Title	A cell wall-associated polysaccharide is required for bacteriophage adsorption to the Streptococcus thermophilus cell surface
Authors	McDonnell, Brian;Hanemaaijer, Laurens;Bottacini, Francesca;Kelleher, Philip;Lavelle, Katherine;Sadovskaya, Irina;Vinogradov, Evgeny;Ver Loren van Themaat, Emiel;Kouwen, Thijs;Mahony, Jennifer;van Sinderen, Douwe
Publication date	2020
Original Citation	McDonnell, B., Hanemaaijer, L., Bottacini, F., Kelleher, P., Lavelle, K., Sadovskaya, I., Vinogradov, E., Ver Loren van Themaat, E., Kouwen, T., Mahony, J. and van Sinderen, D. 'A cell wall-associated polysaccharide is required for bacteriophage adsorption to the Streptococcus thermophilus cell surface', Molecular Microbiology,
Type of publication	Article (peer-reviewed)
Link to publisher's version	https://onlinelibrary.wiley.com/doi/abs/10.1111/mmi.14494 - 10.1111/mmi.14494
Rights	© 2020 John Wiley & Sons Ltd. This is the peer reviewed version of the following article: McDonnell, B, Hanemaaijer, L, Bottacini, F, et al. A cell wall#associated polysaccharide is required for bacteriophage adsorption to the Streptococcus thermophilus cell surface. Mol Microbiol. 2020, which has been published in final form at https://doi.org/10.1111/mmi.14494. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.
Download date	2024-04-27 22:21:48
Item downloaded from	https://hdl.handle.net/10468/9729

University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

MR. BRIAN MCDONNELL (Orcid ID : 0000-0003-1704-8759)

Article type : Research Article

A cell wall-associated polysaccharide is required for bacteriophage adsorption to the *Streptococcus thermophilus* cell surface

Running title: *S. thermophilus* phages target saccharidic receptors.

Brian McDonnell[a], Laurens Hanemaaijer[b], Francesca Bottacini[a], Philip Kelleher[a], Katherine Lavelle[a], Irina Sadovskaya[c], Evgeny Vinogradov[d], Emiel Ver Loren van Themaat[b], Thijs Kouwen[b], Jennifer Mahony[a] and Douwe van Sinderen[a†].

[a] School of Microbiology & APC Microbiome Ireland, University College Cork, Cork, Ireland.

[b] DSM Biotechnology Center, Delft, The Netherlands.

[c] Équipe BPA, Université du Littoral Côte d'Opale, Institut Régional Charles Violette EA 7394, USC Anses-ULCO, Boulogne-sur-Mer, France.

[d] Institute for Biological Sciences, National Research Council of Canada, Ottawa, Ontario, Canada.

†Corresponding author. Email: d.vansinderen@ucc.ie. Phone: +353 (0)21 4901365.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/MMI.14494

Keywords: adsorption, dairy, mutation, polysaccharide, NMR, structure

Summary

Streptococcus thermophilus strain ST64987 was exposed to a member of a recently discovered group of *S. thermophilus* phages (the 987 phage group), generating phage-insensitive mutants, which were then characterised phenotypically and genomically. Decreased phage adsorption was observed in selected bacteriophage-insensitive mutants, and was partnered with a sedimenting phenotype and increased cell chain length or aggregation. Whole genome sequencing of several bacteriophage-insensitive mutants identified mutations located in a gene cluster presumed to be responsible for cell wall polysaccharide production in this strain. Analysis of cell surface-associated glycans by methylation and NMR spectroscopy revealed a complex branched rhamno-polysaccharide in both ST64987 and phage-insensitive mutant BIM3. In addition, a second cell wall-associated polysaccharide of ST64987, composed of hexasaccharide branched repeating units containing galactose and glucose, was absent in the cell wall of mutant BIM3. Genetic complementation of three phage-resistant mutants was shown to restore the carbohydrate and phage resistance profiles of the wild-type strain, establishing the role of this gene cluster in cell wall polysaccharide production and phage adsorption and, thus, infection.

Keywords: adsorption, dairy, mutation, polysaccharide, NMR, structure

Introduction

Streptococcus thermophilus is arguably best-known for its presence in the dairy environment, being widely used in the production of yoghurt and various cheeses (Reinheimer *et al.*, 1997, Mora *et al.*, 2002). Being the only member of the genus *Streptococcus* to enjoy a Qualified Presumptive Safety status as defined by the European Food Safety Authority (Leuschner *et al.*, 2010), *S. thermophilus* is one of the most valuable species to the dairy industry, along with two other lactic acid bacteria - *Lactococcus lactis* (Bolotin *et al.*, 2001), and *Lactobacillus delbrueckii* (Van de Guchte *et al.*, 2006), with which *S. thermophilus* has a proto-cooperative relationship during yoghurt production (Driessen *et al.*, 1982, Herve-Jimenez *et al.*, 2009, Arioli *et al.*, 2017).

Phage predation of *S. thermophilus* commonly leads to slow acidification or even complete failure of industrial fermentations (reviewed by Garneau & Moineau, 2011). Analysis of predatory phages has revealed the global presence of four distinct phage groups, i.e. the *cos*-containing, *pac*-containing, 5093 and 987 phage groups, which have been described in detail elsewhere (McDonnell *et al.*, 2017). A comprehensive understanding of the initial phage-host interactions in this species is lacking, though the involvement of several phage-encoded proteins in host recognition has been investigated (Duplessis & Moineau, 2001, Lavelle *et al.*, 2018). The genes encoding these proteins (such as the tail tape measure protein (TMP) and receptor-binding protein (RBP)) are known to harbour genetically variable regions linked to host recognition (Duplessis & Moineau, 2001, Duplessis *et al.*, 2006). More recently, these variable regions have been shown to contain carbohydrate binding domains similar to those found in BppA (an accessory protein found on the phage Tuc2009 baseplate), accounting for their variability and proposed carbohydrate binding functions (Legrand *et al.*, 2016, Lavelle *et al.*, 2018).

Studies conducted on cell envelope extracts of *S. thermophilus* have concluded that the phage receptor in this species is saccharidic in nature (Quiberoni *et al.*, 2000, Binetti *et al.*, 2002), based on reduced phage adsorption following (i) treatment of cell wall preparations with peptidoglycan and polysaccharide-degrading chemical agents, and (ii) pre-incubation of phage samples with monosaccharides. These findings are consistent with observations in other Gram-positive dairy bacteria, e.g. *L. lactis* (Dupont *et al.*, 2004), for which role of cell wall-associated polysaccharides (CWPSs) in phage adsorption have been identified (Chapot-Chartier *et al.*, 2010, Mahony *et al.*, 2013, Ainsworth *et al.*, 2014, Mahony *et al.*, 2015, Vinogradov *et al.*, 2018a, Vinogradov *et al.*, 2018b). An important distinction when discussing polysaccharides produced by

S. thermophilus should be made between exopolysaccharides (EPS), generally considered to be loosely associated with the cell wall and/or secreted into the surrounding environment, and cell wall (or capsular) polysaccharides (CPS; in the case of *S. thermophilus*, also termed RGP), which are tightly and sometimes covalently associated with the peptidoglycan layer (Stingele *et al.*, 1996, Mozzi *et al.*, 2006, Rodriguez *et al.*, 2008). Mende and colleagues (2014) analysed both released and capsular EPS/CPS 'fractions' in *S. thermophilus* strain ST-143, observing that they differed in (amongst other properties) monosaccharide composition, the most notable being the presence of rhamnose in the CPS fraction, but not in the EPS fraction (N.B. detection of N-acetylated sugars was not possible using the specific technique employed). The distinction between these two fractions may be reflected in the genetic content of clusters associated with polysaccharide synthesis.

Two such clusters have been identified in *S. thermophilus*, termed *eps* (exopolysaccharide) and *rgp* (rhamnose-glucose polysaccharide; Stingele *et al.*, 1996, Hols *et al.*, 2005, Goh *et al.*, 2011). It is presumed that the *eps* cluster is responsible for the production of a loosely-bound or released EPS, as evidenced by the production of EPS by an *L. lactis* strain harbouring a heterologously expressed *eps* cluster (Germond *et al.*, 2001), and disruptions in this cluster being implicated in the loss of a 'ropy' phenotype (Stingele *et al.*, 1996). The *eps* clusters of *S. thermophilus* are known to be highly variable in size and genetic composition (Bourgoin *et al.*, 1999, Pluvinet *et al.*, 2004), as opposed to *rgp* clusters which are generally more conserved, being akin to those identified in *Streptococcus mutans* (Hols *et al.*, 2005, Goh *et al.*, 2011). Herein, these distinct carbohydrate structures will be termed RGP (rhamnose-glucose polysaccharide; to denote the particular polysaccharide layer that is most closely associated with the cell wall) and EPS (exopolysaccharide, representing a polysaccharide layer that is more loosely associated with the cell wall), collectively denoted as CWPSs (cell wall-associated polysaccharides).

The most compelling evidence of CWPS involvement in *S. thermophilus* phage-host interactions was presented recently (Szymczak *et al.*, 2018), and suggests that phages using the *cos*- and *pac*- modes of DNA packaging recognize (a component of) the RGP, which is tightly linked to the PG (peptidoglycan) layer. In contrast, complex baseplate-encoding phages (e.g. the 987 group phages; McDonnell *et al.*, 2016, Szymczak *et al.*, 2017) appear to target the more loosely associated cell wall glycan, or EPS (Szymczak *et al.*, 2018). These findings strongly suggest that genes (or clusters of genes) which have a role in the biosynthesis of *S. thermophilus* CWPS components are also responsible for phage receptor production.

The current study conclusively demonstrates that the *eps* cluster represents a genetic locus which is involved in the adsorption of members of one of the four defined phage groups infecting *S. thermophilus* (McDonnell *et al.*, 2016), namely the 987 group, to the cell surface of their host. Genetic complementation of three selected bacteriophage-insensitive mutants (BIMs) was shown to largely or fully restore phage sensitivity to these BIMs. CWPS structures of the examined strains (one WT and one BIM) were elucidated. Two CWPSs, one being a highly branched rhamnose-rich glycan (RGP) and a second (EPS) possessing a hexasaccharide repeating unit with a lactosyl side chain, were identified, and their chemical structures established. Furthermore, the EPS component of the CWPS was shown to be produced in the WT and complemented BIM only. In summation, the link between a specific cell wall-associated polysaccharide and phage sensitivity is clearly demonstrated.

Results and Discussion

BIM generation & phenotypic characterisation

S. thermophilus strain ST64987 (sensitive to 987 group phages) was employed in this study (Table 1). The genome sequences of the phages used (9871-9874) have been described previously (McDonnell *et al.*, 2016), and are generally conserved with respect to their defined group. Strain ST64987 was selected based on preliminary work which had shown that non-CRISPR-mediated BIMs are readily derived from this strain (at an approximate frequency of 10^{-4} to 10^{-5}) following exposure to phage 9871, with no CRISPR-mediated BIMs having been detected using this strain to date to our knowledge (data not shown). This is in contrast to previously analysed strains from which CRISPR-based BIMs are frequently derived (Deveau *et al.*, 2008). To investigate this, three representative BIMs (B987.1NC3, B987.1NC4 and B987.1NC8; henceforth known as BIM3, BIM4 and BIM8; Table 1) were selected for phage sensitivity analysis (Supplementary Table S1) and CRISPR locus sequencing (Supplementary Table S2). Three CRISPR loci were detected on the genomes of each strain (WT and BIMs; Supplementary Table S2 and S3), a commonly observed characteristic of *S. thermophilus* (Horvath *et al.*, 2008). CRISPR locus sequencing indicated that these BIMs had become resistant to phages 9871-4 by a CRISPR-independent mechanism, owing to the 100 % spacer identity between WT and BIMs. Furthermore, none of these spacers matched (with 100 % identity) any location on the genomes of the four phages (9871-4) used in challenges described herein.

Interestingly, BIMs 3, 4 and 8 exhibited subtle sedimenting phenotypes which were characterised by a slightly larger and apparently thinner cell pellet at the bottom of the tubes following overnight growth (Supplementary Figure S1). In order to investigate this phenotype further, all three BIMs were analysed by phage adsorption assays (Table 2) and microscopy (Table 3; Fig. 1), the latter to determine cell chain length relative to the WT. These analyses revealed that BIMs 3, 4 and 8 formed longer cell chains or aggregates (‘clumps’) than the parent and exhibited a significant reduction in their ability to adsorb phage 9871. Considering the high nucleotide and amino acid identity between the receptor binding genes of 9871 and the other 987 group phages used in this study (McDonnell *et al.*, 2016), it may be inferred that all three BIMs are resistant to phages 9871-4 by adsorption deficiency.

The observed sedimenting phenotype was reminiscent of observed similar phenotypes of phage resistant mutants of *L. lactis* (Dupont *et al.*, 2004, Breum *et al.*, 2007, Ainsworth *et al.*, 2014), as well as an *S. thermophilus* strain in which regions of the cell segregation-involved *cse*

gene were deleted (Layec *et al.*, 2009). Long chain phenotypes have previously been noted in *S. thermophilus* mutants in which the *rodA* gene (encoding the peptidoglycan polymerase RodA) was disrupted (Thibessard *et al.*, 2002). As such, sedimenting and long chain-forming phenotypes may be associated with an alteration in one or more components of the bacterial cell envelope. This phenotype, which confers an obvious advantage to the strain, may have industrial implications which are not currently known. While artificial genetic modification is not involved in this method of BIM generation (which would render the strain unsuitable for food production), a pilot fermentation using a sedimenting, long chain or aggregate forming strain may be required to explore the industrial impact(s) further.

SNP identification

To determine the genetic basis for the observed phenotypic characteristics of the three derived BIMs, the parent ST64987 strain and three BIMs were subjected to whole genome sequencing and subsequent single nucleotide polymorphism (SNP) analysis. Genome sequencing, sequence alignments and SNP calling was performed *in silico* as detailed in the Experimental Procedures. Between 10 and 14 nucleotide variations were initially detected in each BIM, as compared to the genome sequence of the parent strain. Subsequently, PCR and Sanger sequencing was carried out in order to eliminate those SNPs which were also present in the parent strain (i.e., apparent sequencing errors). This resulted in the number of SNPs between WT and BIMs being reduced to either 4 or 5 (Supplementary Table S4), with the majority of these located in an intergenic region containing a number of repeated sequences, which could therefore not be confirmed or ruled out as being either genuine SNPs or sequencing errors by the method described above.

Comparison of the SNPs located outside this region showed that the analysed BIMs carried SNPs in a previously identified gene cluster, designated as the *eps* cluster (Stingele *et al.*, 1996, Pluvinet *et al.*, 2004). Among different wild-type *S. thermophilus* strains, the *eps* cluster is characterized by a high level of variability, the presence (in some cases) of nucleotide identity to lactococcal *eps* clusters, and of multiple transposon/IS elements (Bourgoin *et al.*, 1999, Pluvinet *et al.*, 2004). This location was considered to be most relevant in order to explain changes in cell surface polysaccharide composition, thereby likely causing phage insensitivity due to adsorption deficiency. The *eps* gene cluster in strain ST64987 is schematically represented in Figure 2, and the mutational effects of each selected SNP are provided in Table 4. In the three sequenced

ST64987-derived BIMs, alterations were found in two distinct *eps* cluster-associated genes, both of which are predicted to encode glycosyltransferases (Fig. 2, Table 4; Supplementary Table S4). These mutations caused either amino acid alterations or the introduction of a stop codon leading to a truncated product (Table 4).

In silico analysis of mutated gene products

The protein product of *orf968*, mutated in BIM4 and BIM8, is a member of family 2 glycosyltransferases, with a HHpred search returning a 99.94 % probability match to TarS, a wall teichoic acid (WTA) β-glycosyltransferase of *Staphylococcus aureus* (Sobhanifar *et al.*, 2016). The product of *orf969*, mutated in BIM3, is a polysaccharide polymerase (predicted using BLAST; Altschul *et al.*, 1990), and contains 9 transmembrane helices (detected using TMHMM; Krogh *et al.*, 2001). This protein does not return a match using the Pfam database, and only one significant (probability 95.66 %) match using HHPred (Söding *et al.*, 2005, Zimmermann *et al.*, 2018), which indicates that it resembles the peptidoglycan polymerase RodA of *Thermus thermophilus* (Sjodt *et al.*, 2018). RodA is a member of the SEDS (shape, elongation, division, and sporulation) family of proteins, and is a transmembrane protein found to be a crucial peptidoglycan glycosyltransferase involved in cell wall biogenesis in other species (Cho *et al.*, 2016, Meeske *et al.*, 2016, Emami *et al.*, 2017). In a prior study, the *rodA* gene of *S. thermophilus* CNRZ368 (the product of which is similarly related to RodA of *T. thermophilus* (100 % probability; HHPred)) was disrupted (Thibessard *et al.*, 2002). Those *S. thermophilus* cells harbouring a disruption in the *rodA* gene showed increased chain length and altered morphology relative to WT cells. Taken together, these data point to a role for the products of *orf0968* and *orf969* in polysaccharide biosynthesis in *S. thermophilus*.

Genetic complementation of the eps mutations

To conclusively prove that the identified mutations were responsible for the observed phage adsorption deficiencies (and resulting phage insensitivities) in BIMs 3, 4, and 8, genetic complementation experiments were undertaken. An intact copy of *orf969* on plasmid pNZ44 (including a homolog of the anti-CRISPR gene AcrⅡA6 to promote stability of the plasmid, resulting in a construct designated pCOMP1; Table 1) was introduced into the BIM3 mutant (see Experimental Procedures). Plasmid construct pCOMP2, which carries an intact copy of *orf968* (Table 1), meanwhile, was introduced into BIM4 and BIM8. The introduction of the native genes

into the mutants (*orf969* into BIM3 and *orf968* into BIM4 and BIM8) largely (in the case of BIM3) or fully (in the case of BIM4 and BIM8) restored their sensitivities to all four 987 group phages, while the introduction of pNZ44 alone had no effect on phage sensitivity (Table 5). Interestingly, the observed efficiencies of plaquing of phages 9871-4 on BIM3::pCOMP1 were approximately tenfold lower than those observed on the parent strain ST64987 (though at least one million times greater than those on the non-complemented BIM3). Though the reason for this observation was not determined, and we speculate that this incomplete restoration of sensitivity, or partial complementation, is due to a non-native promoter being present on the complementing plasmid – leading to *orf969* expression differences, or a perturbation of EPS production kinetics, between the WT and complemented BIM.

Plasmid curing was then performed, which resulted in the reversion to the phage resistant phenotype of all three BIMs prior to the introduction of the complementing pCOMP1 or pCOMP2 constructs (Table 4).

Carbohydrate analysis I: Preparation of total CWPS and preliminary structural analysis

In order to verify the suspected difference in cell wall (CW)-associated carbohydrate structures of the WT strain and a representative mutant, the CW-associated polysaccharides of strain ST64987 and its derivative BIM3 were extracted and analysed. CWPSs were first extracted from washed cells with cold TCA, followed by sequential re-extraction with hot HCl (0.01 M and 0.1 M), as described previously (Vinogradov *et al.*, 2016, Vinogradov *et al.*, 2018a, Vinogradov *et al.*, 2018b). HCl extracts were treated with HF and purified on a Sephadex G-50 column to generate CWPS preparations, as described in the Experimental Procedures.

Only small amounts of polysaccharide (PS) material were obtained from TCA extracts of WT and mutant strains. Both TCA extracts contained Rha, Glc, Gal and GalNAc, while their ^{1}H spectra indicated heterogeneous mixtures. The CWPS preparation from HCl extracts contained Rha, Glc, Gal and GlcNAc in an approximate molar ratio of 1.1 : 1.0 : 0.4 : 0.2 (WT strain) and 1.4 : 1.0 : 0.27 : 0.25 (mutant strain).

The NMR spectra of the CWPS preparation of the WT strain were highly heterogeneous. Methylation analysis showed the presence of 2- and 2,3-substituted Rha, terminal and 2,6-substituted Glc, terminal Gal, 3-substituted GlcNAc and large number of mono-substituted Hex residues (Fig. 3 A), indicating a complex branched structure and/or a mixture of different components.

Taking the presence of a 3-linked GlcNAc in the WT CWPS preparation into account, we attempted the *N*-deacetylation and deamination in order to obtain CWPS fragments. This method of selective cleavage of a sugar chain at the 2-deoxy-2-aminosugar residue has previously been successfully applied for structural investigation of *L. lactis* UC509.9 CWPS, a complex branched, rhamnose-rich polysaccharide with GlcNAc residues in branching points (Vinogradov *et al.*, 2018b).

N-Deacetylated polysaccharide was desalted on a Biogel P2 column, where it eluted within the void volume. After deamination with nitric acid and fractionation of the Biogel P2 column, it afforded a deaminated polysaccharide (DPS) and oligosaccharide (OS) fractions, which were then analysed by methylation analysis and 2D NMR.

Methylation analysis of the DPS preparation revealed the presence of 2-substitued Rha, terminal and 2,6-substituted Glc as main components, as well as some minor mono- and di-substituted Hex residues. The structure of the main component was established by 2D NMR. The ^{1}H -^{13}C heteronuclear single-quantum coherence (HSQC) NMR spectra revealed the presence of four major anomeric signals: A, B, C, and D. Complete assignment of 2D NMR spectra (Supplementary Table S5) and analysis of nuclear Overhauser effect spectroscopy (NOESY) led to the identification of a branched tetrasaccharide repeating unit containing a linear backbone of two 2-linked α-Rha residues and a 6-linked α-Glc residue substituted at position 2 with α-Glc (Fig. 4 (i)).

Oligosaccharide fraction after deamination contained 2-linked Rha, terminal Glc, terminal Gal and 2-linked Gal as main components. 2D NMR analysis of this fraction allowed us to identify two oligosaccharides, a tetrasaccharide OS1 and a trisaccharide OS2 (Fig. 4 (ii), Supplementary Table S6). OS1 and OS2 exhibit a common trisaccharide fragment α-Gal-2-α-Rha-3-anhMan, with a 3-substituted 2, 5-anhydro-mannose X (product of the deamination of glucosamine) at the reducing end.

Thus, structural data of deamination products indicate that *S. thermophilus* ST64987 rhamnose rich RGP is composed of a backbone tetrasaccharide repeating units, carrying tri- and tetrasaccharide side chains with GlcNAc at branching points. Presence of the 2,3-linked Rha in the wild-type ST64987 CWPS (Fig. 3) indicates that the OS side chains are linked to the backbone at position 3 of Rha residue B or D (Fig. 4 (i)).

A number of anomeric signals of lower intensity between 4.4 and 4.7 ppm in the NMR spectra of the deaminated and original WT CWPS preparations (data not shown), together with the

data of methylation analysis, pointed to the presence of an additional polysaccharide component EPS, absent in the CWPS preparations of the mutant strain. 2D NMR analysis indicated that this component contained mainly β-Gal*p* and β-Glc.

Carbohydrate analysis II: Purification and structural analysis of RGP and EPS

In order to separate the two CWPS components of *S. thermophilus* ST64987, we took advantage of the presence of GlcNAc in RGP, being absent in EPS. The total CWPS preparation was *N*-deacetylated, thus providing positive charge to RGP due to the presence of free amino groups of $GlcNH_2$ residues. The resulting mixture was separated by cation-exchange chromatography giving a neutral fraction of EPS and a positively charged fraction of the *N*-deacetylated RGP.

Detailed 2D NMR analysis of the *N*-deacetylated RGP (Fig. 5; Supplementary Table S7) confirmed the previously obtained structural data and allowed us to clarify complete structure of this complex polymer. It showed that RGP from *S. thermophilus* ST64987 is composed of a backbone tetrasaccharide repeating unit identical to that of the corresponding DPS (see above) and carrying di-, tri- and tetrasaccharide side chains with the same structural motif and GlcNAc at branching points, on the majority of Rha residue B (Fig. 4 (i)).

The chemical structure of the purified EPS, as established by 2D-NMR techniques (Fig. 6), is presented in Figure 4 (iii). It was shown to be composed of hexasaccharide repeating units containing β-Glc, β-Gal and α-Glc, and to be identical to the EPS produced by *Lactobacillus delbrueckii* subsp. *bulgaricus* 291 (Faber *et al.*, 2001). ^{1}H and ^{13}C chemical shifts of the EPS (Supplementary Table S8) were identical or very close to literature data (Faber *et al.*, 2001). It is remarkable that the EPS repeating unit has a trimeric main chain of lactosyl-Glc and carries lactosyl branches. Curiously, this "lactose-enriched" polysaccharide is produced by two bacterial strains most widely used as yoghurt starters.

In summation, we identified two CWPSs present in the wild type strain ST64987: the major component, a complex branched rhamnose-rich RGP, and EPS, composed of hexasaccharide branched repeating units containing Gal and Glc.

Methylation analysis profile of the CWPS preparation of the mutant strain BIM3 showed the presence of all components characteristic of RGP, but not the 4- and 4,6- linked Glc and 4-linked Gal, that are characteristic of EPS (Fig. 3 B). Anomeric signals of EPS were absent in the ^{1}H-

NMR spectrum of this preparation (Fig. 7). These data clearly indicate that strain BIM3 produces an intact RGP, but that EPS is completely absent.

As mentioned above, complementation of BIM3 with plasmid pCOMP1 carrying the native *orf969* gene largely restored the sensitivity of this BIM to 987 group phages. In order to verify if the introduction of plasmid pCOMP1 into strain BIM3 reverted the carbohydrate profile of this BIM back to that of the WT strain, the extraction and purification of the CWPS was also carried out for the complemented strain BIM3::pCOMP1. Methylation analysis and ^{1}H-NMR data of this preparation, being identical to that of the WT strain (Fig. 3), confirmed that complementation restored EPS production in this complemented mutant.

Conclusions

Bacteriophages of *S. thermophilus* pose a costly risk to dairy fermentations worldwide (Garneau & Moineau, 2011). Despite this problem, the molecular details of the streptococcal phage receptor have yet to be elucidated (Quiberoni *et al.*, 2000, Szymczak *et al.*, 2018).

A recent study (Szymczak *et al.*, 2018) has lent credence to the hypothesis that streptococcal phage receptors are saccharidic in nature (Quiberoni *et al.*, 2000). Here, a BIM analysis approach was adopted to confirm this hypothesis. Three analysed BIMs were found to exhibit (i) identical CRISPR loci to the WT, (ii) insensitivity to four applied phages, (iii) altered phenotypic characteristics, and (iv) a genetic mutation within the *eps* cluster, all of which indicated that a cell surface alteration led to phage insensitivity. Adsorption and chain length studies in the BIMs further supported this hypothesis.

In order to irrefutably prove the essential role of the *eps*-located genes *orf968* and *orf969* in phage sensitivity in this strain, complementation studies were undertaken. While BIMs harbouring a mutated version of *orf968* or *orf969* were resistant to four applied phages, BIMs provided with a WT copy of the applicable mutated gene through complementation were largely susceptible to those phages.

CWPS analysis of ST64987 WT indicated the presence of a complex branched rhamno-polysaccharide (RGP), which was unaltered in BIM3. NMR spectra and methylation analysis indicated that a second polysaccharide structure (EPS), intact in the WT strain, was completely absent in the mutant - representing significant changes to the CWPSs of the BIM. In addition, the introduction of an intact copy of *orf969* into the mutant successfully converted the CWPS profile of BIM3 to that of the WT strain.

The results above confirm that *orf969* is required for the production of the EPS carbohydrate structure in ST64987, which in turn is required for infection by phage 9871. The findings herein represent definitive proof of the role and identity of a CWPS in *S. thermophilus* phage infection.

Experimental procedures

Growth and storage conditions of bacterial strains and bacteriophages

Bacterial strains and bacteriophages used in this study are listed in Table 1. *S. thermophilus* strains were routinely grown as described previously (McDonnell *et al.*, 2016). Bacteriophages were enumerated using a standard method (Lillehaug, 1997, McDonnell *et al.*, 2016), followed by single plaque purification (at least twice) according to a previously described method (Moineau *et al.*, 1994). Phage preparations were filtered (0.45 μm; Sarstedt, Numbrecht, Germany) and stored at 4 °C until required.

BIM generation and PCR screening

BIMs were generated using a standard plaque assay protocol (Lillehaug, 1997) modified as previously described (McDonnell *et al.*, 2018). All BIMs employed in the present study were generated by this method, with the exception of BIM4, which was generated by inoculating (1 % inoculum) 10 ml 10 % RSM medium with both fresh ST64987 overnight culture and phage 9871 (at a titer of approximately 1 x 10^8 pfu/ml). The mixture was incubated overnight at 42 °C whereupon clotting of the medium was observed. The resultant BIMs were twice single colony purified on LM17 and subjected to phenotypic and genotypic assays as described below.

BIMs were screened for phage resistance using the spot and plaque assay methods, as above. Those found to be resistant to phages were subjected to CRISPR sequencing (using the primers listed in Supplementary Table S3) to confirm their relatedness to the parent strain and, where applicable, to verify that no alterations had been made to the CRISPR loci spacer number or content. CRISPR repeat spacer loci were amplified, purified and sequenced as described previously (Horvath *et al.*, 2008; Supplementary Table S3, McDonnell *et al.*, 2018). CRISPR sequences were assembled using the Seqman program (Version 7.1.0 (44.1); DNAStar, Madison, WI, U.S.A) and visualised using the online CRISPR finder program (Grissa *et al.*, 2007).

Phenotypic assays

The generated BIMs were visually assessed for sedimentation in broth and cell chain length as described previously (McDonnell *et al.*, 2018), where microscopy was performed using a confocal microscope (LSM5; Carl Zeiss AG, Germany). Adsorption assays were performed based on a published method (Garvey *et al.*, 1996), and adapted as previously described (McDonnell *et al.*, 2016).

Genome sequencing & in silico analysis

Whole genome sequencing of *S. thermophilus* strain ST64987 was performed using a combination of Illumina and PacBio single molecule real-time (SMRT) sequencing. Briefly, SMRTbell libraries were constructed as previously described (Murray *et al.*, 2012) and SMRT sequencing was performed on a PacBio RSII instrument by GATC Biotech (Konstanz, Germany). Illumina sequencing was performed on a MiSeq instrument using a paired-end library in house by BaseClear (Leiden, The Netherlands). These sequencing data were assembled using the PacBio SMRTPortal platform (version 2.3.0), using MIRA v3.9 (http://www.chevreux.org/projects_mira.html). Open reading frame (ORF) prediction and initial annotation was performed using the PRODIGAL v2.0 (http://prodigal.ornl.gov) which utilised the Blastp program (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Further functional annotation of ORFs within *eps* and *rgp* gene clusters was performed manually using Pfam (Sonnhammer *et al.*, 1997) and HHpred (Soding *et al.*, 2005), where appropriate. Complete genomes were visualised using Artemis v16 (Rutherford *et al.*, 2000). The genomes of a total of three derived BIMs were sequenced using Illumina technology only, as described above. Single nucleotide polymorphisms (SNPs) were detected using the Bowtie2 alignment tool (Langmead *et al.*, 2009) and SAMtools (Li *et al.*, 2009). SNPs (Table 4) were either confirmed or rejected as true SNPs using PCR and Sanger sequencing as described above.

Construction & verification of complementation plasmid vectors

Plasmid preparations, restrictions and ligations of plasmid vectors used in the complementation of mutations in BIMs were performed as described previously (McDonnell *et al.*, 2018).

Plasmid vectors (Table 1), having been introduced into the relevant strain (as described below), were extracted using a GeneJet plasmid Miniprep kit (Thermo Scientific, Waltham, M.A., U.S.A.) and subjected to Sanger sequencing (as described above) using primers designed across the multiple cloning site (MCS) of pNZ44, and (where necessary) internal primers based on (predicted) glycosyltransferase-encoding genes (Supplementary Table S3). To promote the stability of the complementation vector and prevent plasmid elimination through BIM3 CRISPR1 activity, a phage encoded anti-CRISPR protein approach was employed. A homolog of the recently characterised anti-CRISPR protein AcrⅡA6 (Hynes *et al.*, 2018), was amplified from the lytic cos type phage ST907-9 (UCC collection) and cloned into pNZ44 prior to the introduction of

orf969. All primers used for the construction of the plasmid vectors are listed in Supplementary Table S3.

Preparation of competent cells, electrotransformation, transformant selection and plasmid curing

Competent cells were prepared using a method adapted from that of Holo and Nes (1989) and as described previously (McDonnell *et al.*, 2018). Where necessary, electrotransformation of competent cells was preceded by a heat shock step, and followed by recovery in HJL medium (30 g tryptone, 10 g yeast extract, 5 g potassium phosphate monobasic, 2 g beef extract, 5 g lactose, 1 L distilled water), as per the method of El Demerdash and colleagues (2003). Initial screening of transformants was based on PCR amplification using a forward primer designed upstream of the multiple cloning site (MCS) of pNZ44 (McGrath *et al.*, 2001), with the 'reverse' primer being designed based on the appropriate insert (Supplementary Table S3), to reduce the incidence of false positive detection. Plasmid DNA was extracted from selected transformants using a GeneJet plasmid Miniprep kit (Thermo Scientific) and the integrity of the constructs verified by Sanger sequencing as described above. Plasmid curing was performed as described previously (McDonnell *et al.*, 2018), with cured strains subjected to verification by PCR and plasmid profiling, as described above.

Extraction and purification of the CWPS

4 L of *S. thermophilus* culture was grown in M17 broth (Oxoid, U.K.) at 42°C supplemented with 0.5 % lactose. CWPS was extracted with cold 5% TCA and hot diluted HCl, essentially as described previously (Vinogradov *et al.*, 2016, Vinogradov *et al.*, 2018a, Vinogradov *et al.*, 2018b). The cell pellet was washed twice with deionized water, re-suspended in 5 % TCA and stirred for 48 h at 4 °C followed by centrifugation at 12,000 × *g*. The supernatant (TCA extract) was treated separately. The pellet was suspended in 0.01 N HCl, extracted at 100 °C with stirring for 20 min, cooled and centrifuged at 12,000 × *g*. The pellet was re-extracted with 0.1 N HCl as described above. The HCl extracts were de-proteinated by the addition of TCA (to a final concentration of 5 %), dialyzed and lyophilized to generate crude HCl extracts. The crude extract (20 mg) was treated with 48 % HF (100 µl) for 24 h at 4 °C. After evaporation of HF under a stream of nitrogen, the residue was re-suspended in water (1 ml) and fractionated on a Sephadex-G-50 column (1 x 40 cm). The fraction eluting close to the void volume was used for detailed NMR and methylation analysis, as well as *N*-deacetylation/deamination.

Deamination of the CWPS preparation

CWPS preparation of the WT strain ST64987 (60 mg) was deacylated (4 M NaOH, 2 mL, 110 °C, 24 h) and desalted on Biogel P2 column (2,6 x80 cm), to yield deacylated CWPS. Deacylated CWPS was deaminated by $NaNO_2$/AcOH: to the solution of the deacylated CWPS in water (2 mL) $NaNO_2$ (20 mg) and acetic acid (0.1 mL) were added, stirred to dissolve components and kept for 1 h at room temperature. Ammonium sulfamate (12.5 %, 2 ml) was then added to destroy excess of $NaNO_2$. Products were isolated by gel chromatography on a Biogel P2 column yielding the deaminated polysaccharide (DPS) and a mixture of oligosaccharides (OS). The OS fraction was re-purified on the Biogel P2 column, and analysed by methylation analysis and 2D NMR before and after reduction with $NaBH_4$.

Separation of EPS and N-*deacetylated RGP*

CWPS preparation of the WT strain ST64987 was *N*-deacetylated by 4 M NaOH (110 °C, 18 h), desalted, and separated on a cation-exchange Hitrap SP column, equilibrated in water. EPS was recovered in neutral fraction. Positively charged *N*-deacetylated EPS was eluted by linear gradient of 0-1 M NaCl, and desalted. Purified EPS and *N*-deacetylated RGP were used for 2D NMR structural analysis.

General and analytical methods

Gel-permeation chromatography was performed on Sephadex G-50 (GE Health Care, 1 x 40 cm) and BioGel P-2 (Biorad, 2.6 x 80 cm) columns, eluted with 0.1% AcOH. Fractions were assayed for neutral sugars (Dubois *et al.*, 1956).

Monosaccharide and methylation analyses were carried out as described previously (Vinogradov *et al.*, 2018a). GC was performed on a Trace GC ULTRA system (Thermo Scientific) equipped with a capillary column NMTR-5MS (30 m × 0.25 mm) and a FID detector using a temperature gradient of 170 °C (3 min) → 250 °C at 5 °C min^{-1}. GC/MS was carried out on the Trace GC ULTRA system TSQ quantum GC detector (Thermo Scientific), equipped with a capillary column SILGEL1MS (30 m x 0.25 mm) using a temperature gradient of 170 °C → 230 °C at 3 °C min^{-1}→ 270 °C at 10 °C min^{-1} (10 min). NMR spectroscopy experiments were carried out on a Bruker AVANCE III 600 MHz (1H) spectrometer with 5 mm Z-gradient probe with an acetone internal reference (2.225 ppm for 1H and 31.45 ppm for 13C) using standard pulse sequences cosygpprqf (gCOSY), mlevphpr (TOCSY, mixing time 120 ms), roesyphpr (ROESY,

mixing time 500 ms), hsqcedetgp (HSQC), hsqcetgpml (HSQC-TOCSY, 80 ms TOCSY delay) and hmbcgplpndqf (HMBC, 100 ms long range transfer delay). Resolution was kept <3 Hz/pt in F2 in proton-proton correlations and <5 Hz/pt in F2 of H-C correlations. The spectra were processed and analysed using the Bruker Topspin 2.1 program.

Accession numbers

The genome sequence data pertaining to ST64987 have been submitted to the GenBank database (ncbi.nlm.nih.gov/genbank) under accession number CP046859.

Acknowledgements

We thank PAGés platform (Plateforme d'Analyses des Glycoconjugués. UMR8576, Faculté des Sciences, Université de Lille, CNRS, 59000 Lille) for technical contributions. We gratefully acknowledge the financial support of DSM Food Specialities. T.R.H.M.K., L.H. and E.V.L.V.T are employees of DSM Food Specialties. This publication has emanated from research supported in part by a research grant from Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2273. D.V.S. is supported by a Principal Investigator award (Ref. no. 13/IA/1953) and a Spoke grant (Ref. no. 17/SP/4678) through SFI. J.M. is in receipt of a Starting Investigator Research Grant (SIRG; Ref. no. 15/SIRG/3430) funded by Science Foundation Ireland (SFI). B.M. is funded by DSM Food Specialties.

Conflict of interest statement

The data presented above are, in part, the subject of a patent application entitled "Phage insensitive *Streptococcus thermophilus*" (reference number WO2015124718).

Author contributions

BMD performed experimental work and wrote the manuscript. LH performed experimental work and assisted in data analysis. FB and PK performed bioinformatics work and interpreted the data generated by same. KL performed experimental work. IS and EV performed experimental work relating to carbohydrate analysis. EVL and TK reviewed and approved the manuscript. JM and DVS contributed to the concept and design of the study, and reviewed the manuscript.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Abbreviated summary

Bacteriophage-insensitive mutants of *S. thermophilus* were analysed phenotypically and genotypically. Mutants were found to be lacking a cell wall polysaccharide which is thus proven to be essential for phage adsorption and, hence, infection.

References

Ainsworth, S., Sadovskaya, I., Vinogradov, E., Courtin, P., Guerardel, Y., Mahony, J., Grard, T., Cambillau, C., Chapot-Chartier, M.P., and van Sinderen, D. (2014) Differences in lactococcal cell wall polysaccharide structure are major determining factors in bacteriophage sensitivity. *MBio* **5**: e00880-00814.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.

Arioli, S., Della Scala, G., Remagni, M.C., Stuknyte, M., Colombo, S., Guglielmetti, S., De Noni, I., Ragg, E., and Mora, D. (2017) *Streptococcus thermophilus* urease activity boosts *Lactobacillus delbrueckii* subsp. *bulgaricus* homolactic fermentation. *International Journal of Food Microbiology* **247**: 55-64.

Binetti, A., Quiberoni, A., and Reinheimer, J. (2002) Phage adsorption to *Streptococcus thermophilus*. Influence of environmental factors and characterization of cell-receptors. *Food research international* **35**: 73-83.

Bolotin, A., Quinquis, B., Renault, P., Sorokin, A., Ehrlich, S.D., Kulakauskas, S., Lapidus, A., Goltsman, E., Mazur, M., Pusch, G.D., Fonstein, M., Overbeek, R., Kyprides, N., Purnelle, B., Prozzi, D., Ngui, K., Masuy, D., Hancy, F., Burteau, S., Boutry, M., Delcour, J., Goffeau, A., and Hols, P. (2004) Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nature Biotechnology* **22**: 1554-1558.

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

Bourgoin, F., Pluvinet, A., Gintz, B., Decaris, B., and Guedon, G. (1999) Are horizontal transfers involved in the evolution of the *Streptococcus thermophilus* exopolysaccharide synthesis loci? *Gene* **233**: 151-161.

Breum, S.Ø., Neve, H., Heller, K.J., and Vogensen, F.K. (2007) Temperate phages TP901-1 and φLC3, belonging to the P335 species, apparently use different pathways for DNA injection in *Lactococcus lactis* subsp. *cremoris* 3107. *FEMS microbiology letters* **276**: 156-164.

Chapot-Chartier, M.P., Vinogradov, E., Sadovskaya, I., Andre, G., Mistou, M.Y., Trieu-Cuot, P., Furlan, S., Bidnenko, E., Courtin, P., Pechoux, C., Hols, P., Dufrene, Y.F., and Kulakauskas, S. (2010) Cell surface of *Lactococcus lactis* is covered by a protective polysaccharide pellicle. *Journal of biological chemistry* **285**: 10464-10471.

Cho, H., Wivagg, C.N., Kapoor, M., Barry, Z., Rohs, P.D., Suh, H., Marto, J.A., Garner, E.C., and Bernhardt, T.G. (2016) Bacterial cell wall biogenesis is mediated by SEDS and PBP polymerase families functioning semi-autonomously. *Nature Microbiology*: 16172.

Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008) Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J bacteriol* **190**: 1390-1400.

Driessen, F., Kingma, F., and Stadhouders, J. (1982) Evidence that *Lactobacillus bulgaricus* in yogurt is stimulated by carbon dioxide produced by *Streptococcus thermophilus*. *Netherlands Milk and Dairy Journal (Netherlands)*.

Dubois, M., Gilles, K.A., Hamilton, J.K., Rebers, P.t., and Smith, F. (1956) Colorimetric method for determination of sugars and related substances. *Analytical Chemistry* **28**: 350-356.

Duplessis, M., Levesque, C.M., and Moineau, S. (2006) Characterization of *Streptococcus thermophilus* host range phage mutants. *Appl Environ Microbiol* **72**: 3036-3041.

Duplessis, M., and Moineau, S. (2001) Identification of a genetic determinant responsible for host specificity in *Streptococcus thermophilus* bacteriophages. *Molecular microbiology* **41**: 325-336.

Dupont, K., Janzen, T., Vogensen, F.K., Josephsen, J., and Stuer-Lauridsen, B. (2004) Identification of *Lactococcus lactis* genes required for bacteriophage adsorption. *Applied and Environmental Microbiology* **70**: 5825-5832.

El Demerdash, H.A., Heller, K.J., and Geis, A. (2003) Application of the *shsp* gene, encoding a small heat shock protein, as a food-grade selection marker for lactic acid bacteria. *Applied and Environmental Microbiology* **69**: 4408-4412.

Emami, K., Guyet, A., Kawai, Y., Devi, J., Wu, L.J., Allenby, N., Daniel, R.A., and Errington, J. (2017) RodA as the missing glycosyltransferase in *Bacillus subtilis* and antibiotic discovery for the peptidoglycan polymerase pathway. *Nature Microbiology* **2**: 16253.

Faber, E.J., Kamerling, J.P., and Vliegenthart, J.F. (2001) Structure of the extracellular polysaccharide produced by *Lactobacillus delbrueckii* subsp. *bulgaricus* 291. *Carbohydrate Research* **331**: 183-194.

Garneau, J.E., and Moineau, S. (2011) Bacteriophages of lactic acid bacteria and their impact on milk fermentations. *Microbial cell factories* **10 Suppl 1**: S20.

Garvey, P., Hill, C., and Fitzgerald, G.F. (1996) The Lactococcal Plasmid pNP40 Encodes a Third Bacteriophage Resistance Mechanism, One Which Affects Phage DNA Penetration. *Applied and Environmental Microbiology* **62**: 676-679.

Germond, J.E., Delley, M., D'Amico, N., and Vincent, S.J. (2001) Heterologous expression and characterization of the exopolysaccharide from *Streptococcus thermophilus* Sfi39. *The FEBS journal* **268**: 5149-5156.

Goh, Y.J., Goin, C., O'Flaherty, S., Altermann, E., and Hutkins, R. (2011) Specialized adaptation of a lactic acid bacterium to the milk environment: the comparative genomics of *Streptococcus thermophilus* LMD-9. *Microbial cell factories* **10 Suppl 1**: S22.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007) CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Research* **35**: W52-57.

Herve-Jimenez, L., Guillouard, I., Guedon, E., Boudebbouze, S., Hols, P., Monnet, V., Maguin, E., and Rul, F. (2009) Postgenomic analysis of *Streptococcus thermophilus* cocultivated in milk with *Lactobacillus delbrueckii* subsp. *bulgaricus*: involvement of nitrogen, purine, and iron metabolism. *Applied and Environmental Microbiology* **75**: 2062-2073.

Holo, H., and Nes, I.F. (1989) High-Frequency Transformation, by Electroporation, of *Lactococcus lactis* subsp. *cremoris* Grown with Glycine in Osmotically Stabilized Media. *Applied and Environmental Microbiology* **55**: 3119-3123.

Hols, P., Hancy, F., Fontaine, L., Grossiord, B., Prozzi, D., Leblond-Bourget, N., Decaris, B., Bolotin, A., Delorme, C., Dusko Ehrlich, S., Guedon, E., Monnet, V., Renault, P., and Kleerebezem, M. (2005) New insights in the molecular biology and physiology of *Streptococcus thermophilus* revealed by comparative genomics. *FEMS microbiology reviews* **29**: 435-463.

Horvath, P., Romero, D.A., Coute-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., and Barrangou, R. (2008) Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J bacteriol* **190**: 1401-1412.

Hynes, A.P., Rousseau, G.M., Agudelo, D., Goulet, A., Amigues, B., Loehr, J., Romero, D.A., Fremaux, C., Horvath, P., Doyon, Y., Cambillau, C., and Moineau, S. (2018) Widespread anti-CRISPR proteins in virulent bacteriophages inhibit a range of Cas9 proteins. *Nature Communications* **9**: 2919.

Krogh, A., Larsson, B., Von Heijne, G., and Sonnhammer, E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology* **305**: 567-580.

Kuipers, O.P., de Ruyter, P.G., Kleerebezem, M., and de Vos, W.M. (1998) Quorum sensing-controlled gene expression in lactic acid bacteria. *Journal of Biotechnology* **64**: 15-21.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**: R25.

Lavelle, K., Murphy, J., Fitzgerald, B., Lugli, G.A., Zomer, A., Neve, H., Ventura, M., Franz, C.M., Cambillau, C., van Sinderen, D., and Mahony, J. (2018) A Decade of *Streptococcus thermophilus* Phage Evolution in an Irish Dairy Plant. *Appl Environ Microbiol* **84**.

Layec, S., Gerard, J., Legue, V., Chapot-Chartier, M.P., Courtin, P., Borges, F., Decaris, B., and Leblond-Bourget, N. (2009) The CHAP domain of Cse functions as an endopeptidase that acts at mature septa to promote *Streptococcus thermophilus* cell separation. *Molecular microbiology* **71**: 1205-1217.

Legrand, P., Collins, B., Blangy, S., Murphy, J., Spinelli, S., Gutierrez, C., Richet, N., Kellenberger, C., Desmyter, A., Mahony, J., van Sinderen, D., and Cambillau, C. (2016) The Atomic Structure of the Phage Tuc2009 Baseplate Tripod Suggests that Host Recognition Involves Two Different Carbohydrate Binding Modules. *MBio* **7**: e01781-01715.

Leuschner, R.G., Robinson, T.P., Hugas, M., Cocconcelli, P.S., Richard-Forget, F., Klein, G., Licht, T.R., Querol, A., Richardson, M., and Suarez, J.E. (2010) Qualified presumption of safety (QPS): a generic risk assessment approach for biological agents notified to the European Food Safety Authority (EFSA). *Trends in Food Science & Technology* **21**: 425-435.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Lillehaug, D. (1997) An improved plaque assay for poor plaque-producing temperate lactococcal bacteriophages. *J Appl Microbiol* **83**: 85-90.

Mahony, J., Kot, W., Murphy, J., Ainsworth, S., Neve, H., Hansen, L.H., Heller, K.J., Sorensen, S.J., Hammer, K., Cambillau, C., Vogensen, F.K., and van Sinderen, D. (2013) Investigation of the relationship between lactococcal host cell wall polysaccharide genotype and 936 phage receptor binding protein phylogeny. *Applied and Environmental Microbiology* **79**: 4385-4392.

Mahony, J., Randazzo, W., Neve, H., Settanni, L., and van Sinderen, D. (2015) Lactococcal 949 group phages recognize a carbohydrate receptor on the host cell surface. *Applied and Environmental Microbiology* **81**: 3299-3305.

McDonnell, B., Mahony, J., Hanemaaijer, L., Kouwen, T.R., and van Sinderen, D. (2018) Generation of Bacteriophage-Insensitive Mutants of *Streptococcus thermophilus* via an Antisense RNA CRISPR-Cas Silencing Approach. *Applied and Environmental Microbiology* **84**: e01733-01717.

McDonnell, B., Mahony, J., Hanemaaijer, L., Neve, H., Noben, J.P., Lugli, G.A., Ventura, M., Kouwen, T.R., and van Sinderen, D. (2017) Global Survey and Genome Exploration of Bacteriophages Infecting the Lactic Acid Bacterium Streptococcus thermophilus. *Front Microbiol* **8**: 1754.

McDonnell, B., Mahony, J., Neve, H., Hanemaaijer, L., Noben, J.-P., Kouwen, T., and van Sinderen, D. (2016) Identification and analysis of a novel group of bacteriophages infecting the lactic acid bacterium *Streptococcus thermophilus*. *Applied and Environmental Microbiology* **82**: 5153-5165.

McGrath, S., Fitzgerald, G.F., and van Sinderen, D. (2001) Improvement and optimization of two engineered phage resistance mechanisms in *Lactococcus lactis*. *Applied and Environmental Microbiology* **67**: 608-616.

Meeske, A.J., Riley, E.P., Robins, W.P., Uehara, T., Mekalanos, J.J., Kahne, D., Walker, S., Kruse, A.C., Bernhardt, T.G., and Rudner, D.Z. (2016) SEDS proteins are a widespread family of bacterial cell wall polymerases. *Nature* **537**: 634-638.

Mende, S., Dong, T., Rathemacher, A., Rohm, H., and Jaros, D. (2014) Physicochemical characterisation of the exopolysaccharides of *Streptococcus thermophilus* ST-143. *International Journal of Food Science & Technology* **49**: 1254-1263.

Moineau, S., Pandian, S., and Klaenhammer, T.R., (1994) Evolution of a Lytic Bacteriophage via DNA Acquisition from the *Lactococcus lactis* Chromosome. In: Applied and Environmental Microbiology. pp. 1832-1841.

Mora, D., Fortina, M.G., Parini, C., Ricci, G., Gatti, M., Giraffa, G., and Manachini, P.L. (2002) Genetic diversity and technological properties of *Streptococcus thermophilus* strains isolated from dairy products. *J Appl Microbiol* **93**: 278-287.

Mozzi, F., Vaningelgem, F., Hebert, E.M., Van der Meulen, R., Foulquie Moreno, M.R., Font de Valdez, G., and De Vuyst, L. (2006) Diversity of heteropolysaccharide-producing lactic acid bacterium strains and their biopolymers. *Applied and Environmental Microbiology* **72**: 4431-4435.

Murray, I.A., Clark, T.A., Morgan, R.D., Boitano, M., Anton, B.P., Luong, K., Fomenkov, A., Turner, S.W., Korlach, J., and Roberts, R.J. (2012) The methylomes of six bacteria. *Nucleic Acids Research* **40**: 11450-11462.

Pluvinet, A., Charron-Bourgoin, F., Morel, C., and Decaris, B. (2004) Polymorphism of *eps* loci in *Streptococcus thermophilus*: sequence replacement by putative horizontal transfer in *S. thermophilus* IP6757. *Int Dairy J* **14**: 627-634.

Quiberoni, A., Stiefel, J.I., and Reinheimer, J.A. (2000) Characterization of phage receptors in *Streptococcus thermophilus* using purified cell walls obtained by a simple protocol. *J Appl Microbiol* **89**: 1059-1065.

Reinheimer, J.A., Binetti, A.G., Quiberoni, A., Bailo, N.B., Rubiolo, A.C., and Giraffa, G. (1997) Natural milk cultures for the production of Argentinian cheeses. *J Food Prot* **60**: 59-63.

Rodriguez, C., Van der Meulen, R., Vaningelgem, F., Font de Valdez, G., Raya, R., De Vuyst, L., and Mozzi, F. (2008) Sensitivity of capsular-producing *Streptococcus thermophilus* strains to bacteriophage adsorption. *Lett Appl Microbiol* **46**: 462-468.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* **16**: 944-945.

Sjodt, M., Brock, K., Dobihal, G., Rohs, P.D.A., Green, A.G., Hopf, T.A., Meeske, A.J., Srisuknimit, V., Kahne, D., Walker, S., Marks, D.S., Bernhardt, T.G., Rudner, D.Z., and Kruse, A.C. (2018) Structure of the peptidoglycan polymerase RodA resolved by evolutionary coupling analysis. *Nature* **556**: 118-121.

Sobhanifar, S., Worrall, L.J., King, D.T., Wasney, G.A., Baumann, L., Gale, R.T., Nosella, M., Brown, E.D., Withers, S.G., and Strynadka, N.C. (2016) Structure and mechanism of *Staphylococcus aureus* TarS, the wall teichoic acid β-glycosyltransferase involved in methicillin resistance. *PLoS pathogens* **12**: e1006067.

Soding, J., Biegert, A., and Lupas, A.N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research* **33**: W244-248.

Söding, J., Biegert, A., and Lupas, A.N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research* **33**: W244-W248.

Sonnhammer, E.L., Eddy, S.R., and Durbin, R. (1997) Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins* **28**: 405-420.

Stingele, F., Neeser, J.R., and Mollet, B. (1996) Identification and characterization of the *eps* (Exopolysaccharide) gene cluster from *Streptococcus thermophilus* Sfi6. *J bacteriol* **178**: 1680-1690.

Sun, Z., Chen, X., Wang, J., Zhao, W., Shao, Y., Wu, L., Zhou, Z., Sun, T., Wang, L., Meng, H., Zhang, H., and Chen, W. (2011) Complete genome sequence of *Streptococcus thermophilus* strain ND03. *J bacteriol* **193**: 793-794.

Szymczak, P., Filipe, S.R., Covas, G., Vogensen, F.K., Neves, A.R., and Janzen, T. (2018) Cell Wall Glycans Mediate Recognition of the Dairy Bacterium *Streptococcus thermophilus* by Bacteriophages. *Applied and Environmental Microbiology* **84**.

Szymczak, P., Janzen, T., Neves, A.R., Kot, W., Hansen, L.H., Lametsch, R., Neve, H., Franz, C.M., and Vogensen, F.K. (2017) Novel Variants of *Streptococcus thermophilus* Bacteriophages Are Indicative of Genetic Recombination among Phages from Different Bacterial Species. *Applied and Environmental Microbiology* **83**.

Thibessard, A., Fernandez, A., Gintz, B., Leblond-Bourget, N., and Decaris, B. (2002) Effects of rodA and pbp2b disruption on cell morphology and oxidative stress response of *Streptococcus thermophilus* CNRZ368. *J bacteriol* **184**: 2821-2826.

Van de Guchte, M., Penaud, S., Grimaldi, C., Barbe, V., Bryson, K., Nicolas, P., Robert, C., Oztas, S., Mangenot, S., and Couloux, A. (2006) The complete genome sequence of *Lactobacillus bulgaricus* reveals extensive and ongoing reductive evolution. *PNAS* **103**: 9274-9279.

Vinogradov, E., Sadovskaya, I., Courtin, P., Kulakauskas, S., Grard, T., Mahony, J., van Sinderen, D., and Chapot-Chartier, M.-P. (2018a) Determination of the cell wall polysaccharide and teichoic acid structures from *Lactococcus lactis* IL1403. *Carbohydrate Research* **462**: 39-44.

Vinogradov, E., Sadovskaya, I., Grard, T., and Chapot-Chartier, M.-P. (2016) Structural studies of the rhamnose-rich cell wall polysaccharide of *Lactobacillus casei* BL23. *Carbohydrate Research* **435**: 156-161.

Vinogradov, E., Sadovskaya, I., Grard, T., Murphy, J., Mahony, J., Chapot-Chartier, M.-P., and van Sinderen, D. (2018b) Structural studies of the cell wall polysaccharide from *Lactococcus lactis* UC509. 9. *Carbohydrate Research* **461**: 25-31.

Zimmermann, L., Stephens, A., Nam, S.Z., Rau, D., Kubler, J., Lozajic, M., Gabler, F., Soding, J., Lupas, A.N., and Alva, V. (2018) A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology* **430**: 2237-2243.

Tables

Table 1. Bacterial strains, bacteriophages, BIMs and plasmid constructs used in this study.

Bacterial strains (WT)	Description	Source
ST64987	Industrial starter strain sensitive to phages 9871-4	DSM, Delft, The Netherlands
NZ9000	*L. lactis* cloning host	(Kuipers *et al.*, 1998)
S. thermophilus BIMs		
BIM3	Non-CRISPR BIM of ST64987 generated against phage 9871	This study
BIM4	Non-CRISPR BIM of ST64987 generated against phage 9871	This study
BIM8	Non-CRISPR BIM of ST64987 generated against phage 9871	This study
Phages		
9871	Lytic 987 group phage infecting ST64987	(McDonnell *et al.*, 2016)
9872	Lytic 987 group phage infecting ST64987	(McDonnell *et al.*, 2016)
9873	Lytic 987 group phage infecting ST64987	(McDonnell *et al.*, 2016)
9874	Lytic 987 group phage infecting ST64987	(McDonnell *et al.*, 2016)
ST907-9	Lytic cos group phage infecting ST90728	UCC collection
Plasmids		
pNZ44	Cloning vector, plasmid control	(McGrath *et al.*, 2001)
pCOMP1	BIM3 *orf969* complementing vector, containing an AcrIIa6 homolog and *orf0969*	This study
pCOMP2	BIM4 and BIM8 *orf968* complementing vector, containing *orf0968*	

Table 2. Adsorption level of phage 9871 on ST64987 WT and derived BIM BIM3. [1]Where indicated, values were calculated from a total of three sets of triplicate data.

Strain applied	Adsorption level of phage 9871 (%)	p value
ST64987 WT	95.0 ± 0.6[1]	N/A
BIM3	27.9 ± 3.2	0.000004
BIM4	19.9 ± 17.3	0.002
BIM8	17.7 ± 9.1	0.0002

Table 3. Average cells per chain (or clump; CPC), CPC % increase and phage adsorption levels observed in parent *S. thermophilus* strain ST64987 and its derived BIMs. The number of cells counted is represented by 'n'.

Parent Strain / BIM	Average CPC	% average CPC increase	*p* value
ST64987 WT	3.6 ± 2.3 (n=265)	N/A	N/A
BIM3	5.8 ± 6.5 (n=400)	61.1	0.007
BIM4	8.7 ± 7.9 (n=364)	141.7	0.00007
BIM8	8.3 ± 9.2 (n=290)	130.6	0.000001

Table 4. Confirmed SNPs (and downstream genetic effects of same) with proposed (or confirmed) involvement phage insensitivity in derived BIMs. Nucleotide deletion is indicated by '0'; frameshift mutation by 'fs'.

BIM	Base modification	Amino acid effect	*orf no*. (location)	Predicted function of product (BLAST)
BIM3	c → t	$Gln_{173}STOP$	*orf0969* (*eps* cluster)	Polysaccharide polymerase
BIM4	t → 0 (fs)	$Leu_{193}Trp$	*orf0968* (*eps* cluster)	Glycosyltransferase
BIM8	a → g	$Asp_{183}Gly$	*orf0968* (*eps* cluster)	"

Table 5. Relative efficiencies of plaquing (EOPs) of parent *S. thermophilus* strain ST64987 as well as derived and complemented BIMs. All values represent the average of three independent experiments. [1]Visible plaques appeared consistently less defined and smaller than those observed on the WT.

Strain	**EOP of phage applied**			
ST64987	**9871**	**9872**	**9873**	**9874**
WT	1	1	1	1
BIM3	$\leq 3.3 \times 10^{-8}$	$\leq 7.7 \times 10^{-8}$	$\leq 7.6 \times 10^{-8}$	$\leq 4.3 \times 10^{-8}$
BIM3:pNZ44	$\leq 3.3 \times 10^{-8}$	$\leq 7.7 \times 10^{-8}$	$\leq 7.6 \times 10^{-8}$	$\leq 4.3 \times 10^{-8}$
BIM3:pCOMP1	0.11 ± 0.01	0.11 ± 0.03	0.12 ± 0.04	0.12 ± 0.02
BIM3 Cured	$\leq 3.3 \times 10^{-8}$	$\leq 7.7 \times 10^{-8}$	$\leq 7.6 \times 10^{-8}$	$\leq 4.3 \times 10^{-8}$
BIM4	$\leq 9.5 \times 10^{-9}$	$\leq 1.3 \times 10^{-10}$	$\leq 9.0 \times 10^{-9}$	$\leq 1.3 \times 10^{-9}$
BIM4:pNZ44	$\leq 9.5 \times 10^{-9}$	$\leq 1.3 \times 10^{-10}$	$\leq 9.0 \times 10^{-9}$	$\leq 1.3 \times 10^{-9}$
BIM4:pCOMP2	0.9 ± 0.1	1.3 ± 0.2	1.2 ± 0.1	0.7 ± 0.0
BIM4 Cured	$\leq 4.3 \times 10^{-9}$	$\leq 7.4 \times 10^{-9}$	$\leq 5.0 \times 10^{-9}$	$\leq 5.2 \times 10^{-8}$
BIM8	$\leq 2.8 \times 10^{-9}$	[1]$1.27 \times 10^{-6} \pm 9.3 \times 10^{-7}$	$\leq 3.2 \times 10^{-9}$	$\leq 6.9 \times 10^{-8}$

BIM8:pNZ44	$\leq 3.1 \times 10^{-8}$	[13] $3.1 \times 10^{-5} \pm 2.3 \times 10^{-5}$	$\leq 6.6 \times 10^{-8}$	$\leq 4.0 \times 10^{-8}$
BIM8:pCOMP2	1.6 ± 0.7	1.0 ± 0.1	1.5 ± 0.5	1.2 ± 0.3
BIM8 Cured	$\leq 3.1 \times 10^{-8}$	[14] $4.0 \times 10^{-4} \pm 3.3 \times 10^{-4}$	$\leq 6.6 \times 10^{-8}$	$\leq 4.0 \times 10^{-8}$

Figure legends

Figure 1. Representative examples of the observed increase in cell chain length of *S. thermophilus* bacteriophage insensitive mutants BIMs relative to the parent strain from which they were derived. (A) ST64987 WT, (B) BIM3, (C) BIM4, (D) BIM8 (images generated using confocal microscopy as described in the Experimental Procedures).

Figure 2. Predicted ORF schematic of the *eps* gene clusters of *S. thermophilus* strains ST64987 as well as ND03 and LMG18311 for comparative purposes (Bolotin *et al.*, 2004, Sun *et al.*, 2011). Predicted ORFs are each indicated by an arrow and are colour coded according to their predicted function (as indicated in the legend). ORFs harbouring SNP(s) in non-CRISPR-mediated BIM(s) are indicated by an asterisk.

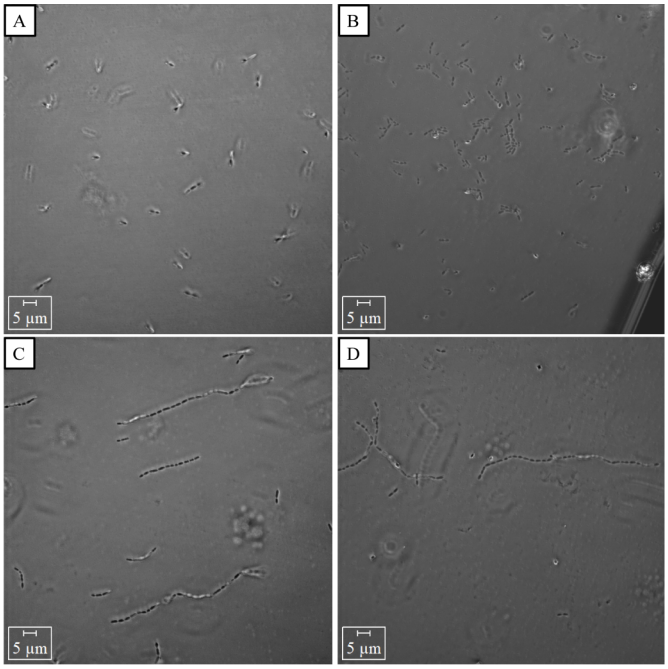
Figure 3. GC profiles corresponding to methylation analysis of CWPS preparations of the WT *S. thermophilus* ST64987 (A), its derivative BIM3 (B) and complemented BIM3::pCOMP1 (C). Residues characteristic for EPS are highlighted.

Figure 4. Structures of repeating units of *N*-deacetylated RGP from *S. thermophilus* ST64987-(i), and oligosaccharides obtained from ST64987 RGP by deamination (ii). GlcN X in the native RGP is *N*-acetylated. Finally, the structure of the repeating unit of the ST64987 EPS is shown in (iii).

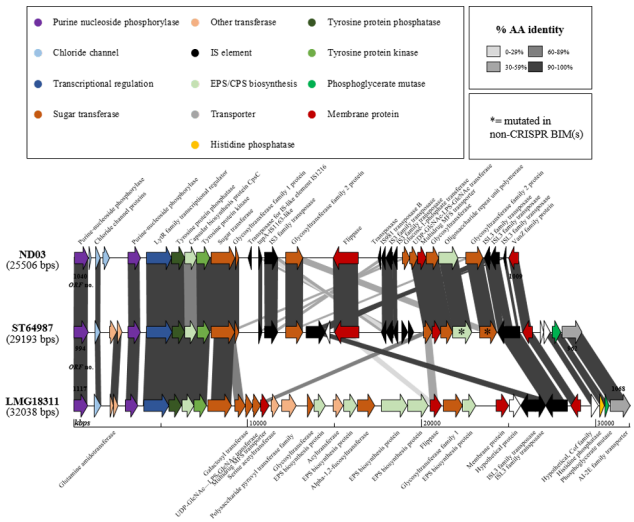
Figure 5. COSY (green), TOCSY (red) and NOESY (black) NMR spectra overlap and structure of repeating units for *N*-deacetylated RGP from *S. thermophilus* ST64987.

Figure 6. ^{1}H-^{13}C HSQC correlation and ^{1}H-NMR spectra of the *S. thermophilus* ST64987 EPS. Signals of CH groups are shown in black, with CH_2 groups shown in green.

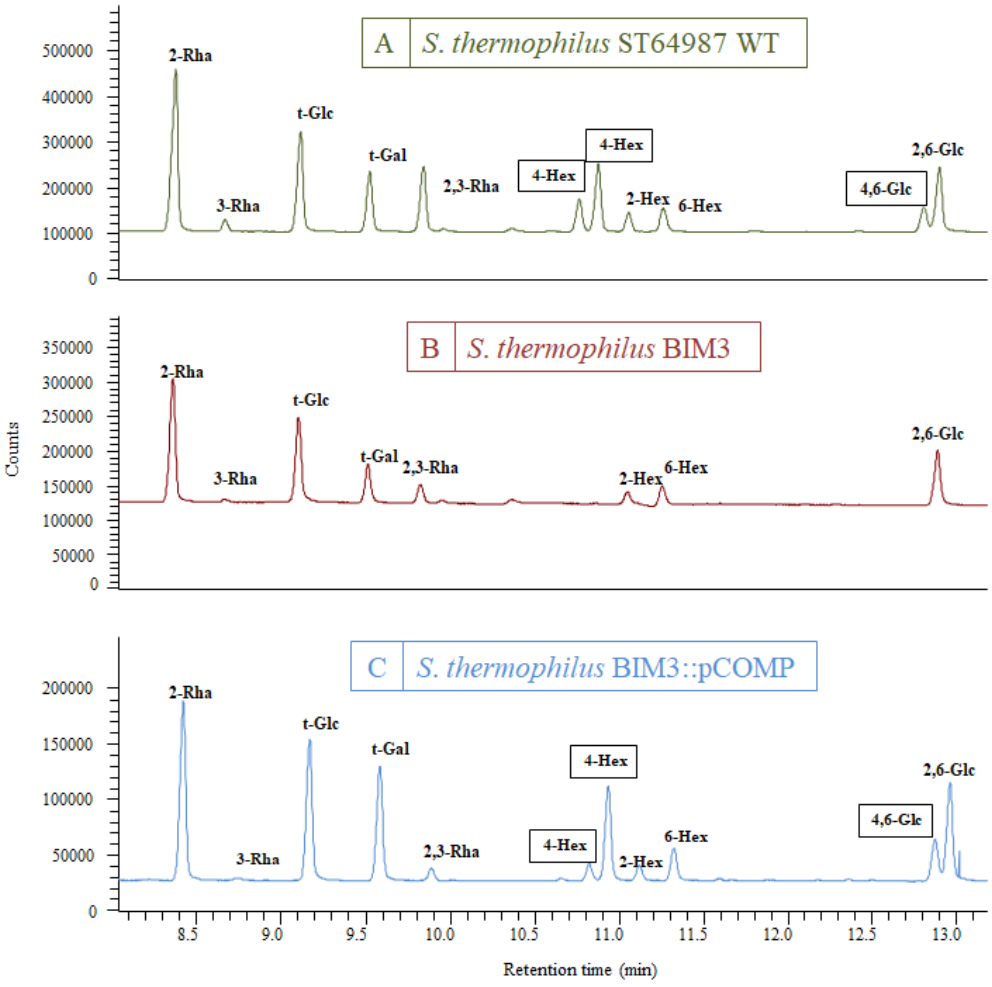
Figure 7. ^{1}H-NMR spectra of the CWPS preparation of the WT *S. thermophilus* ST64987 (bottom) and its derivative BIM3 (top). Anomeric region of EPS is highlighted.



mmi_14494_f1.tif



mmi_14494_f2.tif



mmi_14494_f3.tif

(i) *S. thermophilus* ST64987 RGP:

```
        C
    α-Glc
       |2       B
-6-α-Glc-2-α-Rha-2-α-Rha-
      A          |3        D
                 R
```

R =

α-Glc-2-α-Gal-2-α-Rha-3-β-GlcN-

F G E X

α-Gal-2-α-Rha-3-β-GlcN-

G' E' X

α-Rha-3-β-GlcN-

E" X

H (no side chain, Rha B is B' in this case) = S. thermophilus ST64987 DPS

(ii) Isolated oligosaccharides obtained by deamination of the *S. thermophilus* ST64987 RGP:

α-Glc-2-α-Gal-2-α-Rha-3-anhMan **OS1**

F B E X

α-Gal-2-α-Rha-3-anhMan **OS2**

G H X

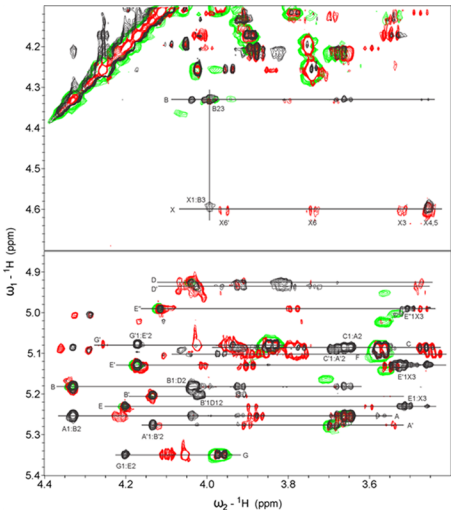
(iii) *S. thermophilus* ST64987 EPS:

```
        B          A        D
-4-β-Glc-4-α-Glc-4-β-Gal-
                |6
   β-Gal-4-β-Glc
```

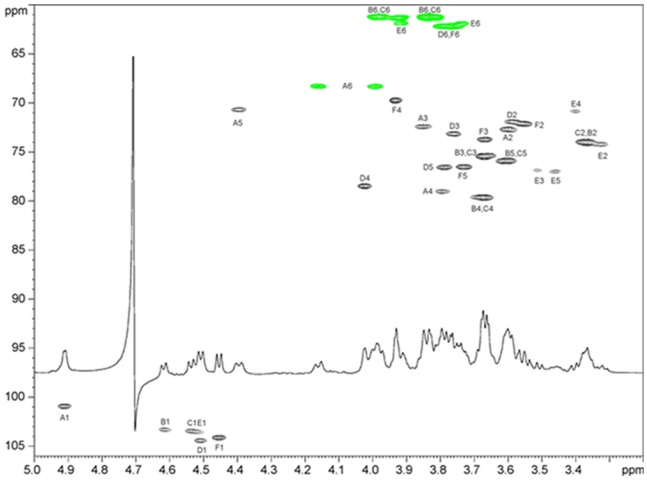
F C

E - terminal β-Glc, probably end of chain or structure with missing Gal F.

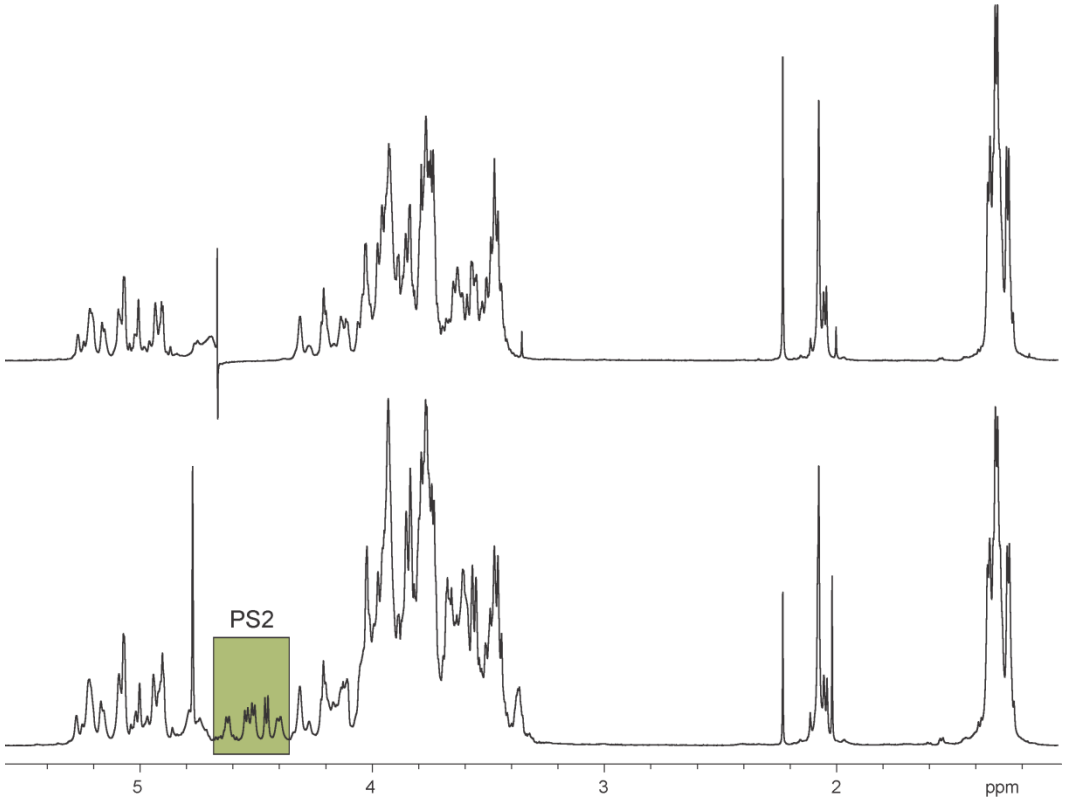
mmi_14494_f4.tif



mmi_14494_f5.tif



mmi_14494_f6.tif



mmi_14494_f7.tif