	4.7×10^{9}	25:1

[a] PPs, phagemid particles containing pDJ01 ssDNA

[b] Helper phage, VCSM13

[c] PPs and helper phage were titrated as described in section 2.2.4.2

The result showed that the ratio of PP to phage titre was 25:1. Therefore, replication and packaging of the new phagemid pDJ01 was as efficient as expected.

3.2.2 The production of "background" wild-type pIII-containing (infectious) PPs in the absence of the signal sequence

In the pDJ01/VCM13d3 system for selective display, phagemid pDJ01 is the only source of the C-terminal domain of pIII, which is necessary for the termination of assembly. However, this truncated pIII lacks the N-terminal domain, which is required for infection. Hence, the PPs obtained using VCSM13d3 should be non-infectious. Any infectious particles obtained would be derived from the rare *gIII*+ helper phage, which may arise during preparation of the VCSM13d3 stock on a complementing strain that carries pIII-producing plasmid by VCSM13d3 - plasmid recombination [202]. Since infectious PPs would constitute a population of undesirable false positives in the secretome selection system, the frequency of their production and production of VCSM13d3 revertants was determined. Comparison of the titres on complementing (K1976) and non-complementing (TG1) hosts showed that the frequency of *gIII*+ phage in the VCSM13d3 stock was less than 8×10^{-6} (Table 3.2). The frequency of infectious pDJ01 phagemid particles generated using the above VCSM13d3 stock was 4.6×10^{-6} (calculated as a ratio of infectious particles to total amount of particles,

according to the method described in Chapter II, section 2.2.4.2. and footnote b of Table 3.2. These frequencies are much lower than the expected frequency of secretome clones in a library (10^{-3} to 10^{-2}), hence these false positives would represent an unsignificant minority in a secretome selection experiment. Therefore VCSM13d3 was deemed suitable for use as a helper phage for the secretome-selective phage display.

Table 3.2. The frequency of production of infectious VCSM13d3 and pDJ01 virions.

Helper phage/PPs	***Frequency of infectious virions***
VCSM13d3 phage[a]	**8×10^{-6}**
pDJ01 PPs[b]	**4.6×10^{-6}**

[a] The frequency of revertants in the stocks grown from single plaques of VCSM13d3 was calculated as a ratio of the titres on the non-complementing TG1 strain to complementing strain K1976. Titres were calculated from the number of the plaques on the lawns of the respective strains as described in 2.2.4.1.

[b] The frequency of the infectious PPs was calculated as a ratio of the number of infectious particles to total number of PPs. The number of the infectious particles was calculated from the Cm^R colonies after infection of the strain TG1. The total number of PPs was deduced from the intensity of the pDJ01 ssDNA band after agarose gel electrophoresis of disassembled virions as described in section 2.2.4.2.

3.2.3 Testing of the signal sequence-dependence of phagemid particle production by pDJ01

Testing of signal sequence-dependent release of stable PPs in the presence of the Δ*gIII* (VCSM13d3) helper phage was essential to establish whether the proposed basic tenet of the project, the selective cloning of secretome proteins, was feasible.

The hallmark of a signal sequence is a hydrophobic α-helix of at least 15 amino acid residues in length at the amino terminus of the protein. In bacteria, this helix is preceded by a few residues, predominantly positively charged, and is followed by either electroneutral or negatively charged residues (Chapter Iв; [159]). pIII has an 18-residue signal sequence, which

is normally processed by the *Escherichia coli* signal peptidase. However, Gram-positive signal sequences are significantly longer than those of Gram-negative bacteria [138]. Hence it was not clear whether the former would be processed with sufficient efficiency in *E. coli* to allow production of functional pIII. We tested this by inserting into pDJ01, in-frame with *gIIIC*, a surface protein from a Gram-positive bacterium - the serum opacity factor of *Streptococcus pyogenes*, M-type 22 (SOF22) [97]. The SOF22 portion of the protein fusion was 963 amino acid residues in length (including the signal sequence), and it lacked the sortase-dependent cell wall-anchoring motif LPATG and the TMH located at the very carboxyl terminus of the protein. Importantly, the signal sequence of SOF22 is 40 residues in length, approximately twice as long as that of pIII. Therefore, this is a typical Gram-positive bacterial secretome protein that might be found, for example, in the intestinal microflora. To establish whether production of SOF22 phagemid particles from its native signal peptide is of the same efficiency as when SOF22 was targeted to the membrane via a Gram negative signal peptide, the mature portion of SOF22 protein was also inserted into the "classical" phage display vector pAK100 carrying the Gram-negative signal sequence *pelB*, but otherwise containing the identical f1 and plasmid origins of replication and antibiotic resistance as in pDJ01 (section 2.1.3, Table 2.2).

Phagemid particles of the pDJ01::SOF22 (named pSOF22) and pAK100::SOF22 (named pAK-SOF22) clones were assembled using the Δ*gIII* helper phage VCSM13d3. It was expected that in the absence of signal sequence (vector pDJ01 without an insert) a low amount of unstable (defective) PPs would be produced, whereas inserting SOF22 carrying a signal sequence into pDJ01 (or fused to a Gram-negative signal peptide as in pAK-SOF22) was expected to result in release of a significantly larger amount of structurally stable virions. Because the obtained PPs are non-infectious, the efficiency of PP production was determined by quantifying the phagemid ssDNA in the particles after heating the concentrated phage stocks in SDS as described in 2.2.4.2.

Densitometry analysis showed that the amount of PPs in the absence of signal sequence was not significantly lower than that obtained in the presence of Gram-positive or Gram-negative signal sequence (Figure 3.3). This could be explained by the fact that in the absence of functional pIII, during a 4 h incubation, very long cell-associated phage filaments (expected to be >10 μm in length) break off the cells spontaneously [197]. These cell-associated phage or polyphage are made of multiple single stranded genomes (in our case pDJ01) whose assembly

has not been terminated by adding pIII as a result of absence of functional pIII (no signal sequence). Due to excessive length, these filaments are broken off the host cells by mechanical shearing during the handling and aeration of the cell culture by rotatory agitation. These defective virions had to be efficiently eliminated to achieve the selection for virions which display the signal sequence-containing proteins. The latter virions contain pIII fusion derived from the signal sequence-containing proteins which forms stabilizing the virion cap and therefore are more stable than pIII-deficient filaments.

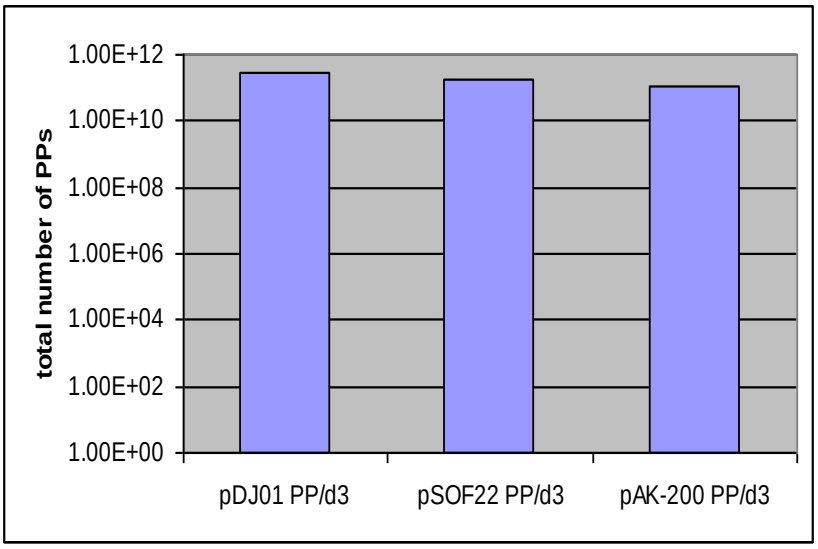


Figure 3.3. Total number of phagemid particles generated using VCSM13d3.

An opportunity for "chemical" selection of pIII-containing particles was provided by differential sensitivity to the ionic detergent sarcosyl. Defective phage particles, which lack pIII, are disassembled by sarcosyl. In contrast, pIII-containing particles are resistant to this detergent [198]. The ssDNA released from incomplete virions can then be eliminated by treatment with DNase I (Figure 3.4).

This protocol was tested using pSOF22 and pDJ01 phagemid particles. Sarcosyl disassembly and DNase I-mediated removal of pIII-negative PPs was monitored by agarose gel electrophoresis (Figure 3.4). After sarcosyl/DNase I treatment ssDNA from unstable phage particles was released and removed (Figure 3.4; lanes 1, 2, 4 and 5). The remaining structurally stable virions were disassembled using SDS (Figure 3.4; lanes 3 and 6).
Together these results show that the sarcosyl/DNase I protocol is functional and also that the

cap structure on pSOF22 PP was formed, implying that SOF22-pIII fusion was correctly targeted to the virion and that the Gram-positive signal sequence of the SOF22 protein was functional in the *E. coli* host.

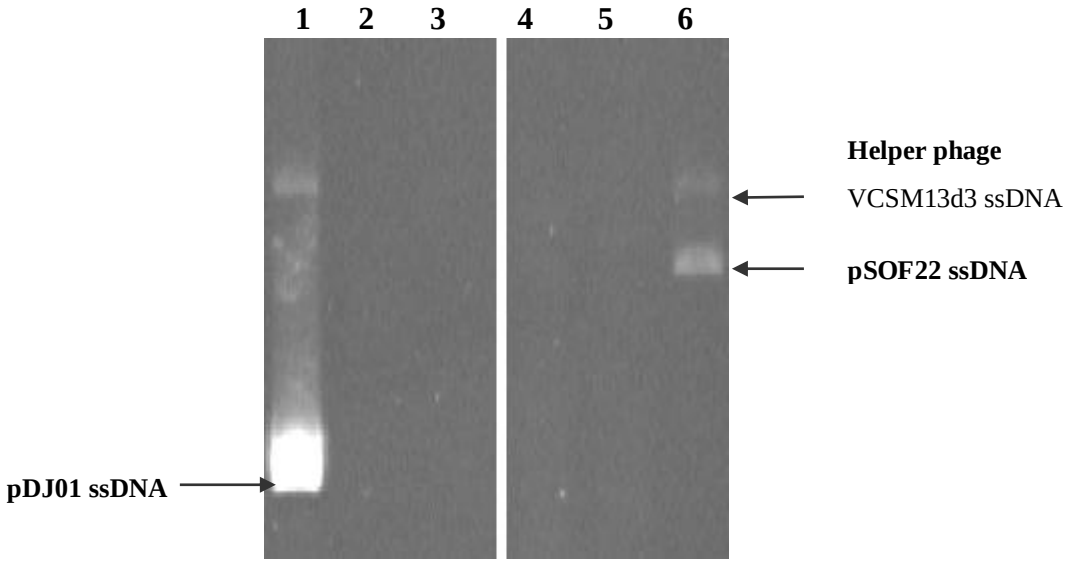


Figure 3.4. The effect of sarcosyl/DNaseI protocol on the ss- (pDJ01) and ss+ (pSOF22) PPs.

Agarose gel electrophoresis of free ssDNA. Lanes: 1, pDJ01 PP sample after sarcosyl treatment, showing ssDNA released from signal sequence-negative (defective) PPs; 2, sample 1 after the DNase I treatment (free ssDNA is removed); 3, purified ssDNA from the sample shown in 2; 4, pSOF22 PP sample after sarcosyl treatment; 5, sample 4 after DNase I treatment; 6, purified ssDNA from the sample shown in 5. Please note that only free ssDNA is detected in this experiment, whereas virion-encapsulated ssDNA remains unstained.

3.2.4 Biological activity of displayed protein SOF22

The ultimate aim of this thesis was to find and functionally characterise some of the proteins responsible for host-microbe interactions, thus the remaining question that needed to be addressed was whether a Gram-positive secretome protein displayed on the phage surface in pDJ01/VCSM13d3 system using its native sequence for membrane targeting and processing retained its biological activity.

Serum opacity factor is a bifunctional cell surface protein comprised of a C-terminal domain

with highly conserved tandem repeats that bind fibronectin [97] and an N-terminal domain (residues 148-843) that mediates opacification of mammalian sera through the binding to high density lipoprotein [251]. Purified pSOF22 PPs were examined for both biological activities. SOF22 was displayed using either the *gIII*-deleted helper phage VCSM13d3 as described previously (section 2.2.7), or *gIII*-positive helper phage, VCSM13. The former resulted in occupancy of all pIII positions in the phagemid particles with the SOF22-pIII fusions (polyvalent display), and the latter in a mixture of the SOF22-pIII fusion and the wild-type pIII (monovalent display) from the *gIII*+ helper phage VSCM13.

3.2.4.1 Serum opacity assay

Purified pSOF22 PPs were subjected to a serum opacity assay, the standard indicator for serum opacity factor (Figure 3.5) [252]. The assay was positive, demonstrating that SOF activity located at the N-terminal domain was retained. Consistent with the expected higher copy number of SOF22-pIII fusions when VCSM13d3 is used as the helper phage, serum opacity generation was greater for the phagemid particles produced by infection with the Δ*gIII* helper phage VCSM13d3.

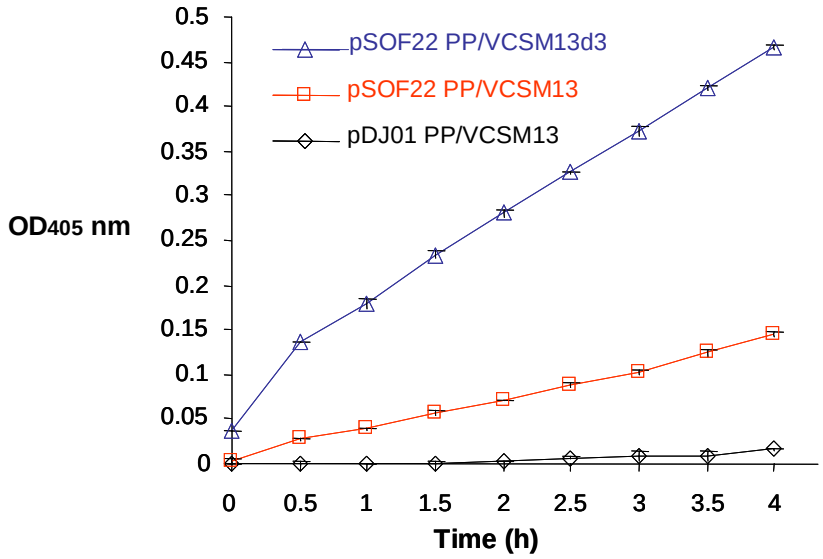


Figure 3.5. Serum opacity activity of the pSOF22 phagemid particles displaying SOF22.

Samples: pSOF22 PP/d3 and pSOF22 PP/wt, phagemid particles displaying the SOF from *S. pyogenes* M22, generated using VCSM13d3 and VCSM13 helper phage, respectively; pDJ01 PP/wt, the vector phagemid particles, generated using the VCSM13 helper phage. Each data point is an average of three replicas of the SOF assay; error bars represent standard deviation.

3.2.4.2 Fibronectin-binding activity assays

Fibronectin-binding activity of pSOF22 phage particles was assessed *in vitro* using phage ELISA and affinity enrichment assays.

Phage ELISA

Phage ELISA was performed as described in Material and Methods (section 2.2.7). The pSOF22 PP sample exhibited dose-dependent binding to immobilised fibronectin that began to plateau when the number of phage particles per assay reached ~10^8. No ELISA signal over background value was detected in control wells in which vector only, pDJ01, phage particles were applied, nor in the wells from which fibronectin was omitted (Figure 3.6). As expected, PP binding was lower for monovalent (pSOF22 PP/wt) relative to polyvalent (pSOF22 PP/d3) display similar to the observation for the serum opacity assay.

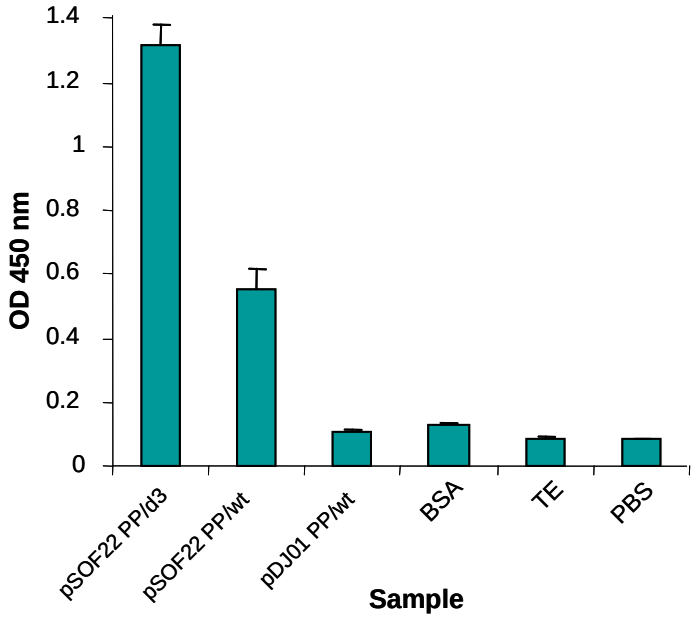


Figure 3.6. Binding of the SOF22-displaying phagemid particles to human fibronectin detected by phage ELISA.

A total of 10^8 phagemid particles were used per assay, each carried out in a well of a 96-well plate. Samples: pSOF22 PP/d3 and pSOF22 PP/wt, phagemid particles displaying the SOF from *S. pyogenes* M22, generated using VCSM13d3 and VCSM13 helper phage, respectively; pDJ01 PP/wt, the vector derived phagemid particles, generated using the VCSM13 helper phage. BSA, TE and PBS, buffer controls. Each data point is an average of three replicas; error bars represent standard deviation.

Affinity binding (Biopanning)

A big advantage of phage display is the convenience with which libraries can be affinity-screened for target-specific binders. Affinity enrichment of library clones encoding binding polypeptides is achieved by incubation of the phage library with an immobilised target, followed by removal of the non-binding phage and amplification of binders to the host bacteria. To establish the conditions for screening of the secretome library clones for binders to various targets involved in adhesion and aggregation and to confirm earlier experimental results that biological function of displayed SOF22 was retained, binding of SOF22 PP/wt to immobilised fibronectin was assayed. Recovery of the pSOF22 phagemid particles were

compared to that of a parallel binding assay of pDJ01 PP/wt sample. pSOF22 PP/wt showed 54000-fold higher recovery relative to that of pDJ01 PP (Table 3.3). This level of enrichment of SOF relative to the non-binder is consistent with the high affinity binding of FnBP domain to Fn. This also implicated that conditions used in biopanning assay could be used successfully for screening of the secretome library for proteins responsible for interaction with Fn.

Table 3.3. Total number of eluted PPs from fibronectin after affinity enrichment assay.

Phagemid particles	*Input titre*[a]	*Output titre*	*Output/Input titre ratio*	*Binding above background*
pSOF22	$2x10^{10}$	$1.3x10^{9}$	$6.5x10^{-2}$	$5.4x10^{4}$
pDJ01	$2x10^{10}$	$2.4x10^{4}$	$1.2x10^{-6}$	

[a] Phagemid particles were titrated as described in section 2.2.4.1

In summary, preliminary experiments suggested that this system could be used for a Gram-positive bacterium at a genome scale.

3.3 Selection of the *Lactobacillus rhamnosus* HN001 secretome

Efficiency of selection of a secretome phage display library was examined at genomic scale. The organism of choice was the Gram-positive probiotic bacterium *L. rhamnosus* strain HN001. A draft genome sequence of this bacterium has been determined by Fonterra; however, the complete genome sequence has not been determined as yet. In this thesis work, a small-insert shot-gun genomic library of HN001 was created in the pDJ01 vector (Figure 3.7). The insert size ranged from 0.3 to 4 Kbp and the primary size of the library was 10^6 clones. The library was first amplified using the plasmid origin of replication (in the absence of a helper phage). In the next step, the amplified library was mass-infected with the Δ*gIII* helper phage VCSM13d3 to initiate replication of the phagemid from the f1 origin and packaging into phagemid particles. Based on the preliminary experiment described in the previous section (3.2.3), inserts encoding the signal sequence-containing proteins in-frame with pIII were expected to restore its function and allow assembly of the terminal cap of the virions, rendering them resistant to sarcosyl. These resistant phagemid particles were expected to display the pIII-secretome protein fusions on the surface and contain the corresponding DNA

sequence inside the phagemid particle. In contrast, defective phagemid particles that lack an insert encoding a signal sequence-containing protein that is translationally fused to gIII were expected to be disassembled in the presence of sarcosyl. Thus, sarcosyl treatment would release the recombinant phagemid ssDNA encapsidated in the defective phagemid particles; the released DNA would then be digested by DNase I and eliminated in the selection step.

After infection with VCSM13d3 helper phage, the library was incubated on a solid medium to minimize growth competition among the library clones. Phagemid particles released from the infected library were collected and purified by PEG precipitation. Sarcosyl-induced release of phagemid DNA was monitored by agarose gel electrophoresis and staining with ethidium bromide (Figure 3.8A, compare lanes 1 and 2). The sarcosyl-released ssDNA was eliminated by DNase I (Figure 3.8A, lane 3). The total DNA in the virions (both encapsulated and free) was detected by disassembling all virions, both defective and pIII-containing, with SDS at 70°C, prior to electrophoresis. The electrophoresis of SDS-disassembled virions detected a weak signal in the post-DNase treatment samples compared to the signal from the sarcosyl-sensitive phagemid particles. This indicated that, as expected, the majority of the inserts were packaged into sarcosyl-sensitive phagemid particles, most likely because they lacked in-frame signal sequence fusions to the vector pIII.

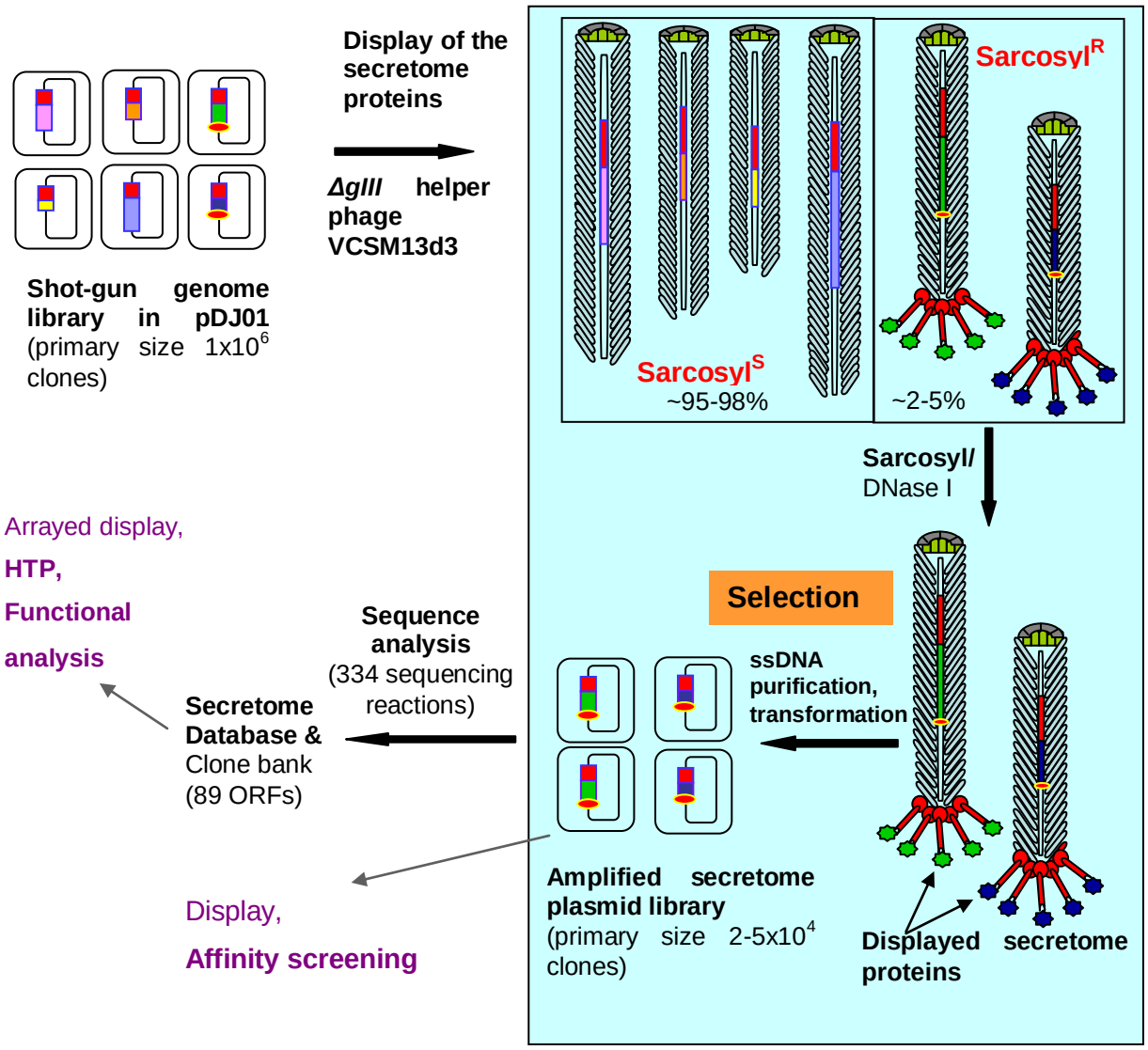


Figure 3.7. The secretome selection diagram
The key selection steps are boxed. Rounded squares represent *E. coli* cells and rounded rectangles represent recombinant phagemids replicating as plasmids inside the cells. pIII is shown as a red rectangle on the plasmid backbone. Inserts are represented as rectangles of various colors and lengths. Small orange ovals represent the signal sequences. The pipe-cleaner-like shapes represent phagemid particles obtained after infection of the library with the helper phage VCSM13d3. The elongated rectangles along the axes represent packaged DNA of the library clones. The top ends of the phagemid particles contain pVII and pIX proteins. The bottom ends of the phagemid particles are either open (signal sequence-negative clones) or capped by protein-pIII fusions (signal sequence-positive clones; popsicle shapes). Sarcosyl^S, phagemid particles sensitive to sarcosyl; Sarcosyl^R, the secretome protein-displaying phagemid particles, resistant to sarcosyl. Numbers in brackets refer to data obtained in the *L. rhamnosus* HN001 secretome selection experiment in this work. Steps denoted in purple indicate downstream applications of the secretome library.HTP, high-throughput screening

A minority of inserts was packaged into sarcosyl-resistant virions and, therefore, probably contained in-frame signal sequence fusions with the vector pIII (Figure 3.8B, lane 3). Densitometry analysis indicated that approximately 2-5% of the total phagemid particles were sarcosyl-resistant. This matches the expected frequency of 3.3% or 1/30 [~1/5 (frequency of secretome- encoding ORFs) × 1/2 (probability of correct insert orientation) × 1/3 (probability of the correct frame fusion of the inserts to pIII)].

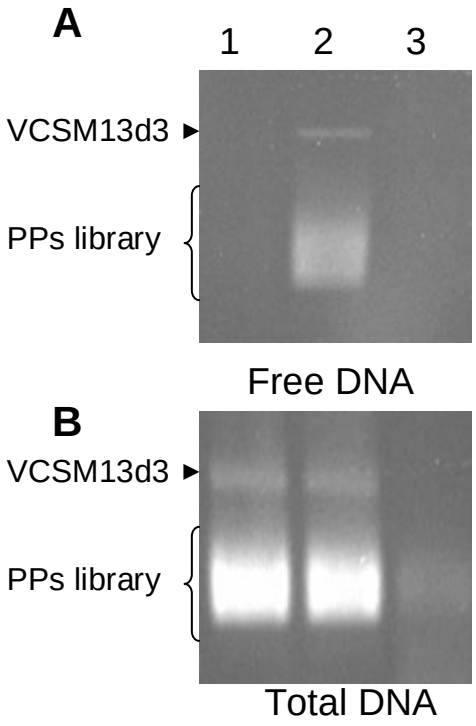


Figure 3.8. Demonstration of the sarcosyl resistance selection step.

(A) Free phagemid DNA (samples were loaded directly on a 0.8 % agarose gel);

(B) Total DNA, the sum of the free DNA and DNA encapsulated in the phagemid particles (samples were heated at 70°C in 1.2 % SDS for 10 minutes before loading, to disassemble the sarcosyl-resistant phagemid particles).

Lanes: 1, library phagemid particles (PPs) before incubation with sarcosyl; 2, after incubation with sarcosyl; 3, after incubation with sarcosyl and DNase I, followed by inactivation of DNase I.

3.4 Efficiency of the secretome library selection

DNA from the sarcosyl-resistant phagemid particles, expected to contain the secretome protein inserts, was purified and transformed into a new culture of *E. coli* TG1 host to amplify the secretome library clones from the plasmid origin of replication. The resulting double-stranded recombinant phagemid DNA was purified from individual colonies and the library inserts were sequenced.

Initially 192 inserts were sequenced and a few "promiscuous" recombinant phagemids that appeared in more than 5 independent transformants were identified. To avoid repeated sequencing of these inserts, a mixture of probes derived from them was used to screen a further 299 transformants by dot-blot hybridisation (Figure 3.9). All plasmids that gave a signal were excluded from sequence analysis. This revealed 157 recombinant phagemids containing promiscuous inserts and 142 non-promiscuous phagemids that were analyzed by sequencing. In total, 491 library inserts were characterized: 334 by sequencing and 157 by hybridisation only. For the inserts that were sequenced, one sequencing reaction was done using a reverse primer complementary to the *gIII* sequence of the vector. If the 5' end of the secretome ORF was not reached, an additional sequencing reaction was done using the forward primer complementary to the vector sequence upstream of the insert, followed by internal primer sequencing reactions if a contig could not be assembled from the forward and reverse reaction.

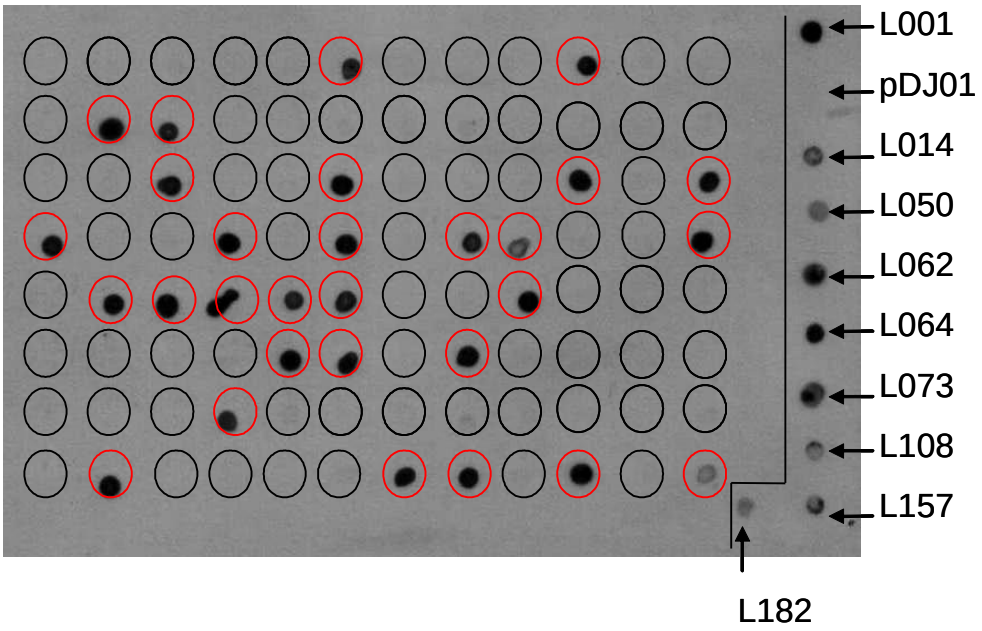


Figure 3.9. Subtraction of promiscuous clones by dot blot hybridisation

Dot-blot hybridisation of recombinant phagemids number L200 to L296. Section of the membrane separated by a line corresponds to the recombinant phagemids from which the probes were derived and empty vector pDJ01 (negative control). Another 288 recombinant phagemid clones were analysed in the same manner. Clones circled with red were excluded from sequencing.

The insert sequences whose translated products in-frame with pIII were longer than 24 residues were analyzed by SignalP 3.0, TMHMM 2.0 and Lip-Pred to predict whether they contained any membrane-targeting signals. This revealed that 411 (84 %) of the 491 inserts analyzed (sequenced or screened by dot-blot hybridisation) contained 87 distinct ORFs predicted to encode secretome proteins in-frame with pIII. Of the remaining 80 non-secretome inserts, 52 contained inserts encoding very short peptides in-frame with pIII (< 24 residues), 12 were empty vector and the remaining 16 inserts encoded peptides longer than 24 residues in-frame with pIII, but these peptides lacked typical membrane-targeting sequences. When infected with Δ*gIII* helper phage VCSM13d3, 14 of these 16 recombinant phagemids failed to assemble sarcosyl-resistant phagemid particles. However, the remaining two recombinant phagemids with no detectable in-frame membrane targeting signals were still able to generate sarcosyl-resistant phagemid particles, and predicted ORF-pIII fusions (Figure 3.10; lanes 6 and 12). This strongly suggests that the two inserts contained concealed or perhaps Sec-independent sequences that allowed proper targeting of pIII in the inner membrane of *E. coli*.

These two inserts contained ORFs encoding putative folding enzyme disulfide isomerase (*lrh88*) and a Cof-like hydrolase (*lrh89*). The subcellular location of homologues of these two enzymes has been reported as in either the periplasm or the cytoplasm. However, the two ORFs that have been selected did not encode the signal sequences normally present in the family members that are targeted to the membrane. Hence, the mechanism of the targeting of these two fusions remains unresolved and could potentially involve a conserved Sec/Tat-independent mechanism.

In summary, most of the non-secretome clones (50 out of 52) were most likely obtained due to the incomplete digestion of released ssDNA by DNase I in the selection step, rather than mistargeting of the pIII fusions.

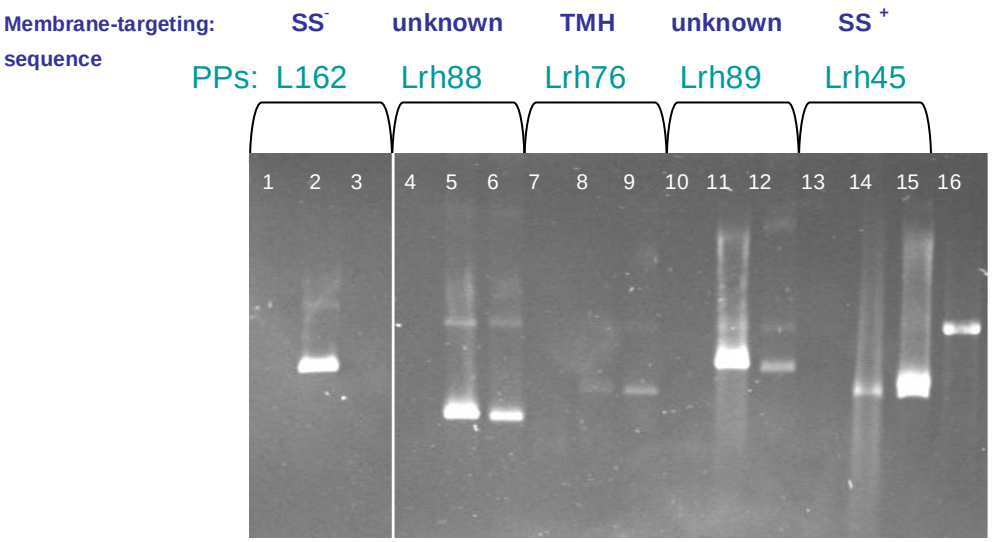


Figure 3.10. Packaging of clones carrying various membrane-targeting sequences into phagemid particles.

Please note that free ssDNA only was detected in this figure**.** Secretome clones containing various types of membrane-targeting sequences were subjected to sarcosyl/DNase I treatment to test packaging into PPs. PPs generated from secretome clone L162, containing 237 residues fused to pIIIC but lacking any detectable signal sequence, were used as a negative control. PPs from Lrh45 secretome clone containing a typical type I signal peptide were used as a positive control. Lrh76 secretome clone encoding for a polytopic (multiple TMH) membrane protein was used as an example of processing of a proximal TMH by *E. coli* secretion machinery.

SS-, no signal sequence; unknown, non-recognisable membrane targeting signal; TMH, transmembrane helix; SS+, type I signal peptide.

Lanes 1,4,7,10,13, untreated PPs; lanes 2,5,8,11,14, sarcosyl-treated samples from lanes 1,4,7,10,13 showing ssDNA released from sarcosyl-resistant PPs; lanes 3,6,9,12,16, DNase I-treated ssDNA purified from sarcosyl-resistant PPs of samples from 2,5,8,11,14. The ssDNA of sarcosyl-sensitive virions was eliminated using sarcosyl/DNase I/EDTA protocol which was applied for secretome selection (section 3.3); lane 16, VCSM13d3 ssDNA.

3.5 Membrane-targeting sequences and distribution

Of the 87 ORFs that encoded proteins with predicted membrane-targeting sequences, 45 contained a type I signal sequence (Table 3.4; see Appendix II, Table A.2 for complete list of targeting sequences and the secretome ORF annotation).

Table 3.4. Types of *L. rhamnosus* HN001 membrane-targeting sequences and distribution.

Membrane-targeting signal in the insert	***Bitopic or extracellular proteins***	***Polytopic integral membrane proteins***	***Total***
Type I signal sequence	45[a]		45
Lipoprotein signal sequence	13		13
Amino-terminal transmembrane helix	13	6	19
Internal transmembrane helix	2	1	3
Multiple transmembrane helices		7	7
No recognisable membrane-targeting signals	2		2
Total	75	14	89

[a] Numbers refer to the number of secretome ORFs predicted to contain a particular type of membrane-targeting signal.

Thirteen ORFs encoded proteins with a predicted lipoprotein signal sequence and 19 with a predicted amino-terminal membrane anchor. Ten ORFs encoded proteins with predicted internal transmembrane α-helices; of those, three have a predicted single transmembrane α-helix and seven have predicted multiple transmembrane α-helices. Notably, 44 out of 89 putative membrane-targeting sequences that have been selected by our method are not type I signal sequences. Given that the type I pIII signal sequence must be cleaved off by the *E. coli* signal peptidase in order to release its amino terminus from the membrane, the non-type I membrane-targeting sequences fused to pIII appear to have been successfully processed in the *E. coli* periplasm, either by the signal peptidase or by some other membrane or periplasmic protease. In the case of the polytopic proteins it is most likely that the first TMH served as a membrane-targeting signal and that the rest of the polypeptide containing additional

membrane spanning domains was cleaved away from the pIIIC portion of the fusion. This was confirmed by comparison between the sizes of the expressed polytopic protein Lrh76-pIII fusion in the total cell to that of the same fusion protein displayed on the surface of the phagemid particles, using Western blotting with anti-pIII antibody (data not shown).

No inserts containing predicted Tat signal sequences were identified by the available software or manual inspection. This is consistent with other *Lactobacillus* species, none of which contain the Tat translocon.

The enrichment of the secretome insert-containing recombinant phagemids was approximately 210-fold (from approximately 1:40 to 5.26:1), suggesting that the stringency of selection was high and that most recombinant phagemids containing non-secretome inserts were eliminated. Of the 89 secretome ORFs identified, over half (49) were present multiple times (between 2 and 5) as distinct recombinant phagemids with different points of fusion to pIII. Analysis of DNA sequence contigs, obtained by assembly of individual sequence reads, indicated that some of these ORFs were organized into operons encoding secretome proteins. For example, one contig encoded two secretome ORFs (*lrh31* and *lrh30*) that were located adjacent to each other within a larger operon (Figure 3.11). A clone bank and a database of the *L. rhamnosus* HN001 secretome clones were generated from the sequence data and were used for bioinformatic characterisation of the secretome.

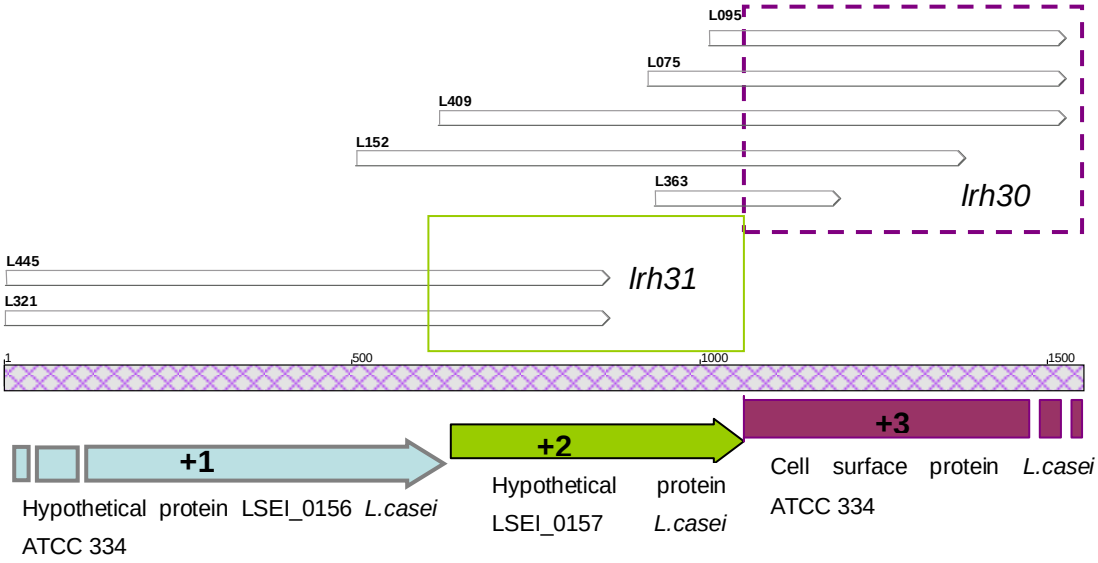


Figure 3.11. Contig corresponding to ORFs in a secretome protein operon.
Top, white arrow-shaped boxes represent individual sequence reads, each from a different transformant. Middle, grey cross-hatched box is the assembled contig. Bottom, predicted ORFs with indicated frame and annotation are represented by the coloured arrows. The first and the third ORFs are partial. The first ORF was not assigned an *lrh* number because it was not directly selected in our screen as a secretome protein.

3.6 Annotation of *Lactobacillus rhamnosus* HN001 secretome proteins

Of the 89 identified ORFs, functions were predicted for 48, comprising 7 functional categories (Table 3.5; see Appendix II, Table A.2 for detailed annotation and the GenBank accession numbers of the secretome ORFs). The largest functional category comprised 22 ORFs encoding putative transport proteins, with 13 of these having similarity to extracellular substrate-binding domains of ABC transporters and each containing a predicted amino-terminal lipoprotein signal sequence. The remaining nine ORFs in the transport protein category were predicted to encode polytopic transmembrane proteins, with one or more internal transmembrane α-helices. ORFs encoding predicted enzymes were the second-largest category. This diverse class included predicted proteases, hydrolases, enzymes involved in the cell wall turnover, autolysins and a dithiol-disulfide isomerase (Table 3.5). One ORF, *lrh15*, had similarity to a sensor histidine protein kinase of *Lactobacillus casei* for which the signal/substrate specificity has not yet been determined. Several ORFs had significant

sequence similarity with known surface proteins. For example, ORF *lrh51* encodes a predicted protein that is similar to a predicted LPxTG-anchored adhesin exoprotein from *L. casei* ATCC 334. The protein family to which Lrh51 belongs appears to be unique to the *L. casei-Pediococcus* group and may play a role in adaptation to the common environment(s) of these two groups. Another ORF, *lrh35*, encodes a predicted protein homologous to a putative collagen adhesin of *Bacillus clausii* KSM-K16. One ORF, *lrh17*, encodes a predicted protein containing a pilin motif and partial E-box motif, which are motifs present in the major pilin proteins of Gram-positive bacteria. Analysis of the putative full-length *lrh17* ORF identified in the draft genome sequence of *L. rhamnosus* HN001 revealed the complete E-box and the cell wall sorting signal; therefore, *lrh17* is likely to encode the major pilin protein of putative *L. rhamnosus* HN001 pili. One of the ORFs, *lrh08*, had sequence similarity to conserved hypothetical proteins that are similar to cell wall-anchored proteins, but appeared to be truncated due to a TAG stop codon. This ORF was translated through the TAG stop codon and displayed as pIII fusion because the *E. coli* host strain that we have used contains a *supE* mutation that reads the TAG stop codon as glutamic acid.

Database searches did not reveal any sequences similar to seven of the ORFs. Putative proteins encoded by these ORFs seem to be unique to *L. rhamnosus* HN001 and, therefore, might potentially be involved in strain-specific interactions between this bacterium and its environment that might be associated with its probiotic effects. One of these ORFs, *lrh62*, encodes a putative serine- and alanine-rich extracellular protein. The insert in the recombinant phagemid encodes 807 residues, but the protein encoded by this ORF is predicted to be 3275 amino acids in length and to contain an LPXTG carboxy-terminal cell wall anchoring motif (as deduced from the draft *L. rhamnosus* HN001 genome sequence). The presence of many alanines (965/3275) and serines (496/3275) and the overall protein size is reminiscent of large serine-rich repeat-containing adhesins of Lactobacilli and Streptococci. However, these adhesins typically contain hundreds of copies of a short and highly conserved serine/alanine-rich motif, whereas the alanine and serine residues of ORF *lrh62*, although highly repetitive throughout the protein due to their large numbers, do not appear to form conserved and regularly repeating motifs that could be revealed by self-alignment matrix analysis.

Table 3.5. Categorisation of *L. rhamnosus* HN001 secretome ORFs.

Category	Number of ORFs
Enzymes	20
Signal transduction components	1
Transport	22
Host/microbial interactions	3
Conserved, unknown function	28
Unique hypothetical proteins	7
Miscellaneous functions	2
Unclassified [a]	6
Total	89

[a] Short fragments of ORFs (encoding 27-57 amino acids) were fused to *gIII*; no hits above the threshold (e^{-10}) were detected using BlastP with automatic detection of short sequences. These ORFs were not classified as unique because the short length has prevented identifying potentially significant hits.

Summary of sections 3.3-3.6

The new secretome selection system was tested at a genomic scale, on the probiotic Gram-positive bacterium *Lactobacillus rhamnosus* HN001. The primary library size before selection was 10^6 inserts. After a single round of selection for the secretome protein-encoding recombinant clones, 491 inserts were analyzed, 84% of which were found to carry a secretome open reading frame. Of the 89 distinct ORFs identified, 87 had membrane targeting signals of various types (type I and type II signal peptides, N-terminal anchors and polytopic TMH) in frame with pIII. However, two library inserts encoding ORFs without any detectable membrane-targeting sequences in frame with pIII were still able to generate the sarcosyl-resistant phagemid particles, suggesting that the two inserts encoded for concealed or perhaps Sec-independent targeting sequences that allowed assembly of pIII into the virion.

For termination of phage assembly, the pIII type I signal sequence must be cleaved off by the *E. coli* signal peptidase, to allow the release of pIII from the membrane. In the HN001 secretome library seven ORFs encoded proteins with predicted multiple transmembrane α-helices which were exported to the membrane via proximal transmembrane segment.

However, in the *E. coli* periplasm the polypeptide was cleaved away from the pIII portion of the fusion, very likely by periplasmic proteases of *E. coli*.

Seven functional categories of secretome proteins were predicted: enzymes, transport proteins, conserved proteins of unknown function, signal transduction components, host/microbial interactions, miscellaneous functions, unique hypothetical proteins and unclassified (due to short fragments of ORFs present in the recombinant phagemids). Seven hypothetical proteins unique to *L. rhamnosus* are of special interest because they might potentially be involved in strain-specific interactions between this bacterium and its environment that might be associated with its probiotic effects.

Chapter IV

Affinity screening of *Lactobacillus rhamnosus* HN001 shot-gun phage display library for host-microbe and microbe-microbe interaction-mediating proteins

Adhesion to intestinal cells and exclusion or reduction of pathogen adhesion via aggregation are very important traits of probiotic bacteria. In animal models and *in vitro* assays using human epithelial cell lines, *Lactobacillus rhamnosus* HN001 demonstrated ability to adhere to the cells in culture. In the *in vivo* experiments HN001 protected mice against an oral challenge of *Salmonella* and *E. coli* O157:H7 [56, 60]. The mechanistic basis for these effects of HN001, and of probiotics in general, is under investigation. HN001 aggregates in laboratory culture, however the molecular basis of cell-cell interactions leading to aggregation has not been revealed as yet. Thus the second aim of this project was to screen *L. rhamnosus* HN001 phage display library for the proteins potentially involved in adhesion or aggregation processes. For that purpose a new phagemid, pYW01, was designed and a large shot-gun (random) phage display library (primary size 10^8 clones) was constructed and screened against appropriate targets.

4.1 Construction of pYW01 phagemid and a *L. rhamnosus* HN001 shot-gun phage display library

The secretome phage display system described in Chapter III is a powerful tool for cloning and identification of secreted, surface and membrane proteins. However, the majority of the clones have only the N-terminal domain displayed on the surface of the phage. According to the literature, proteins involved in host microbial interactions, especially adhesion, often bind to the receptors via C-terminal domain [97, 226]. Furthermore, there are a few examples of "moonlighting" proteins in Gram-positive bacteria that do not contain any membrane targeting Sec-dependent sequences, but are nevertheless secreted into the environment. To be able to display these types of adhesion proteins or domains, a HN001 phage display library was constructed in which proteins were displayed irrespective of their intrinsic targeting signals.

Initially, phagemid pYW01 was constructed, which contained a PelB signal peptide fused to the carboxyterminal domain of the phage pIII (Figure 4.1), a typical or "classical" phage display cassette. The rest of the phagemid features were based on the pDJ01 backbone, with exception of the pDJ01 plasmid origin of replication, colEI, which was replaced with the colD *ori* from pGZ119EH (section 2.2.6.3) [235]. This change was introduced to enable use of the pYW01 phagemid in development of a new "oligovalent" (two or more copies of protein displayed on the surface of infectious phagemid particles) phage display system by supplying

full length pIII from a compatible (colEI ori) plasmid (data not shown). A feature different from most phage display vectors constructed by other research groups was the insertion of *psp* promoter, which is induced by phage infection, and therefore requires no inducer.

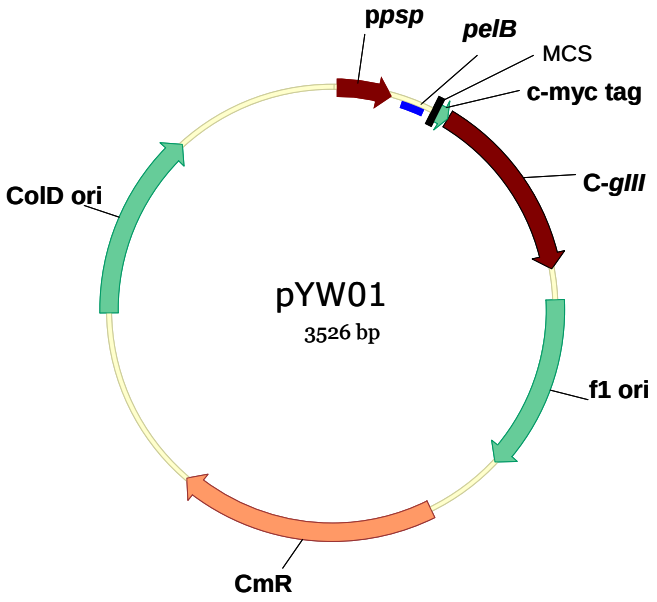


Figure 4.1. Phagemid vector used in construction of *L. rhamnosus* HN001 shot-gun phage display library

gIIIC, C-domain of *gIII*; **CmR**, chloramphenicol resistance cassette; **colD ori**, the colD plasmid origin of replication; ***ppsp***, phage shock protein promoter; **MCS**, multiple cloning cassette; **c-myc**, a common peptide tag followed by a single amber stop codon; **f1 ori**, the f1 phage origin of replication for generation of ssDNA for packaging into the phagemid particles.

A shot-gun phage display library of HN001 was constructed as described in Material and Methods (section 2.2.11). The inserts ranging in size from 0.3 to 4 Kbp were cloned into the pYW01 phagemid. The primary size of the obtained library was 3 x 10^8 clones. The library was first amplified using the plasmid origin of replication (in the absence of a helper phage). In the next step, the amplified library was mass-infected with the $gIII^+$ helper phage VCSM13 to initiate replication from the phagemid f1 origin and packaging into phagemid particles. These phagemid particles were expected to display the whole HN001 proteome and contain

corresponding DNA sequences inside of the virions. Phagemid particles released from the infected library were collected, purified by PEG precipitation and the phagemid particles were titrated as described in section 2.2.4.1. This master library titre was 2.2 x 10^{13} PPs/ml; the library was aliquoted and stored in 7 % DMSO at 80ºC for later use in affinity screening procedures.

4.2 Affinity screening of the *L. rhamnosus* HN001 shot-gun library for cell-cell interacting proteins

Formation of multicellular aggregates through cell-cell adherence is often observed in probiotic bacteria. This phenomenon often correlates with a stronger ability of probiotic strains to adhere and colonise the intestine, and to produce immunomodulatory effects *in vitro* and *in vivo* [74, 78, 253]. The aggregates could be formed between bacteria from the same strain/species (autoaggregation) or between different bacterial species (coaggregation). Under the laboratory growth conditions, *L. rhamnosus* HN001 autoaggregates, therefore surfaces of HN001 cells do interact. It is possible that aggregation is mediated by surface protein(s) and if so, they could be identified by phage display. Therefore the shot-gun HN001 library was affinity screened using the HN001 cells as bait.

For the enrichment procedure, strain HN001 was cultivated to late exponential phase in MRS medium supplemented with 1 mM $CaCl_2$. The Ca^{2+} ions were added based on the evidence that these ions are essential in binding of some proteins to the cell surface via lectin domains [101, 102, 254]. The HN001 cells were harvested, washed and used as bait to enrich for self-adhering phagemid clones from the library. Approximately 3 x 10^9 phagemid particles of the HN001 shot-gun library were used in each round of the affinity enrichment. The number of phagemid particles in eluate after the first round of panning and two consequent cycles of panning is presented in Table 4.1. After the first panning, the number of recovered phagemids was 8.3 x 10^3. In the third cycle of panning 7.6 x 10^6 phagemids were eluted. This 916x increase in the number of eluted phage particles after second and third round of panning compared to eluate after the first panning indicated that specific enrichment had taken place.

Table 4.1. Number of phagemid particles isolated in the screening of the phage display library on *L. rhamnosus* HN001 cells.

1st panning [a]	2nd panning [a]	3rd panning [a]	Enrichment [b]
8.3×10^3	1.3×10^5	7.6×10^6	916 x

[a] The total number of phagemid particles in the eluate (0.1 mL) after affinity selection was determined by titration (section 2.2.4.1).

[b] Enrichment was calculated as the ratio of number of eluted particles in the 3rd round of panning relative to the 1st round of panning. The starting number of particles (input) in each round of panning was 10^9.

Plasmid profiles of enriched library pools after each round of panning were monitored by agarose gel electrophoresis as another indicator of enrichment of selected clones (Figure 4.2). Disappearance of a smear of recombinant plasmids due to random insert size distribution and appearance of discrete bands indicated enrichment of particular recombinant phagemids relative to the rest of the library. The most prominent three bands seen on the agarose gel after third panning were each excised, purified and used to separately transform *E. coli* host strain TG1.

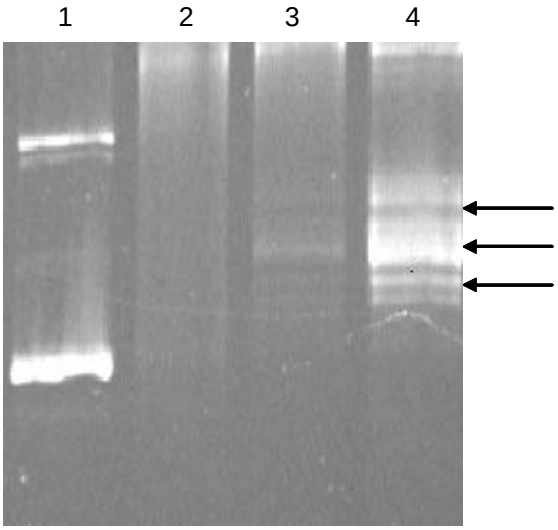


Figure 4.2. Phagemid profiles of the library over three rounds of affinity panning on HN001 cells as bait.

Recombinant phagemid pool replicating from plasmid *ori* was purified from aliquots of overnight amplified HN001 library after each round of panning. Phagemid DNA bands were visualised using EtBr after agarose gel electrophoresis. Lanes: 1, vector pYW01; 2, 1st round of panning; 3, 2nd round of panning; 4, 3rd round of panning. Arrows indicate phagemid DNA bands excised from the agarose gel.

Colony PCR of 30 transformants was carried out (10 colonies from each of the three transformations) to examine insert size distribution (Figure 4.3). The insert sizes ranged from 1.5 to 3 Kbp. Phagemid dsDNA from ten transformants whose insert sizes differed from each other were purified and sequenced with the sense and antisense primers complementary to the pYW01 vector sequence upstream and downstream of the inserts.

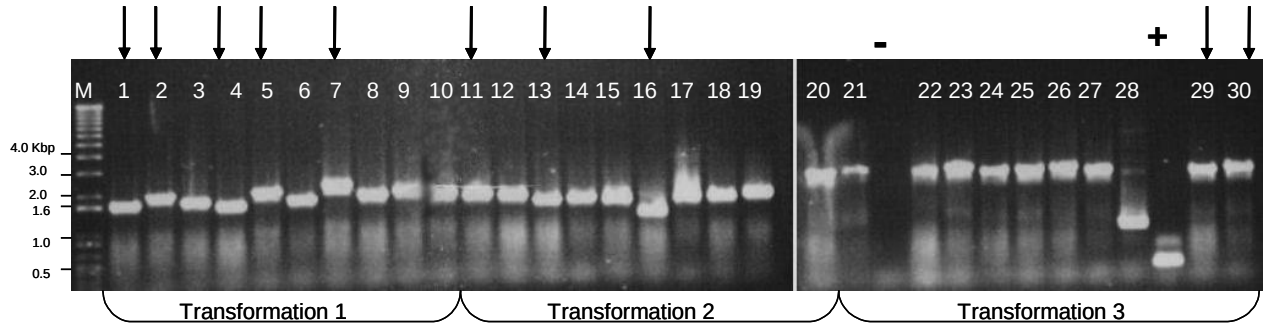


Figure 4.3. Colony PCR of 30 clones selected from the *L. rhamnosus* HN001 phage display library by affinity screening against HN001 cells.

Primers, pDJ01F03 and pDJ01R02, were used for bacterial colony PCR. Arrows indicate sequenced inserts. M, 1 kb+ DNA ladder; +, PCR positive control (vector pYW01); -, PCR negative control (untransformed *E. coli* TG1 colony). Lanes 1-10, 10-20 and 20 -30 corresponded to three different transformations with three eluted plasmid bands indicated in Figure 4.2.

4.2.1 Sequence analysis of affinity selected clones

Sequence analyses revealed that all 10 inserts contained the 5'-moiety of a single open reading frame (ORF) of a secretome protein, in frame with c-myc tag and *gIIIC*. The ORF sequence was fused to pIII, preceded by a putative ribosomal site and initiated by an ATG start codon, followed by an N-terminal TMH. Upstream of the transcriptional site a -10 box (ACATAAAAT) and -35 box (TTGATT) were identified in the putative promoter region using BPROM promoter prediction program from Softberry server (www.softberry.com/berry.phtml). Translated products of the inserts encoding selected ORF ranged in size from 334 to 376 amino acids. Search of the *L. rhamnosus* HN001 secretome database identified this ORF as *lrh33* (Chapter III). However, all inserts selected for binding to HN001 cells were longer than the fragment present in the secretome database/clone bank (Table A.3). To distinguish between the (shorter) secretome insert, *lrh33*, and affinity-selected larger insert, the latter one was named *lrh33*A (for Affinity-selected), and former *lrh33*S (for Short).

Subcellular localisation and membrane-targeting prediction algorithms pSORT and SignalP revealed, as for the original secretome library-encoded insert, that the *lrh33*A encodes for an extracellular protein with an amino-terminal membrane anchor. Furthermore, Pfam, CDART and SMART database searches detected possible functional domains, two bacterial Ig-like type 3 domains (Big-3; E-values 4e-04 and 3e-05). The Ig-like domains are commonly found in the surface and secreted proteins in both prokaryotes and eukaryotes and are typically involved in extracellular adhesion and binding processes.

Given that phage-display cloning requires that the 3'-terminally truncated ORFs are fused to *gIII*, the 3'-end of the affinity-selected ORF was missing from the recombinant phagemids and was identified in the HN001 draft genome sequence. The translated complete ORF *lrh33* was 503 residues in length (1512 bp) with predicted molecular mass of 54.47 kDa (see Appendix III, Figure A.3 for *lrh33* nucleotide sequence and translation, BlastP aligments and Pfam domains search). In addition to the Big-3 domains, Lrh33 also contains at the C-terminus a domain frequently found in Ca^{2+} chelating serine proteases (SCP-like extracellular protein domain). Analysis of the sequence upstream of the *lrh33* revealed two ORFs, unique to *L. rhamnosus* HN001, encoding secretome proteins Lrh62 and Lrh04 (also retrieved in the secretome library in Chapter III; Table A.3). Five out of ten *lrh33*-encoding inserts retrieved from the shot-gun library and the *lrh33* secretome library insert also contained the 3' portion of *lrh04* (Figure 4.4). The two upstream ORFs are orientated in the same transcriptional direction as *lrh33* separated by long intergenic spacers: 343 bp between *lrh62* and *lrh04*, and 385 bp between *lrh04* and *lrh33*. Sequence analysis of downstream sequence from ORF *lrh33* revealed the presence of an ORF, transcribed in the opposite direction, encoding a putative transposase similar to one found in *L. casei* ATCC 334. This enzyme is encoded within a predicted transposable element, possibly belonging to insertional element of IS5 family (IS*5*). Using the FindTerm bacterial terminators prediction algorithm (http://www.softberry.ru/berry.phtml) two terminator sequences were detected: one upstream of putative transcriptional start of *lrh62* and one bordering the putative IS5 element. No terminator sequences were detected using the same prediction software in between *lrh62*, *lrh04* and *lrh33* ORFs. All three proteins, Lrh62, Lrh04 and Lrh33 were annotated in Chapter III (Table 3.5) as unique for *L. rhamnosus* HN001, due to lower than cut-off matches for the fragments encoded by the corresponding secretome library inserts, as determined by BLAST search.

When BLAST search was performed with the Lrh33A fragment, a hypothetical cell surface protein precursor (lp_2795; NCBI Accession number NP_786170.1) from *Lactobacillus plantarum* WCFS1 was the only protein that showed significant sequence similarity (cut off e^{-13}). However, this similarity is limited to a short segment of the amino acids sequence (~150 residues in length, 10 % gaps) covering the Big-3 domains of Lrh33A. Furthermore, lp_2795 is twice the size of Lrh33A (1096 residues) and contains a LPXTG cell-wall anchor and leucine-rich repeats, both of which are absent from Lrh33 ORF.

Together, these observations indicated that *lrh33* ORF encodes a protein unique to the *L. rhamnosus* HN001. This protein, via its Ig-like domains, may potentially be involved in strain-specific interactions between *L. rhamnosus* cells.

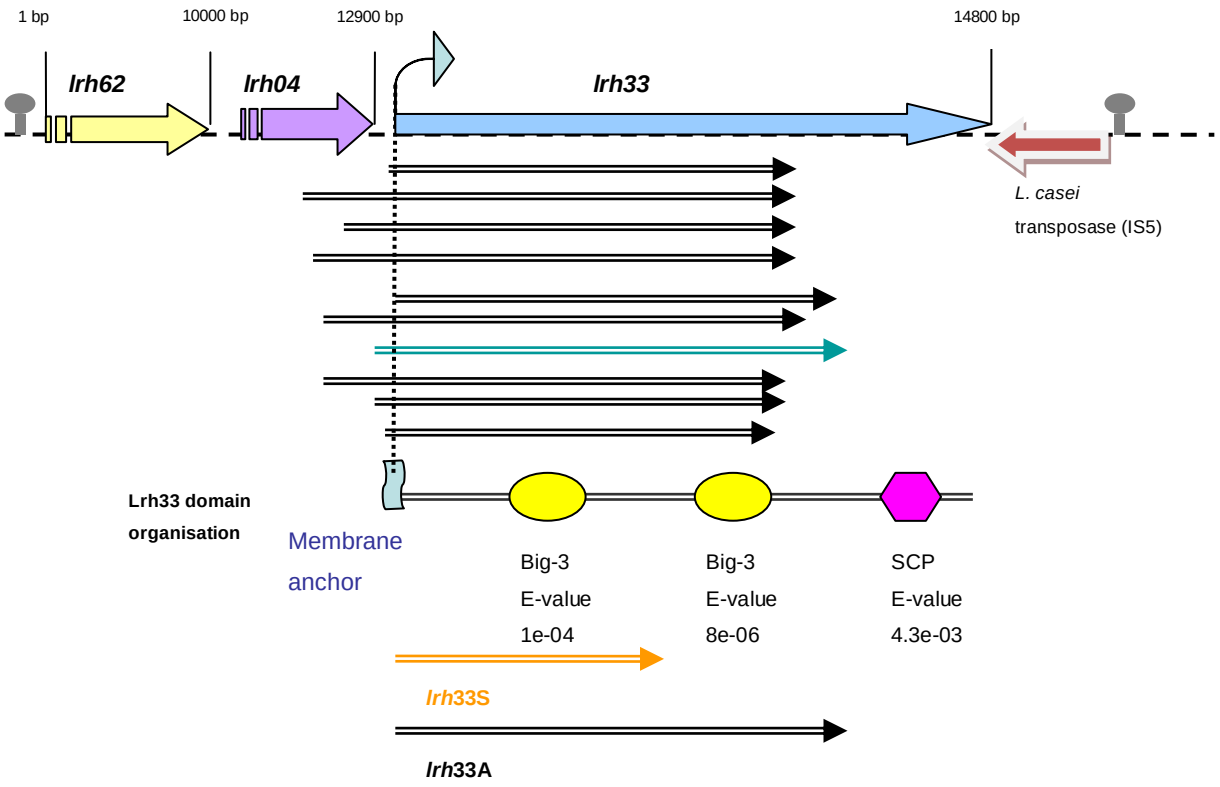


Figure 4.4. Schematic representation of the Lrh33A-encoding phagemid inserts and organisation of HN001 unique ORF set.

Bottom, schematic representation of putative functional domains of Lrh33A predicted by the Pfam, CDART and SMART algorithms: Big-3 (Bacterial Ig-like domain-group 3) and SCP (Ca^{2+}-chelating serine protease). N-terminal membrane anchor was detected by SignalP 3.0 software. Arrows represent 10 Lrh33A-encoding phagemid inserts. The clone used for affinity-binding assay is marked with green arrow. *lrh33*S, short fragment of *lrh33* identified by secretome phage display. The two terminators flanking the *lrh62-lrh04-lrh33* gene cluster, are indicted (mushroom-like grey shapes).

4.2.2 Mapping of HN001-interacting domain of Lrh33

Affinity-binding assay was carried out to compare the recovery of the Lrh33A-displaying phage particles to that of PPs displaying the vector only (pYW01) using HN001 cells as bait. The pLrh33A phagemid particles were generated from the clone with the longest *lrh33*A insert fused to *gIIIC* (Figure 4.4). Purified pLrh33A PP sample was assayed using a panning procedure similar to the one used in library screening (section 2.2.12.1 and 4.2). Panning

resulted in an average 450-fold higher recovery of pLrh33A-binding PPs compared to the negative control (vector only PPs), indicating strong binding to the *L. rhamnosus* HN001 cells (Figure 4.5). Similar increase of recovery of the pLrh33A PPs was observed when they were panned against HN001 cells, compared to the number of recovered PPs from the negative buffer control (no HN001 cells). Interestingly, if the HN001 cells, used as a ligand, were washed with buffer (followed by quick spin and removal of supernatant) prior to the affinity enrichment assay, binding of pLrh33A PPs was greatly reduced (data not shown), suggesting that Lrh33 interacts with cell surface molecules/structures loosely associated with cell wall peptidoglycan or releasable by incubation in buffer.

In addition, phagemid particles derived from the secretome library clone *lrh33*S were included to map the possible cell-cell interacting domain. The *lrh33*S ORF from the secretome library (Chapter III) encodes 199 N-terminal amino acids of Lrh33 fused to pIIIC and contains only one (N-terminal) bacterial Ig-like type 3 domain. The number of recovered Lrh33S-displaying PPs was approximately the same as the number of recovered vector control PPs, suggesting that for interaction with the HN001 surface structures, sequence downstream of the N-terminal Big-3 domain, containing the C-terminal Ig-like domain, is required.

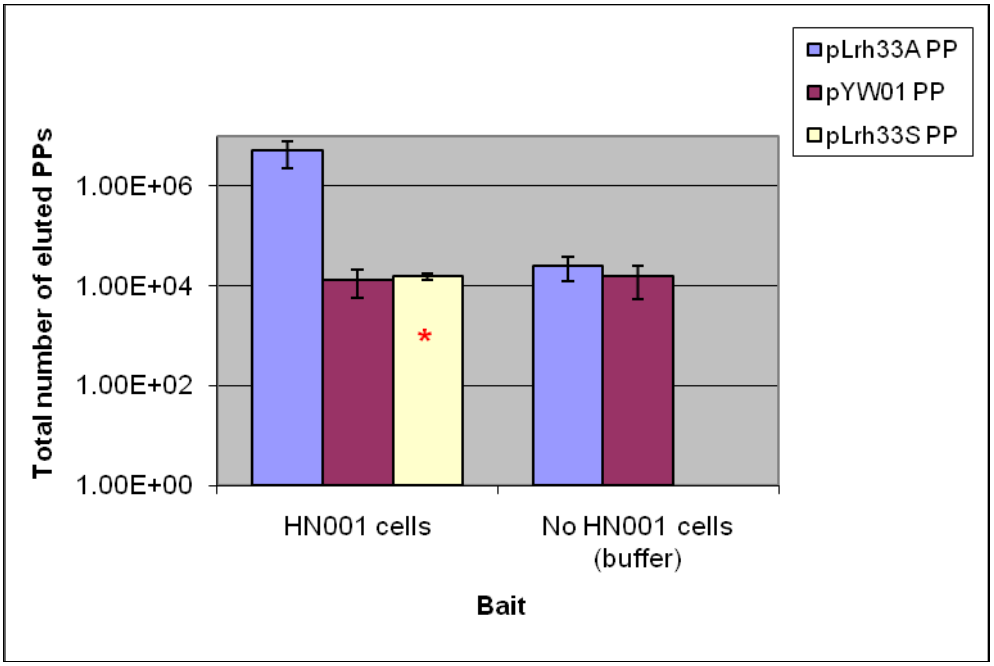


Figure 4.5. Mapping of the Lrh33 HN001-binding domain.
Samples: pLrh33A PP, pYW01 PP and pLrh33 PP, phagemid particles generated using VCSM13 helper phage. **X-axis (Bait)**: HN001 cells, HN001 cells in buffer (1X TBS, 2 mM $CaCl_2$); No HN001 cells, buffer control (1X TBS; 2 mM $CaCl_2$). Each data point is an average of three replicas; error bars represent standard deviation. * , average of 2 replicas.

4.3 Affinity screening of *L. rhamnosus* HN001 shot-gun phage display library for fibronectin-binding proteins

The extracellular matrix (ECM) and the basement membrane of the gastrointestinal tract epithelium lining can be utilised by microorganisms for attachment [69]. One of the major components of ECM and basement membranes is fibronectin. Fibronectin is a large multifunctional glycoprotein that is found in most body fluids as soluble dimer or in an insoluble form in the ECM [255]. It has multiple domains that are capable of binding to a wide variety of polymers such as collagens, fibrin, heparin, and actin, as well as eukaryotic cell surface proteins integrins. Since binding to fibronectin on the surface of the intestinal epithelial cells is the main adhesion route for many different enteropathogenic bacterial species it is possible that probiotic bacteria may also use this protein as an adhesion target [108]. Furthermore, *L. rhamnosus* HN001 was found to bind to fibronectin indicating that cognate binding molecule is likely encoded by HN001 genome [121]. To identify putative

fibronectin-binding proteins the *L. rhamnosus* HN001 shot-gun phage display library was screened for phagemid particles expressing proteins with ability to bind to human fibronectin.

Approximately $3x10^{11}$ phagemid particles were used in each round of affinity enrichment. A total of three panning cycles were carried out, with the change of fibronectin solid immobilisation matrix from polystyrene microtiter plate to streptavidin magnetic beads in the third round of panning. The change of solid support was introduced to reduce or eliminate the phagemid particles that may bind to the immobilisation matrix (e.g. plastic of the microtiter plates).

Increase of 96-fold (12-fold between round I and II and 8-fold between round II and III) in the number of eluted phagemid particles after the third repanning compared to the first panning was observed, indicating that the specific interaction occurred but that affinity of interaction is low. Enriched library plasmid profiles monitored over three rounds of panning suggested that at least two clones of different size had been selected (Figure 4.6). The DNA from these two phagemid bands was purified from the agarose gel and used to transform *E. coli* strain TG1. Five colonies from each transformation were picked at random and the DNA sequence of the inserts was determined.

Out of ten selected colonies, four contained an empty vector with rearrangement in the colD origin of replication and three contained inserts out of frame with pIII. The sequencing of the remaining three colonies identified two library clones (two colonies contained identical clone) that made a functional fusion with *pelB* signal peptide and pIIIC. Sequence analysis revealed that the DNA inserts encoded for 291 amino acid and 233 amino acids long ORFs (Fn01and Fn02), respectively. BLASTP search of Fn01 and Fn02 showed the highest similarity to the hypothetical proteins from *Lactobacillus casei*, LSEI_0735, and LSEI_2459 respectively. According to the SignaP algorithm both homologs contain signal sequences.

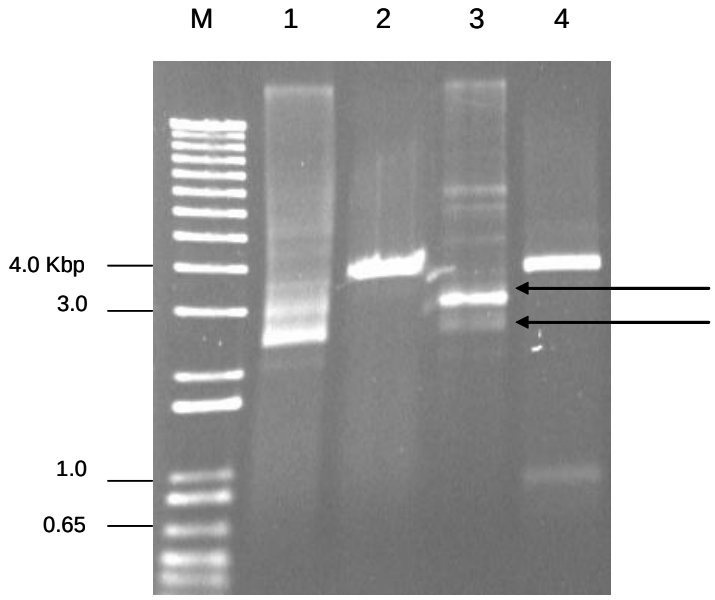


Figure 4.6. Phagemid distribution in HN001 library pools eluted after the 2nd and 3rd rounds of affinity screening against human fibronectin.

Recombinant phagemid supercoiled dsDNA was purified from the aliquots of amplified HN001 library pools after each round of panning. Phagemid DNA was visualised using EtBr after agarose gel electrophoresis. Lanes: 1 Kb+ dsDNA ladder (Invitrogen); 1, 2nd round uncut DNA; 2, *Eco*RI/*Sph*I restriction endonuclease digestion of sample from 1. 3, 3rd round uncut DNA; 4, *Eco*RI/*Sph*I restriction endonuclease digestion of sample from 3. Arrows indicate phagemid DNA excised from the agarose gel.

To examine binding of purified PPs displaying the two ORFs, phagemid particles displaying Fn01 and Fn02 were purified and assayed for binding to microtiter plate-immobilised fibronectin and unspecific ligands, BSA (bovine serum albumin; used in the blocking step of affinity panning of the library) and streptavidin (used for immobilisation of biotinylated Fn onto the magnetic beads), to determine the specificity of the binding for the ligand of interest, Fn. Biopanning resulted in an average increase of PP recovery from Fn-coated microtiter plates relative to BSA by a factor of 5 and 12, respectively, for Fn01 and Fn02 PPs. Given a titration error of typically ± 25%, low increase of binding to Fn relative to BSA indicated possible lack of specific Fn binding for Fn01 and Fn02. However, 26-fold higher recovery was observed for Fn02 PPs from streptavidin compared to BSA-coated wells. This finding indicated that Fn02 might have been selected as low affinity streptavidin-binding protein as a result of the third round of biopanning which was carried out on fibronectin-coated

streptavidin magnetic beads. Interestingly, both Fn01 and Fn02 demonstrated a 10-20-fold increase in binding to BSA when compared to the vector-derived phagemid particles, suggesting that both proteins may be nonspecific "sticky" proteins that bind weakly to a number of protein ligands. The hypothesis that Fn02 protein is a streptavidin-specific binder was not further tested due to the lack of evidence that streptavidin binding is involved in the bacterial adhesion process.

Summary

In search for proteins that mediate adhesion and aggregation, implicated in the initial steps of bacterial colonisation and in exclusion or reduction of pathogen adhesion, a large *L. rhamnosus* HN001 shot-gun phage display library ($3x10^8$ inserts) was created in a new phagemid, pYW01. Recombinant library phagemid particles were screened for aggregation-mediating proteins, using HN001 cells as bait, and for proteins involved in adhesion to extracellular matrix using fibronectin as bait.

After three rounds of panning of *L. rhamnosus* HN001 shot-gun library against HN001 cells, secretome protein, Lrh33, was selected. Binding assay showed 450-fold increase in binding of purified recombinant Lrh33 phagemid particles to HN001 cells compared to the vector only control, indicating strong binding. All inserts included the putative Lrh33 N-terminal transmembrane helix, hence the Lrh33-pIIIC fusion was targeted to the virion by the intrinsic Lrh33 signal, rather than the vector-encoded PelB signal sequence. This suggests that the HN001-binding Lrh33-pIIIC fusion could have been selected for by affinity-screening of the secretome library. This is in agreement with the fact that, with a few exceptions, adhesins of Gram-positive bacteria are bitopic (N- or C-terminal TM anchored) or cell-wall anchored proteins that carry Sec-dependent membrane-targeting signals [110, 112, 138].

Comparison of HN001 binding of a phage-displayed Lrh33 clone obtained from the library panning and a clone with a shorter Lrh33 insert from the secretome clone bank indicated that Lrh33 interacts with surface of HN001 cells either via both of its Ig-like type 3 domains or the C-terminal one.

Screening of *L. rhamnosus* HN001 phage display library against fibronectin resulted in 96-fold increase of phagemid particle recovery after 3 rounds of panning. Two recovered clones, Fn01 and Fn02, encoded each a secretome protein in frame with pIIIC; however in addition to these expected types of recombinants, clones representing a rearranged vector and clones encoding for out-of-frame inserts were also recovered., Individual binding assays of purified Fn01 and Fn02-displaying recombinant phagemid particles showed that these two surface proteins demonstrated weak binding to avidin and BSA, in addition to binding to fibronectin. Protein, Fn02 demonstrated significant increase in binding to streptavidin relative to BSA. Therefore, Fn01 and Fn02 may be unspecific "sticky" proteins rather than specific fibronectin-binding proteins

Chapter V

Discussion

5.1 New secretome phage display technology

A new system for direct selection, expression and display of the secretome, based on the requirement of a signal sequence for assembly of sarcosyl-resistant filamentous phage virions, has been described in this thesis. While a phage display system for cloning secretome proteins has been previously reported [256] it is not efficient for enrichment and display of Gram-positive secretome proteins. That system uses *gIII*-positive helper phage, and the signal sequence-encoding inserts are affinity-enriched based on the presence of a vector-encoded affinity tag incorporated into the fusion. Therefore, the secretome-pIII fusions must successfully compete with the helper phage-derived wild-type pIII for incorporation into the virion. The efficiency of that system for recovery of Gram-positive secretome proteins is poor, with two successive rounds of affinity selection and amplification resulting in identification of only 52 secretome ORFs from a library of the primary size of 10^7 clones [185]. Our system resulted in 89 secretome ORFs from a library of only 10^6 clones in one round of selection, hence it performed at least 20-fold more efficiently than the previously reported enrichment method.

A much lower efficiency of the previously published system could be explained by low efficiency of processing the Gram-positive signal sequences compared to the wild-type pIII signal sequence. As a consequence, a significant number of secretome proteins would be out-competed by the native pIII of the helper phage and would fail to be incorporated into the phagemid particles, preventing their affinity selection. The much higher efficiency of our method is due to direct selection for the release of the correctly assembled phagemid particles. Wild-type pIII is not present in the system; hence, the recombinant fusions cannot be outcompeted by native pIII. Furthermore, the previously reported system [256] uses a vector with a very strong constitutive promoter that likely confers toxic effects to the host *E. coli*, known to be sensitive to overexpression of pIII fusions [234, 257]. As a result, many clones that impair growth of the host *E. coli* and phage assembly would have been lost. Our display system has the advantage of using the very tightly regulated *psp* promoter. This promoter is induced by infection of individual cells with helper phage; it does not require addition of an inducer compound or washing away of an inhibitor [250] and has also been shown to improve display of pIII fusion proteins that are toxic to *E. co*li when overexpressed [258]. This promoter allows the expression of ORFs that do not contain their own transcriptional signals,

such as those located within operons and distal to the promoter in genomic libraries, as well as expression of coding sequences in cDNA libraries. Bioinformatic elucidation of the meta-secretome of complex microbial communities, such as those that colonize the human gastrointestinal tract, is impractical with current sequencing technologies because of the poor coverage of the metagenome gene pool, even in large-scale projects [1, 211].The high efficiency of secretome selection system developed in this thesis would allow selective cloning, sequencing, and functional analyses of surface and secreted proteins on a metagenomic scale, where the limiting factor is the initial size of the library [1, 2].

Based on the estimated size of the *L. rhamnosus* genome (approximately 3 Mb; W Kelly, personal communication) and the percentage of the secretome clones in lactobacilli [110], the coverage of the secretome that we achieved using secretome phage display system is likely to be about 44%. The small size of the primary library ($1x10^6$ clones) was most likely the limiting factor to achieve better coverage of the secretome of *L. rhamnosus* HN001. A larger primary library would likely be required to examine whether an increase in the coverage will be achieved, and thus test the limits of described selection method in coverage of the secretome.

Several reporter fusion systems and cell surface display screening methods have been used to identify secretome proteins and even to systematically analyze the topology of membrane proteins [213-216, 218, 259, 260]. However, a distinct advantage of phage display is in that the protein is automatically purified by association with the virion, simplifying functional characterisation. In this thesis, it is shown that phagemid particles assembled by incorporation of the 963-residue surface protein SOF of the Gram-positive bacterium *S. pyogenes*, targeted by its intrinsic signal sequence, demonstrated two biological activities of this protein corresponding to two independently folding domains. Hence, display and folding of this protein in the context of the phage virion must be reasonably efficient and accurate. Therefore, presumably proteins with an activity of interest could be identified by arraying the secretome clone bank and using high-throughput activity screening. Alternatively, the “raw” secretome phage display library pool, obtained after the selection step, could be screened for activities of interest by well-established phage display library screening protocols. Applied to microbial communities at a metagenomic scale, these methods would allow functional capture

of proteins involved in microbe-microbe and microbe-host interactions, including proteins from yet uncultivated bacteria.

5.2 Secretome of *L. rhamnosus* HN001

Bacteria of the *Lactobacillus* genus are found in diverse environments. Some are found in the gastrointestinal tract and thus comprise part of the gut microbial community that numbers hundreds of bacterial species, whereas others are found on plant material or in fermented foods [261]. Lactobacilli secrete bacteriocins, which kill other Gram-positive bacteria, including pathogens [25, 84, 262]. Furthermore, several *Lactobacillus* surface and secreted proteins have been implicated in intra-species aggregation and co-aggregation with pathogenic bacteria [63, 83, 84, 86] and in one case have been reported to have had an impact on the expression of virulence factors of a pathogenic bacterium [263]. It has been demonstrated that probiotic lactobacilli can modulate activation of dendritic cells [15, 264-266], but the proteins mediating these effects have not yet been identified. In recent years several *Lactobacillus* genomes have been sequenced [38, 47, 109, 267, 268]. Comparative and functional analyses of these bacteria have revealed several proteins involved in colonisation or adhesion [34, 107, 108, 110, 269, 270].

L. rhamnosus HN001 is a probiotic bacterium that transiently colonises the human gut, stabilises the gut microflora, and enhances parameters of both innate and acquired immunity [55, 56, 271]. Bioinformatic analysis of the *L. rhamnosus* HN001 secretome in this thesis revealed a number of features in common with other probiotic bacteria, but also some distinct secretome proteins unique to *L. rhamnosus* HN001. A total of 89 ORFs encoding seven functional classes of extracellular and transmembrane proteins were identified. *In silico* secretome analyses of the completely sequenced genomes of other lactobacilli revealed a similar distribution of categories of predicted secretome proteins. For example, in the *L. plantarum* and *L. reuteri* secretomes the largest classes with assigned function were enzymes (30-35%) and transport proteins (10-15%), while for approximately 45% of total secretome ORFs the function of encoded proteins could not be predicted [111, 186, 272]. Furthermore, ORFs encoding substrate-binding domains of ABC transporters predominated among predicted *L. reuteri* transport proteins (15%) and the same was found in *L. plantarum* (14%) [38] and *L. johnsonii* (17%) [268]. A large proportion of transport proteins, enzymes and hypothetical proteins identified in these studies is consistent with our observations for *L.*

rhamnosus, although compared to the other lactobacilli, HN001 did have a somewhat higher proportion of transport proteins (25% versus 10-15%) and lower proportion of enzymes (23% versus 30-35%). These differences could be due to only partial sequencing of the HN001 secretome or may be the consequence of experimentally derived secretome data for *L. rhamnosus* HN001 versus *in silico* prediction for *L. plantarum* and *L. johnsonii*. The proportion of HN001 secretome ORFs-encoding proteins that are part of the signalling system and host-microbial interaction groups (2%) was similar to observations for other species of the *Lactobacillus* genus (5%). Within this class, only one ORF, *lrh15*, encoded a protein with similarity to a histidine kinase and three ORFs (*lrh51*, *lrh35* and *lrh62*) encoded proteins with predicted adhesin properties.

Only one report has been published thus far that describes an experimentally derived secretome of a lactobacillus, *L. reuteri* DSM 20016^{T} [185]; however, only 52 proteins were retrieved in that report. Comparison between different functional classes from *L. reuteri* DSM 20016^{T} and *L. rhamnosus* HN001 showed similar trends; the same classes of proteins were detected and the relative proportion corresponding to each class was similar.

Seven unique secretome ORFs are identified: *lrh01*, *lrh04*, *lrh07*, *lrh12*, *lrh31*, *lrh33* and *lrh62*. The ORF *lrh62* encodes a large (3275 residues) Ala/Ser-rich surface protein with irregular repeats and low complexity regions. Some of these attributes are typical for many of lactobacilli adhesins [106, 111]. For example, the *L. plantarum* genome encodes 16 proteins with similar features and one of them, lp_1303a, is of similar size (3300 residues) to Lrh62, with a 1600 residues long regularly repeating Ser/Asp motif possibly acting as a spacer between two functional domains (an Ig-like domain and an LPTXG peptidoglycan anchor) [111]. However, Lrh62 Ala/Ser repeats are irregular and this is the unique characteristic that may have a strain-specific function distinguishing *L. rhamnosus* HN001 from other lactobacilli. The other two unique ORFs, *lrh04* and *lrh33*, encode for proteins that contain bacterial Ig-like domains type 3 (Big-3). Given that Ig-like domains are most often involved in cell-cell adhesion [92, 99, 102, 103] there is also possibility that both Lrh04 and Lrh33 are involved in host-HN001 specific interactions. Bacterial cell adhesion property has been confirmed for Lrh33 by its selection for binding to *L. rhamnosus* HN001 from shot-gun phage display library of HN001 (Chapter IV).

Clustered ORFs *lrh62*, *lrh04* and *lrh33* are ORFs arranged in the same transcriptional direction and separated by intergenic spacers of 360 and 385 bp. Prokaryotic promoter prediction analysis indicated putative promoter regions upstream of *lrh62* and *lrh33*. This suggests that all three ORFs are transcribed as separate units. Nevertheless, *in silico* analysis revealed that there are no predicted termination sequences in between the three ORFs. Although in general the intergenic distances are relatively small (<20 bp) between ORFs of the same operon, operons in which the members are highly expressed are the exceptions to this rule as they can contain wider intergenic spaces (>100 bp) [273, 274]. Thus we can speculate that in spite of the 360/385 bp intergenic distances, *lrh62*, *lrh04* and *lrh33* could be transcribed coordinately. Additional arguments in favour of this hypothesis are (i) absence of transcriptional terminators between the ORFs (ii) hypothetical functional unity, i.e. involvement in adhesion process via Ig-like domains (Lrh33 and Lrh04) or low-complexity regions (Lrh62). Identification of IS5 element immediately downstream of *lrh33* ORF indicates a possibility that this whole gene cluster could have been transferred horizontally into HN001 as a unit.

5.3 Lrh33 binds to the *L. rhamnosus* HN001 cell surface

Development of an efficient method for the identification of secretomes of Gram-positive bacteria was an important step towards functional characterisation of secretome proteins. Adhesion to intestinal mucosal surface is the crucial step in initiation of probiotic effects and in some lactobacilli adhesion is correlated with aggregation phenotypes [80, 81, 92, 112]. Given that HN001 aggregates under laboratory growth conditions, a search for proteins involved in cell-cell interactions was undertaken. Intact HN001 cells were used as an affinity substrate (bait or ligand) for enrichment of recombinant phagemid particles displaying putative cell-cell binding protein, from a shot-gun phage display library of *L. rhamnosus* HN001 strain. The secretome phage display library could have been screened, however due to the small size of the primary library and incomplete set of secretome proteins recovered by sequencing the secretome library inserts, it was recognised that a significant number of proteins of interest may be missing from the library. Furthermore, the secretome library was lacking secreted or surface proteins without Sec-dependent targeting signals, and possibly C-terminally located domains of secretome proteins. Therefore a large shot-gun phage display library ($3x10^8$ primary clones) was constructed in a new vector pYW01. Screening of this library using HN001 cells as bait, selected ten clones that contained inserts of different sizes.

However, sequence analysis revealed that these ten inserts all carried the same ORF, *lrh33*, already identified in the HN001 secretome library.

Monoclonal phagemid particles displaying Lrh33 were shown to bind to HN001 as much as 450-fold stronger than did the vector only control, indicating relatively high affinity interaction between the HN001 cell surface and Lrh33. Given that the bacterial cell surface is a multifaceted structure, one could expect to find several interaction proteins, although this may not be the case. If there is more than one HN001-binding protein, a possible reason why only one gene was selected might be due to differences in binding affinity, with high-affinity interactions between Lrh33-expressing phage and HN001 cells, outcompeting putative low-affinity binders in the pannings. Alternatively, cell-binding proteins may not have been folded correctly on the surface of the phage, or could have been toxic to *E. coli* and therefore absent from the library. Finally, additional HN001-interacting entities could be non-protein surface molecules.

No prediction concerning the exact function of Lrh33 could be derived from its primary sequence using bioinformatics. Therefore, the affinity-screening approach in this case was the only means for its functional assignment. Only weak homology to a hypothetical cell surface protein precursor, Lp_2795 of *L. plantarum* WCFS1 [38, 111], was detected. The homology region spans putative bacterial Ig-like domains type 3 (Big-3) of Lrh33. Lp_2795 is a unique protein of *L. plantarum* and contains, in addition to three Big-3 domains, novel repeated domains designated repeat_5. The function of repeat_5 is possibly in protein-protein interactions, based on similarity of these repeats to leucine-rich repeat regions from *L. monocytogenes* internalin [111]. Manual and *in silico* search of the Lrh33 sequence did not reveal any similar repeats. Instead, Lrh33 contains a C-terminal SCP-like extracellular protein domain (Ca^{2+} chelating serine protease), a domain found in many proteins, including plant pathogenesis-related protein 1 (PR-1) and cysteine-rich secretory proteins (CRISPs) [275-277]. Many other SCP-like domain-containing proteins are found in bacteria and archaea but little is known about their biological roles [278]. It has been proposed that SCP domain may function as a Ca^{2+}-chelating serine protease [279]. The presence of functionally different sets of domains, together with the differences in size and mode of anchoring to the cell envelope, suggest that Lp_2795 and Lrh33 are not homologues. The only proteins with similar domain architecture to the Lrh33, but with no similarity at a primary sequence, are five putative surface proteins of *Clostridium difficile* [280]. They contain three type-2 cell wall-binding

domains at the N-terminus followed by one Big-3 and one SCP-like domain. No function was assigned for either of these proteins. In summary, Lrh33 is a unique *Lactobacillus* cell surface-binding protein, distinct from previously identified *Lactobacillus* aggregation proteins.

Lrh33 appears to bind to the HN001 cells is via its two bacterial immunoglobulin-like domains type 3 (Big-3). The immunoglobulin protein superfamily is common in all kingdoms of life and most commonly serves as a scaffold displaying interaction loops. It is very often involved in cell-cell adhesion and extracellular glycohydrolysis [99]. Ig-like domains in bacteria can have direct role in binding to the receptors as in the case of *E. coli* pili adhesin F17-G [101] or a structural role, separating an adhesin domain of a surface protein from the cell surface, thus providing better accessibility of the receptor-binding domain for interaction with host receptors as shown in *E. coli* intimin [102]. Mapping of the Lrh33 putative binding domain in Chapter IV demonstrated that the absence of the C-terminal Big-3 domain almost completely abolishes binding of the Lrh33 to the HN001 cells. This finding suggests that both Big-3 domains or only the C-terminal Big-3 domain are necessary for binding to the receptors at HN001 cell surface.

A notable decrease in binding of the Lrh33-displaying phage to HN001 cells was observed if the latter were washed prior to use in affinity selection procedure and binding assay. This could be indication that some of the molecules or structures loosely associated or releasable by incubation in the buffer are receptors for Lrh33. Many of the proteins that contain Ig-like domains are lectins that interact with diverse carbohydrate structures [101, 281]. If we speculate that Lrh33 acts as a lectin then the candidate receptors might be bacterial cell surface carbohydrates such as HN001 EPS. However, surface protease-cleavable or unstable protein structures can not be excluded as possible Lrh33 receptors.

Together these results suggest that Lrh33 is a HN001-binding protein that interacts with receptors located on the cell surface. This interaction might influence (i) auto-aggregation ability of HN001, (ii) co-aggregation ability of HN001 and (iii) enhancement of adhesion through aggregation and these remain to be tested in the future work.

5.4 Search for fibronectin-binding polypeptides

As previously demonstrated through the series of *in vitro* assays and human feeding trials, HN001 appears to be relatively adhesive strain which *in vivo* transiently colonises the intestinal surface and modulates the host immunity [55, 121, 271]. It may be speculated that HN001 exerts its beneficial immunomodulatory effect through the adhesion to the mucosal surface [282]. Numerous enteropathogenic bacteria use binding to intestinal extracellular matrix (ECM) components as a primary mechanism by which they invade intestinal cells and subsequently disseminate [92]. It has also been demonstrated by many studies that probiotic bacteria, including many *Lactobacillus* species, are able to bind ECM components [76, 108, 110, 119]. Fibronectin (Fn), one of the major ECM glycoproteins, is often used by pathogenic bacteria as an adhesion target. For example, fibronectin-binding proteins of *Staphylococcus aureus* and *Streptococcus pyogenes* have been reported to mediate bacterial adhesion to and invasion of the host cells [131]. In lactobacilli, several fibronectin-binding proteins are identified and their contribution to adhesion to intestinal cell lines described [48, 108].

HN001 was found to bind soluble and immobilised fibronectin and this interaction is most likely mediated by cell surface protein(s) [121]. Thus, panning of the shot-gun HN001 phage display library against immobilised Fn was used to search for Fn-binding protein(s). After three rounds of affinity selection, two putative Fn-binders were isolated. Sequence analysis revealed that both are secreted proteins, similar to two hypothetical proteins of *L. casei*. Selection of binding clones is usually manifested by two to three orders of magnitude increase in the number of eluted phage between the first and the third round of library panning. Increased phage recovery of ~ 100-fold in the third round of library panning indicated that binding clones have been enriched for in the library. This binding was shown however, to be non-specific by a binding assay in which phagemid particles displaying putative Fn-binding proteins, Fn01 and Fn02, showed only 5-fold and 12-fold, respectively, increase in binding to Fn immobilised on microtiter plates in comparison to the binding to unrelated ligand (BSA). Interestingly, in the same binding assay Fn02 showed 26-fold higher binding to streptavidin in comparison to BSA, indicating that Fn02 binds significantly to streptavidin. This finding suggested one possible explanation for selection of Fn02 protein. This protein might have been selected by binding to streptavidin-coated magnetic beads on which Fn was immobilised between the second and third rounds of panning. A change of solid support for immobilisation

of ligands is common practice in phage library screening procedure to eliminate "background" of phagemid clones that bind to the matrix used to immobilise the ligand [193, 283]. However, in this case there is possibility that Fn02 is a streptavidin-binding protein which was present in the library pool after 2nd round of panning (on plastic) and was enriched for the streptavidin in the 3rd round. Significance of the streptavidin-binding in adhesion process of HN001 is unclear. It could be speculated that Fn02 might, by binding to bacterial streptavidin and formation of a non-digestible complex, have a role in enhancement of biotin availability for the host. Both Fn01 and Fn02-displaying PPs demonstrated increased binding to BSA in comparison to vector-derived PPs. In combination with their binding to streptavidin and fibronectin, both of which were present during the library panning protocol, these binding properties of Fn01 and Fn02 indicate that they may be non-specific "sticky" proteins that may have some role in weak interactions with host and bacterial proteins during colonisation of GIT.

There are a few reasons that might explain why specific fibronectin-binding proteins were not selected using panning against immobilised Fn. Firstly, the system that we used (monovalent system) displays only one copy of fusion, allowing selection of only high-affinity binders [59, 226]. Thus proteins with low affinity for fibronectin would be missed. Experiments described in chapter II showed that the fibronectin-binding domain of streptococcal/staphylococcal family (contained within *S. pyogenes* SOF22 protein) folds correctly and binds to human Fn 10,000-fold over the vector control when displayed on the surface of the phage. Furthermore, SOF22-displaying recombinant PPs, when spiked into the HN001 shot-gun library, were enriched to homogeneity over two rounds of panning (data not shown). Failure to identify specific FnBP by panning of the HN001 library strongly suggests that a high-affinity fibronectin-binding protein either does not exist in HN001, or if it does, it belongs to a different protein family that cannot fold correctly when fused to pIII. Secondly, the putative Fn-binding protein fusion to pIII may be toxic to *E. coli*, and hence lead to lack of assembly of fusion-displaying phagemid particles and effectively be missing from the library.

Chapter VI

Conclusion and Future directions

6.1 Conclusion

This research shows that it is possible to select, with a high efficiency, the secretome of Gram-positive bacteria, by using a system consisting of a phage display phagemid vector that does not contain a signal sequence and a *gIII*-deleted helper phage. Gram-positive secretome proteins, targeted to the virion by their signal sequences, can be directly purified and functionally characterised. The method developed in this research work is sufficiently efficient to identify and display 44% of the secretome of Gram-positive bacterium *L. rhamnosus* HN001 by analyzing fewer than 500 clones from a primary library of 10^6 clones.

In search for cell-cell adhesion proteins, which may be involved in aggregation and adhesion processes, *L. rhamnosus* HN001 shot-gun phage display library was screened against HN001 cells. A secretome protein Lrh33 was identified in this screen as an HN001-binding protein. Further, binding region of Lrh33 was located to bacterial Ig-like domains type 3 (Big 3). The receptors for Lrh33 on the HN001 cell surface appear to be loosely attached or easily degradable molecules or structures on the surface of HN001 cells.

Two ORFs (Fn01 and Fn02) were identified in the screening of *L. rhamnosus* HN001 shot-gun phage display library against human ECM protein fibronectin. However, binding assays of purified Fn01 and Fn02 phagemid particles showed that they are most likely non-specific "sticky" proteins that bind many ligands, including BSA, Milk casein, streptavidin and fibronectin. Fn02 shows strong preference for binding to streptavidin, which was used to immobilise biotinylated fibronectin in the library panning procedure.

6.2 Future directions

L. rhamnosus HN001 is a probiotic bacterium that adheres to the intestinal epithelial cells and plays a role in immunomodulation of the mucosal immunity. However, thus far mechanisms or mediators of these probiotic effects have not been defined. As seen in the example of the successful identification and characterisation of secretome protein Lrh33, the HN001 shot-gun phage display library could be used in identification of proteins that mediate adherence/colonisation and immunomodulation of the host GIT. Although fibronectin-binding proteins were not identified, the HN001 phage display library could be subjected to

biopanning to select for proteins that bind to (i) other extracellular matrix proteins such as collagen IV, laminin and tenascin, (ii) mucin and (iii) various GIT cell lines (Caco-2, HT29-MTX). Another area of exploration is to identify HN001 proteins that modulate immune response by assaying purified phagemid particles displaying individual secretome proteins from the secretome clone bank generated in this work. Assays could include analysis of induction of cytokines (e.g. IL-12, TGFβ1, IL-15) and upregulation of cell surface activation markers (e.g. CD4,CD56, ICAM1) in human gut epithelial cell lines [55, 284, 285], or expression of IL-10, IFN-γ, and markers of activation or proliferation in human immune cells (e.g.monocytes and NK cells) [286-288].

In Chapter IV Lrh33 was characterised as a HN001-binding protein that interacts with an HN001 cell surface receptor(s) via either both of or the C-terminal Ig-like domain type 3. Further mapping, using HN001-binding assay and a recombinant PPs displaying only the C-terminal Big-3 domain, could be performed to establish whether the N-terminal Big-3 domain is necessary for binding to HN001 cell surface.

Due to the complex nature of the HN001 cell surface, it is hard to predict the Lrh33 receptor. The chemical character of the receptor could be characterised by performing various chemical (proteases and/or sodium periodate treatment) and physical (washing, fractionations) treatments. Subsequently, chemically/physically-treated HN001 cells could be used in an Lrh33-binding assay to determine whether the Lrh33 receptor is a carbohydrate or proteinaceous in nature and whether it is in the cell wall or membrane fraction. A series of HN001 mutants of major surface proteins and EPS has been constructed in Fonterra (M. Collett, unpublished observations) and those can be tested for binding to Lrh33, to investigate whether they are receptors for Lrh33. Since genetic tools for *L. rhamnosus* HN001 are available, it will be possible to directly examine whether Lrh33 is indeed an aggregation protein, and to determine its specificity, loss- and gain-of-function mutations could be constructed - a Δ*lrh33*HN001 “knock-out” mutant, a trans-complemented strain and a strain with overexpressed Lrh33.

The SignalP algorithm predicted that the N-terminal hydrophobic TMH of Lrh33 is an N-terminal membrane anchor rather than a signal sequence due to a poor signal peptidase cleavage consensus. However, it is necessary to test this prediction by investigation of the

subcellular localisation of Lrh33 experimentally. This will be achieved by expression of C-terminally tagged Lrh33 and determining its subcellular location by cell fractionation and fluorescent microscopy, combined with detection of the tagged Lrh33.

The presence of a cluster of unique ORFs, *lrh62, lrh04* and *lrh33* is intriguing, as it may correspond to a functional cluster. Lrh62 and Lrh04 secretome clones were identified in the secretome library, but were significantly truncated in the periplasm of *E. coli* (data not shown). The knock-out approach could therefore be used to test whether these two proteins contribute to binding of Lrh33 to HN001 or aggregation-mediating proteins, either as receptors or accessory proteins. Knockout mutations of each *lrh62* and *lrh04* could be constructed and tested for aggregation and binding to phage-displayed Lrh33.

It was hypothesised that *lrh62, lrh04* and *lrh33* may form an operon. This could be examined by mapping of the transcripts corresponding to this gene cluster and monitoring the relative expression of the three ORFs using standard methods such as primer extension and northern blotting.

In recent years, metagenomics or microbial community shot-gun sequencing has enabled cultivation-independent gene discovery as exemplified in recent metagenomics of human GIT microbial community [1]. However, even large-scale projects that use next-generation sequencing methods such as 454 FLX sequencers capture only a fraction of genetic information in complex microbial communities. More importantly, shot-gun sequencing does not provide the means for functional analysis of identified genes. The secretome-selective phage display developed in this research work can be therefore used as novel technology, metasecretome display, to functionally screen the entire microbial community for the secretome proteins that interact with host surfaces, immune system and other microbes in community.

When extrapolated to the metagenome scale, a comparable coverage of the meta-secretome of a complex microbial community of up to 100 species is achievable with a primary library size of 10^8 clones and analysis of approximately 50,000 sequencing reactions, both of which are easily achievable using standard techniques. Furthermore, Gram-positive *Firmicutes* (*Clostridiales, Bacilliales* and *Lactobacilliales*) and *Actinobacteria* (*Actinomycetales* and

Bifidobacteriales) are prominent groups of bacteria in the human gut microbial community [1, 2]. Hence, the highly efficient selection of Gram-positive bacterial secretome ORFs achieved by the direct selection method described in this thesis is crucial to avoid the secretome library being dominated by Gram-negative secretome proteins [289]. Bioinformatic studies of archaeal signal sequences suggest that they closely resemble those of bacteria. It is therefore expected that archaeal signal sequences would be selected using this method [158, 290]. In contrast, proteins exported via Tat and Sec-independent translocation pathways of Gram-negative bacteria (type I and III secretion systems) would presumably be absent due to the fundamentally different mechanisms of translocation through the bacterial envelope [138, 139, 291].

The information on the secretome proteins coding sequences in the secretome protein database could be also used for DNA microarrays to examine the expression of the secretome proteins under different growth conditions.

REFERENCES

1. Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraser-Liggett CM, Nelson KE: **Metagenomic analysis of the human distal gut microbiome**. *Science* 2006, **312**(5778):1355-1359.
2. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA: **Diversity of the human intestinal microbial flora**. *Science* 2005, **308**(5728):1635-1638.
3. Hooper LV, Wong MH, Thelin A, Hansson L, Falk PG, Gordon JI: **Molecular analysis of commensal host-microbial relationships in the intestine**. *Science* 2001, **291**(5505):881-884.
4. Ouwehand AC, Salminen S, Isolauri E: **Probiotics: an overview of beneficial effects**. *Antonie Van Leeuwenhoek* 2002, **82**(1-4):279-289.
5. Stappenbeck TS, Hooper LV, Gordon JI: **Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells**. *Proc Natl Acad Sci U S A* 2002, **99**(24):15451-15455.
6. Ley RE, Turnbaugh PJ, Klein S, Gordon JI: **Microbial ecology: human gut microbes associated with obesity**. *Nature* 2006, **444**(7122):1022-1023.
7. Favier CF, Vaughan EE, De Vos WM, Akkermans AD: **Molecular monitoring of succession of bacterial communities in human neonates**. *Appl Environ Microbiol* 2002, **68**(1):219-226.
8. Palmer C, Bik EM, Digiulio DB, Relman DA, Brown PO: **Development of the Human Infant Intestinal Microbiota**. *PLoS Biol* 2007, **5**(7):e177.
9. Penders J, Thijs C, Vink C, Stelma FF, Snijders B, Kummeling I, van den Brandt PA, Stobberingh EE: **Factors influencing the composition of the intestinal microbiota in early infancy**. *Pediatrics* 2006, **118**(2):511-521.
10. Artis D: **Epithelial-cell recognition of commensal bacteria and maintenance of immune homeostasis in the gut**. *Nat Rev Immunol* 2008, **8**(6):411-420.
11. Bouma G, Strober W: **The immunological and genetic basis of inflammatory bowel disease**. *Nat Rev Immunol* 2003, **3**(7):521-533.
12. Karin M, Lawrence T, Nizet V: **Innate immunity gone awry: linking microbial infections to chronic inflammation and cancer**. *Cell* 2006, **124**(4):823-835.

13. Macdonald TT, Monteleone G: **Immunity, inflammation, and allergy in the gut**. *Science* 2005, **307**(5717):1920-1925.

14. Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR: **Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases**. *Proc Natl Acad Sci U S A* 2007, **104**(34):13780-13785.

15. Manichanh C, Rigottier-Gois L, Bonnaud E, Gloux K, Pelletier E, Frangeul L, Nalin R, Jarrin C, Chardon P, Marteau P *et al*: **Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach**. *Gut* 2006, **55**(2):205-211.

16. Havenaar R, B. T. Brink, and J. H. J. Huis In'T Veld: **Selection of strains for probiotics use**. *In R Fuller (ed), Probiotics The scientific basis Chapman & Hall, London, England* 1992:209-223.

17. Rachmilewitz D, Katakura K, Karmeli F, Hayashi T, Reinus C, Rudensky B, Akira S, Takeda K, Lee J, Takabayashi K *et al*: **Toll-like receptor 9 signaling mediates the anti-inflammatory effects of probiotics in murine experimental colitis**. *Gastroenterology* 2004, **126**(2):520-528.

18. Johnson-Henry KC, Hagen KE, Gordonpour M, Tompkins TA, Sherman PM: **Surface-layer protein extracts from *Lactobacillus helveticus* inhibit enterohaemorrhagic *Escherichia coli* O157:H7 adhesion to epithelial cells**. *Cell Microbiol* 2007, **9**(2):356-367.

19. McDonough F. WP, Wong N., Hotchins A. and Bodwell C. : **Role of vitamins and minerals in growth-stimulation of rats fed yoghurt**. *Fed Proc* 1983, **42**:556-558.

20. Deeth HC, Tamime AY: **Yoghurt: nutritive and therapeutic aspects**. *J Food Prot* 1981, **44**:78-86

21. Mazmanian SK, Round JL, Kasper DL: **A microbial symbiosis factor prevents intestinal inflammatory disease**. *Nature* 2008, **453**(7195):620-625.

22. Pereira DI, Gibson GR: **Cholesterol assimilation by lactic acid bacteria and bifidobacteria isolated from the human gut**. *Appl Environ Microbiol* 2002, **68**(9):4689-4693.

23. Savino F, Pelle E, Palumeri E, Oggero R, Miniero R: ***Lactobacillus reuteri* (American Type Culture Collection Strain 55730) versus simethicone in the treatment of infantile colic: a prospective randomized study**. *Pediatrics* 2007, **119**(1):e124-130.

24. Weizman Z, Asli G, Alsheikh A: **Effect of a probiotic infant formula on infections in child care centers: comparison of two probiotic agents**. *Pediatrics* 2005, **115**(1):5-9.

25. Corr SC, Li Y, Riedel CU, O'Toole PW, Hill C, Gahan CG: **Bacteriocin production as a mechanism for the antiinfective activity of *Lactobacillus salivarius* UCC118**. *Proc Natl Acad Sci U S A* 2007, **104**(18):7617-7621.

26. Klaenhammer TR, Kullen MJ: **Selection and design of probiotics**. *Int J Food Microbiol* 1999, **50**(1-2):45-57.

27. FAO/WHO: **Guidelines for the Evaluation of Probiotics in Food**. *Report of a Joint FAO/WHO Working Group on Drafting Guidelines for the Evaluation of Probiotics in Food* 2002, **London, Ontario, Canada:FAO/WHO**

28. Leahy SC, Higgins DG, Fitzgerald GF, van Sinderen D: **Getting better with bifidobacteria**. *J Appl Microbiol* 2005, **98**(6):1303-1315.

29. Ventura M, O'Connell-Motherway M, Leahy S, Moreno-Munoz JA, Fitzgerald GF, van Sinderen D: **From bacterial genome to functionality; case bifidobacteria**. *Int J Food Microbiol* 2007, **120**(1-2):2-12.

30. Holzapfel WH, Haberer P, Geisen R, Bjorkroth J, Schillinger U: **Taxonomy and important features of probiotic microorganisms in food and nutrition**. *Am J Clin Nutr* 2001, **73(**2 Suppl):365S-373S.

31. Tannock GW: **Probiotics and Prebiotics: Scientific Aspects**: Caister Academic Press, Wymondham; 2005.

32. Felis GE, Dellaglio F: **Taxonomy of Lactobacilli and Bifidobacteria**. *Curr Issues Intest Microbiol* 2007, **8**(2):44-61.

33. Collins MD, Rodrigues U, Ash C, Aguirre M, Farrow JAE, Martinez-Murcia A, Phillips BA, Williams AM, Wallbanks S.:**Phylogenetic analysis of the genus *Lactobacillu*s and related lactic acid bacteria as determined by reverse transcriptase sequencing of 16S rRNA**. *FEMS Microbiol Lett* 1991, **77**(1):5-12.

34. Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, Koonin E, Pavlov A, Pavlova N, Karamychev V, Polouchine N *et al*: **Comparative genomics of the lactic acid bacteria**. *Proc Natl Acad Sci U S A* 2006, **103**(42):15611-15616.

35. **Lactic Acid Bacteria - Microbiology and Functional Aspects**, Second edn: Marcel Dekker Inc; 1998.

36. McKay LL, Baldwin KA: **Applications for biotechnology: present and future improvements in lactic acid bacteria**. *FEMS Microbiol Rev* 1990, **7**(1-2):3-14.

37. van de Guchte M, Penaud S, Grimaldi C, Barbe V, Bryson K, Nicolas P, Robert C, Oztas S, Mangenot S, Couloux A *et al*: **The complete genome sequence of *Lactobacillus bulgaricus* reveals extensive and ongoing reductive evolution**. *Proc Natl Acad Sci U S A* 2006, **103**(24):9274-9279.

38. Kleerebezem M, Boekhorst J, van Kranenburg R, Molenaar D, Kuipers OP, Leer R, Tarchini R, Peters SA, Sandbrink HM, Fiers MW *et al*: **Complete genome sequence of *Lactobacillus plantarum* WCFS1**. *Proc Natl Acad Sci U S A* 2003, **100**(4):1990-1995.

39. Canchaya C, Claesson MJ, Fitzgerald GF, van Sinderen D, O'Toole P W: **Diversity of the genus *Lactobacillus* revealed by comparative genomics of five species**. *Microbiology (SGM)* 2006, **152**(Pt 11):3185-3196.

40. McCracken A, Turner MS, Giffard P, Hafner LM, Timms P: **Analysis of promoter sequences from *Lactobacillus* and *Lactococcus* and their activity in several *Lactobacillus* species**. *Arch Microbiol* 2000, **173**(5-6):383-389.

41. Rud I, Jensen PR, Naterstad K, Axelsson L: **A synthetic promoter library for constitutive gene expression in *Lactobacillus plantarum***. *Microbiology (Reading, England)* 2006, **152**(Pt 4):1011-1019.

42. Delcour J, Ferain T, Deghorain M, Palumbo E, Hols P: **The biosynthesis and functionality of the cell-wall of lactic acid bacteria**. *Antonie Van Leeuwenhoek* 1999, **76**(1-4):159-184.

43. Grangette C, Nutten S, Palumbo E, Morath S, Hermann C, Dewulf J, Pot B, Hartung T, Hols P, Mercenier A: **Enhanced antiinflammatory capacity of a *Lactobacillus plantarum* mutant synthesizing modified teichoic acids**. *Proc Natl Acad Sci U S A* 2005, **102**(29):10321-10326.

44. Makino S, Ikegami S, Kano H, Sashihara T, Sugano H, Horiuchi H, Saito T, Oda M: **Immunomodulatory effects of polysaccharides produced by *Lactobacillus delbrueckii* ssp. *bulgaricus* OLL1073R-1**. *J Dairy Sci* 2006, **89**(8):2873-2881.

45. Chabot S, Yu H, Deleseleuc L, Cloutier M, Van Calsteren M, Lessard D, Roy D, Lacroix M, Oth D: **Exopolisaccharides from *Lactobacillus rhamnosus* RW-9595W stimulate TNF, IL-6 and IL-12 in human and mouse cultured immunocompetent cells, and IFN-γ in mouse splenocytes**. *Lait* 2001, **81**:683-697.

46. Avall-Jaaskelainen S, Palva A: ***Lactobacillus*** **surface layers and their applications**. *FEMS Microbiol Rev* 2005, **29**(3):511-529.

47. Altermann E, Russell WM, Azcarate-Peril MA, Barrangou R, Buck BL, McAuliffe O, Souther N, Dobson A, Duong T, Callanan M *et al*: **Complete genome sequence of the probiotic lactic acid bacterium *Lactobacillus acidophilus* NCFM**. *Proc Natl Acad Sci U S A* 2005, **102**(11):3906-3912.

48. Hynonen U, Westerlund-Wikstrom B, Palva A, Korhonen TK: **Identification by flagellum display of an epithelial cell- and fibronectin-binding function in the SlpA surface protein of *Lactobacillus brevis***. *J Bacteriol* 2002, **184**(12):3360-3367.

49. McGroarty JA: **Cell surface appendages of lactobacilli**. *FEMS Microbiol Lett* 1994, **124**(3):405-409.

50. Rendon MA, Saldana Z, Erdem AL, Monteiro-Neto V, Vazquez A, Kaper JB, Puente JL, Giron JA: **Commensal and pathogenic *Escherichia coli* use a common pilus adherence factor for epithelial cell colonization**. *Proc Natl Acad Sci U S A* 2007, **104**(25):10637-10642.

51. Sauer FG, Mulvey MA, Schilling JD, Martinez JJ, Hultgren SJ: **Bacterial pili: molecular mechanisms of pathogenesis**. *Curr Opin Microbiol* 2000, **3**(1):65-72.

52. Ton-That H, Marraffini LA, Schneewind O: **Sortases and pilin elements involved in pilus assembly of *Corynebacterium diphtheriae***. *Mol Microbiol* 2004, **53**(1):251-261.

53. Scott JR, Zahner D: **Pili with strong attachments: Gram-positive bacteria do it differently**. *Mol Microbiol* 2006, **62**(2):320-330.

54. Prasad J, Gill H, Smart J, Gopal P: **Selection and characterisation of *Lactobacillus* and *Bifidobacterium* strains for use as probiotics**. *Int Dairy J* 1998, **8**:993-1002.

55. Gill HS, Rutherfurd KJ, Prasad J, Gopal PK: **Enhancement of natural and acquired immunity by *Lactobacillus rhamnosus* (HN001), *Lactobacillus acidophilus* (HN017) and *Bifidobacterium lactis* (HN019)**. *Br J Nutr* 2000, **83**(2):167-176.

56. Gopal P, Prasad J, Smart J, Gill H: **In vitro adherence properties of *Lactobacillus rhamnosus* DR20 and *Bifidobacterium lactis* DR10 strains and their antagonistic activity against an enterotoxigenic *Escherichia coli*.** *Int J Food Microbiol* 2001, **67**:207-216.

57. Zhou JS, Gopal PK, Gill HS: **Potential probiotic lactic acid bacteria *Lactobacillus rhamnosus* (HN001), *Lactobacillus acidophilus* (HN017) and *Bifidobacterium lactis* (HN019) do not degrade gastric mucin in vitro**. *Int J Food Microbiol* 2001, **63**(1-2):81-90.

58. Zhou JS, Shu Q, Rutherfurd KJ, Prasad J, Birtles MJ, Gopal PK, Gill HS: **Safety assessment of potential probiotic lactic acid bacterial strains *Lactobacillus rhamnosus* HN001, *Lb. acidophilus* HN017, and *Bifidobacterium lactis* HN019 in BALB/c mice**. *Int J Food Microbiol* 2000, **56**(1):87-96.

59. Beckmann C, Waggoner JD, Harris TO, Tamura GS, Rubens CE: **Identification of novel adhesins from Group B streptococci by use of phage display reveals that C5a peptidase mediates fibronectin binding**. *Infect Immun* 2002, **70**(6):2869-2876.

60. Gill HS, Shu Q, Lin H, Rutherfurd KJ, Cross ML: **Protection against translocating *Salmonella typhimurium* infection in mice by feeding the immuno-enhancing probiotic *Lactobacillus rhamnosus* strain HN001**. *Med Microbiol Immunol* 2001, **190**(3):97-104.

61. Bernet-Camard MF, Lievin V, Brassart D, Neeser JR, Servin AL, Hudault S: **The human *Lactobacillus acidophilus* strain LA1 secretes a nonbacteriocin antibacterial substance(s) active *in vitro* and *in vivo***. *Appl Environ Microbiol* 1997, **63**(7):2747-2753.

62. Drago L, Gismondo MR, Lombardi A, de Haen C, Gozzini L: **Inhibition of *in vitro* growth of enteropathogens by new *Lactobacillus* isolates of human intestinal origin**. *FEMS Microbiol Lett* 1997, **153**(2):455-463.

63. Roos S, Lindgren S, Jonsson H: **Autoaggregation of *Lactobacillus reuteri* is mediated by a putative DEAD-box helicase**. *Mol Microbiol* 1999, **32**(2):427-436.

64. Reid G: **The scientific basis for probiotic strains of *Lactobacillus***. *Appl Environ Microbiol* 1999, **65**(9):3763-3766.

65. Cross ML, Stevenson LM, Gill HS: **Anti-allergy properties of fermented foods: an important immunoregulatory mechanism of lactic acid bacteria?** *Int Immunopharmacol* 2001, **1**(5):891-901.

66. Sheih YH, Chiang BL, Wang LH, Liao CK, Gill HS: **Systemic immunity-enhancing effects in healthy subjects following dietary consumption of the lactic acid bacterium *Lactobacillus rhamnosus* HN001**. *J Am Coll Nutr* 2001, **20**(2 Suppl):149-156.

67. Antikainen J, Anton L, Sillanpaa J, Korhonen TK: **Domains in the S-layer protein CbsA of *Lactobacillus crispatus* involved in adherence to collagens, laminin and lipoteichoic acids and in self-assembly**. *Mol Microbiol* 2002, **46**(2):381-394.

68. Kato I, Tanaka K, Yokokura T: **Lactic acid bacterium potently induces the production of interleukin-12 and interferon-gamma by mouse splenocytes**. *Int J Immunopharmacol* 1999, **21**(2):121-131.

69. Lussier C, Basora N, Bouatrouss Y, Beaulieu JF: **Integrins as mediators of epithelial cell-matrix interactions in the human small intestinal mucosa**. *Microsc Res Tech* 2000, **51**(2):169-178.

70. Magalhaes JG, Tattoli I, Girardin SE: **The intestinal epithelial barrier: how to distinguish between the microbial flora and pathogens**. *Semin Immunol* 2007, **19**(2):106-115.

71. Agerberth B, Gudmundsson GH: **Host antimicrobial defence peptides in human disease**. *Curr Top Microbiol Immunol* 2006, **306**:67-90.

72. Martini F: **Fundamentals of anatomy and physiology**. New Jersey: Prentice-Hall, Inc.; 1992.

73. Pamer EG: **Immune responses to commensal and environmental microbes**. *Nat Immunol* 2007, **8**(11):1173-1178.

74. Kolenbrander PE: **Oral microbial communities: biofilms, interactions, and genetic systems**. *Annu Rev Microbiol* 2000, **54**:413-437.

75. Reid G, Bruce AW, McGroarty JA, Cheng KJ, Costerton JW: **Is there a role for lactobacilli in prevention of urogenital and intestinal infections?** *Clin Microbiol Rev* 1990, **3**(4):335-344.

76. Sillanpaa J, Martinez B, Antikainen J, Toba T, Kalkkinen N, Tankka S, Lounatmaa K, Keranen J, Hook M, Westerlund-Wikstrom B *et al*: **Characterization of the collagen-binding S-layer protein CbsA of *Lactobacillus crispatus***. *J Bacteriol* 2000, **182**(22):6440-6450.

77. Boris S, Suarez JE, Barbes C: **Characterization of the aggregation promoting factor from *Lactobacillus gasseri*, a vaginal isolate**. *J Appl Microbiol* 1997, **83**(4):413-420.

78. Cesena C, Morelli L, Alander M, Siljander T, Tuomola E, Salminen S, Mattila-Sandholm T, Vilpponen-Salmela T, von Wright A: ***Lactobacillus crispatus*** **and its nonaggregating mutant in human colonization trials**. *J Dairy Sci* 2001, **84**(5):1001-1010.

79. Jankovic I, Ventura M, Meylan V, Rouvet M, Elli M, Zink R: **Contribution of aggregation-promoting factor to maintenance of cell shape in *Lactobacillus gasseri* 4B2**. *J Bacteriol* 2003, **185**(11):3288-3296.

80. Collado MC, Meriluoto J, Salminen S: **Adhesion and aggregation properties of probiotic and pathogen strains** *Eur Food Res Technol* 2008, **226**(5):1065-1073.

81. Del Re B, Sgorbati B, Miglioli M, Palenzona D: **Adhesion, autoaggregation and hydrophobicity of 13 strains of *Bifidobacterium longum***. *Lett Appl Microbiol* 2000, **31**(6):438-442.

82. Rickard AH, Gilbert P, High NJ, Kolenbrander PE, Handley PS: **Bacterial coaggregation: an integral process in the development of multi-species biofilms**. *Trends Microbiol* 2003, **11**(2):94-100.

83. Schachtsiek M, Hammes WP, Hertel C: **Characterization of *Lactobacillus coryniformis* DSM 20001T surface protein Cpf mediating coaggregation with and aggregation among pathogens**. *Appl Environ Microbiol* 2004, **70**(12):7078-7085.

84. Lozo J, Jovcic B, Kojic M, Dalgalarrondo M, Chobert JM, Haertle T, Topisirovic L: **Molecular characterization of a novel bacteriocin and an unusually large aggregation factor of *Lactobacillus paracasei* subsp. *paracasei* BGSJ2-8, a natural isolate from homemade cheese**. *Curr Microbiol* 2007, **55**(3):266-271.

85. Siciliano RA, Cacace G, Mazzeo MF, Morelli L, Elli M, Rossi M, Malorni A: **Proteomic investigation of the aggregation phenomenon in *Lactobacillus crispatus***. *Biochim Biophys Acta* 2008, **1784**(2):335-342.

86. Marcotte H, Ferrari S, Cesena C, Hammarstrom L, Morelli L, Pozzi G, Oggioni MR: **The aggregation-promoting factor of *Lactobacillus crispatus* M247 and its genetic locus**. *J Appl Microbiol* 2004, **97**(4):749-756.

87. Walter J, Schwab C, Loach DM, Ganzle MG, Tannock GW: **Glucosyltransferase A (GtfA) and inulosucrase (Inu) of *Lactobacillus reuteri* TMW1.106 contribute to cell aggregation, in vitro biofilm formation, and colonization of the mouse gastrointestinal tract**. *Microbiology (Reading, England)* 2008, **154**(Pt 1):72-80.

88. Ofek I, Doyle RJ: **Bacterial adhesion to cells and tissues**. New York: Chapman & Hall; 1994.

89. Nassif X, Beretti JL, Lowy J, Stenberg P, O'Gaora P, Pfeifer J, Normark S, So M: **Roles of pilin and PilC in adhesion of *Neisseria meningitidis* to human epithelial and endothelial cells**. *Proc Natl Acad Sci U S A* 1994, **91**(9):3769-3773.

90. Yasukawa K, Martin P, Tinsley CR, Nassif X: **Pilus-mediated adhesion of *Neisseria meningitidis* is negatively controlled by the pilus-retraction machinery**. *Mol Microbiol* 2006, **59**(2):579-589.

91. Jones CH, Pinkner JS, Roth R, Heuser J, Nicholes AV, Abraham SN, Hultgren SJ: **FimH adhesin of type 1 pili is assembled into a fibrillar tip structure in the *Enterobacteriaceae***. *Proc Natl Acad Sci U S A* 1995, **92**(6):2081-2085.

92. Pizarro-Cerda J, Cossart P: **Bacterial adhesion and entry into host cells**. *Cell* 2006, **124**(4):715-727.

93. Imberty A, Mitchell EP, Wimmerova M: **Structural basis of high-affinity glycan recognition by bacterial and fungal lectins**. *Curr Opin Struct Biol* 2005, **15**(5):525-534.

94. Westerlund B, Korhonen TK: **Bacterial proteins binding to the mammalian extracellular matrix**. *Mol Microbiol* 1993, **9**(4):687-694.

95. Patti JM, Allen BL, McGavin MJ, Hook M: **MSCRAMM-mediated adherence of microorganisms to host tissues**. *Annu Rev Microbiol* 1994, **48**:585-617.

96. Joh D, Wann ER, Kreikemeyer B, Speziale P, Hook M: **Role of fibronectin-binding MSCRAMMs in bacterial adherence and entry into mammalian cells**. *Matrix Biol* 1999, **18**(3):211-223.

97. Rakonjac JV, Robbins JC, Fischetti VA: **DNA sequence of the serum opacity factor of group A streptococci: identification of a fibronectin-binding repeat domain**. *Infect Immun* 1995, **63**(2):622-631.

98. Courtney HS, Hasty DL, Dale JB: **Molecular mechanisms of adhesion, colonization, and invasion of group A streptococci**. *Ann Med* 2002, **34**(2):77-87.

99. Halaby DM, Mornon JP: **The immunoglobulin superfamily: an insight on its tissular, species, and functional diversity**. *J Mol Evol* 1998, **46**(4):389-400.

100. Bateman A, Eddy SR, Chothia C: **Members of the immunoglobulin superfamily in bacteria**. *Protein Sci* 1996, **5**(9):1939-1941.

101. Buts L, Bouckaert J, De Genst E, Loris R, Oscarson S, Lahmann M, Messens J, Brosens E, Wyns L, De Greve H: **The fimbrial adhesin F17-G of enterotoxigenic *Escherichia coli* has an immunoglobulin-like lectin domain that binds N-acetylglucosamine**. *Mol Microbiol* 2003, **49**(3):705-715.

102. Luo Y, Frey EA, Pfuetzner RA, Creagh AL, Knoechel DG, Haynes CA, Finlay BB, Strynadka NC: **Crystal structure of enteropathogenic *Escherichia coli* intimin-receptor complex**. *Nature* 2000, **405**(6790):1073-1077.

103. Hamburger ZA, Brown MS, Isberg RR, Bjorkman PJ: **Crystal structure of invasin: a bacterial integrin-binding protein**. *Science* 1999, **286**(5438):291-295.

104. Conway PL, Kjelleberg S: **Protein-mediated adhesion of *Lactobacillus fermentum* strain 737 to mouse stomach squamous epithelium**. *J Gen Microbiol* 1989, **135**(5):1175-1186.

105. Mukai T, Arihara K, Itoh H: **Lectin-like activity of *Lactobacillus acidophilus* strain JCM 1026**. *FEMS Microbiol Lett* 1992, **77**(1-3):71-74.

106. Roos S, Jonsson H: **A high-molecular-mass cell-surface protein from *Lactobacillus reuteri* 1063 adheres to mucus components**. *Microbiology (Reading, England)* 2002, **148**(Pt 2):433-442.

107. Pretzer G, Snel J, Molenaar D, Wiersma A, Bron PA, Lambert J, de Vos WM, van der Meer R, Smits MA, Kleerebezem M: **Biodiversity-based identification and functional characterization of the mannose-specific adhesin of *Lactobacillus plantarum***. *J Bacteriol* 2005, **187**(17):6128-6136.

108. Buck BL, Altermann E, Svingerud T, Klaenhammer TR: **Functional analysis of putative adhesion factors in *Lactobacillus acidophilus* NCFM**. *Appl Environ Microbiol* 2005, **71**(12):8344-8351.

109. Claesson MJ, Li Y, Leahy S, Canchaya C, van Pijkeren JP, Cerdeno-Tarraga AM, Parkhill J, Flynn S, O'Sullivan GC, Collins JK *et al*: **Multireplicon genome architecture of *Lactobacillus salivarius***. *Proc Natl Acad Sci U S A* 2006, **103**(17):6718-6723.

110. van Pijkeren JP, Canchaya C, Ryan KA, Li Y, Claesson MJ, Sheil B, Steidler L, O'Mahony L, Fitzgerald GF, van Sinderen D *et al*: **Comparative and functional analysis of sortase-dependent proteins in the predicted secretome of *Lactobacillus salivarius* UCC118**. *Appl Environ Microbiol* 2006, **72**(6):4143-4153.

111. Boekhorst J, Wels M, Kleerebezem M, Siezen RJ: **The predicted secretome of *Lactobacillus plantarum* WCFS1 sheds light on interactions with its environment**. *Microbiology (Reading, England)* 2006, **152**(Pt 11):3175-3183.

112. Velez MP, De Keersmaecker SC, Vanderleyden J: **Adherence factors of *Lactobacillus* in the human gastrointestinal tract**. *FEMS Microbiol Lett* 2007, **276**(2):140-148.

113. Granato D, Bergonzelli GE, Pridmore RD, Marvin L, Rouvet M, Corthesy-Theulaz IE: **Cell surface-associated elongation factor Tu mediates the attachment of *Lactobacillus johnsonii* NCC533 (La1) to human intestinal cells and mucins**. *Infect Immun* 2004, **72**(4):2160-2169.

114. Bergonzelli GE, Granato D, Pridmore RD, Marvin-Guy LF, Donnicola D, Corthesy-Theulaz IE: **GroEL of *Lactobacillus johnsonii* La1 (NCC 533) is cell surface associated: potential role in interactions with the host and the gastric pathogen *Helicobacter pylori***. *Infect Immun* 2006, **74**(1):425-434.

115. Roos S, Aleljung P, Robert N, Lee B, Wadstrom T, Lindberg M, Jonsson H: **A collagen binding protein from *Lactobacillus reuteri* is part of an ABC transporter system?** *FEMS Microbiol Lett* 1996, **144**(1):33-38.

116. Miyoshi Y, Okada S, Uchimura T, Satoh E: **A mucus adhesion promoting protein, MapA, mediates the adhesion of *Lactobacillus reuteri* to Caco-2 human intestinal epithelial cells**. *Biosci Biotechnol Biochem* 2006, **70**(7):1622-1628.

117. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL *et al*: **The Pfam protein families database**. *Nucleic Acids Res* 2004, **32**(Database issue):D138-141.

118. Rojas M, Ascencio F, Conway PL: **Purification and characterization of a surface protein from *Lactobacillus fermentum* 104R that binds to porcine small intestinal mucus and gastric mucin**. *Appl Environ Microbiol* 2002, **68**(5):2330-2336.

119. Lorca G, Torino MI, Font de Valdez G, Ljungh AA: **Lactobacilli express cell surface proteins which mediate binding of immobilized collagen and fibronectin**. *FEMS Microbiol Lett* 2002, **206**(1):31-37.

120. Aleljung P, Shen W, Rozalska B, Hellman U, Ljungh A, Wadstrom T: **Purification of collagen-binding proteins of *Lactobacillus reuteri* NCIB 11951**. *Curr Microbiol* 1994, **28**(4):231-236.

121. Authier A: **Characterisation of adhesion of a probiotic bacterium *Lactobacillus rhamnosus* HN001 to extracellular matrix proteins and intestinal cell line Caco-2**. *MSc thesis* Palmerston North: Massey University; 2003.

122. Dallo SF, Kannan TR, Blaylock MW, Baseman JB: **Elongation factor Tu and E1 beta subunit of pyruvate dehydrogenase complex act as fibronectin binding proteins in *Mycoplasma pneumoniae***. *Mol Microbiol* 2002, **46**(4):1041-1051.

123. Kinoshita H, Uchida H, Kawai Y, Kawasaki T, Wakahara N, Matsuo H, Watanabe M, Kitazawa H, Ohnuma S, Miura K *et al*: **Cell surface *Lactobacillus plantarum* LA 318 glyceraldehyde-3-phosphate dehydrogenase (GAPDH) adheres to human colonic mucin**. *J Appl Microbiol* 2008, **104**(6):1667-1674.

124. Hurmalainen V, Edelman S, Antikainen J, Baumann M, Lahteenmaki K, Korhonen TK: **Extracellular proteins of *Lactobacillus crispatus* enhance activation of human plasminogen**. *Microbiology (Reading, England)* 2007, **153**(Pt 4):1112-1122.

125. Pancholi V, Fischetti VA: **A major surface protein on group A streptococci is a glyceraldehyde-3-phosphate-dehydrogenase with multiple binding activity**. *J Exp Med* 1992, **176**(2):415-426.

126. Gozalbo D, Gil-Navarro I, Azorin I, Renau-Piqueras J, Martinez JP, Gil ML: **The cell wall-associated glyceraldehyde-3-phosphate dehydrogenase of *Candida albicans* is also a fibronectin and laminin binding protein**. *Infect Immun* 1998, **66**(5):2052-2059.

127. Granato D, Perotti F, Masserey I, Rouvet M, Golliard M, Servin A, Brassart D: **Cell surface-associated lipoteichoic acid acts as an adhesion factor for attachment of *Lactobacillus johnsonii* La1 to human enterocyte-like Caco-2 cells**. *Appl Environ Microbiol* 1999, **65**(3):1071-1077.

128. Ruas-Madiedo P, Gueimonde M, Margolles A, de los Reyes-Gavilan CG, Salminen S: **Exopolysaccharides produced by probiotic strains modify the adhesion of probiotics and enteropathogens to human intestinal mucus**. *J Food Prot* 2006, **69**(8):2011-2015.

129. Lamont RJ, Jenkinson HF: **Life below the gum line: pathogenic mechanisms of *Porphyromonas gingivalis***. *Microbiol Mol Biol Rev* 1998, **62**(4):1244-1263.

130. Orth K, Xu Z, Mudgett MB, Bao ZQ, Palmer LE, Bliska JB, Mangel WF, Staskawicz B, Dixon JE: **Disruption of signaling by *Yersinia* effector YopJ, a ubiquitin-like protein protease**. *Science* 2000, **290**(5496):1594-1597.

131. Schwarz-Linek U, Hook M, Potts JR: **Fibronectin-binding proteins of gram-positive cocci**. *Microbes Infect* 2006, **8**(8):2291-2298.

132. Lipsitch M, O'Hagan JJ: **Patterns of antigenic diversity and the mechanisms that maintain them**. *J R Soc Interface* 2007, **4**(16):787-802.

133. Bayliss CD, Field D, Moxon ER: **The simple sequence contingency loci of *Haemophilus influenzae* and *Neisseria meningitidis***. *J Clin Invest* 2001, **107**(6):657-662.

134. Areschoug T, Carlsson F, Stalhammar-Carlemalm M, Lindahl G: **Host-pathogen interactions in *Streptococcus pyogenes* infections, with special reference to puerperal fever and a comment on vaccine development**. *Vaccine* 2004, **22 Suppl 1**:S9-S14.

135. Moxon ER, Rainey PB, Nowak MA, Lenski RE: **Adaptive evolution of highly mutable loci in pathogenic bacteria**. *Curr Biol* 1994, **4**(1):24-33.

136. Maione D, Margarit I, Rinaudo CD, Masignani V, Mora M, Scarselli M, Tettelin H, Brettoni C, Iacobini ET, Rosini R *et al*: **Identification of a universal Group B streptococcus vaccine by multiple genome screen**. *Science* 2005, **309**:148-150.

137. Economou A: **Bacterial secretome: the assembly manual and operating instructions (Review)**. *Mol Membr Biol* 2002, **19**(3):159-169.

138. Tjalsma H, Bolhuis A, Jongbloed JD, Bron S, van Dijl JM: **Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome**. *Microbiol Mol Biol Rev* 2000, **64**(3):515-547.

139. Economou A, Christie PJ, Fernandez RC, Palmer T, Plano GV, Pugsley AP: **Secretion by numbers: Protein traffic in prokaryotes**. *Mol Microbiol* 2006, **62**(2):308-319.

140. Xie K, Dalbey RE: **Inserting proteins into the bacterial cytoplasmic membrane using the Sec and YidC translocases**. *Nat Rev Microbiol* 2008, **6**(3):234-244.

141. Samuelson JC, Chen M, Jiang F, Moller I, Wiedmann M, Kuhn A, Phillips GJ, Dalbey RE: **YidC mediates membrane protein insertion in bacteria**. *Nature* 2000, **406**(6796):637-641.

142. Lee PA, Tullman-Ercek D, Georgiou G: **The bacterial twin-arginine translocation pathway**. *Annu Rev Microbiol* 2006, **60**:373-395.

143. Wang IN, Smith DL, Young R: **Holins: The protein clocks of bacteriophage infections**. *Annu Rev Microbiol* 2000, **54**:799-825.

144. Desvaux M, Hebraud M: **The protein secretion systems in *Listeria*: inside out bacterial virulence**. *FEMS Microbiol Rev* 2006, **30**(5):774-805.

145. Gerlach RG, Hensel M: **Protein secretion systems and adhesins: the molecular armory of Gram-negative pathogens**. *Int J Med Microbiol* 2007, **297**(6):401-415.

146. Nouwen N, Ranson N, Saibil H, Wolpensinger B, Engel A, Ghazi A, Pugsley AP: **Secretin PulD: Association with pilot PulS, structure, and ion-conducting channel formation**. *Proc Natl Acad Sci U S A* 1999, **96**(14):8173-8177.

147. Klauser T, Pohlner J, Meyer TF: **The Secretion Pathway of Iga Protease-Type Proteins in Gram-Negative Bacteria**. *Bioessays* 1993, **15**(12):799-805.

148. Forman S, Wulff CR, Myers-Morales T, Cowan C, Perry RD, Straley SC: **yadBC of *Yersinia pestis*, a new virulence determinant for bubonic plague**. *Infect Immun* 2008, **76**(2):578-587.

149. Backert S, Selbach M: **Role of type IV secretion in *Helicobacter pylori* pathogenesis**. *Cell Microbiol* 2008, **10**(8):1573-1581.

150. Li JX, Wolf SG, Elbaum M, Tzfira T: **Exploring cargo transport mechanics in the type IV secretion systems**. *Trends Microbiol* 2005, **13**(7):295-298.

151. Farizo KM, Fiddner S, Cheung AM, Burns DL: **Membrane localization of the S1 subunit of pertussis toxin in *Bordetella pertussis* and implications for pertussis toxin secretion**. *Infect Immun* 2002, **70**(3):1193-1201.

152. Christie PJ, Atmakuri K, Krishnamoorthy V, Jakubowski S, Cascales E: **Biogenesis, architecture, and function of bacterial type IV secretion systems**. *Annu Rev Microbiol* 2005, **59**:451-485.

153. Pukatzki S, Ma AT, Sturtevant D, Krastins B, Sarracino D, Nelson WC, Heidelberg JF, Mekalanos JJ: **Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* host model system**. *Proc Natl Acad Sci U S A* 2006, **103**(5):1528-1533.

154. Abdallah AM, Gey van Pittius NC, Champion PA, Cox J, Luirink J, Vandenbroucke-Grauls CM, Appelmelk BJ, Bitter W: **Type VII secretion--mycobacteria show the way**. *Nat Rev Microbiol* 2007, **5**(11):883-891.

155. Schneewind O, Model P, Fischetti VA: **Sorting of protein A to the staphylococcal cell wall**. *Cell* 1992, **70**(2):267-281.

156. Pallen MJ: **The ESAT-6/WXG100 superfamily -- and a new Gram-positive secretion system?** *Trends Microbiol* 2002, **10**(5):209-212.

157. Gey Van Pittius NC, Gamieldien J, Hide W, Brown GD, Siezen RJ, Beyers AD: **The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C Gram-positive bacteria**. *Genome Biol* 2001, **2**(10):RESEARCH0044.

158. Bardy SL, Eichler J, Jarrell KF: **Archaeal signal peptides--a comparative survey at the genome level**. *Protein Sci* 2003, **12**(9):1833-1843.

159. Nielsen H, Engelbrecht J, Brunak S, von Heijne G: **Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites**. *Protein Eng* 1997, **10**(1):1-6.

160. Palmer T, Berks BC: **Moving folded proteins across the bacterial cell membrane**. *Microbiology (Reading, England)* 2003, **149**(Pt 3):547-556.

161. van Roosmalen ML, Geukens N, Jongbloed JD, Tjalsma H, Dubois JY, Bron S, van Dijl JM, Anne J: **Type I signal peptidases of Gram-positive bacteria**. *Biochim Biophys Acta* 2004, **1694**(1-3):279-297.

162. Luirink J, von Heijne G, Houben E, de Gier JW: **Biogenesis of inner membrane proteins in *Escherichia coli***. *Annu Rev Microbiol* 2005, **59**:329-355.

163. Perlman D, Halvorson HO: **A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides**. *J Mol Biol* 1983, **167**(2):391-409.

164. Tjalsma H, Antelmann H, Jongbloed JD, Braun PG, Darmon E, Dorenbos R, Dubois JY, Westers H, Zanen G, Quax WJ *et al*: **Proteomics of protein secretion by *Bacillus subtilis*: separating the "secrets" of the secretome**. *Microbiol Mol Biol Rev* 2004, **68**(2):207-233.

165. Chang CN, Rey M, Bochner B, Heyneker H, Gray G: **High-level secretion of human growth hormone by *Escherichia coli***. *Gene* 1987, **55**(2-3):189-196.

166. Smith GP: **Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface**. *Science* 1985, **228**:1315-1317.

167. Papanikou E, Karamanou S, Economou A: **Bacterial protein secretion through the translocase nanomachine**. *Nat Rev Microbiol* 2007, **5**(11):839-851.

168. Zhou M, Boekhorst J, Francke C, Siezen RJ: **LocateP: genome-scale subcellular-location predictor for bacterial proteins**. *BMC Bioinformatics* 2008, **9**:173.

169. Tjalsma H, van Dijl JM: **Proteomics-based consensus prediction of protein retention in a bacterial membrane**. *Proteomics* 2005, **5**(17):4472-4482.

170. Desvaux M, Dumas E, Chafsey I, Hebraud M: **Protein cell surface display in Gram-positive bacteria: from single protein to macromolecular protein structure**. *FEMS Microbiol Lett* 2006, **256**(1):1-15.

171. Maier KS, Hubich S, Liebhart H, Krauss S, Kuhn A, Facey SJ: **An amphiphilic region in the cytoplasmic domain of KdpD is recognized by the signal recognition particle and targeted to the *Escherichia coli* membrane**. *Mol Microbiol* 2008, **68**(6):1471-1484.

172. Tjalsma H, Kontinen VP, Pragai Z, Wu H, Meima R, Venema G, Bron S, Sarvas M, van Dijl JM: **The role of lipoprotein processing by signal peptidase II in the Gram-positive eubacterium *Bacillus subtilis*. Signal peptidase II is required for the efficient secretion of alpha-amylase, a non-lipoprotein**. *J Biol Chem* 1999, **274**(3):1698-1707.

173. Sutcliffe IC, Russell RR: **Lipoproteins of gram-positive bacteria**. *J Bacteriol* 1995, **177**(5):1123-1128.

174. Navarre WW, Schneewind O: **Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope**. *Microbiol Mol Biol Rev* 1999, **63**(1):174-229.

175. Comfort D, Clubb RT: **A comparative genome analysis identifies distinct sorting pathways in gram-positive bacteria**. *Infect Immun* 2004, **72**(5):2710-2722.

176. Dramsi S, Magnet S, Davison S, Arthur M: **Covalent attachment of proteins to peptidoglycan**. *FEMS microbiology reviews* 2008, **32**(2):307-320.

177. Fernandez-Tornero C, Lopez R, Garcia E, Gimenez-Gallego G, Romero A: **A novel solenoid fold in the cell wall anchoring domain of the pneumococcal virulence factor LytA**. *Nat Struct Biol* 2001, **8**(12):1020-1024.

178. Rosenow C, Ryan P, Weiser JN, Johnson S, Fontan P, Ortqvist A, Masure HR: **Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of *Streptococcus pneumoniae***. *Mol Microbiol* 1997, **25**(5):819-829.

179. Waligora AJ, Hennequin C, Mullany P, Bourlioux P, Collignon A, Karjalainen T: **Characterization of a cell surface protein of *Clostridium difficile* with adhesive properties**. *Infect Immun* 2001, **69**(4):2144-2153.

180. Joris B, Englebert S, Chu CP, Kariyama R, Daneo-Moore L, Shockman GD, Ghuysen JM: **Modular design of the *Enterococcus hirae* muramidase-2 and *Streptococcus faecalis* autolysin**. *FEMS Microbiol Lett* 1992, **70**(3):257-264.

181. Goward CR, Scawen MD, Murphy JP, Atkinson T: **Molecular evolution of bacterial cell-surface proteins**. *Trends Biochem Sci* 1993, **18**(4):136-140.

182. Jonquieres R, Bierne H, Fiedler F, Gounon P, Cossart P: **Interaction between the protein InlB of *Listeria monocytogenes* and lipoteichoic acid: a novel mechanism of protein association at the surface of gram-positive bacteria**. *Mol Microbiol* 1999, **34**(5):902-914.

183. Buist G, Kok J, Leenhouts KJ, Dabrowska M, Venema G, Haandrikman AJ: **Molecular cloning and nucleotide sequence of the gene encoding the major peptidoglycan hydrolase of *Lactococcus lactis*, a muramidase needed for cell separation**. *J Bacteriol* 1995, **177**(6):1554-1563.

184. Leenhouts K, Buist G, Kok J: **Anchoring of proteins to lactic acid bacteria**. *Antonie Van Leeuwenhoek* 1999, **76**(1-4):367-376.

185. Wall T, Roos S, Jacobsson K, Rosander A, Jonsson H: **Phage display reveals 52 novel extracellular and transmembrane proteins from *Lactobacillus reuteri* DSM 20016(T)**. *Microbiology (SGM)* 2003, **149**(Pt 12):3493-3505.

186. Bath K, Roos S, Wall T, Jonsson H: **The cell surface of *Lactobacillus reuteri* ATCC 55730 highlighted by identification of 126 extracellular proteins from the genome sequence**. *FEMS Microbiol Lett* 2005, **253**(1):75-82.

187. Hartford O, Francois P, Vaudaux P, Foster TJ: **The dipeptide repeat region of the fibrinogen-binding protein (clumping factor) is required for functional expression of the fibrinogen-binding domain on the *Staphylococcus aureus* cell surface**. *Mol Microbiol* 1997, **25**(6):1065-1076.

188. Sturme MH, Francke C, Siezen RJ, de Vos WM, Kleerebezem M: **Making sense of quorum sensing in lactobacilli: a special focus on *Lactobacillus plantarum* WCFS1**. *Microbiology (Reading, England)* 2007, **153**(Pt 12):3939-3947.

189. Frost LS, Ippen-Ihler K, Skurray RA: **Analysis of the sequence and gene products of the transfer region of the F sex factor**. *Microbiol Rev* 1994, **58**(2):162-210.

190. Marvin DA, Hohn B: **Filamentous bacterial viruses**. *Bacteriol Rev* 1969, **33**(2):172-209.

191. Russel M, Model P: **Genetic analysis of the filamentous bacteriophage packaging signal and of the proteins that interact with it**. *J Virol* 1989, **63**(8):3284-3295.

192. Kehoe JW, Kay BK: **Filamentous phage display in the new millennium**. *Chem Rev* 2005, **105**(11):4056-4072.

193. Russel M, Lowman HB, Clackson T: **Introduction to phage biology and phage display**. In: *Practical Approach to Phage Display.* Edited by Clackson T, Lowman HB. New York: Oxford University Press, Inc.; 2004: 1-26.

194. Rakonjac J, Conway JF: **Bacteriophages: self-assembly and applications**. In: *Molecular Bionanotechnology.* Edited by Rehm B. Norwich, UK: Horizon Scientific Press; 2006: 153-190.

195. Russel M, Model P: **Filamentous Phage**. In: *The Bacteriophages (Second Edition).* Edited by Calendar RC. New York: Oxford University Press, Inc.; 2006: 146-160.

196. Parmley SF, Smith GP: **Antibody-selectable filamentous fd phage vectors: affinity purification of target genes**. *Gene* 1988, **73**(2):305-318.

197. Russel M: **Moving through the membrane with filamentous phages**. *Trends Microb* 1995, **3**(6):223-228.

198. Rakonjac J, Feng J-n, Model P: **Filamentous phage are released from the bacterial membrane by a two-step mechanism involving a short C-terminal fragment of pIII**. *J Mol Biol* 1999, **289**:1253-1265.

199. Rakonjac J, Model P: **The roles of pIII in filamentous phage assembly**. *J Mol Biol* 1998, **282**:25-41.

200. Russel M: **Protein-protein interactions during filamentous phage assembly**. *J Mol Biol* 1993, **231**(3):689-697.

201. Rodi DJ, Makowski L: **Phage-display technology--finding a needle in a vast molecular haystack**. *Curr Opin Biotechnol* 1999, **10**(1):87-93.

202. Rakonjac J, Jovanovic G, Model P: **Filamentous phage infection-mediated gene expression: construction and propagation of the *gIII* deletion mutant helper phage R408d3**. *Gene* 1997, **198**:99-103.

203. Lei SP, Lin HC, Wang SS, Callaway J, Wilcox G: **Characterization of the *Erwinia carotovora* pelB gene and its product pectate lyase**. *J Bacteriol* 1987, **169**(9):4379-4383.

204. Bendtsen JD, Nielsen H, von Heijne G, Brunak S: **Improved prediction of signal peptides: SignalP 3.0**. *J Mol Biol* 2004, **340**:783-795.

205. Chen Y, Yu P, Luo J, Jiang Y: **Secreted protein prediction system combining CJ-SPHMM, TMHMM, and PSORT**. *Mamm Genome* 2003, **14**(12):859-865.

206. Gardy JL, Brinkman FS: **Methods for predicting bacterial protein subcellular localization**. *Nat Rev Microbiol* 2006, **4**(10):741-751.

207. Gardy JL, Laird MR, Chen F, Rey S, Walsh CJ, Ester M, Brinkman FS: **PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis**. *Bioinformatics (Oxford, England)* 2005, **21**(5):617-623.

208. Juncker AS, Willenbrock H, Von Heijne G, Brunak S, Nielsen H, Krogh A: **Prediction of lipoprotein signal peptides in Gram-negative bacteria**. *Protein Sci* 2003, **12**(8):1652-1662.

209. Nakai K, Horton P: **PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization**. *Trends Biochem Sci* 1999, **24**(1):34-36.

210. Roosild TP, Greenwald J, Vega M, Castronovo S, Riek R, Choe S: **NMR structure of Mistic, a membrane-integrating protein for membrane protein expression**. *Science* 2005, **307**(5713):1317-1321.

211. Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, Wu D, Eisen JA, Hoffman JM, Remington K *et al*: **The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific**. *PLoS Biol* 2007, **5**(3):e77.

212. Benhar I: **Biotechnological applications of phage and cell display**. *Biotechnol Adv* 2001, **19**(1):1-33.

213. Wernerus H, Stahl S: **Biotechnological applications for surface-engineered bacteria**. *Biotechnol Appl Biochem* 2004, **40**(Pt 3):209-228.

214. Manoil C, Mekalanos JJ, Beckwith J: **Alkaline phosphatase fusions: sensors of subcellular location**. *J Bacteriol* 1990, **172**(2):515-518.

215. Broome-Smith JK, Tadayyon M, Zhang Y: **Beta-lactamase as a probe of membrane protein assembly and protein export**. *Mol Microbiol* 1990, **4**(10):1637-1644.

216. Avall-Jaaskelainen S, Lindholm A, Palva A: **Surface display of the receptor-binding region of the *Lactobacillus brevis* S-layer protein in *Lactococcus lactis* provides nonadhesive lactococci with the ability to adhere to intestinal epithelial cells**. *Appl Environ Microbiol* 2003, **69**(4):2230-2236.

217. Downing KJ, McAdam RA, Mizrahi V: ***Staphylococcus aureus* nuclease is a useful secretion reporter for mycobacteria**. *Gene* 1999, **239**(2):293-299.

218. Poquet I, Ehrlich SD, Gruss A: **An export-specific reporter designed for gram-positive bacteria: application to *Lactococcus lactis***. *J Bacteriol* 1998, **180**(7):1904-1912.

219. Lunder M, Bratkovic T, Doljak B, Kreft S, Urleb U, Strukelj B, Plazar N: **Comparison of bacterial and phage display peptide libraries in search of target-binding motif**. *Appl Biochem Biotechnol* 2005, **127**(2):125-131.

220. Irving MB, Pan O, Scott JK: **Random-peptide libraries and antigen-fragment libraries for epitope mapping and the development of vaccines and diagnostics**. *Current opinion in chemical biology* 2001, **5**(3):314-324.

221. Mullen LM, Nair SP, Ward JM, Rycroft AN, Henderson B: **Phage display in the study of infectious diseases**. *Trends Microbiol* 2006, **14**(3):141-147.

222. Pasqualini R, Ruoslahti E: **Organ targeting *in vivo* using phage display peptide libraries**. *Nature* 1996, **380**(6572):364-366.

223. Forrer P, Jung S, Pluckthun A: **Beyond binding: using phage display to select for structure, folding and enzymatic activity in proteins**. *Curr Opin Struct Biol* 1999, **9**(4):514-520.

224. Jacobsson K, Frykberg L: **Phage display shot-gun cloning of ligand-binding domains of prokaryotic receptors approaches 100% correct clones**. *Biotechniques* 1996, **20**(6):1070-1076, 1078, 1080-1071.

225. Jacobsson K, Rosander A, Bjerketorp J, Frykberg L: **Shotgun Phage Display - Selection for Bacterial Receptins or other Exported Proteins**. *Biol Proced Online* 2003, **5**:123-135.

226. Ausmees N, Jacobsson K, Lindberg M: **A unipolarly located, cell-surface-associated agglutinin, RapA, belongs to a family of Rhizobium-adhering proteins (Rap) in *Rhizobium leguminosarum* bv. *trifolii***. *Microbiology (Reading, England)* 2001, **147**(Pt 3):549-559.

227. Benedek O, Salam Khan A, Schneider G, Nagy G, Autar R, Pieters RJ, Emody L: **Identification of laminin-binding motifs of *Yersinia pestis* plasminogen activator by phage display**. *Int J Med Microbiol* 2005, **295**(2):87-98.

228. Lauterbach SB, Lanzillotti R, Coetzer TL: **Construction and use of *Plasmodium falciparum* phage display libraries to identify host parasite interactions**. *Malaria J* 2003, **2**:1-4.

229. Antonara S, Chafel RM, LaFrance M, Coburn J: ***Borrelia burgdorferi* adhesins identified using *in vivo* phage display**. *Mol Microbiol* 2007, **66**(1):262-276.

230. Lindmark H, Jacobsson K, Frykberg L, Guss B: **Fibronectin-binding protein of *Streptococcus equi* subsp. *zooepidemicus***. *Infect Immun* 1996, **64**(10):3993-3999.

231. Zhang L, Jacobsson K, Strom K, Lindberg M, Frykberg L: ***Staphylococcus aureus* expresses a cell surface protein that binds both IgG and beta2-glycoprotein I**. *Microbiology (Reading, England)* 1999, **145 (Pt 1)**:177-183.

232. Sambrook J, Fritsch EF, Maniatis T: **Molecular Cloning - A Laboratory Manual**, 2nd edn; 1989.

233. Gibson TJ: **Studies on the Epstein-Barr virus genome**. *PhD thesis, University of Cambridge, UK* 1984.

234. Krebber A, Burmester J, Pluckthun A: **Inclusion of an upstream transcriptional terminator in phage display vectors abolishes background expression of toxic fusions with coat protein g3p**. *Gene* 1996, **178**(1-2):71-74.

235. Lessl M, Balzer D, Lurz R, Waters VL, Guiney DG, Lanka E: **Dissection of IncP conjugative plasmid transfer: definition of the transfer region Tra2 by mobilization of the Tra1 region in trans**. *J Bacteriol* 1992, **174**(8):2493-2500.

236. Birnboim HC, Doly J: **A rapid alkaline extraction procedure for screening recombinant plasmid DNA**. *Nucleic Acids Res* 1979, **7**(6):1513-1523.

237. Nelson FK, Friedman SM, Smith GP: **Filamentous phage DNA cloning vectors: a noninfective mutant with a nonpolar deletion in gene III**. *Virology* 1981, **108**(2):338-350.

238. Schagger H, von Jagow G: **Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa**. *Anal Biochem* 1987, **166**(2):368-379.

239. Tesar M, Beckmann C, Rottgen P, Haase B, Faude U, Timmis KN: **Monoclonal antibody against pIII of filamentous phage: an immunological tool to study pIII fusion protein expression in phage display systems**. *Immunotechnology* 1995, **1**(1):53-64.

240. Harlow E, Lane D: **Using antibodies : a laboratory manual**. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press; 1999.

241. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL: **Improved microbial gene identification with GLIMMER**. *Nucleic Acids Res* 1999, **27**(23):4636-4641.

242. Voltan S, Castagliuolo I, Elli M, Longo S, Brun P, D'Inca R, Porzionato A, Macchi V, Palu G, Sturniolo GC *et al*: **Aggregating phenotype in *Lactobacillus crispatus* determines intestinal colonization and TLR2 and TLR4 modulation in murine colonic mucosa**. *Clin Vaccine Immunol* 2007, **14**(9):1138-1148.

243. Sonnhammer EL, von Heijne G, Krogh A: **A hidden Markov model for predicting transmembrane helices in protein sequences**. *Proceedings / International Conference on Intelligent Systems for Molecular Biology ; ISMB* 1998, **6**:175-182.

244. Dotto GP, Enea V, Zinder ND: **Functional analysis of bacteriophage f1 intergenic region**. *Virology* 1981, **114**(2):463-473.

245. Taylor PD, Attwood TK, Flower DR: **BPROMPT: A consensus server for membrane protein prediction**. *Nucleic Acids Res* 2003, **31**(13):3698-3700.

246. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool**. *J Mol Biol* 1990, **215**(3):403-410.

247. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN *et al*: **The COG database: an updated version includes eukaryotes**. *BMC Bioinformatics* 2003, **4**:41.

248. Letunic I, Copley RR, Pils B, Pinkert S, Schultz J, Bork P: **SMART 5: domains in the context of genomes and networks**. *Nucleic Acids Res* 2006, **34**(Database issue):D257-260.

249. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R: **InterProScan: protein domains identifier**. *Nucleic Acids Res* 2005, **33**(Web Server issue):W116-120.

250. Model P, Jovanovic G, Dworkin J: **The *Escherichia coli* phage shock protein operon**. *Mol Microbiol* 1997, **24**:255-261.

251. Courtney HS, Zhang YM, Frank MW, Rock CO: **Serum opacity factor, a streptococcal virulence factor that binds to apolipoproteins A-I and A-II and disrupts high density lipoprotein structure**. *J Biol Chem* 2006, **281**(9):5515-5521.

252. Maxted WR, Widdowson JP, Fraser CA: **Antibody to streptococcal opacity factor in human sera**. *J Hyg (Lond)* 1973, **71**(1):35-42.

253. Castagliuolo I, Galeazzi F, Ferrari S, Elli M, Brun P, Cavaggioni A, Tormen D, Sturniolo GC, Morelli L, Palu G: **Beneficial effect of auto-aggregating *Lactobacillus crispatus* on experimentally induced colitis in mice**. *FEMS Immunol Med Microbiol* 2005, **43**(2):197-204.

254. Smit G, Swart S, Lugtenberg BJ, Kijne JW: **Molecular mechanisms of attachment of *Rhizobium* bacteria to plant roots**. *Mol Microbiol* 1992, **6**(20):2897-2903.

255. Pankov R, Yamada KM: **Fibronectin at a glance**. *J Cell Sci* 2002, **115**(Pt 20):3861-3863.

256. Rosander A, Bjerketorp J, Frykberg L, Jacobsson K: **Phage display as a novel screening method to identify extracellular proteins**. *J Microbiol Methods* 2002, **51**(1):43-55.

257. Bradbury A: **Diversity by design**. *Trends Biotechnol* 1998, **16**:99-102.

258. Beekwilder J, Rakonjac J, Jongsma M, Bosch D: **A phagemid vector using the *E. coli* phage shock promoter facilitates phage display of toxic proteins**. *Gene* 1999, **228**(1-2):23-31.

259. Lee SY, Choi JH, Xu Z: **Microbial cell-surface display**. *Trends Biotechnol* 2003, **21**(1):45-52.

260. Georgiou G, Stathopoulos C, Daugherty PS, Nayak AR, Iverson BL, Curtiss R, 3rd: **Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines**. *Nat Biotechnol* 1997, **15**(1):29-34.

261. Vaughan EE, Heilig HG, Ben-Amor K, de Vos WM: **Diversity, vitality and activities of intestinal lactic acid bacteria and bifidobacteria assessed by molecular approaches**. *FEMS Microbiol Rev* 2005, **29**(3):477-490.

262. Stern NJ, Svetoch EA, Eruslanov BV, Perelygin VV, Mitsevich EV, Mitsevich IP, Pokhilenko VD, Levchuk VP, Svetoch OE, Seal BS: **Isolation of a *Lactobacillus salivarius* strain and purification of its bacteriocin, which is inhibitory to *Campylobacter jejuni* in the chicken gastrointestinal system**. *Antimicrob Agents Chemother* 2006, **50**(9):3111-3116.

263. Medellin-Pena MJ, Wang H, Johnson R, Anand S, Griffiths MW: **Probiotics affect virulence-related gene expression in *Escherichia coli* O157:H7**. *Appl Environ Microbiol* 2007, **73**(13):4259-4267.

264. O'Hara AM, O'Regan P, Fanning A, O'Mahony C, Macsharry J, Lyons A, Bienenstock J, O'Mahony L, Shanahan F: **Functional modulation of human intestinal epithelial cell responses by *Bifidobacterium infantis* and *Lactobacillus salivarius***. *Immunology* 2006, **118**(2):202-215.

265. Christensen HR, Frokiaer H, Pestka JJ: **Lactobacilli differentially modulate expression of cytokines and maturation surface markers in murine dendritic cells**. *J Immunol* 2002, **168**(1):171-178.

266. Mohamadzadeh M, Olson S, Kalina WV, Ruthel G, Demmin GL, Warfield KL, Bavari S, Klaenhammer TR: **Lactobacilli activate human dendritic cells that skew T cells toward T helper 1 polarization**. *Proc Natl Acad Sci U S A* 2005, **102**(8):2880-2885.

267. Boekhorst J, Siezen RJ, Zwahlen MC, Vilanova D, Pridmore RD, Mercenier A, Kleerebezem M, de Vos WM, Brussow H, Desiere F: **The complete genomes of *Lactobacillus plantarum* and *Lactobacillus johnsonii* reveal extensive differences in chromosome organization and gene content**. *Microbiology (SGM)* 2004, **150**(Pt 11):3601-3611.

268. Pridmore RD, Berger B, Desiere F, Vilanova D, Barretto C, Pittet AC, Zwahlen MC, Rouvet M, Altermann E, Barrangou R *et al*: **The genome sequence of the probiotic intestinal bacterium *Lactobacillus johnsonii* NCC 533**. *Proc Natl Acad Sci U S A* 2004, **101**(8):2512-2517.

269. Walter J, Chagnaud P, Tannock GW, Loach DM, Dal Bello F, Jenkinson HF, Hammes WP, Hertel C: **A high-molecular-mass surface protein (Lsp) and methionine sulfoxide reductase B (MsrB) contribute to the ecological performance of *Lactobacillus reuteri* in the murine gut**. *Appl Environ Microbiol* 2005, **71**(2):979-986.

270. Klaenhammer TR, Barrangou R, Buck BL, Azcarate-Peril MA, Altermann E: **Genomic features of lactic acid bacteria effecting bioprocessing and health**. *FEMS Microbiol Rev* 2005, **29**(3):393-409.

271. Tannock GW, Munro K, Harmsen HJ, Welling GW, Smart J, Gopal PK: **Analysis of the fecal microflora of human subjects consuming a probiotic product containing *Lactobacillus rhamnosus* DR20**. *Appl Environ Microbiol* 2000, **66**(6):2578-2588.

272. Bolotin A, Wincker P, Mauger S, Jaillon O, Malarme K, Weissenbach J, Ehrlich SD, Sorokin A: **The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403**. *Genome Res* 2001, **11**(5):731-753.

273. Brouwer RW, Kuipers OP, Hijum SA: **The relative value of operon predictions**. *Brief Bioinform* 2008.

274. Price MN, Arkin AP, Alm EJ: **The life-cycle of operons**. *PLoS Genet* 2006, **2**(6):e96.

275. Guo M, Teng M, Niu L, Liu Q, Huang Q, Hao Q: **Crystal structure of the cysteine-rich secretory protein stecrisp reveals that the cysteine-rich domain has a K+ channel inhibitor-like fold**. *J Biol Chem* 2005, **280**(13):12405-12412.

276. Szyperski T, Fernandez C, Mumenthaler C, Wuthrich K: **Structure comparison of human glioma pathogenesis-related protein GliPR and the plant pathogenesis-related protein P14a indicates a functional link between the human immune system and a plant defense system**. *Proc Natl Acad Sci U S A* 1998, **95**(5):2262-2266.

277. Fernandez C, Szyperski T, Bruyere T, Ramage P, Mosinger E, Wuthrich K: **NMR solution structure of the pathogenesis-related protein P14a**. *J Mol Biol* 1997, **266**(3):576-593.

278. Yeats C, Bentley S, Bateman A: **New knowledge from old: *in silico* discovery of novel protein domains in *Streptomyces coelicolor***. *BMC Microbiol* 2003, **3**:3.

279. Milne TJ, Abbenante G, Tyndall JD, Halliday J, Lewis RJ: **Isolation and characterization of a cone snail protease with homology to CRISP proteins of the pathogenesis-related protein superfamily**. *J Biol Chem* 2003, **278**(33):31105-31110.

280. Sebaihia M, Wren BW, Mullany P, Fairweather NF, Minton N, Stabler R, Thomson NR, Roberts AP, Cerdeno-Tarraga AM, Wang H *et al*: **The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome**. *Nat Genet* 2006, **38**(7):779-786.

281. Fraser JS, Yu Z, Maxwell KL, Davidson AR: **Ig-like domains on bacteriophages: a tale of promiscuity and deceit**. *J Mol Biol* 2006, **359**(2):496-507.

282. Borja Sánchez PBMCU: **Exported proteins in probiotic bacteria: adhesion to intestinal surfaces, host immunomodulation and molecular cross-talking with the host**. *FEMS Immunol Med Microbiol* 2008, **54**(1):1-17.

283. Barbas III CF, Burton DR, Scott JK, Silverman GJ: **Phage display: a laboratory manual**. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press; 2001.

284. Kagnoff MF, Eckmann L: **Epithelial cells as sensors for microbial infection**. *J Clin Invest* 1997, **100**(1):6-10.

285. Wallace TD, Bradley S, Buckley ND, Green-Johnson JM: **Interactions of lactic acid bacteria with human intestinal epithelial cells: effects on cytokine production**. *J Food Prot* 2003, **66**(3):466-472.

286. Haller D, Blum S, Bode C, Hammes WP, Schiffrin EJ: **Activation of human peripheral blood mononuclear cells by nonpathogenic bacteria *in vitro*: evidence of NK cells as primary targets**. *Infect Immun* 2000, **68**(2):752-759.

287. Haller D, Bode C, Hammes WP: **Cytokine secretion by stimulated monocytes depends on the growth phase and heat treatment of bacteria: a comparative study between lactic acid bacteria and invasive pathogens**. *Microbiol Immunol* 1999, **43**(10):925-935.

288. Haller D, Serrant P, Granato D, Schiffrin EJ, Blum S: **Activation of human NK cells by staphylococci and lactobacilli requires cell contact-dependent costimulation by autologous monocytes**. *Clin Diagn Lab Immunol* 2002, **9**(3):649-657.

289. Rosander A, Frykberg L, Ausmees N, Muller P: **Identification of extracytoplasmic proteins in *Bradyrhizobium japonicum* using phage display**. *Mol Plant Microbe Interact* 2003, **16**(8):727-737.

290. Albers SV, Szabo Z, Driessen AJ: **Protein secretion in the *Archaea*: multiple paths towards a unique cell surface**. *Nat Rev Microbiol* 2006, **4**(7):537-547.

291. Paschke M, Hohne W: **A twin-arginine translocation (Tat)-mediated phage display system**. *Gene* 2005, **350**(1):79-88.

APPENDIX I

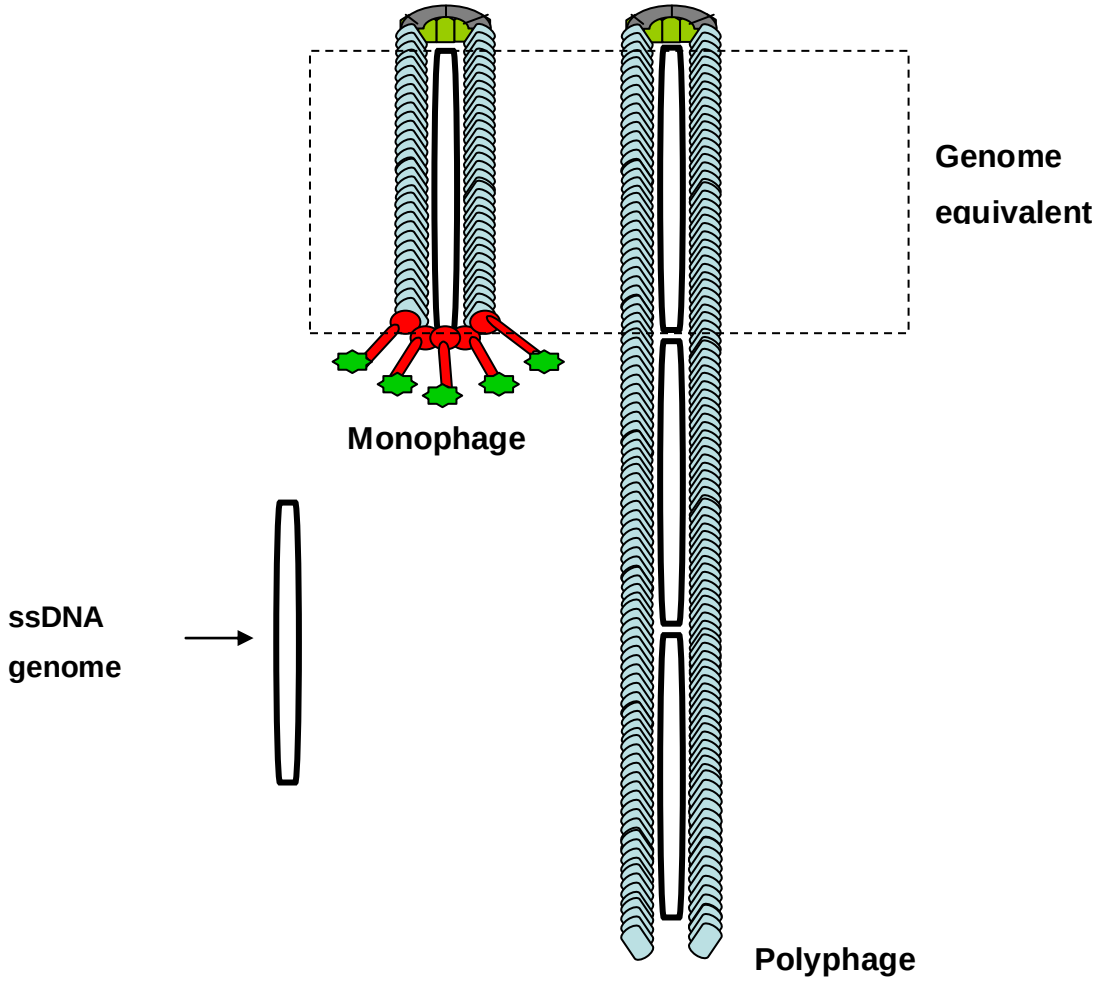


Figure A.1. Schematic diagram showing a genome equivalent

A monophage is a virion which contains only one ssDNA genome. A polyphage is a virion particle which carries multiple copies of ssDNA genome. A genome equivalent is a portion of a phage or phage like filament that contains one ssDNA genome

APPENDIX II

Table A.2. The pIII-targeting signals from *L. rhamnosus* HN001 and annotation of putative ORFs

Secretome ORF[a]	Targeting signal[b]	Length of displayed peptide (amino acids)	Signal P (NN) D value	Signal P (HMM)	LipPred (probability score)	Predicted signal sequence/anchor[f]	Protein description based on BLAST /InterProScan/SMART/Glimmer analyses	Pfam or COG classification	Accession number
Lrh01	SS	259	0.692	1		Mykagkhwlftfstltflaggagf vfsgnqavha/dsn	Unique hypothetical extracellular protein		EU232015
Lrh02	SS	436	0.55	0.997		Mkifkskrstnahparnrkffgigl igftllvlggfvlr/fsf	Cell division protein FtsI/penicillin-binding protein 2 *Lactobacillus casei* ATCC 334	(pfam00905), Transpeptidase, Penicillin binding protein transpeptidase domain; (pfam031717), Penicillin-binding protein dimerisation domain.	EU232016
Lrh03	SS	190	0.851	1		Mkkkvfiitclavmismvlgplq avga/ada	Conserved hypothetical extracellular protein		EU232017
Lrh04	SS	399	0.644	1		Mrrthfrswkkgkrwlfasslivtv igagalesgktvka/dsv	Unique hypothetical extracellular protein		EU232018
Lrh05	N-terminal TMH 5-27	386	0.631			Mklss*wilrvsliamvilslvftsli wr*np	Conserved hypothetical protein *Lactobacillus casei* ATCC 334		EU232019
Lrh06	SS	199	0.797	1		Mrkkllllgllaclsflwfgthqqvh a/aeg	Conserved hypothetical extracellular protein *Lactobacillus casei* ATCC 334		EU232020
Lrh07	SS	74	0.686	1		Mlmllnhallsikkmianicvlvlif pgqaaa/sts	Unique hypothetical extracellular protein		EU232021
Lrh08	N-terminal TMH 21-43	513	0.444			Mlynrvkehvvndkkkmyk*rg sqwvvvsafaialggvaftqvq*p vs	Possible cell surface protein *Lactobacillus casei* ATCC 334		EU232022
Lrh09	Internal TMH 112-134	164	0.163			amlsgifipiipviaatglflgl	Sucrose PTS, EIIBCA *Lactobacillus casei* ATCC 334	(COG1263) PtsG, Phosphotransferase system IIC components, glucose/maltose/N-acetylglucosamine-specific	EU232023

Lrh10	SS	210	0.826	1		Mkklikwlglaaaglalitavpqivra/dda	Predicted secreted protein *Lactobacillus casei* ATCC 334	(COG4086) Predicted secreted protein [Function unknown]	EU232024
Lrh11	SS	64	0.506	1		Mrhnnrhagylfranlkqnfkfsgiwlvilimmiisga/akl	Putative exporter of polyketide antibiotics *Lactobacillus reuteri*		EU232025
Lrh12	N-terminal TMH 20-39	152	0.505			Mkqtsesrrrqrppkrkwm*aplavlilglviiggyfyi*aqq	Unique hypothetical protein		EU232026
Lrh13	Internal TMH1-22	201	0.544			fllgcvfisgltlglawlssa	Membrane associated transglutaminase-like enzyme *Lactobacillus casei* ATCC 334		EU232027
Lrh14	LSS	269	0.626		0.99	Mkrrmimavmafltmagivisg/car	Extracellular solute-binding protein, family 3 *Lactobacillus reuteri* 100-23	(pfam00497) Bacterial extracellular solute-binding proteins, family 3	EU232028
Lrh15	N-terminal TMH 15-37	96	0.368			Mnadkgikrqqrrlf*igqllsfaglflvlgvivfffl*yer	Signal transduction histidine kinase *Lactobacillus casei* ATCC 334		EU232029
Lrh16	SS	166	0.885	1		Mkfnkammtlvaavtlagsvsavtpvfa/dts	Surface antigen YP_805328.1 *Lactobacillus casei* ATCC 334	(COG3883) Uncharacterized protein conserved in bacteria [Function unknown]	EU232030
Lrh17	SS	446	0.849	1		Mqvtfkkighsllaalmlmsfllpllsagktvha/att	Surface protein from Gram-positive cocci, anchor region: Cna B-type *Enterococcus faecium* DO	(pfam05738), Cna_B, Cna protein B-type domain; E-box motif (major pilin protein)	EU232031
Lrh18	N-terminal TMH 21-43	45	0.336			Mhsslgrfvrleiktnvlrt*sliavlgwglipivftltpqsl*aqa	Short sequence, no significant hits.		EU232032
Lrh19	SS	167	0.488	0.991		Mkaqeaihinkhkvplwavivlflvalgggvagg/yal	Carboxy-terminal processing proteinase *Lactobacillus casei* ATCC 334	(COG0793), Prc, Periplasmic protease [Cell envelope biogenesis, outer membrane]	EU232033
Lrh20	SS	412	0.942	1		Mklktklitlvvaflaaisfalpsqvna/akg	Putative lysin [Bacteriophage Lc-Nu]	(pfam01476), LysM	EU232034
Lrh21	SS	130	0.851	1		Mkkklfcfitlivmlimpatqvfa/sel	Cell surface protein *Lactobacillus casei* ATCC 334		EU232035
Lrh22	N-terminal TMH 17-39	355	0.476			Mlwkdirksfttswgr*ffsilllmmlgsfalvglwvag*pdr	Putative ABC transporter permease protein *Lactococcus lactis* subsp. *cremoris* MG1363		EU232036
Lrh23	SS	470	0.668	1		Mvrkrwtvigslllmaalmggdva/apv	N-acetylmuramoyl-L-alanine amidase/putative S-layer protein *Clostridium tetani* E88		EU232037
Lrh24	SS	142	0.653	0.94		Maaivvifglgfffthqqsasa/adn	ABC-type sugar transport system, substrate binding-component *Lactobacillus casei* ATCC 334		EU232038

Lrh25	SS	130	0.793	1		Mrlkqlvagaitvatlagvgvsgfaattvha/add	Surface antigen YP_807226.1 *Lactobacillus casei* ATCC 334		EU232039
Lrh26	SS	118	0.8	1		Mnhrswwllpllllgfglaglqpvqa/asg	Beta-propeller domains of methanol dehydrogenase type *Lactobacillus casei* ATCC 334	(COG1512) Beta-propeller domains of methanol dehydrogenase type	EU232040
Lrh27	N-terminal TMH 4-26	77	0.559	0.101		Mat*ltywlktywyapliiliggfffw*qq	DNA uptake protein related DNA-binding protein *Lactobacillus casei* ATCC 334		EU232041
Lrh28	LSS	228	0.562		1	Mkrlisiavailav/clg	Spermidine/putrescine ABC transporter, substrate binding protein *Lactobacillus plantarum* WCFS1	(pfam01547) SBP_bac_1, Bacterial extracellular solute-binding protein	EU232042
Lrh29	SS	316	0.561	1		Mkksiwivgllgwvmvlfsgmwqqsaqa/age	Conserved hypothetical protein EF0711 *Enterococcus faecalis* V583		EU232043
Lrh30	SS	163	0.848	1		Mkkfyvmasmlglgmlglagvtptlvsa/sdv	Cell surface protein *Lactobacillus casei* ATCC 334		EU232044
Lrh31	SS	52	0.842	1		Mkkwlllvvlligftcppsqala/vet	Unique hypothetical extracellular protein		EU232045
Lrh32	SS	63	0.85	1		Mkrkwllsllialplmlgalipgktvaa/aqk	Putative glutamine/glutamate ABC transporter, membrane-spanning/substrate-binding subunit precursor *Lactobacillus sakei* subsp. *sakei* 23K	(pfam00497), SBP_bac_3, Bacterial extracellular solute-binding proteins, family 3	EU232046
Lrh33	N-terminal TMH 18-42	199	0.384	0.992		Mmrqkqlqsalldrhgk*fktakvrgkwmkasevvlsfsvlva*fs	Unique hypothetical extracellular protein		EU232047
Lrh34	N-terminal TMH 33-55	381	0.225			Mdngshnfqetpqdqkpqrsaekpkkngllkqtaiv*avvaaligggagggtaywai*nht	Trypsin-like serine protease with PDZ domain *Lactobacillus casei* ATCC 334		EU232048
Lrh35	SS	266	0.662	1		Mprkwihmlmlllmlvtqigsa/avp	Collagen adhesion protein *Bacillus clausii* KSM-K16		EU232049
Lrh36	LSS	241	0.567		0.99	Mktfrkllltllipvllgg/cgf	Amino acid ABC transporter, amino acid-binding protein (*Streptococcus agalactiae* COH1)	(pfam04069), OpuAC, Substrate binding domain of ABC-type glycine betaine transport system	EU232050
Lrh37	SS	291	0.549	0.998		Mrlfgeektryrlyksgklwlvaligffvlaighqpnqvka/ssm	Hypothetical protein LSEI_2320 *Lactobacillus casei* ATCC 334	(pfam06458), DUF1085, Repeat of unknown function	EU232051
Lrh38	N-terminal TMH 9-31	27	0.234			Mkkvlnkq*sigfligiavglivyla*pl	Transport protein *Lactobacillus plantarum* WCFS1	(pfam00939) Na_sulph_symp, Sodium:sulfate symporter transmembrane region	EU232052

Lrh39	Multiple TMHs	370				N-terminal TMH: wtldwwligilvlaa	4-amino-4-deoxy-L-arabinose transferase related glycosyltransferase of PMT family *Lactobacillus casei* ATCC 334	(pfam0236), PMT, Dolichyl-phosphate-mannose-protein mannosyltransferase	EU232053
Lrh40	SS	165	0.696	1		Mskkywigsvafagllvgsfsa gslvla/ade	Hypothetical protein LSEI_2084 *Lactobacillus casei* ATCC 334		EU232054
Lrh41	SS	255	0.712	1		Mqkffsnkklialmiallvslglva gsiya/snn	Cell shape-determining protein, MreC *Lactobacillus casei* ATCC 334	(pfam04085), MreC, rod shape-determining protein MreC	EU232055
Lrh42	SS	274	0.616	1		Mkkpigkilwrglgplliaiilvaal mlvpfkfgrssqatirqa/ass	DltD - Protein involved in D-alanine esterification of lipoteichoic acid and wall teichoic acid *Lactobacillus rhamnosus*	(pfam04918) DltD_M, DltD central region	EU232056
Lrh43	LSS	288	0.662		0.99	Mtfykkiiaalamvatvgflta/cs s	Putative glycine/betaine/carnitine ABC transporter, substrate binding lipoprotein precursor *Lactobacillus sakei* subsp. *sakei* 23K	(pfam04069), OpuAC, Substrate binding domain of ABC-type glycine betaine transport system	EU232057
Lrh44	SS	261	0.787	1		Mqrrkrisgilitalmiavvfliww vpsla/psk	Predicted esterase *Lactobacillus casei* ATCC 334	(pfam0075), Esterase	EU232058
Lrh45	SS	241	0.903	1		Mkkllstvlfsavalsavalskpg hvna/atk	Cell wall-associated hydrolase *Lactobacillus casei* ATCC 334		EU232059
Lrh46	SS	56	0.561	0.929		Mrvkqlimkiliaslamtavvglp rgqvkp/ktm	Signal sequence only, no significant hits		EU232060
Lrh47	N-terminal TMH 13-35	155	0.473			Mkllwmefkrqt*rsltfviytflvva fivmnvwpl*lsr	hypothetical protein LSEI_0463 *Lactobacillus casei* ATCC 334		EU232061
Lrh48[c]	SS	26	0.408			Mkhrnrlfwliillvagvflvys/tes	Predicted Zn-dependent protease *Lactobacillus casei* ATCC 334		EU232062
Lrh49	Multiple TMHs	278				N-terminal TMH: lgpllalvvliilvtvmsptfvs	Ribose/xylose/arabinose/galactos ide ABC-type transport system, permease component *Lactobacillus casei* ATCC 334	(COG1172), AraH, Ribose/xylose/arabinos e/galactoside ABC-type transport systems, permease components [Carbohydrate transport and metabolism]	EU232063
Lrh50	N-terminal TMH 4-23	149	0.378			Mkk*lflidllwiglailaglvws*qh d	Uncharacterized conserved membrane protein *Lactobacillus casei* ATCC 334	(COG1172), AraH, Ribose/xylose/arabinos e/galactoside ABC-type transport systems, permease components [Carbohydrate transport and metabolism]	EU232064
Lrh51	SS	69	0.764	1		Mykakkrwliagtalllmpaffq pgqevha/dsk	Adhesion exoprotein *Lactobacillus casei* ATCC 334	(COG1172), AraH, Ribose/xylose/arabinos e/galactoside ABC-type transport systems, permease components	EU232065

Lrh52	LSS	84	0.54		1	Mtnhhrlwlwvlllicsllglss/cha	Taurine-binding periplasmic protein precursor *Clostridium tetani* E88	(smart00062), PBPb, Bacterial periplasmic substrate-binding proteins	EU232066
Lrh53	SS	62	0.52	0.99		Mthsgkkktilfvcatgiatstvva/ekt	Putative phosphotransferase system enzyme IIB *Shigella dysenteriae* Sd197	(COG3414), SgaB, Phosphotransferase system, galactitol-specific IIB component [Carbohydrate transport and metabolism]	EU232067
Lrh54	SS	97	0.587	0.712		Mmtkkwakvfliigiflvvsggawlwlnrga/tps	Alpha/beta hydrolase superfamily protein *Lactobacillus casei* ATCC 334	(COG4814), Uncharacterized protein with an alpha/beta hydrolase fold [General function prediction only]	EU232068
Lrh55	N-terminal TMH 24-46	163	0.441			Marrregkqskspkppkklwrri*lkwtllglllifiagvglfswya*kda	Multimodular transpeptidase-transglycosylase PBP 1A *Lactobacillus salivarius subsp. salivarius* UCC118	(pfam00912), Transgly, Transglycosylase	EU232069
Lrh56	SS	201	0.579	0.686		Mkrnnwiwlgllgmvlvvsa/gwy	Hypothetical protein SMU.1071c *Streptococcus mutans* UA159	(COG4814), Uncharacterized protein with an alpha/beta hydrolase fold [General function prediction only]	EU232070
Lrh57	SS	35	0.467			Mtlsrsqrsimkmtrkrrwliiltvlgvlvvagg/v	Signal sequence only, no significant hits		EU232071
Lrh58	N-terminal TMH 26-46	68	0.269			Mnkvfdklkpifeaiaankyvsairdg*fiacmpiiifssifmmiay*vpn	LacE *Lactobacillus rhamnosus* TCELL-1	(COG1455) CelB, Phosphotransferase system cellobiose-specific component IIC [Carbohydrate transport and metabolism]	EU232072
Lrh59	Multiple TMHs	255				N-terminal TMH: iiqniltqaaitialiamlglll	Putative sugar-specific permease, SgaT/UlaA *Enterococcus faecium* DO	(pfam04215), SgaT_UlaA, Putative sugar-specific permease, SgaT/UlaA	EU232073
Lrh60	SS	64	0.567	0.632		Mnrlwdkpwvnrllalllaiglfa/yvnvesinntrqs/and	Hypothetical protein LSEI_1017 *Lactobacillus casei* ATCC 334		EU232074
Lrh61	Multiple TMHs	729				N-terminal TMH: llfisfgvtfvislivfaal	Predicted membrane protein *Lactobacillus casei* ATCC 334		EU232075
Lrh62[d]	SS	807	0.725	0.954		Mhapassaapvassyavaassdaafa/kqa	Uncharacterized hypothetical extracellular protein		EU232076
Lrh63	SS	246	0.605	1		Mrrfywwlapllligivlgstphwvha/adq	Hypothetical protein LSEI_2364 *Lactobacillus casei* ATCC 334		EU232077
Lrh64	LSS	141	0.626		0.99	Mkkrlivvilmllalvatg/css	Thiamine biosynthesis membrane-associated lipoprotein *Lactobacillus casei* ATCC 334	(pfam02424), ApbE family. ApbE is involved in thiamine synthesis	EU232078

Lrh65	Multiple TMHs	214				N-terminal TMH: lgvliilqntimnllpwggpta	H+/citrate symporter *Lactobacillus casei* ATCC 334	(pfam03600), CitMHS, Citrate transporter	EU232079
Lrh66	SS	85	0.592	1		Makkrrrftlkkvplaawlviaffa vgsmlvvssywsqrqa/evs	N-acetylmuramidase *Lactobacillus casei* ATCC 334		EU232080
Lrh67	LSS	172	0.681		0.99	Mkkqtwvkallamftvmlvailt a/cgs	ABC-type sugar transport system, periplasmic component *Lactobacillus casei* ATCC 334	(pfam01547), SBP_bac_1, Bacterial extracellular solute-binding protein	EU232081
Lrh68	LSS	192	0.779		1	Mrswkklvlvvaallalvsvta/c gk	ABC-type metal ion transport system, periplasmic component/surface antigen *Lactobacillus casei* ATCC 334	(pfam03180), Lipoprotein_9, NLPA lipoprotein	EU232082
Lrh69	LSS	72	0.523		1	Mmryfrfqlhqlltnrknlavigial vali/cqf	Hypothetical protein LSEI_2331 *Lactobacillus casei* ATCC 334		EU232083
Lrh70	Multiple TMHs	285				N-terminal TMH: ylflslliivlliltllqyl	Hypothetical protein LSEI_1227 *Lactobacillus casei* ATCC 334		EU232084
Lrh71	LSS	106	0.63		1	Mkkfftltlltgaivllaa/cgk	Lipoprotein, pheromone precursor *Lactobacillus salivarius* subsp. *salivarius* UCC118		EU232085
Lrh72	SS	57	0.552	1		Mkkaatifaillgidlfapqqaah a/qtq	Signal sequence only, no significant hits		EU232086
Lrh73	Multiple TMHs	353				N-terminal TMH: nfgyafdafypslllpyvwlrl	Hypothetical protein LSEI_1109 *Lactobacillus casei* ATCC 334		EU232087
Lrh74[c]	SS	33	0.505			Mkkllrglvaamvavlmiggaa pavsvaaa/ess	Signal sequence only, no significant hits		EU232088
Lrh75	SS	235	0.525	0.65		Mqvlsrvwqsirwffrrfqvirwlil lgltvilifsawftyka/kta	Bifunctional glycolsyltransferase/transpeptida se penicillin binding protein 2A *Lactobacillus sakei* subsp. *sakei* 23K	(pfam00912), Transgly, Transglycosylase.	EU232089
Lrh76	Multiple TMHs	233				N-terminal TMH: wllivdqslffmailavmygala	Uncharacterized membrane-bound protein conserved in bacteria *Lactobacillus casei* ATCC 334	(pfam06570), DUF1129, Protein of unknown function	EU232090
Lrh77	Internal TMH 56-78	249				lvknlvlillpllllgvfgyfas	Cell division septal protein *Lactobacillus casei* ATCC 334	(pfam03799), FtsQ, Cell division protein	EU232091
Lrh78	N-terminal TMH 21-43	68	0.586			Mrltelirisfralaankrrs*fltmlgi iigiasvvtilslgn*gik	ABC-type antimicrobial peptide transport system, permease component *Lactobacillus casei* ATCC 334		EU232092
Lrh79	LSS	128	0.695		0.99	Mkkhlvltiaatvligltlig/cgk	ABC transporter substrate binding/membrane-spanning protein *Streptococcus thermophilus* LMG 18311	(pfam04392), DUF534, Protein of unknown function	EU232093
Lrh80	N-terminal TMH 15-37	146	0.546			Miksewlsvlkrpmv*ivvlffitlip aisgvylssm*wnt	Predicted membrane protein *Lactobacillus brevis* ATCC	(COG1511), Predicted membrane protein [Function unknown]	EU232094

Lrh81	LSS	201	0.662		0.99	Mmkrrwllgiatglmalalta/cs s	ABC-type sugar transport system, periplasmic component *Lactobacillus casei* ATCC 334	(pfam01547), SBP_bac_1, Bacterial extracellular solute-binding protein	EU232095
Lrh82	SS	256	0.574	1		Mpsrfknkrqvikkiacsaligvs iiltla/vlg	Hypothetical protein LSEI_2692 *Lactobacillus casei* ATCC 334		EU232096
Lrh83	SS	239	0.568	0.996		Mkkstrtilisvvaaivvivggfga /fhw	Cell division protein FtsI/penicillin-binding protein 2 *Lactobacillus casei* ATCC 334	(pfam05223), MecA_N, NTF2-like N-terminal transpeptidase domain	EU232097
Lrh84	N-terminal TMH 7-29	231	0.335			Mdfrr*iewiflvvfvglniflgisyfq aq*qvd	Hypothetical protein LSEI_2805 *Lactobacillus casei* ATCC 334	(COG4853), Uncharacterized protein conserved in bacteria [Function unknown]	EU232098
Lrh85	N-terminal TMH 12-34	89	0.436			Mkklhdrhips*lifwlvvilvalltm pdvsgivr*dkg	Predicted membrane protein *Lactobacillus casei* ATCC 334	(COG2409), Predicted drug exporters of the RND superfamily [General function prediction only]	EU232099
Lrh86	N-terminal TMH 4-23	200	0.527			Mk*rmlgfmiiifgflgiayfqsg*tt a	Putative ABC transporter substrate binding protein *Lactobacillus acidophilus* NCFM	(pfam04392), DUF534, Putative secreted protein of unknown function	EU232100
Lrh87	LSS	33	0.761		0.99	Mrrnlimkkgalwtsaaiassfl/ cst	Signal sequence only, no significant hits		EU232101
Lrh88[e]	No targeting signal	53					Dithiol-disulfide isomerase *Lactobacillus casei* ATCC 334	(COG2761), FrnE, Predicted dithiol-disulfide isomerase involved in polyketide biosynthesis [Secondary metabolites biosynthesis, transport, and catabolism].	EU232102
Lrh89[e]	No targeting signal	170					Predicted hydrolase of the HAD superfamily *Oenococcus oeni* PSU-1		EU232103

[a]The open reading frames are numbered in the order of identification. Skipped numbers represent the open reading frames that did not carry any recognisable membrane-targeting sequence.

[b]SS, signal sequence (Type I); LSS, lipoprotein signal sequence (Type II), TMH, transmembrane helix. Numbers denote position relative to the N-terminus of the protein.

[c]ORFs with SignalP C-score bellow threshold (0.52) due to the short sequence. These were assigned as signal sequence type I rather than N terminal membrane anchor based on sequence similarity to the nearest homolog that already carries a type I signal sequence.

[d]Vector pDJ01 provided start codon for Lrh62 (in bold)

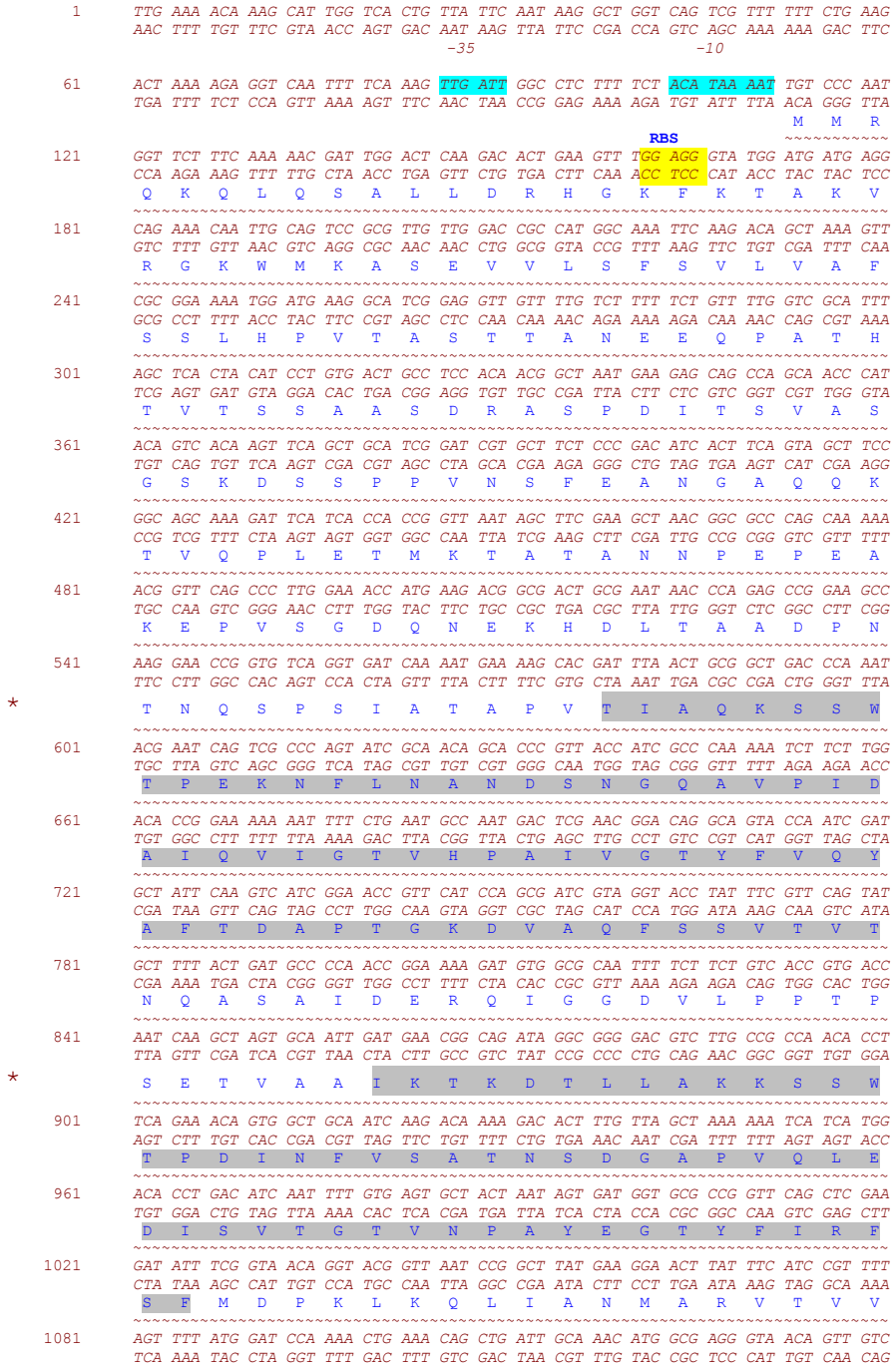
[e]ORFs with no signal sequence or transmembrane helices. Experimentally demonstrated to provide signal sequence-like activity that results in

assembly of the fusion into a sarcosyl-resistant virion.
[f]Amino acids in italics correspond to the sequence of predicted transmembrane α helix.

APPENDIX III

Figure A.3

1) Nucleotide sequence and translation of Lrh33G ORF



```
          S   S   L   T   S   K   P   N   P   G   A   S   E   M   L   K   V   K   N   L
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1141    AGT AGT CTG ACC TCG AAA CCG AAT CCG GGT GCT TCG GAG ATG TTG AAG GTG AAA AAT CTA
        TCA TCA GAC TGG AGC TTT GGC TTA GGC CCA CGA AGC CTC TAC AAC TTC CAC TTT TTA GAT
         S   S   F   F   P   S   T   S   N   Y   P   S   V   K   N   E   Y   L   L   D
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1201    TCA TCT TTT TTT CCT TCA ACA TCC AAT TAT CCA AGC GTT AAA AAT GAG TAT CTG CTA GAT
        AGT AGA AAA AAA GGA AGT TGT AGG TTA ATA GGT TCG CAA TTT TTA CTC ATA GAC GAT CTA
         D   P   N   F   K   P   D   E   A   E   I   V   S   Y   F   H   Q   Y   V   N
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1261    GAT CCG AAC TTC AAA CCG GAT GAG GCT GAG ATT GTC AGC TAC TTT CAT CAA TAT GTC AAT
        CTA GGC TTG AAG TTT GGC CTA CTC CGA CTC TAA CAG TCG ATG AAA GTA GTT ATA CAG TTA
         E   L   R   A   L   N   G   Q   S   A   L   T   I   S   E   R   D   T   S   R
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1321    GAA TTG CGG GCT TTG AAC GGA CAG TCG GCG TTG ACA ATT TCG GAA CGC GAT ACA AGT CGG
        CTT AAC GCC CGA AAC TTG CCT GTC AGC CGC AAC TGT TAA AGC CTT GCG CTA TGT TCA GCC
         A   Q   Q   R   V   R   A   I   V   E   D   F   N   H   E   A   V   S   H   A
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1381    GCA CAA CAA CGC GTG CGG GCG ATT GTT GAA GAT TTT AAT CAT GAA GCC GTT AGT CAC GCA
        CGT GTT GTT GCG CAC GCC CGC TAA CAA CTT CTA AAA TTA GTA CTT CGG CAA TCA GTG CGT
         T   E   N   I   G   T   N   T   G   I   T   D   H   M   R   S   N   Q   E   I
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1441    ACG GAG AAC ATC GGT ACG AAC ACT GGC ATT ACG GAT CAT ATG CGG TCT AAT CAG GAG ATC
        TGC CTC TTG TAG CCA TGC TTG TGA CCG TAA TGC CTA GTA TAC GCC AGA TTA GTC CTC TAG
         A   Y   Y   M   V   M   A   W   Y   D   E   T   D   N   P   E   P   L   G   N
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1501    GCA TAT TAC ATG GTG ATG GCC TGG TAT GAC GAG ACC GAT AAC CCG GAG CCG CTT GGT AAC
        CGT ATA ATG TAC CAC TAC CGG ACC ATA CTG CTC TGG CTA TTG GGC CTC GGC GAA CCA TTG
         G   H   Y   G   H   R   A   N   L   I   Y   G   G   P   S   M   G   V   S   F
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1561    GGA CAT TAT GGG CAT CGG GCG AAT TTA ATT TAT GGC GGG CCG TCA ATG GGC GTG AGT TTT
        CCT GTA ATA CCC GTA GCC CGC TTA AAT TAA ATA CCG CCC GGC AGT TAC CCG CAC TCA AAA
         Y   R   P   P   A   G   S   A   E   F   S   D   Y   Y   A   F   E   A   P   I
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1621    TAC CGT CCC CCT GCA GGT AGC GCT GAG TTT AGC GAT TAT TAT GCC TTT GAA GCG CCT ATT
        ATG GCA GGG GGA CGT CCA TCG CGA CTC AAA TCG CTA ATA ATA CGG AAA CTT CGC GGA TAA

1681    TAG
        ATC
```

** Two Big-3 domains are highlighted grey*

2) Lrh33 BLASTP hits and alignments to the two top hits

```
Query ID: Lrh33
Query Length: 503 AA
Database Name: nr - All non-redundant GenBank CDS
translations+PDB+SwissProt+PIR+PRF excluding
environmental samples from WGS
Projects Program: BLASTP 2.2.19+
```

Sequences producing significant alignments:		Score (Bits)	E Value
NP_786170.1|	cell surface protein precursor [Lb. plantarum)	80.1	5e-13
YP_001986344.1|	Beta-fructosidase [Lb.casei)	78.6	1e-12
gb|ABD57319.1|	beta-fructosidase precursor [Lb. paracasei)	77.8	2e-12
NP_784643.1|	cell surface protein precursor [Lb.plantarum)	76.3	7e-12
YP_848283.1|	cell wall surface anchor family protein [Listeria)	74.7	2e-11
BAD88632.1|	fructan hydrolase [Lb.casei]	72.8	7e-11
NP_816151.1|	cell wall surface anchor family protein [Enteroco..	71.2	2e-10
NP_470145.1|	cell wall anchor domain-containing protein [List..	68.6	1e-09
ZP_03056851.1|	cell wall surface anchor protein [Enteroco...	65.9	8e-09
ZP_00232462.1|	cell wall surface anchor family protein [Lister..	65.9	9e-09
NP_464368.1|	peptidoglycan binding protein [Listeria mono...	64.7	2e-08
ZP_00229842.1|	cell wall surface anchor family protein [Lister.	63.5	5e-08
YP_013462.1|	cell wall surface anchor family protein [Listeria.	63.5	5e-08
YP_848986.1|	hypothetical protein lwe0785 [Listeria welsh...	61.2	2e-07

NP_786170.1
GENE ID: 1063145 lp_2795 | cell surface protein precursor [Lactobacillus plantarum WCFS1] Length=1039

```
Score = 80.1 bits (196), Expect = 5e-13,
Identities = 59/178 (33%), Positives = 91/178 (51%), Gaps = 18/178 (10%)

Lrh33  160  SSWTPEKNFLNANDSNGQAVPIDAIQVIGTVHPAIVGTYFVQYAFTDAPTGKDVAQFSSV  219
            S+W P+ NF NA DS+GQ +   AI+V+GT      G Y + Y FTD  TG+     +++
Sbjct  716  STWQPQSNFQNATDSDGQTLDWSAIEVVGTPDWTTAGDYRLTYQFTD-KTGQ--LVTATM  772

Lrh33  220  TVTNQASAIDERQIGGDVLPPTPSETVAAIKTKDTLLAKKSSWTPDINFVSATNSDGAPV  279
            TVT      DE+           +E+ + ++  D+ +    SW P  N V AT+ +G  +
Sbjct  773  TVTVVIEEADEQ-----------AESQSDLQIHDSTITVGESWQPSDNLVLATDVNGGEL  821

Lrh33  280  QLEDISVTGTVNPAYEGTYFIRFSFMDPKLKQLIANMARVTVVSSLTSKPN---PGAS  334
             L D+ VTGTV+    G Y + + + D    Q+   +A VTVV++     N   PGA+
Sbjct  822  SLADLVVTGTVDTNQAGVYQVTYQYTDAS-GQVFTRVATVTVVAASDGDTNTEQPGAT  878



Score = 48.9 bits (115), Expect = 0.001,
Identities = 42/155 (27%), Positives = 65/155 (41%), Gaps = 28/155 (18%)

Lrh33  161  SWTPEKNFLNANDSNGQAVPIDAIQVIGTVHPAIVGTYFVQYAFTDAPTGKDVAQFSSVT  220
            +W  +  F  A D +G A+  + + + G V+  + G Y + Y +         +Q  ++T
Sbjct  638  TWHAQDGFDGATDKDGHAIDFNDVTITGEVNTMVPGDYQITYTYG--------SQTQTIT  689

Lrh33  221  VT---NQASAIDERQIGGDVLPPTPSETVAAIKTKDTLLAKKSSWTPDINFVSATNSDGA  277
            VT   NQAS                +  A + T        S+W P  NF +AT+SDG
Sbjct  690  VTVKENQASL-------------NLYQNHATVHTDG---QGTSTWQPQSNFQNATDSDGQ  733

Lrh33  278  PVQLEDISVTGTVNPAYEGTYFIRFSFMDPKLKQL  312
             +    I V GT +    G Y + + F D K  QL
Sbjct  734  TLDWSAIEVVGTPDWTTAGDYRLTYQFTD-KTGQL  767



Score = 42.0 bits (97), Expect = 0.15,
Identities = 25/67 (37%), Positives = 35/67 (52%), Gaps = 1/67 (1%)

Lrh33  155  TIAQKSSWTPEKNFLNANDSNGQAVPIDAIQVIGTVHPAIVGTYFVQYAFTDAPTGKDVA  214
            TI    SW P  N + A D NG  + +  + V GTV     G Y V Y +TDA +G+
Sbjct  797  TITVGESWQPSDNLVLATDVNGGELSLADLVVTGTVDTNQAGVYQVTYQYTDA-SGQVFT  855

Lrh33  215  QFSSVTV  221
            + ++VTV
Sbjct  856  RVATVTV  862



Score = 38.1 bits (87), Expect = 2.0,
Identities = 17/57 (29%), Positives = 28/57 (49%), Gaps = 0/57 (0%)

Lrh33  248  AIKTKDTLLAKKSSWTPDINFVSATNSDGAPVQLEDISVTGTVNPAYEGTYFIRFSF  304
            ++   DT L    +W     F  AT+ DG  +   D+++TG VN    G Y I +++
Sbjct  625  SLTVHDTELHAGGTWHAQDGFDGATDKDGHAIDFNDVTITGEVNTMVPGDYQITYTY  681
```

YP_001986344.1
GENE ID: 6404417 fosE | Beta-fructosidase [Lactobacillus casei BL23]
Length=493

```
Score = 78.6 bits (192), Expect = 1e-12,
Identities = 53/168 (31%), Positives = 84/168 (50%), Gaps = 25/168 (14%)

Lrh33  162  WTPEKNFLNANDSNGQAVPIDAIQVIGTVHPAIVGTYFVQYAFTDAPTGKDVAQFSSVTV  221
            W    NF++A D++G  +   ++ V G+V P   G Y V Y++TDA  G  V   +++TV
Sbjct  39   WNAADNFVSATDADGNGIDFKSVNVSGSVDPTKPGKYEVTYSYTDAG-GNQVGAKATITV  97

Lrh33  222  TNQASAIDERQIGGDVLPPTPSETVAAIKTKDTLLAK--KSSWTPDINFVSATNSDGAPV  279
                                  T A+IK KD+ L     + W    NFVSAT++DG  +
Sbjct  98   V---------------------STKASIKAKDSTLVAGPDTKWNAADNFVSATDADGNGI  136

Lrh33  280  QLEDISVTGTVNPAYEGTYFIRFSFMDPKLKQLIANMARVTVVSSLTS  327
              + ++V+G+V+P   G Y + +S+ D    Q+ A  A +TVVS+  S
Sbjct  137  DFKSVNVSGSVDPTKPGDYEVTYSYTDAGGNQVSA-KATITVVSTKAS  183


Score = 78.2 bits (191), Expect = 2e-12,
Identities = 52/165 (31%), Positives = 83/165 (50%), Gaps = 25/165 (15%)

Lrh33  162  WTPEKNFLNANDSNGQAVPIDAIQVIGTVHPAIVGTYFVQYAFTDAPTGKDVAQFSSVTV  221
            W    NF++A D++G  +   ++ V G V P   G Y V Y++TD   G  V+  +++TV
Sbjct  199  WNAADNFVSATDADGNGIDFKSVNVSGNVDPTKPGKYEVTYSYTDVG-GNQVSAKATITV  257

Lrh33  222  TNQASAIDERQIGGDVLPPTPSETVAAIKTKDTLLAK--KSSWTPDINFVSATNSDGAPV  279
                                T A+IK KD+ L     + W    NFVSAT++DG  +
Sbjct  258  V---------------------STKASIKAKDSTLVAGPDTKWNAADNFVSATDADGNGI  296


Lrh33  280  QLEDISVTGTVNPAYEGTYFIRFSFMDPKLKQLIANMARVTVVSS  324
              ++++V+G+V+P   G Y + +S+ D    Q+ A  A +TVVSS
Sbjct  297  DFKNVNVSGSVDPTKPGKYEVTYSYTDAGGNQVSA-KATITVVSS  340


Score = 77.4 bits (189), Expect = 3e-12,
Identities = 53/168 (31%), Positives = 84/168 (50%), Gaps = 25/168 (14%)

Lrh33  162  WTPEKNFLNANDSNGQAVPIDAIQVIGTVHPAIVGTYFVQYAFTDAPTGKDVAQFSSVTV  221
            W    NF++A D++G  +   ++ V G+V P   G Y V Y++TDA  G  V+  +++TV
Sbjct  119  WNAADNFVSATDADGNGIDFKSVNVSGSVDPTKPGDYEVTYSYTDAG-GNQVSAKATITV  177

Lrh33  222  TNQASAIDERQIGGDVLPPTPSETVAAIKTKDTLLAK--KSSWTPDINFVSATNSDGAPV  279
                                T A+IK KD+ L     + W    NFVSAT++DG  +
Sbjct  178  V---------------------STKASIKAKDSTLVAGPDTKWNAADNFVSATDADGNGI  216

Lrh33  280  QLEDISVTGTVNPAYEGTYFIRFSFMDPKLKQLIANMARVTVVSSLTS  327
              + ++V+G V+P   G Y + +S+ D    Q+ A  A +TVVS+  S
Sbjct  217  DFKSVNVSGNVDPTKPGKYEVTYSYTDVGGNQVSA-KATITVVSTKAS  263
```

3) Lrh33 conserved domains [CDART search]

	Multi-dom	E-value
fam07523, Big_3, Bacterial Ig-like domain (group 3)	no	8e-06

acterial Ig-like domain (group 3)

CD Length: 68 **Bit Score:** 46.85 **E-value:** 8e-06

```
                    10        20        30        40        50
                ....*....|....*....|....*....|....*....|....*....|....*.
Lrh33      249 IKTKDTLLAKKSSWTPDINFVSATNSDGAPVQLEDISVTGTVNPAYEGTYFIRFSF 304
pfam07523    2 IEAHDSTIYVGDSWDAEDNFVSATDKDGKAVDFSDVTVSGTVDTTKAGTYEVTYTY 57
```

fam07523, Big_3, Bacterial Ig-like domain (group 3)	no	1e-04

acterial Ig-like domain (group 3)

CD Length: 68 **Bit Score:** 42.99 **E-value:** 1e-04

```
                      10        20        30        40        50        60
            ....*....|....*....|....*....|....*....|....*....|....*....|....*...
Lrh33 155 TIAQKSSWTPEKNFLNANDSNGQAVPIDAIQVIGTVHPAIVGTYFVQYAFTDAPTgkdvaqfsSVTVT 222
pfam07523 8TIYVGDSWDAEDNFVSATDKDGKAVDFSDVTVSGTVDTTKAGTYEVTYTYDGVSK--------TITVT 67
```