

# SCIENTIFIC REPORTS



OPEN

## 4,3- $\alpha$ -Glucanotransferase, a novel reaction specificity in glycoside hydrolase family 70 and clan GH-H

Received: 01 August 2016  
Accepted: 25 November 2016  
Published: 06 January 2017

Joana Gangoit<sup>1</sup>, Sander S. van Leeuwen<sup>1</sup>, Gerrit J. Gerwig<sup>1</sup>, Stéphane Duboux<sup>2</sup>, Christina Vafiadi<sup>2</sup>, Tjaard Pijning<sup>3</sup> & Lubbert Dijkhuizen<sup>1</sup>

Lactic acid bacteria possess a diversity of glucansucrase (GS) enzymes that belong to glycoside hydrolase family 70 (GH70) and convert sucrose into  $\alpha$ -glucan polysaccharides with ( $\alpha 1 \rightarrow 2$ )-, ( $\alpha 1 \rightarrow 3$ )-, ( $\alpha 1 \rightarrow 4$ )- and/or ( $\alpha 1 \rightarrow 6$ )-glycosidic bonds. In recent years 3 novel subfamilies of GH70 enzymes, inactive on sucrose but using maltodextrins/starch as substrates, have been established (e.g. GtfB of *Lactobacillus reuteri* 121). Compared to the broad linkage specificity found in GSs, all GH70 starch-acting enzymes characterized so far possess 4,6- $\alpha$ -glucanotransferase activity, cleaving ( $\alpha 1 \rightarrow 4$ )-linkages and synthesizing new ( $\alpha 1 \rightarrow 6$ )-linkages. In this work a gene encoding a putative GH70 family enzyme was identified in the genome of *Lactobacillus fermentum* NCC 2970, displaying high sequence identity with *L. reuteri* 121 GtfB 4,6- $\alpha$ -glucanotransferase, but also with unique variations in some substrate-binding residues of GSs. Characterization of this *L. fermentum* GtfB and its products revealed that it acts as a 4,3- $\alpha$ -glucanotransferase, converting amylose into a new type of  $\alpha$ -glucan with alternating ( $\alpha 1 \rightarrow 3$ )/( $\alpha 1 \rightarrow 4$ )-linkages and with ( $\alpha 1 \rightarrow 3,4$ ) branching points. The discovery of this novel reaction specificity in GH70 family and clan GH-H expands the range of  $\alpha$ -glucans that can be synthesized and allows the identification of key positions governing the linkage specificity within the active site of the GtfB-like GH70 subfamily of enzymes.

Glycoside hydrolase family 70 (<http://www.cazy.org/>) initially was established to accommodate glucansucrase enzymes converting sucrose into  $\alpha$ -glucan polysaccharides, exclusively found in lactic acid bacteria, especially in the genera *Leuconostoc*, *Streptococcus*, *Lactobacillus*, *Weissella* and *Oenococcus*<sup>1,2</sup>. Depending on the GS specificity, structurally different  $\alpha$ -glucans are formed. Initially mostly glucansucrase enzymes synthesizing  $\alpha$ -glucans with ( $\alpha 1 \rightarrow 3$ )- (mutan) or ( $\alpha 1 \rightarrow 6$ )-linkages (dextran) were characterized<sup>3-6</sup>; in recent years also various  $\alpha$ -glucans with ( $\alpha 1 \rightarrow 2$ )- or ( $\alpha 1 \rightarrow 4$ )-linkages have been identified as glucansucrase products<sup>7-9</sup>. A few glucansucrase enzymes have been characterized that produce  $\alpha$ -glucans with alternating ( $\alpha 1 \rightarrow 3$ )/( $\alpha 1 \rightarrow 6$ )-linkages (alternan) (e.g. alternansucrase of *Leuconostoc mesenteroides* NRRL B-1355)<sup>10</sup> or ( $\alpha 1 \rightarrow 4$ )/( $\alpha 1 \rightarrow 6$ )-linkages (reuteran) (e.g. reuteransucrase of *Lactobacillus reuteri* 121)<sup>8,9</sup>. Also, branching points may be introduced, either by the same enzyme<sup>11</sup> or by separate ( $\alpha 1 \rightarrow 2$ )- or ( $\alpha 1 \rightarrow 3$ )-branching enzymes of *L. mesenteroides* strains<sup>12,13</sup>. GSs thus are able to synthesize all four possible linkage types of  $\alpha$ -glycosidic bonds [( $\alpha 1 \rightarrow 2$ ), ( $\alpha 1 \rightarrow 3$ ), ( $\alpha 1 \rightarrow 4$ ) and ( $\alpha 1 \rightarrow 6$ )], but certain combinations of glycosidic linkages have been never found within the same  $\alpha$ -glucan. For example, no wild-type glucansucrase enzymes synthesizing  $\alpha$ -glucans with ( $\alpha 1 \rightarrow 3$ ) plus ( $\alpha 1 \rightarrow 4$ )-linkages have been reported so far. The  $\alpha$ -glucans produced by GSs also differ in their degree of branching, molecular mass, and conformation<sup>2</sup>. Such differences result in  $\alpha$ -glucans showing different functional properties with diverse and promising industrial applications.

Family GH70 GS enzymes (acting on sucrose) are evolutionary related to family GH13  $\alpha$ -amylase enzymes (acting on maltodextrins/starch), constituting clan GH-H<sup>14-17</sup>. Due to their evolutionary relatedness, GH70 and GH13 family enzymes display similarities in their sequence and structure, and use an  $\alpha$ -retaining double displacement catalytic mechanism<sup>14,18</sup>. Both families have a catalytic ( $\beta/\alpha$ )<sub>8</sub> barrel structure in their proteins, with domains A, B and C, and an active site with 4 conserved regions regarded as sequence fingerprints for the

<sup>1</sup>Microbial Physiology, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands. <sup>2</sup>Nestlé Research Center, Vers-chez-Les-Blanc, Lausanne, Switzerland. <sup>3</sup>Biophysical Chemistry, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands. Correspondence and requests for materials should be addressed to L.D. (email: L.Dijkhuizen@rug.nl)

individual enzyme specificities<sup>18,19</sup>. However, GSs also exhibit unique features. In family GH13 the order of these 4 conserved regions is I-II-III-IV. In contrast, in GS enzymes this  $(\beta/\alpha)_8$  barrel is circularly permuted, which results in the conserved region order II-III-IV-I. Moreover, GSs possess two extra and unique domains IV and V<sup>12,20–22</sup>. Besides, GSs present a “U-fold” domain structure in which 4 (domains A, B, IV and V) of the 5 domains are built up from two discontinuous segments of the polypeptide chain<sup>12,20,21</sup>. During their evolution from GH13, the GH70 enzymes appear to have undergone a sequence of gene rearrangements that resulted in this unusual, circularly permuted domain organization<sup>23</sup>.

In recent years several maltodextrins/starch converting enzymes have been identified within the GH70 family supporting the evolutionary relatedness of GH13 and GH70 families<sup>23–26</sup>. Firstly, it was found that *L. reuteri* 121 produced a GS-like enzyme that was inactive on sucrose. The gene encoding this enzyme was found upstream of the gene encoding the glucansucrase GtfA and was designated as *gtfB*. Instead of sucrose, the *L. reuteri* 121 GtfB acts on maltodextrins and starch substrates, cleaving  $(\alpha 1 \rightarrow 4)$ -linkages from the non-reducing end of the donor substrate, and synthesizing new  $(\alpha 1 \rightarrow 6)$ -linkages on the non-reducing end of the product (4,6- $\alpha$ -glucanotransferase activity, 4,6- $\alpha$ -GTase)<sup>24,27,28</sup>. This results in the synthesis of products with linear chains of  $(\alpha 1 \rightarrow 6)$ - and  $(\alpha 1 \rightarrow 4)$ -linkages (isomalto/maltopolysaccharides, IMMP)<sup>18</sup>. Later on, 2 more GtfB homologues were characterized displaying the same substrate and product specificity<sup>25</sup>. A total of 46 related GtfB type of enzymes are currently available in databases, constituting a new GH70 subfamily; with 4 exceptions they are all found in the genus *Lactobacillus*. Following the annotation of many new genome sequences, a few family GH70 enzymes also were found in non-LAB genera. Previously we have characterized the GtfC enzyme of *Exiguobacterium sibiricum* 255–15<sup>23</sup> and the GtfD enzyme of *Azotobacter chroococcum* NCIMB 8003<sup>26</sup>. Both of these enzymes are inactive with sucrose and active with maltodextrins/starch, displaying 4,6- $\alpha$ -GTase activity that resulted in synthesis of isomalto/malto oligosaccharides (IMMO) (GtfC)<sup>23</sup> and in a reuteran type of  $\alpha$ -glucan (GtfD)<sup>26</sup>. Surprisingly, the domain organization in GtfC and GtfD resembles that of GH13 enzymes, with a non-permuted order of conserved regions I-II-III-IV, and lacking domain V found in other GH70 enzymes. GtfC and GtfD represent 2 additional GH70 subfamilies and structurally appear to be very interesting evolutionary intermediates between GH13  $\alpha$ -amylase and GH70 glucansucrase enzymes, allowing further analysis of the evolutionary origins and differentiation of the (sub) families in clan GH-H (<http://www.cazy.org/>).

Based on the discovery of GtfB, GtfC and GtfD type of enzymes, it appears that 4,6- $\alpha$ -GTase is a common activity within the GH70 family. However, the lack of 3D structural information and of  $\alpha 1 \rightarrow 2$  and  $\alpha 1 \rightarrow 3$  synthesizing maltodextrins/starch utilizing type of GH70 enzymes has limited our understanding of the structural features determining this different substrate and product specificity. For GSs, the crystal structure of GTF180- $\Delta$ N with a bound maltose acceptor combined with site-directed mutagenesis experiments revealed that the linkage specificity is controlled by residues forming the acceptor substrate binding subsites +1/+2<sup>20,29–32</sup>. Despite their different domain organization, GtfB, GtfC, and GtfD 4,6- $\alpha$ -GTase enzymes display high conservation in their motifs I to IV, particularly in residues forming the acceptor substrate binding subsites in GSs<sup>23,26</sup>. Interestingly, a large number of these residues conserved in these 4,6- $\alpha$ -GTases differ in the GSs group of enzymes allowing the 4,6- $\alpha$ -GTase activity to be predicted by the analysis of the signature motifs I to IV.

In this work, annotation of the *Lactobacillus fermentum* NCC 2970 genome sequence resulted in identification of a GtfB-like GH70 enzyme displaying unique variations in some of the residues in conserved regions II and IV, contributing to the active site donor/acceptor substrate binding subsites. We therefore decided to biochemically characterize this *L. fermentum* GtfB enzyme, including a detailed characterization of its products. This revealed that it possesses 4,3- $\alpha$ -glucanotransferase (4,3- $\alpha$ -GTase) activity with maltodextrins/starch, a novel reaction specificity in family GH70 and clan GH-H. The *L. fermentum* GtfB thus represents a very interesting enzyme for investigation of structural features determining the linkage specificity in the GtfB-like GH70 subfamily of enzymes.

## Materials And Methods

***Lactobacillus fermentum* NCC 2970 genome sequence analysis.** The *Lactobacillus fermentum* NCC 2970 genome sequence has been obtained previously (GenBank accession no. CP017151) and was analyzed for the presence of carbohydrate active enzymes using the dbCAN database for automated Carbohydrate-active enzyme Annotation<sup>33</sup>. Hits having an E-Value below  $10^{-18}$  and a coverage above 0.35 were considered. Exploration of the genome and its gene content was performed using the WallGene software<sup>34</sup>.

***L. fermentum* NCC 2970 GtfB protein sequence analysis.** BLAST analysis of the *L. fermentum* NCC 2970 genome using the *L. reuteri* 121 *gtfB* gene (GenBank accession no. AAU08014.2) as query sequence revealed the presence of a single gene encoding a GH70 protein (CDS0277). Sequences showing similarity to the *L. fermentum* NCC 2970 GH70 protein sequence (GenBank accession no. AOR73699) were found using NCBI BLASTp searches (<http://www.ncbi.nlm.nih.gov/BLAST/>) against the non-redundant protein sequence database. Multiple amino acid sequence alignments were made with Clustal W2 using the Jalview 2 desktop application. The presence of a signal peptide was analyzed using the Signal P4 server. Subcellular localization of the *L. fermentum* GtfB protein was predicted using CELLO v.2.5: subCELLular LOcalization predictor (<http://cello.life.nctu.edu.tw/>). The theoretical  $M_w$  (molecular weight) of the GtfB protein was predicted on ExpASY Compute pI/ $M_w$  ([http://web.expasy.org/compute\\_pi/](http://web.expasy.org/compute_pi/)).

Phylogenetic analysis was performed using MEGA version 6<sup>35</sup> with a total of 72 amino acid sequences corresponding to representative characterized GH70 proteins indexed in CAZy and GtfB-like protein sequences identified via BLASTp. Sequences were aligned by MUSCLE, using default parameters. A phylogenetic tree was constructed by the Maximum Likelihood method based on the JTT matrix model using MEGA6. Partial deletion of the positions containing alignment gaps and missing data was conducted. Statistical confidence of the inferred phylogenetic relationships was assessed by performing 1,000 bootstrap replicates.

**Cloning of the *L. fermentum* NCC 2970 *gtfB* gene.** The DNA fragment coding for an N-terminally truncated version of the GtfB protein (amino acids 616–1593) was amplified from *L. fermentum* NCC 2970 chromosomal DNA using Phusion DNA polymerase (Finnzyme, Helsinki, Finland) and cloned into a modified pET15b vector using ligation-independent cloning (LIC). The primers used for amplifying the N-terminally truncated *gtfB* gene derivative incorporated 5' extensions (underlined) to facilitate the LIC cloning, and were: Forward CAGGGACCCGGTTTTGGTAAAGATGGTCGGATTG and Reverse CGAGGAGAAGCCCGTTAATTGTCTTCAATATTAGCATAATAATC. The resulting PCR product was purified from the agarose band and digested in the presence of dATP, with the 3' to 5' exonuclease activity of the T4 DNA polymerase (New England Biolabs). In parallel, the pET15b/LIC vector was digested with KpnI, isolated from gel and then treated with T4 DNA polymerase (New England Biolabs) in the presence of dTTP. The treated pET15b/LIC vector and the amplicon were mixed together in a 1:4 molar ratio, and the mixture was used to transform *Escherichia coli* DH5 $\alpha$  cells (Phabagen). This resulted in a *gtfB*- $\Delta$ N construct containing an N-terminal His6-tag cleavable by a 3C protease. The constructed expression vector pET15b/*gtfB*- $\Delta$ N was transformed into host *E. coli* BL21 Star (DE3). The gene sequence was verified by nucleotide sequencing (GATC, Cologne, Germany).

**Enzyme expression and purification.** For expression of the *L. fermentum* NCC 2970 GtfB- $\Delta$ N enzyme, an *E. coli* star BL21 (DE3) overnight culture, transformed with pET15b/*gtfB*- $\Delta$ N was diluted 1:100 into fresh LB broth supplemented with ampicillin (100  $\mu$ g ml<sup>-1</sup>) and grown at 37 °C and 220 rpm until the optical density at 600 nm reached approximately 0.4. The temperature for culturing was then decreased to 16 °C, and the inducer isopropyl- $\beta$ -D-1-thiogalactopyranoside was added to a final concentration of 0.1 mM. After 20 h, cells were harvested by centrifugation (10,000 g  $\times$  20 min), and subsequently disrupted with B-PER lysis reagent in accordance to the protocol described by the manufacturer (Thermo Scientific, Pierce). The recombinant GtfB protein was isolated from the cell-free extract by His-tag affinity chromatography using Ni<sup>2+</sup>-nitrilotriacetate (Ni-NTA) as column material (Sigma-Aldrich). After washing the column with 25 mM Tris-HCl (pH 8.0), 1 mM CaCl<sub>2</sub>, bound proteins were eluted with 200 mM imidazole in the same buffer and the imidazole was removed by use of a stirred ultrafiltration unit (Amicon, Beverly, MA) with a 30,000 molecular weight cut off. For further purification, the protein was loaded on a 1 ml-HiTrap column (GE Healthcare) and eluted (at a 1 ml/min flow rate) using a linear gradient of NaCl (from 0 to 1 M) in 20 mM Tris-HCl buffer (pH 8.0), containing 1 mM CaCl<sub>2</sub>. Fractions of 1 ml were collected using an Äkta fast protein liquid chromatograph (FPLC; GE Healthcare, Uppsala, Sweden). The buffer was exchanged by ultrafiltration (YM30 membranes; Millipore, Billerica, MA). The purification progress was assessed by SDS-PAGE analysis of the fractions, and the protein concentrations were determined using a Nanodrop 2000 spectrophotometer (Isogen Life Science, De Meern, The Netherlands).

**Enzyme activity assays.** The initial total activity of the *L. fermentum* NCC 2970 GtfB- $\Delta$ N enzyme was determined by the amylose-iodine method as described before<sup>23,27</sup>. The decrease in absorbance of the  $\alpha$ -glucan-iodine complex resulting from transglycosylation and/or hydrolytic activity was monitored at 660 nm for 14 min at 40 °C. The reaction mixture contained 0.125% (w v<sup>-1</sup>) amylose V (AVEBE, Foxhol, The Netherlands), 2.5  $\mu$ g ml<sup>-1</sup> of enzyme, 25 mM sodium acetate (pH 5.5) and 1 mM CaCl<sub>2</sub>. One unit of activity was defined as the amount of enzyme converting 1 mg of substrate per min.

The effect of pH on enzyme activity was determined at 40 °C by varying the pH between 3.5 and 8.0. Sodium citrate buffer (25 mM) was used for pH values between 3.5 and 7.0, and sodium phosphate buffer (25 mM) for pH values between 7.0 and 8.0. The optimal temperature was determined in 25 mM sodium citrate buffer (pH 5.5), 1 mM CaCl<sub>2</sub>, at temperatures ranging from 30 to 65 °C. The effect of temperature on GtfB- $\Delta$ N stability was determined by incubating the enzyme at a concentration of 0.1 mg ml<sup>-1</sup> in 20 mM Tris-HCl buffer (pH 5.5) with 1 mM CaCl<sub>2</sub> at temperatures from 30 to 55 °C for 10 min. Samples were then immediately cooled to 4 °C, and the residual activity was measured under the standard reaction conditions in 25 mM sodium citrate (pH 5.5) with 1 mM CaCl<sub>2</sub> at 40 °C.

**Substrate utilization by *L. fermentum* NCC 2970 GtfB.** *L. fermentum* NCC 2970 GtfB- $\Delta$ N (25  $\mu$ g ml<sup>-1</sup>) was separately incubated with 25 mM sucrose (Acros), nigerose (Sigma-Aldrich), panose (Sigma-Aldrich), isomaltose (Sigma-Aldrich), isomaltotriose (Sigma-Aldrich), isomaltopentaose (Carbosynth), malto-oligosaccharides (MOS) with degrees of polymerization (DP) 2–7, and 0.6% (w v<sup>-1</sup>) amylose V (AVEBE, Foxhol, The Netherlands), potato starch (Sigma-Aldrich) and amylopectin (Sigma-Aldrich). All reactions were performed in 25 mM sodium acetate buffer (pH 5.5) with 1 mM CaCl<sub>2</sub> at 37 °C for 24 h. Reactions were stopped by 6 min of incubation at 100 °C. The progress of the reactions was monitored by thin-layer chromatography (TLC) and/or high-performance-anion-exchange chromatography (HPAEC).

**Thin Layer Chromatography and High Performance Anion Exchange Chromatography with pulsed amperometric detection analysis.** Thin layer chromatography (TLC) was performed on silica gel 60 F254, 20  $\times$  20 cm TLC sheets (Merck, Darmstadt, Germany). The TLC plates were developed in *n*-butanol:acetic acid:water (2:1:1, v v<sup>-1</sup>) solvent system for 6 h. The carbohydrates were visualized with orcinol/sulfuric acid staining and compared with a simultaneous run of a mixture of glucose and MOS (DP2 to DP7).

Carbohydrate samples were diluted 3:100 in DMSO and analysed by HPAEC-PAD on an ICS3000 workstation (Thermo Scientific, Amsterdam, The Netherlands), equipped with a CarboPac PA-1 column (Thermo Scientific; 250  $\times$  2 mm) and an ICS3000 electrochemical detection module. A gradient of sodium acetate from 10 to 240 mM in 100 mM NaOH was applied over 57 min at 0.25 ml min<sup>-1</sup> flow rate. The injection volume of each sample was 5  $\mu$ l. The identity of the peaks was assigned using commercial oligosaccharide standards.

**HPSEC analysis.** HPSEC analyses of the product mixtures were performed using a size exclusion chromatography system (Agilent Technologies 1260 Infinity) equipped with a multi angle laser light scattering detector (SLD 7000 PSS, Mainz), a viscometer (ETA-2010 PSS, Mainz) and a differential refractive index detector (G1362A 1260 RID Agilent Technologies), as described before<sup>26</sup>. Separation was carried out by using three PFG-SEC columns with porosities of 100, 300 and 4000 Å, coupled with a PFG guard column. DMSO-LiBr (0.05 M) was used as eluent at a flow rate of 0.5 ml min<sup>-1</sup>. The system was calibrated and validated using a standard pullulan kit (PSS, Mainz, Germany) with  $M_w$  ranging from 342 to 805 000 Da. The specific RI increment value  $dn/dc$  was also measured by PSS and was 0.072 ml g<sup>-1</sup> (private communication with PSS). The multiangle laser light scattering signal was used to determine the molecular masses of the amylose V and the polymer generated by the *L. fermentum* GtfB- $\Delta$ N enzyme. The specific RI increment value,  $dn/dc$  for these polysaccharides in this system was taken to be the same as for pullulan. The molecular mass of the *L. reuteri* GtfB- $\Delta$ N polymer was determined by universal calibration method. WinGPC Unity software (PSS, Mainz) was used for data processing. Measurements were performed in duplicate.

**Synthesis, isolation and characterization of the products synthesized by *L. fermentum* NCC 2970 GtfB from amylose.** Purified GtfB- $\Delta$ N (0.25 mg) was incubated with amylose V for 48 h at 37 °C under the conditions described above in “Substrate utilization by *L. fermentum* NCC 2970 GtfB”. The product mixture obtained was fractionated by size-exclusion chromatography on a BioGel P2 column (2.5 × 50 cm; Bio-Rad, Veenendaal, The Netherlands) using 10 mM NH<sub>4</sub>HCO<sub>3</sub> as eluent at a flow rate of 48 ml h<sup>-1</sup>. The poly-/oligo-saccharides present in the different Biogel P2 pools were analyzed by matrix-assisted laser-desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS), nuclear magnetic resonance (NMR) spectroscopy and methylation analysis. The polysaccharide was also subjected to Smith degradation analysis.

**MALDI-TOF mass spectrometry.** MALDI-TOF-MS measurements were recorded on an Axima™ mass spectrometer (Shimadzu Kratos Inc., Manchester, UK) equipped with a nitrogen laser (337 nm, 3 ns pulse width). Sample solutions (1  $\mu$ l) were spotted on a MALDI target and mixed immediately with 1  $\mu$ l 10 mg ml<sup>-1</sup> 2,5-dihydroxybenzoic acid matrix solution in 40% aqueous acetonitrile. Positive-ion mode spectra were recorded using the reflector mode at a resolution of 5000 Full Width at Half Maximum (FWHM) and delayed extraction optimised at 800  $m/z$  by software control. Mass spectra were generally acquired from 1 to 5000  $m/z$  with ion-gate blanking at 200  $m/z$ .

**NMR spectroscopy.** One- and two- dimensional <sup>1</sup>H nuclear magnetic resonance (NMR) spectra were recorded on a Varian Inova 500 spectrometer (NMR Center, University of Groningen), using D<sub>2</sub>O as solvent and at a probe temperature of 298 K. Prior to analysis, samples were exchanged twice in D<sub>2</sub>O (99.9 atom% D, Cambridge Isotope Laboratories, Inc., Andover, MA) with intermediate lyophilization, and then dissolved in 0.6 ml of D<sub>2</sub>O. One-dimensional 500-MHz <sup>1</sup>H NMR spectra were recorded at a 4 000 Hz spectral width and 16k complex points, using a WET1D pulse to suppress the HOD signal. Two-dimensional <sup>1</sup>H-<sup>1</sup>H spectra (COSY, TOCSY MLEV17 30, 50, and 150 ms, and ROESY 300 ms) were recorded with 4 000 Hz spectral width, collecting 200 increments. In case of TOCSY spectra 2 000 complex data points were collected, for COSY and ROESY spectra 4000 complex data points were used. 2D <sup>13</sup>C-<sup>1</sup>H NMR spectra were recorded in 128 increments of 2000 complex points with 4000 Hz spectral width in  $t_2$  and 10 000 Hz in  $t_1$ . All NMR spectra were processed with MestReNova 5.3 (Mestrelabs Research SL, Santiago de Compostella, Spain). Manual phase correction was performed and a Whittaker smoother baseline correction was applied. Chemical shifts ( $\delta$ ) were expressed in ppm and calibrated with the internal standard acetone ( $\delta$  2.225 ppm for <sup>1</sup>H and  $\delta$  31.08 ppm for <sup>13</sup>C). The percentage of different linkages was estimated by integration of the respective signal peak areas.

**Methylation analysis.** Methylation analysis was performed as described previously<sup>36</sup>. Briefly, the polymer and oligosaccharides samples (~5 mg) were per-methylated using CH<sub>3</sub>I and solid NaOH in DMSO, and subsequently hydrolyzed with trifluoroacetic acid. Partially methylated monosaccharides were reduced with NaBD<sub>4</sub>. The resulting partially methylated alditols were per-acetylated using pyridine:acetic anhydride (1:1 v/v) at 120 °C yielding mixtures of partially-methylated alditol acetates (PMAAs). PMAAs were analysed by GLC-EI-MS and GLC-FID as described<sup>36</sup>.

**Smith degradation.** Samples of 1–2 mg polysaccharide were dissolved in 1 ml 50 mM NaIO<sub>4</sub> in 100 mM NaOAc (pH 4.1) and stirred for 112 h at 4 °C in the dark. Excess IO<sub>4</sub><sup>-</sup> was neutralised by adding 300  $\mu$ l ethylene glycol. The degraded product was dialyzed in a SpectraPor 1000 Da cut-off dialysis floater against running tap water for 48 h. After dialysis the sample was reduced with NaBH<sub>4</sub> at room temperature overnight, followed by dialysis as described above. The reduced polysaccharide sample was lyophilized and hydrolyzed in 1 ml 90% formic acid at 90 °C for 30 min. After cooling to room temperature formic acid was evaporated by N<sub>2</sub> stream. The dried polysaccharide fragments were analyzed as TMS derivatives on GLC-EI-MS and GLC-FID, and by HPAEC-PAD as described previously.

***L. fermentum* NCC 2970 GtfB product analysis with hydrolytic enzymes.** The polymer produced by GtfB- $\Delta$ N was dissolved at a concentration of 5 mg ml<sup>-1</sup> in 50 mM sodium acetate buffer pH 5.0, and incubated separately with an excess of  $\alpha$ -amylase (*Aspergillus oryzae*; Megazyme), dextranase (*Chaetomium erraticum*; Sigma-Aldrich), and pullulanase M1 (*Klebsiella planticola*; Megazyme) for 48 h at 37 °C. Starch, *L. reuteri* 121 GtfB IMMP polymer and *A. chroococcum* reuteran-like GtfD polymer, were used as positive controls for the  $\alpha$ -amylase, dextranase and pullulanase treatments, respectively, resulting in complete degradation under these conditions. The degree of hydrolysis was examined by TLC analysis.

**Structural modeling of the *L. fermentum* NCC 2970 GtfB protein.** The Phyre server<sup>37</sup> was used to obtain a homology model for the *L. fermentum* NCC 2970 GtfB protein. Querying the GtfB sequence against the PDB database suggested that the structure of DSR-E  $\Delta N_{123}$ -GBD-CD2 from *L. mesenteroides* NRRL B-1299 DSRE (PDB: 3TTO<sup>12</sup>) is the highest scoring structure with the highest number of residues aligned (residues 537–1500, covering domains A, B, C, IV and V). However, since we wanted to focus on the active site at the interface of domains A and B, we chose the structure of *L. reuteri* 180 Gtf180- $\Delta N$  (PDB: 3KLLK<sup>20</sup>), which has the highest homology with *L. fermentum* GtfB in these domains, for the Phyre one-to-one threading modeling protocol. The resulting model contained residues 674–1593 (40% homology to Gtf180- $\Delta N$ ). The model was superposed with complex of *L. reuteri* 180 Gtf180- $\Delta N$  with maltose (PDB: 3KLL<sup>20</sup>), for acceptor substrate modelling.

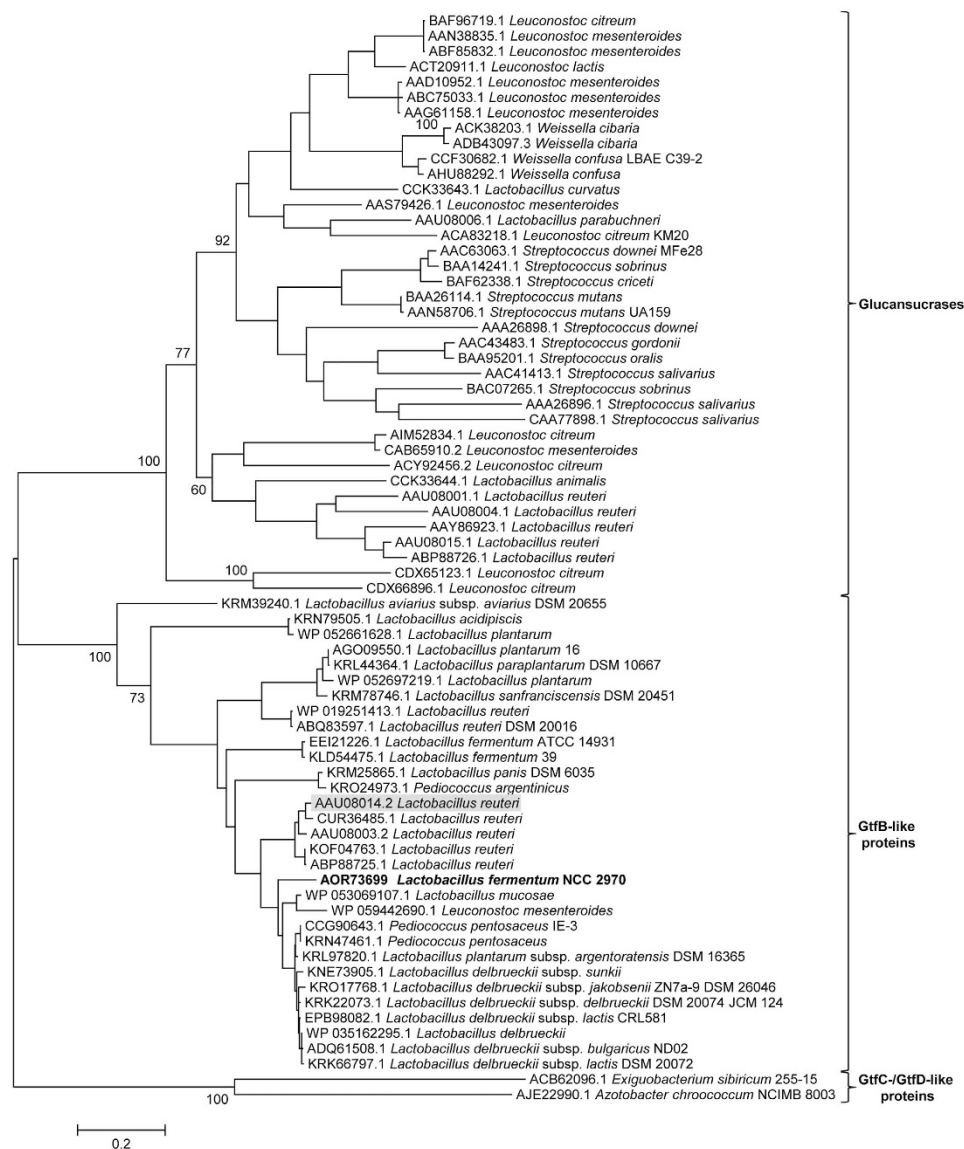
## Results and Discussion

***L. fermentum* NCC 2970 carbohydrate-active enzymes.** The genome sequences of more than 2500 isolates from the Nestlé Culture Collection have been subjected to analysis with a focus on the discovery and characterization of novel GH70 enzymes. The *L. fermentum* NCC 2970 genome was found to encode a total of 32 putative carbohydrate-active enzymes (CAZymes), as analyzed using dbCAN (Table S1). Of these, a putative GH70 family member (CDS0221) annotated as a dextransucrase (EC 2.4.1.5) was found located next to a transposase suggesting the potential acquisition of this gene via horizontal gene transfer. Interestingly, the *L. fermentum* NCC 2970 genome also encodes CAZymes enabling degradation of prebiotic compounds, i.e. compounds that are not digested by human enzymes<sup>38–40</sup>. These are (1) a member of the GH13 family (CDS1589), annotated as a  $\alpha$ -glucosidase, putatively involved in isomalto-oligosaccharide degradation; (2) a member of the GH32 family (CDS0603) annotated as a  $\beta$ -fructofuranosidase, putatively involved in degradation of fructo-oligosaccharides; (3) 3 different GH2 family members annotated as  $\beta$ -galactosidases (CDS0592, CDS1127, CDS1128), putatively involved in lactose, galacto-oligosaccharide or galactan degradation. Furthermore, a cluster of 5 genes was shown to be relatively unique to *L. fermentum* NCC 2970 and contained three proteins without any function attributed (CDS0595, CDS0596 and CDS0599) as well as two glycoside hydrolases: a member of the GH3 family (CDS0597) annotated as a  $\beta$ -N-acetylhexosaminidase (EC 3.2.1.52), and a GH78 family member (CDS0598) annotated as an  $\alpha$ -L-rhamnosidase (EC 3.2.1.40). Interestingly, adjacent to this cluster, also a putative xylan 1,3- $\beta$ -xylosidase (EC 3.2.1.72) belonging to the GH43 family (CDS594) was found. Overall this region of the *L. fermentum* NCC 2970 genome may be involved in the degradation of plant or fungi derived substrates as enzymes present were previously shown to hydrolyze chitin derived substrates<sup>41</sup>, flavonoid rhamnoglycosides (i.e. nagirin, rutin and hesperidin)<sup>42</sup> or  $\beta$ -1,3-xylan<sup>43</sup>.

***L. fermentum* NCC 2970 GtfB protein sequence analysis.** Annotation of the *L. fermentum* NCC 2970 genome thus resulted in identification of a single GH70 family protein. BLASTp analysis of this *L. fermentum* NCC 2970 GH70 protein revealed that the closest homologs of this protein all were members of the GtfB-like 4,6- $\alpha$ -GTase GH70 subfamily with more than 47% identical amino acid sequences, including 3 further *L. fermentum* strains, namely 39, ATCC 14931 and 28-3-CHN (See supplementary Table S2). The phylogenetic relationship of the *L. fermentum* NCC 2970 GH70 protein to other characterized GH70 enzymes and (putative) GH70 4,6- $\alpha$ -GTases identified by this BLASTp search is depicted in Fig. 1. Glucansucrases (GSs) are present in the genomes of lactic acid bacteria of the genera *Leuconostoc*, *Streptococcus*, *Lactobacillus*, *Oenococcus* and *Weissella*, but most genes encoding putative GtfB homologs are currently found in *Lactobacillus* strains (except 4 GtfB-like proteins present in *Pediococcus* and *Leuconostoc* strains). The *L. fermentum* NCC 2970 GH70 enzyme clustered together with the *L. reuteri* 121 GtfB 4,6- $\alpha$ -GTase and its homologs, indicating that this protein is a member of the GtfB-type of GH70 subfamily. The GtfB-like proteins display a domain organization resembling that of GH70 GSs with a circularly permuted ( $\beta/\alpha$ )<sub>8</sub> barrel<sup>24</sup>, however, biochemically they are more related to the *E. sibiricum* 255-15 GtfC and *A. chroococcum* NCIMB 8003 GtfD enzymes<sup>23,26</sup>. In agreement with these observations, the GtfB-like GH70 subfamily is evolutionary more closely related to GSs, but it has an intermediate position between the GSs and the GtfC and GtfD 4,6- $\alpha$ -GTases encoded by *E. sibiricum* 255-15 and *A. chroococcum* NCIMB 8003, respectively.

The newly identified GtfB gene sequence from *L. fermentum* NCC 2970 encodes a polypeptide consisting of 1593 amino acids with a calculated molecular mass of 180 kDa. In agreement with the extracellular location of GH70 enzymes, this *L. fermentum* GtfB was predicted to function as an extracellular protein by the CELLO program. Analysis of the N-terminus of the amino acid sequence of *L. fermentum* NCC 2970 GtfB revealed that this protein contains the classical Gram-positive N-terminal signal peptide of 39 amino acids. The *L. fermentum* NCC 2970 GtfB shares significant amino acid identity (77% identity, 57% coverage) with the *L. reuteri* 121 GtfB, the first characterized GH70 member with 4,6- $\alpha$ -glucanotransferase activity<sup>24</sup>. The *L. fermentum* 39, ATCC14931 and 28-3-CHN GtfB proteins are 99% identical to each other, but only 79%, 78% and 73% with the NCC 2970 GtfB protein.

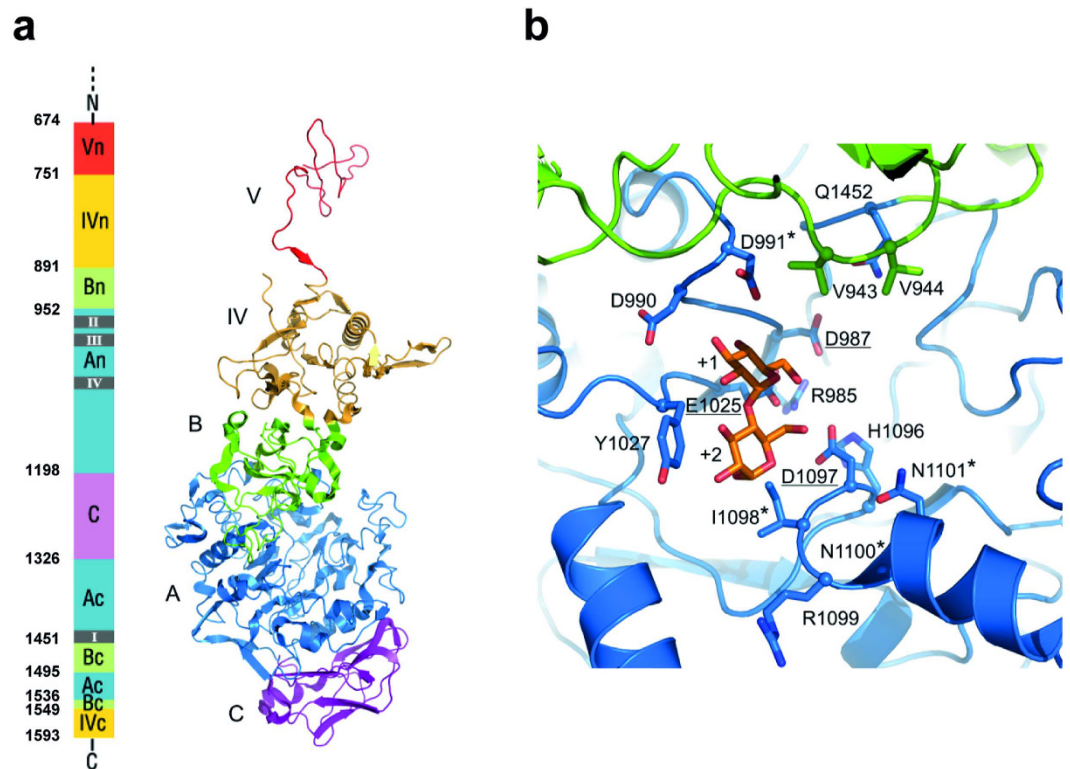
The three-dimensional (3D) model structure of the *L. fermentum* NCC 2970 GtfB protein was deduced by homology modeling using *L. reuteri* 180 Gtf180- $\Delta N$  GS as the template (Fig. 2). This homology model contained residues 674–1593, and revealed that the *L. fermentum* GtfB displays the typical U-fold domain organization characteristic of most GH70 proteins consisting of five domains (A, B, C, IV and V), with a circularly permuted catalytic ( $\beta/\alpha$ )<sub>8</sub> barrel (Fig. 2a). Similar to *L. reuteri* 121 GtfB<sup>27</sup>, domains A, B, C and IV of the *L. fermentum* NCC 2970 GtfB protein are built up from discontinuous N- and C-terminal stretches of the polypeptide chain. Notably, sequence alignment predicted that *L. fermentum* GtfB has no C-terminal domain V segment; while only part of domain V (residues 674–750) is present in the homology model obtained from one-to-one threading with *L. reuteri* 180 Gtf180- $\Delta N$ . It is likely that other N-terminal residues further complete this domain. Indeed, a homology model based on the DSR-E  $\Delta N_{123}$ -GBD-CD2 structure contained a complete domain V (residues 537–750) (not shown). Besides, the *L. fermentum* NCC 2970 GtfB has a variable N-terminal domain of ~500 residues present



**Figure 1.** Phylogenetic tree constructed on the basis of complete amino acid sequence alignment of some of the characterized GH70 proteins annotated in the CAZy database, and of (putative) GH70 GtfB-like proteins identified by a BLAST search using the *L. fermentum* NCC 2970 GH70 protein as query sequence (shown in bold). The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model. The bar represents a genetic distance of 0.2 substitutions per position (20% amino acid sequence difference). The bootstrap values adjacent to the main nodes represent the probabilities based on 1000 replicates. The protein sequences are annotated by their Genbank accession number and bacterial origin. The *L. reuteri* 121 GtfB 4,6- $\alpha$ -GTase is highlighted with a grey background.

in many glucansucrases and GtfB-like proteins of *Lactobacillus* species that is believed to be involved in cell wall attachment<sup>44</sup>.

Aiming to predict the reaction and/or product specificity of the *L. fermentum* NCC 2970 GtfB enzyme, the homology regions I-IV of this protein were identified by sequence alignments with other GH70 family proteins (Fig. 3) and further inspected in the Gtf180- $\Delta$ N-based homology model (Fig. 2b). These four homology motifs contain the catalytic and substrate-binding residues, and thus are regarded as sequence fingerprints for the individual enzyme specificities in both GH70 and GH13 family enzymes<sup>15,45</sup>. The order of the conserved regions I-IV in the *L. fermentum* NCC 2970 GtfB is II-III-IV-I and corresponds to the order found in GSs and GtfB-like 4,6- $\alpha$ -glucanotransferases, reflecting its circularly permuted domain organization. The seven residues strictly conserved in motifs I-IV of GH70 family enzymes<sup>45</sup> are present in the *L. fermentum* GtfB. Among these seven residues, the three putative catalytic residues Asp987, Glu1025 and Asp1097 (*L. fermentum* NCC 2970 GtfB numbering), were identified as the nucleophile, the general acid/base catalyst and the transition state stabilizer, respectively. The arrangement of the catalytic residues is identical to that in known GS structures, but other residues lining the active site show differences. For example, similar to other GtfB-like enzymes, a Tyr residue in motif III



**Figure 2. Homology model for *L. fermentum* NCC 2970 GtfB.** Tertiary structure prediction was accomplished by using the Phyre2 server<sup>37</sup>. Among the different templates found, the *L. reuteri* 180 Gtf180- $\Delta$ N (PDB: 3KLL<sup>20</sup>), was chosen as it displayed the highest homology with *L. fermentum* NCC 2970 GtfB in domains A and B. (a) Schematic representation of the *L. fermentum* GtfB 3D model structure. Domains A, B, C, IV and V are colored in blue, green, magenta, yellow and red, respectively. Only part of the N-terminal half of domain V is present in the homology model. (b) Close-up of the active site at the interface of domains A (blue) and B (green). The catalytic residues (underlined) and other residues lining the active site, present in motifs I-IV are shown in stick representation. Residues that are different from previously identified 4,6- $\alpha$ -GTases are indicated with an asterisk. A maltose acceptor substrate bound in subsites +1 and +2 (orange carbon atoms) is shown based on the Gtf180/maltose complex (PDB: 3KLL<sup>20</sup>).

replaces the Trp residue (W1065 in *L. reuteri* Gtf180 GS) that is conserved in all GSs, in which it provides an aromatic stacking interaction at subsites +1/+2 with acceptor substrates<sup>20,22</sup>; in branching sucrases the tryptophan is absent<sup>12,13</sup>. It is noteworthy that *E. sibiricum* GtfC and *A. chroococcum* GtfD 4,6- $\alpha$ -GTase enzymes also have a Tyr at this position, which therefore represents one of the main differences discriminating a GH70 enzyme active on starch/maltodextrin substrates from “classical” GSs. A second difference is Asp991 of *L. fermentum* NCC 2970 GtfB; it replaces the conserved Asn present in GSs and most GtfB-like 4,6- $\alpha$ -GTases. In the GS of *L. reuteri* 180 this residue was found to be essential for activity and linkage specificity<sup>31</sup>. The preceding Asp is conserved in *L. fermentum* GtfB (Asp 990). Finally, residues in motif IV were previously shown to be involved in determining linkage specificity; in particular the first and fourth residue downstream the transition state stabilizer<sup>20,29,46,47</sup>. Glucansucrases show variation in this region, while GtfB-like 4,6- $\alpha$ -GTases have an almost completely conserved QRKN motif (note that the alignment predicts a one-residue gap, Fig. 3). The *L. fermentum* NCC 2970 GtfB instead has IRNN, also different from the GtfC and GtfD enzymes. Thus, despite the fact that the *L. fermentum* NCC 2970 GtfB exhibits clear sequence similarity with the *L. reuteri* 121 GtfB 4,6- $\alpha$ -GTase protein, it contains unique sequence features. The homology regions I-IV of the GtfB proteins of *L. fermentum* strains 39, ATCC 14931 and 28-3-CHN are similar to that of the *L. reuteri* 121 GtfB (Fig. 3). Inspection of the *L. fermentum* NCC 2970 GtfB homology model and the likely binding mode of an acceptor substrate (maltose) confirmed that the residues mentioned above indeed cluster around the active site and substrate-binding regions (Fig. 2b), and that the residues in *L. fermentum* NCC 2970 GtfB that are different from previously identified 4,6- $\alpha$ -GTases likely are in functionally important positions for determining the substrate- and product specificity of the enzyme. Therefore, we investigated the *L. fermentum* NCC 2970 GtfB substrate and product specificity in detail.

**Purification and biochemical characterization of the *L. fermentum* NCC 2970 GtfB.** The *L. fermentum* NCC 2970 GtfB- $\Delta$ N protein was successfully expressed in *E. coli* BL21(DE3) Star. Under the growth and inductions conditions used, the expression level of GtfB- $\Delta$ N was relatively low in both the soluble and insoluble fractions (Supplementary Fig. S1). The enzyme was purified from the soluble fraction by two chromatographic steps consisting of metal chelate chromatography and anion-exchange chromatography, as described in the Experimental Section. SDS-PAGE analysis of the purified enzyme revealed a single protein band with an

Bacterial strain	NCBI accession numbers	Specificity	Motif II	Motif III	Motif IV	Motif I
<b>a</b>			<b>3 4</b>	<b>5</b>	<b>67</b>	<b>1 2</b>
<i>L. fermentum</i> NCC 2970	AOR73699	4,3- $\alpha$ -GT	983 <b>GFR</b> <b>IDA</b> <b>ADD</b> <b>M</b> <b>D</b>	1020 <b>HLSY</b> <b>NE</b> <b>GY</b> <b>G</b> <b>P</b> <b>G</b>	1092 <b>YVTN</b> <b>HD</b> <b>IR</b> <b>-</b> <b>NN</b> <b>L</b> <b>I</b> <b>N</b> <b>G</b>	1446 <b>ED</b> <b>IV</b> <b>M</b> <b>N</b> <b>Q</b>
<b>b</b>						
<i>Lactobacillus reuteri</i> 121 (GtfB)	AAU08014.2	4,6- $\alpha$ -GT	1011 <b>GFR</b> <b>VD</b> <b>AA</b> <b>DN</b> <b>ID</b>	1048 <b>HLSY</b> <b>NE</b> <b>GY</b> <b>H</b> <b>S</b> <b>G</b>	1120 <b>FVTN</b> <b>HD</b> <b>QR</b> <b>-</b> <b>KN</b> <b>L</b> <b>I</b> <b>N</b> <b>R</b>	1478 <b>ED</b> <b>IV</b> <b>M</b> <b>N</b> <b>Q</b>
<i>Lactobacillus reuteri</i> ML1 (ML4)	AAU08003.2	4,6- $\alpha$ -GT	1012 <b>GFR</b> <b>VD</b> <b>AA</b> <b>DN</b> <b>ID</b>	1049 <b>HLSY</b> <b>NE</b> <b>GY</b> <b>H</b> <b>S</b> <b>G</b>	1121 <b>FVTN</b> <b>HD</b> <b>QR</b> <b>-</b> <b>KN</b> <b>L</b> <b>I</b> <b>N</b> <b>R</b>	1479 <b>ED</b> <b>IV</b> <b>M</b> <b>N</b> <b>Q</b>
<i>Lactobacillus reuteri</i> DSM 20016 (GtfW)	ABQ83597.1	4,6- $\alpha$ -GT	748 <b>GFR</b> <b>VD</b> <b>AA</b> <b>DN</b> <b>ID</b>	785 <b>HLVY</b> <b>NE</b> <b>GY</b> <b>H</b> <b>S</b> <b>G</b>	858 <b>FVTN</b> <b>HD</b> <b>QR</b> <b>-</b> <b>KN</b> <b>V</b> <b>I</b> <b>N</b> <b>Q</b>	1215 <b>ED</b> <b>L</b> <b>V</b> <b>M</b> <b>N</b> <b>Q</b>
<i>Lactobacillus salivarius</i> GJ-24	EGM52218.1	ND	1026 <b>GFR</b> <b>IDA</b> <b>AD</b> <b>H</b> <b>I</b> <b>D</b>	1063 <b>HLSY</b> <b>NE</b> <b>GY</b> <b>R</b> <b>S</b> <b>G</b>	1134 <b>YVTN</b> <b>HD</b> <b>QR</b> <b>-</b> <b>AN</b> <b>L</b> <b>I</b> <b>N</b> <b>G</b>	1485 <b>V</b> <b>D</b> <b>I</b> <b>V</b> <b>M</b> <b>N</b> <b>Q</b>
<i>Pediococcus pentosaceus</i> IE-3	CCG90643.1	ND	380 <b>GFR</b> <b>IDA</b> <b>AD</b> <b>H</b> <b>I</b> <b>D</b>	417 <b>HLSY</b> <b>NE</b> <b>GY</b> <b>H</b> <b>S</b> <b>G</b>	489 <b>FVTN</b> <b>HD</b> <b>QR</b> <b>-</b> <b>KN</b> <b>L</b> <b>I</b> <b>N</b> <b>S</b>	841 <b>ED</b> <b>IV</b> <b>M</b> <b>N</b> <b>Q</b>
<i>Lactobacillus plantarum</i> 16	WP_016526729.1	ND	749 <b>GFR</b> <b>VD</b> <b>AA</b> <b>DN</b> <b>ID</b>	786 <b>HLVY</b> <b>NE</b> <b>GY</b> <b>N</b> <b>Y</b> <b>G</b>	859 <b>FVTN</b> <b>HD</b> <b>QR</b> <b>-</b> <b>NN</b> <b>L</b> <b>V</b> <b>N</b> <b>R</b>	1201 <b>ED</b> <b>L</b> <b>V</b> <b>M</b> <b>N</b> <b>Q</b>
<i>Lactobacillus fermentum</i> 39	KLD54475.1	ND	870 <b>GFR</b> <b>IDA</b> <b>AD</b> <b>H</b> <b>I</b> <b>D</b>	907 <b>HLIY</b> <b>NE</b> <b>GY</b> <b>H</b> <b>L</b> <b>G</b>	979 <b>FVTN</b> <b>HD</b> <b>QR</b> <b>-</b> <b>KN</b> <b>L</b> <b>I</b> <b>N</b> <b>Q</b>	1331 <b>ED</b> <b>L</b> <b>V</b> <b>M</b> <b>N</b> <b>Q</b>
<i>Lactobacillus fermentum</i> ATCC 14931	EEl21226.1	ND	406 <b>GFR</b> <b>IDA</b> <b>AD</b> <b>H</b> <b>I</b> <b>D</b>	443 <b>HLIY</b> <b>NE</b> <b>GY</b> <b>H</b> <b>S</b> <b>G</b>	512 <b>FVTN</b> <b>HD</b> <b>QR</b> <b>-</b> <b>KN</b> <b>L</b> <b>I</b> <b>N</b> <b>Q</b>	862 <b>ED</b> <b>L</b> <b>V</b> <b>M</b> <b>N</b> <b>Q</b>
<i>Lactobacillus fermentum</i> 28-3-CHN	EEZ25696.1	ND	378 <b>GFR</b> <b>IDA</b> <b>AD</b> <b>H</b> <b>I</b> <b>D</b>	415 <b>HLIY</b> <b>NE</b> <b>GY</b> <b>H</b> <b>L</b> <b>G</b>	487 <b>FVTN</b> <b>HD</b> <b>QR</b> <b>-</b> <b>KN</b> <b>L</b> <b>I</b> <b>N</b> <b>Q</b>	839 <b>ED</b> <b>L</b> <b>V</b> <b>M</b> <b>N</b> <b>Q</b>
<i>Lactobacillus acidipiscis</i> KCTC 13900	WP_035631372.1	ND	296 <b>GFR</b> <b>ND</b> <b>AA</b> <b>DN</b> <b>ID</b>	333 <b>HLVY</b> <b>NE</b> <b>GY</b> <b>H</b> <b>S</b> <b>G</b>	406 <b>FVTN</b> <b>HD</b> <b>QR</b> <b>-</b> <b>KN</b> <b>V</b> <b>I</b> <b>N</b> <b>Q</b>	765 <b>V</b> <b>D</b> <b>M</b> <b>V</b> <b>M</b> <b>N</b> <b>Q</b>
<i>Lactobacillus delbrueckii</i> JCM 17838	WP_050952694.1	ND	265 <b>GFR</b> <b>IDA</b> <b>AD</b> <b>H</b> <b>I</b> <b>D</b>	302 <b>HLSY</b> <b>NE</b> <b>GY</b> <b>H</b> <b>L</b> <b>G</b>	374 <b>FVTN</b> <b>HD</b> <b>QR</b> <b>-</b> <b>KN</b> <b>L</b> <b>I</b> <b>N</b> <b>R</b>	732 <b>ED</b> <b>IV</b> <b>M</b> <b>N</b> <b>Q</b>
<i>Lactobacillus panis</i> DSM 6035	KRM25865.1	ND	988 <b>GFR</b> <b>VD</b> <b>AA</b> <b>DN</b> <b>V</b> <b>D</b>	1025 <b>HLVY</b> <b>NE</b> <b>GY</b> <b>H</b> <b>S</b> <b>D</b>	1097 <b>FVTN</b> <b>HD</b> <b>QR</b> <b>-</b> <b>KN</b> <b>L</b> <b>I</b> <b>N</b> <b>Q</b>	1455 <b>ED</b> <b>L</b> <b>V</b> <b>M</b> <b>N</b> <b>Q</b>
<i>Leuconostoc mesenteroides</i>	WP_059442690.1	ND	252 <b>GFR</b> <b>IDA</b> <b>AD</b> <b>H</b> <b>I</b> <b>D</b>	289 <b>HLIY</b> <b>NE</b> <b>GY</b> <b>R</b> <b>S</b> <b>G</b>	360 <b>FVTN</b> <b>HD</b> <b>QR</b> <b>-</b> <b>AN</b> <b>L</b> <b>I</b> <b>N</b> <b>G</b>	711 <b>ED</b> <b>IV</b> <b>M</b> <b>N</b> <b>Q</b>
<b>c</b>						
<i>Exiguobacterium sibiricum</i> 255-15	ACB62096.1	4,6- $\alpha$ -GT	403 <b>GFR</b> <b>IDA</b> <b>AS</b> <b>H</b> <b>Y</b> <b>D</b>	433 <b>HLSY</b> <b>IE</b> <b>SY</b> <b>K</b> <b>S</b> <b>E</b>	504 <b>FVNN</b> <b>HD</b> <b>QE</b> <b>-</b> <b>KN</b> <b>R</b> <b>V</b> <b>N</b> <b>Q</b>	138 <b>M</b> <b>D</b> <b>L</b> <b>V</b> <b>P</b> <b>N</b> <b>Q</b>
<b>d</b>						
<i>Azotobacter chroococcum</i> NCIMB 8003	AJE22990.1	4,6- $\alpha$ -GT	467 <b>GFR</b> <b>IDA</b> <b>AS</b> <b>H</b> <b>I</b> <b>N</b>	500 <b>HLSY</b> <b>IE</b> <b>SY</b> <b>V</b> <b>T</b> <b>Q</b>	567 <b>FVNN</b> <b>HD</b> <b>QE</b> <b>-</b> <b>HN</b> <b>I</b> <b>L</b> <b>V</b> <b>T</b>	202 <b>V</b> <b>D</b> <b>V</b> <b>V</b> <b>P</b> <b>N</b> <b>Q</b>
<b>e</b>						
<i>Lactobacillus reuteri</i> 180 (Gtf180)	AAU08001.1	Dextranucrase	1021 <b>GIR</b> <b>VD</b> <b>AV</b> <b>DN</b> <b>V</b> <b>D</b>	1058 <b>HIN</b> <b>I</b> <b>L</b> <b>E</b> <b>D</b> <b>W</b> <b>G</b> <b>W</b> <b>D</b>	1131 <b>FVRA</b> <b>HD</b> <b>S</b> <b>NA</b> <b>Q</b> <b>D</b> <b>Q</b> <b>I</b> <b>R</b> <b>Q</b>	1503 <b>AD</b> <b>W</b> <b>V</b> <b>P</b> <b>D</b> <b>Q</b>
<i>Lactobacillus reuteri</i> 121 (GtfA)	AAU08015.1	Reuteransucrase	1020 <b>SVR</b> <b>VD</b> <b>AP</b> <b>DN</b> <b>I</b> <b>D</b>	1056 <b>HIN</b> <b>I</b> <b>L</b> <b>E</b> <b>D</b> <b>W</b> <b>N</b> <b>H</b> <b>A</b>	1128 <b>FVRA</b> <b>HD</b> <b>S</b> <b>NS</b> <b>Q</b> <b>D</b> <b>Q</b> <b>I</b> <b>Q</b> <b>N</b>	1508 <b>AD</b> <b>W</b> <b>V</b> <b>P</b> <b>D</b> <b>Q</b>
<i>Streptococcus mutans</i> SI (GtfSI)	BAA26114.1	Mutansucrase	473 <b>SIR</b> <b>VD</b> <b>AV</b> <b>DN</b> <b>V</b> <b>D</b>	510 <b>HLS</b> <b>I</b> <b>L</b> <b>E</b> <b>A</b> <b>W</b> <b>S</b> <b>Y</b> <b>N</b>	583 <b>FIRA</b> <b>HD</b> <b>S</b> <b>EV</b> <b>Q</b> <b>D</b> <b>L</b> <b>I</b> <b>R</b> <b>D</b>	954 <b>AD</b> <b>W</b> <b>V</b> <b>P</b> <b>D</b> <b>Q</b>
<i>Leuconostoc mesenteroides</i> NRRL-1355 CAB65910.2		Alternansucrase	631 <b>GIR</b> <b>VD</b> <b>AV</b> <b>DN</b> <b>V</b> <b>D</b>	668 <b>HLS</b> <b>I</b> <b>L</b> <b>E</b> <b>D</b> <b>W</b> <b>N</b> <b>G</b> <b>K</b>	762 <b>FVRA</b> <b>HD</b> <b>Y</b> <b>D</b> <b>A</b> <b>Q</b> <b>D</b> <b>P</b> <b>I</b> <b>R</b> <b>K</b>	1168 <b>AD</b> <b>W</b> <b>V</b> <b>P</b> <b>D</b> <b>Q</b>
<i>Leuconostoc citreum</i> NRRL B-1299	CDX66820.1	(1 $\rightarrow$ 2) Branching sucrose	2206 <b>SIR</b> <b>IDA</b> <b>VD</b> <b>FI</b> <b>H</b>	2243 <b>HIS</b> <b>L</b> <b>V</b> <b>E</b> <b>A</b> <b>G</b> <b>L</b> <b>D</b> <b>A</b>	2317 <b>I</b> <b>I</b> <b>H</b> <b>A</b> <b>H</b> <b>D</b> <b>K</b> <b>G</b> <b>V</b> <b>Q</b> <b>E</b> <b>K</b> <b>V</b> <b>G</b> <b>A</b>	2688 <b>AD</b> <b>V</b> <b>V</b> <b>D</b> <b>N</b> <b>Q</b>
<i>Leuconostoc citreum</i> NRRL B-742	CDX65123.1	(1 $\rightarrow$ 3) Branching sucrose	667 <b>SMR</b> <b>IDA</b> <b>I</b> <b>S</b> <b>F</b> <b>V</b> <b>D</b>	704 <b>HIS</b> <b>I</b> <b>V</b> <b>E</b> <b>A</b> <b>P</b> <b>K</b> <b>G</b> <b>E</b>	783 <b>I</b> <b>V</b> <b>H</b> <b>A</b> <b>H</b> <b>D</b> <b>K</b> <b>I</b> <b>D</b> <b>Q</b> <b>T</b> <b>V</b> <b>I</b> <b>H</b>	1182 <b>AD</b> <b>F</b> <b>V</b> <b>A</b> <b>N</b> <b>Q</b>

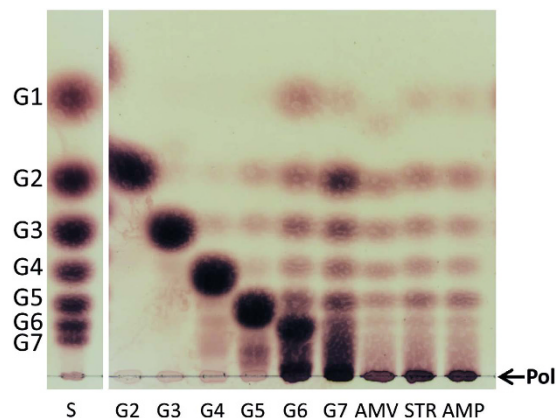
**Figure 3.** Sequence alignment of conserved motifs I-IV in the catalytic domains of the *L. fermentum* NCC 2970 4,3- $\alpha$ -GTase GtfB enzyme (a), (putative) GtfB-like 4,6- $\alpha$ -GTase enzymes (b), *E. sibiricum* GtfC 4,6- $\alpha$ -GTase enzyme (c), *A. chroococcum* GtfD 4,6- $\alpha$ -GTase enzyme (d), and glucansucrase enzymes (e). The seven strictly conserved amino acid residues in GH70 enzymes (indicated by the numbers 1 to 7 above the sequences) are also conserved in the novel *L. fermentum* 4,3- $\alpha$ -GT GtfB protein. Amino acids that constitute the catalytic triad are shown in bold and lightly shaded. Residues forming acceptor substrate binding subsites -1, +1 and +2 in glucansucrase Gtf180- $\Delta$ N<sup>20</sup> are highlighted in green, red and blue, respectively. Symbols: NU = nucleophile, A/B = general acid/base, TS = transition state stabilizer.

apparent molecular mass of ~110 kDa (Supplementary Fig. S1) which fits the theoretical value deduced from the sequence. As a result of this purification process, a total of 0.4 mg of pure GtfB- $\Delta$ N protein per L of *E. coli* culture was obtained. A domain V truncated variant (with amino acids 751 to 1593) was constructed with the aim of increasing protein expression. This truncation approach did not significantly improve the expression level of the *L. fermentum* NCC 2970 GtfB protein (Data not shown).

The effects of pH and temperature on the enzyme activity were determined by the iodine-staining assay using amylose V as substrate. The purified recombinant *L. fermentum* NCC 2970 GtfB exhibited the highest activity at pH 5.5 and retained more than 80% of this activity over a pH from 4 to 6 (Supplementary Fig. S2a). The enzyme was active between 30 and 60 °C, showing its maximal activity at 50 °C, but the activity decreased drastically when the reaction was performed at 65 °C (Supplementary Fig. S2b). In addition, the enzyme showed low stability at temperatures above 45 °C in 20 mM Tris-HCl buffer (pH 8.0) (Supplementary Fig. S2c). Similar pH and temperature optimum values were reported for the *L. reuteri* 121 GtfB 4,6- $\alpha$ -GTase enzyme<sup>27</sup>. The specific total activity of the purified *L. fermentum* NCC 2970 GtfB enzyme on 0.125% (w v<sup>-1</sup>) amylose V in sodium acetate buffer (pH 5.5), containing 1 mM CaCl<sub>2</sub> was determined using an iodine-staining assay to be 22 ± 0.36 U mg<sup>-1</sup> of protein. This value is around 10 times higher than the one reported for the GtfB from *L. reuteri* 121 (under its optimal conditions of pH 5.0 and 40 °C), namely, 2.8 U mg<sup>-1</sup><sup>27</sup>. The activity of the *L. fermentum* NCC 2970 GtfB was also assayed in the presence of the chelating agent EDTA at a final concentration of 10 mM, resulting in a slight (20%) inhibition.

**Substrate and product specificity of the *L. fermentum* NCC 2970 GtfB.** The differences observed in the homology motifs suggested that the *L. fermentum* NCC 2970 GtfB may possess a new enzymatic reaction and/or product specificity. Therefore the *L. fermentum* GtfB enzyme was incubated with different oligosaccharides and polysaccharides at 37 °C for 24 h, and the reaction products were analysed by TLC (Fig. 4). Similar to *L. reuteri* GtfB, the *L. fermentum* NCC 2970 GtfB enzyme failed to act on sucrose, panose, nigerose, and isomalto-oligosaccharides with DP2, DP3, and DP5 (Data not shown). Instead, the *L. fermentum* GtfB enzyme showed both hydrolysis and transglycosylase (disproportionation) activity on malto-oligosaccharides (MOS) of DP 6 and 7, evident from the accumulation of products with lower- and higher-molecular-mass than the MOS substrates. *L. fermentum* GtfB was not active on MOS of DP below 4 and showed low disproportionating activity with maltopentaose (DP5); for *L. reuteri* 121 GtfB activity was already observed with maltotriose (DP3)<sup>23</sup>.





**Figure 4.** TLC analysis of the product mixtures synthesized by the *L. fermentum* NCC 2970 GtfB enzyme in incubations with malto-oligosaccharides (DP2–DP7), amylose V, starch, and amylopectin. The reaction mixtures containing 25 mM malto-oligosaccharides or 0.6% (w/v<sup>-1</sup>) polysaccharides and 25 μg ml<sup>-1</sup> of enzyme were incubated at 37 °C and pH 5.5 during 24 h. S, standard; G1, glucose; G2–G7, maltose to maltoheptaose; AMV, amylose V; STR, starch; AMP, amylopectin; Pol, polymer.

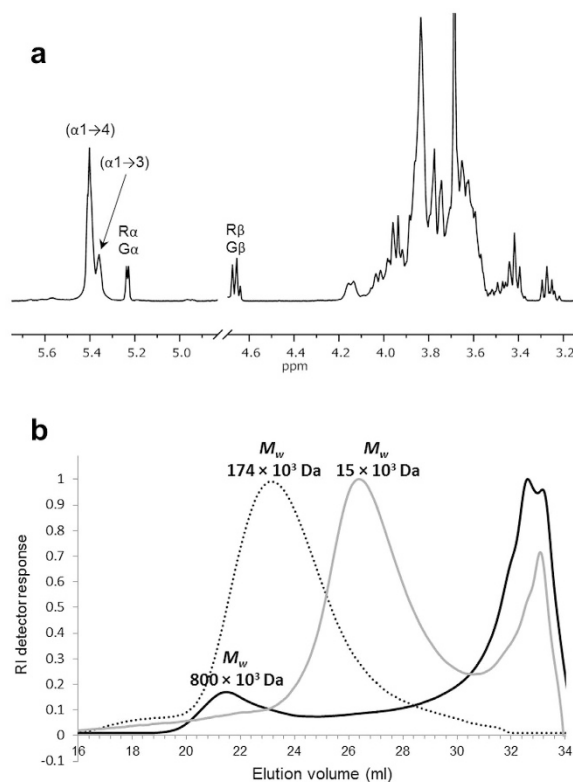
Incubation of amylose V, potato starch and amylopectin with the *L. fermentum* NCC 2970 GtfB enzyme resulted in the appearance of a range of lower molecular mass products indicating that the enzyme was also active on these polymers. It should be noted that the *L. fermentum* GtfB enzyme produced significantly larger amounts of oligosaccharide products than the *L. reuteri* 121 GtfB, which mainly synthesized a (modified) polymer from amylose V<sup>23</sup>. As observed by TLC, MOS of DP 2 to 5 accumulated from the various polymeric substrates, reflecting that *L. fermentum* GtfB requires maltohexaose or longer MOS as glucose donor substrates.

Aiming to study the product specificity of the *L. fermentum* NCC 2970 GtfB, the maltoheptaose and amylose V derived product mixtures were analyzed by one-dimensional <sup>1</sup>H NMR spectroscopy. As an example, the <sup>1</sup>H-NMR spectrum of the product mixture generated from amylose V is depicted in Fig 5a. <sup>1</sup>H-NMR analysis of both product mixtures revealed the presence of two groups of anomeric overlapping signals at δ ~5.40 and 5.35 corresponding to (α1 → 4) linkages and newly synthesized (α1 → 3) linkages. The presence of (α1 → 3) linkages was corroborated by the structural-reporter-group signal for (α1 → 3) linkages at δ<sub>H-5</sub> 4.16 ppm<sup>46</sup>. The spectra also showed signals corresponding to free glucose units (Gα H-1, δ 5.225; Gβ H-1, δ 4.637) and 4-substituted reducing-end glucose residues (Rα H-1, δ 5.225; Rβ H-1, δ 4.652). Small signals corresponding to trace amounts (less than 1%) of (α1 → 6) linkages (H-1, δ~4.97) were detected as well. The molar ratios of the (α1 → 4)-linked, (α1 → 3)-linked glucose residues for maltoheptaose and amylose V products were 86:14 and 81:19, respectively. Methylation analysis of the product mixture generated by the *L. fermentum* NCC 2970 GtfB from amylose V showed the presence of terminal, 3-substituted, 4-substituted, and 3,4-disubstituted glucopyranose residues in a molar percentage of 25, 16, 55, and 4%, which is in agreement with the linkage ratios determined by <sup>1</sup>H NMR and confirms that the *L. fermentum* GtfB exhibits (α1 → 3) linkage specificity. In contrast to the *L. reuteri* GtfB enzyme that only generates linear (α1 → 6) glucan chains (4,6-α-GTase), the *L. fermentum* NCC 2970 GtfB acts as a 4,3-α-glucanotransferase (4,3-α-GTase) catalyzing the cleavage of (α1 → 4) linkages and the formation of new (α1 → 3) in linear or branched orientations. The high relative amount of terminal glucopyranose residues (25%) detected in the reaction product mixture correlates with the presence of glucose and small MOS obtained as side products of the *L. fermentum* GtfB hydrolase/transglycosylase activity.

The action of the *L. fermentum* NCC 2970 GtfB enzyme on amylose V was also characterized by HPSEC with multidetection (Fig. 5b). The HPSEC profile of the starting amylose V substrate exhibited a single peak eluting at 23.0 ml, with an average  $M_w$  of  $174 \times 10^3$  Da, whereas *L. fermentum* GtfB-treated amylose V showed two different molecular mass distribution peaks: An early peak eluting at 21.5 ml corresponding to a high molecular mass polymer with an average  $M_w$  of  $800 \times 10^3$  Da and, a second peak corresponding to oligosaccharides with an average  $M_w$  of 1400 Da. In case of the *L. reuteri* 121 GtfB 4,6 α-GT, the peak corresponding to the IMMP eluted at 26.5 ml and had an average  $M_w$  of  $15 \times 10^3$  Da. Thus, the *L. fermentum* GtfB is capable of synthesizing a polymer with a  $M_w$  value about 4 and 50 times higher than those of the starting amylose V substrate and the IMMP product, respectively. However, based on the refractive index signal this polymer represents less than 20% of the total product mixture of the *L. fermentum* GtfB, while a significant proportion of low molecular mass glucans are present.

#### Structural characterization of the products synthesized by the *L. fermentum* NCC 2970 GtfB from amylose V.

For a more detailed structural characterization, the products produced by the *L. fermentum* NCC 2970 GtfB from amylose V were purified by size-exclusion chromatography on Biogel P2. The seven fractions obtained, designated as F1–F7, were analysed separately by MALDI-TOF MS and <sup>1</sup>H NMR spectroscopy (Table 1). Fractions F6 and F7 contained the hydrolysis products maltotriose (DP3) and maltose (DP2), respectively, and were not further studied. The 1D <sup>1</sup>H NMR spectrum (Fig. 5a) showed α-anomeric signals at δ 5.41 and 5.37; this region may contain anomeric signals of both (α1 → 4) and (α1 → 3) linked α-D-Glcp residues<sup>48</sup>. Therefore, fractions F1–F5 were also subjected to methylation analysis in order to determine the presence of



**Figure 5. Structural analysis of the products synthesized by the incubation of 0.6% ( $w v^{-1}$ ) amylose V with  $25 \mu g ml^{-1}$  of *L. fermentum* NCC 2970 GtfB for 24 h at 37 °C and pH 5.5. (a)  $^1H$  NMR spectra ( $D_2O$ , 300 K) of the product mixtures formed. The anomeric signals indicated as  $G\alpha/\beta$  and  $R\alpha/\beta$  correspond to free glucose and reducing  $-(1 \rightarrow 4)$ -D-Glcp units, respectively. Chemical shifts are given in parts per million relative to the signal of internal acetone ( $\delta$  2.225). (b) HPSEC molecular mass distribution of the reaction products generated from amylose V by the *L. fermentum* NCC 2970 GtfB enzyme. The elution profiles of the starting amylose V substrate (dashed line) and of the products synthesized by *L. reuteri* 121 GtfB from amylose V (grey line) are also included.**

( $\alpha 1 \rightarrow 4$ ) and ( $\alpha 1 \rightarrow 3$ ) linkages and to determine the degree of branching. The methylation analysis data fit with the peak at  $\delta$  5.37 corresponding to ( $\alpha 1 \rightarrow 3$ )-linked residues (Table 1). The  $^1H$  NMR and methylation analysis of the Biogel P2 fractions demonstrated that the percentage of ( $\alpha 1 \rightarrow 3$ ) linkages increased with increasing DP of the products (Table 1). The highest molecular mass product **F1** containing polymeric material of DP >30 showed an increased percentage of  $\rightarrow 3$ -GlcP-(1  $\rightarrow$  glucosyl units (28%) and branching  $\rightarrow 3,4$ -GlcP-(1  $\rightarrow$  glucosyl units (8%) over those in the total reaction product mixture, 16 and 4%, respectively.

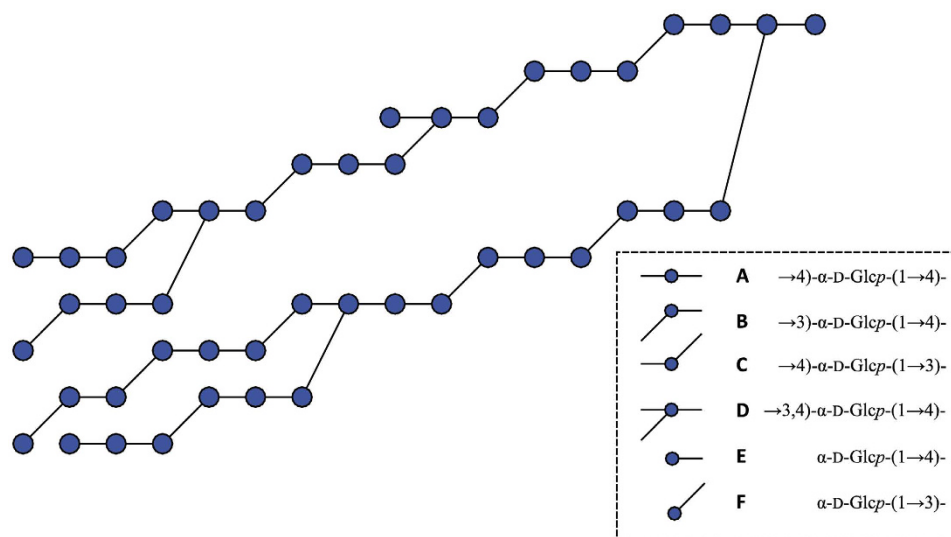
To unravel the structural elements present in fraction **F1**, 2D NMR ( $^{13}C$ - $^1H$  HSQC,  $^1H$ - $^1H$  TOCSY, and  $^1H$ - $^1H$  ROESY) experiments were performed (Supplementary Fig. S3). For a detailed description of the structural characterization of fraction **F1**, see supplementary information. Most notably, the 2D NMR data fit with the occurrence of alternating ( $\alpha 1 \rightarrow 3$ )/( $\alpha 1 \rightarrow 4$ ) linkages and ( $\alpha 1 \rightarrow 3,4$ ) branching points in the *L. fermentum* NCC 2970 GtfB polymer, whereas signals corresponding to sequential ( $\alpha 1 \rightarrow 3$ )-linkages were not identified. The absence of consecutive ( $\alpha 1 \rightarrow 3$ )-linkages in fraction **F1** was further confirmed by Smith degradation analysis (Supplementary Fig. S4). Taking together all data from 1D and 2D NMR spectroscopy analysis, methylation analysis and Smith degradation analysis, a composite model for **F1** was constructed, showing all identified structural elements in the correct relative abundance (Fig. 6).

#### Oligosaccharides formed in time from maltoheptaose and amylose V by the *L. fermentum* NCC 2970 GtfB enzyme.

To gain a better understanding of the action pattern of the *L. fermentum* NCC 2970 GtfB enzyme, time course experiments were performed with the maltoheptaose and amylose V substrates. The oligosaccharide products formed after 10 min, 1 h and 24 h of reaction were analyzed by HPAEC. At the early stage of the reaction with maltoheptaose (slightly contaminated with G5 and G6), G2 was identified as the main reaction product (Fig. 7). Small peaks corresponding to glucose (G1) and maltotriose (G3) were also identified, whereas a peak with an unknown structure but with a high DP eluted at 58 min. The release of G2, together with the synthesis of a compound with a higher DP, suggests that the *L. fermentum* GtfB enzyme catalyzes the transfer of a maltopentaosyl unit to a MOS acceptor substrate. Later in time, the amounts of glucose and MOS with DP2 to 5 increased, whereas after 24 h G6 and G7 had become depleted. In addition, as a result of the *L. fermentum* GtfB disproportionating activity, peaks that did not fit the MOS retention times and probably corresponding to oligosaccharides containing ( $\alpha 1 \rightarrow 3$ ) linkages, were detected at shorter and longer retention times than that of the

Sample	DP	Chemical shift (%) <sup>a</sup>			Methylation analysis (%) <sup>b</sup>		
		( $\alpha 1 \rightarrow 4$ )	( $\alpha 1 \rightarrow 3$ )	GlcP (1 $\rightarrow$ )	$\rightarrow 3$ -GlcP-(1 $\rightarrow$ )	$\rightarrow 4$ -GlcP-(1 $\rightarrow$ )	$\rightarrow 3,4$ -GlcP-(1 $\rightarrow$ )
F1	>30	60	40	6	28	58	8
F2	~8–20	71	29	12	21	61	6
F3	~6–8	80	20	14	15	66	5
F4	~5–6	88	12	21	10	67	2
F5	~4–5	89	11	21	10	64	5
F6	~3	100	0	ND	ND	ND	ND
F7	~2	100	0	ND	ND	ND	ND

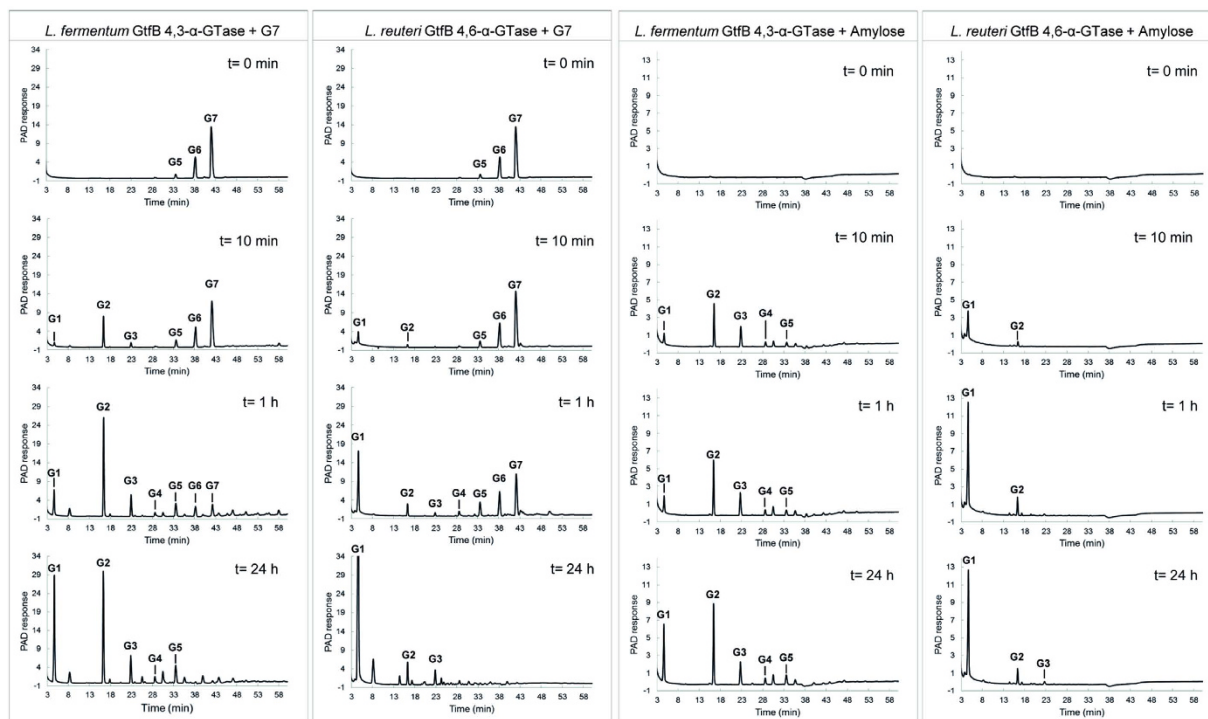
**Table 1. Linkage composition of the fractions (F1–F7) yielded by size-exclusion chromatography on Biogel P2 of the product mixture obtained from the incubation of 0.6% (w v<sup>-1</sup>) amylose V with 25  $\mu\text{g ml}^{-1}$  of *L. fermentum* NCC 2970 GtfB.** <sup>a</sup>The data represent the ratios of integration of the peak areas of the ( $\alpha 1 \rightarrow 4$ ) linkage signal at 5.41 ppm and the ( $\alpha 1 \rightarrow 3$ ) linkage signal at 5.37 ppm in the <sup>1</sup>H NMR spectra. <sup>b</sup>The linkage distribution data are shown in molar percentages based on GLC intensities.



**Figure 6. Composite model of the *L. fermentum* NCC 2970 GtfB polysaccharide product.** The composite model includes all structural elements identified by 1D and 2D NMR spectroscopy analysis, methylation analysis and Smith degradation analysis. Residue labels (A–F) presented on the right correspond with those in Table S3.

G7 starting substrate. Incubation with the amylose V yielded G2 as the first clear reaction product, together with G1, G3, G4, G5 (Fig. 7). These profiles significantly differed from the ones obtained with the *L. reuteri* 121 GtfB enzyme when both G7 and amylose V are used as substrates (Fig. 7). With both substrates the *L. reuteri* 121 GtfB releases glucose (instead of G2) as the main first product at the beginning of the reaction. The pronounced accumulation of MOS with DP2 to DP5 by the *L. fermentum* NCC 2970 GtfB was not seen for the *L. reuteri* 121 GtfB. These results suggest that the *L. fermentum* and *L. reuteri* 121 GtfB enzymes differ in their mechanism of polymerization. The novel *L. fermentum* GtfB enzyme preferentially catalyzes the transfer of MOS of low DP, showing a mode of action more resembling that of the *A. chroococcum* GtfD enzyme<sup>26</sup>. On the other hand, the *L. fermentum* GtfB 4,3- $\alpha$ -GT seems to require MOS of a certain minimum length (at least DP6) to act on and to produce an  $\alpha$ -glucan with an ( $\alpha 1 \rightarrow 3$ )/( $\alpha 1 \rightarrow 4$ ) alternating structure and with ( $\alpha 1 \rightarrow 3$ ) branching points. Finally, the data indicate that the *L. fermentum* NCC 2970 GtfB 4,3- $\alpha$ -GT active site may present more than one donor substrate binding subsite, as observed for enzymes belonging to the evolutionary related GH13 and GH77 families<sup>49,50</sup>.

**Enzymatic treatment of the *L. fermentum* NCC 2970 GtfB product.** To further characterize the polymer synthesized by the *L. fermentum* NCC 2970 GtfB enzyme (Biogel P2 fraction F1), this  $\alpha$ -glucan was treated with a high activity dose of  $\alpha$ -amylase, dextranase and pullulanase enzymes. The polymer produced by *L. fermentum* GtfB was resistant to the endo-( $\alpha 1 \rightarrow 4$ )-hydrolase activity of  $\alpha$ -amylase. As revealed by TLC analysis, only trace amounts of glucose, maltose and high molecular mass oligosaccharides were detected after 48 h of  $\alpha$ -amylase digestion (Fig. 8). Under the same reaction conditions, the starch control substrate was completely hydrolyzed, indicating that the presence of  $\alpha$ -(1  $\rightarrow$  3) linkages makes the *L. fermentum* GtfB polymer resistant to  $\alpha$ -amylase digestion. The *L. fermentum* GtfB polymer was also subjected to dextranase and pullulanase enzymatic



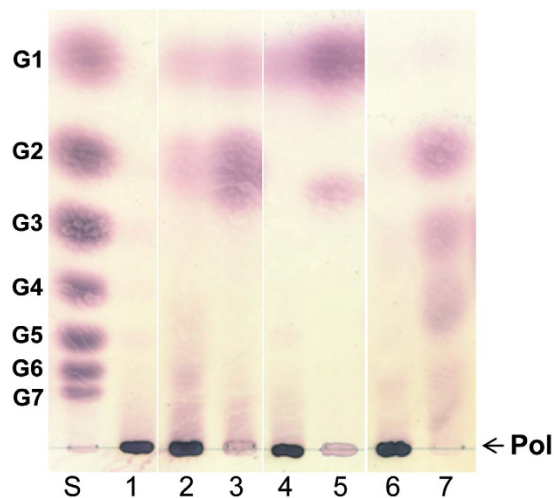
**Figure 7.** HPAEC-PAD profile of the oligosaccharide mixtures formed upon incubation of maltoheptaose and amylose V with  $25 \mu\text{g ml}^{-1}$  of *L. fermentum* NCC 2970 GtfB and *L. reuteri* 121 GtfB for  $t = 10$  min, 1 h, and 24 h (pH 5.5,  $37^\circ\text{C}$ ). The identity of peaks was assigned using commercial oligosaccharide standards. G1, glucose; G2–G7, maltose to maltoheptaose.

treatments. Dextranase catalyzes the endohydrolysis of consecutive (1 → 6) linkages in dextran, while pullulanase cleaves the (1 → 6) linkages present in pullulan, amylopectin and  $\alpha$ - and  $\beta$ -limit dextrins of starch. Thus, for the dextranase and the pullulanase treatments, the polymers produced from amylose V by the *L. reuteri* GtfB and the *A. chroococcum* GtfD were included as positive controls, respectively. As expected, dextranase efficiently hydrolyzed the IMMP produced by the *L. reuteri* 121 GtfB, whereas pullulanase completely digested the reuteran-like polymer synthesized by the *A. chroococcum* GtfD enzyme. In contrast, no hydrolysis was observed in the case of the *L. fermentum* NCC 2970 GtfB polymer, which is in agreement with the absence of (1 → 6) linkages (less than 1%) deduced from the  $^1\text{H}$  NMR analysis.

## Conclusions

This paper describes the identification and biochemical characterization of the first GH70 enzyme cleaving ( $\alpha$ 1 → 4)-linkages and synthesizing ( $\alpha$ 1 → 3)-linkages, encoded by *L. fermentum* NCC 2970. We propose to name this enzyme ( $\alpha$ 1 → 4)- $\alpha$ -D-glucan: ( $\alpha$ 1 → 4), ( $\alpha$ 1 → 3)- $\alpha$ -D-glucan  $\alpha$ -glucanotransferase, in short 4,3- $\alpha$ -glucanotransferase. Regarding its primary sequence, this protein clearly belongs to the novel GtfB-like GH70 subfamily which originally only comprised 4,6- $\alpha$ -glucanotransferases, however, it also possesses unique variations in residues forming the donor and acceptor binding subsites of GSs in particular in signature motifs II and IV. This suggests that the *L. fermentum* NCC 2970 GtfB active site displays distinctive features that may lead to its different specificity compared to other GtfB-like enzymes. Indeed, the *L. fermentum* NCC 2970 GtfB resembles previously characterized GtfB enzymes in using maltodextrins and starch as substrates, but instead of catalyzing the synthesis of consecutive ( $\alpha$ 1 → 6) linkages, it displays ( $\alpha$ 1 → 3) linkage specificity. The *L. fermentum* GtfB activity results in the synthesis of a unique  $\alpha$ -glucan built up with different lengths of malto-oligosaccharides, interconnected by single ( $\alpha$ 1 → 3) linkages in linear and branched orientations.

Even though GSs have a wide product spectrum and synthesize various types of linkages, none of the GSs characterized so far produce an  $\alpha$ -glucan consisting of ( $\alpha$ 1 → 4)- and ( $\alpha$ 1 → 3)-linkages. As far as we are aware, the structure of the *L. fermentum* NCC 2970 GtfB product also differs from that corresponding to  $\alpha$ -glucans containing both ( $\alpha$ 1 → 4)- and ( $\alpha$ 1 → 3)-linkages produced by different lichens and fungi<sup>51,52</sup>. Indeed, most of these polysaccharides have linear structures<sup>53–57</sup>, whereas others are mainly composed of ( $\alpha$ 1 → 3)-linkages and/or do not present ( $\alpha$ 1 → 3,4)-branching points<sup>58,59</sup>. Owing to the novel structure of the polymer produced by *L. fermentum* NCC 2970 GtfB, it will be of interest to study the physico-chemical and functional properties of this  $\alpha$ -glucan and to explore its applications. For example, the direct action of the *L. fermentum* GtfB on the starch present in food matrices is expected to result in the synthesis of slowly digestible starch derivatives. Overall, this study expands the diversity of the GH70  $\alpha$ -glucans that can be produced, and provides new insights in the understanding of the features that determine the reaction/product specificity of GH70 family proteins.



**Figure 8.** TLC analysis of the *L. fermentum* NCC 2970 GtfB polymer after treatment with *Aspergillus oryzae*  $\alpha$ -amylase, *Chaetomium erraticum* dextranase and *Klebsiella planticola* pullulanase M1. Lane 1, *L. fermentum* GtfB polymer before the enzymatic treatments; Lanes 2–3, product mixtures generated by the  $\alpha$ -amylase enzymatic treatment of the *L. fermentum* GtfB polymer and starch, respectively. Lanes 4–5, product mixtures generated by the dextranase enzymatic treatment of the *L. fermentum* GtfB polymer and IMMP, respectively. Lanes 6–7, product mixtures generated by the pullulanase enzymatic treatment of the *L. fermentum* GtfB polymer and the *A. chroococcum* GtfD polymer, respectively. Lane S, standard; G1, glucose; G2–G7, maltose to maltoheptaose; Pol, polymer.

## References

- Monchois, V., Willemot, R. M. & Monsan, P. Glucansucrases: mechanism of action and structure function relationships. *FEMS Microbiol. Rev.* **23**, 131–151 (1999).
- Leemhuis, H. *et al.* Glucansucrases: three-dimensional structures, reactions, mechanism, alpha-glucan analysis and their implications in biotechnology and food applications. *J. Biotechnol.* **163**, 250–272 (2013).
- Shiroza, T., Ueda, S. & Kuramitsu, H. K. Sequence analysis of the gtfB gene from *Streptococcus mutans*. *J. Bacteriol.* **169**, 4263–4270 (1987).
- Ueda, S., Shiroza, T. & Kuramitsu, H. K. Sequence analysis of the gtfC gene from *Streptococcus mutans* GS-5. *Gene* **69**, 101–109 (1988).
- Monchois, V., Remaud-Simeon, M., Russell, R. R., Monsan, P. & Willemot, R. M. Characterization of *Leuconostoc mesenteroides* NRRL B-512F dextranase (DSRS) and identification of amino-acid residues playing a key role in enzyme activity. *Appl. Microbiol. Biotechnol.* **48**, 465–472 (1997).
- Monchois, V., Remaud-Simeon, M., Monsan, P. & Willemot, R. M. Cloning and sequencing of a gene coding for an extracellular dextranase (DSRB) from *Leuconostoc mesenteroides* NRRL B-1299 synthesizing only an alpha (1-6) glucan. *FEMS Microbiol. Lett.* **159**, 307–315 (1998).
- Fabre, E. *et al.* Role of the two catalytic domains of DSR-E dextranase and their involvement in the formation of highly alpha-1,2 branched dextran. *J. Bacteriol.* **187**, 296–303 (2005).
- Kralj, S. *et al.* Molecular characterization of a novel glucosyltransferase from *Lactobacillus reuteri* strain 121 synthesizing a unique, highly branched glucan with alpha-(1 → 4) and alpha-(1 → 6) glucosidic bonds. *Appl. Environ. Microbiol.* **68**, 4283–4291 (2002).
- Kralj, S., Stripling, E., Sanders, P., van Geel-Schutten, G. H. & Dijkhuizen, L. Highly hydrolytic reuteransucrase from probiotic *Lactobacillus reuteri* strain ATCC 55730. *Appl. Environ. Microbiol.* **71**, 3942–3950 (2005).
- Côté, G. L. & Robyt, J. F. Isolation and partial characterization of an extracellular glucansucrase from *Leuconostoc mesenteroides* NRRL B-1355 that synthesizes an alternating (1 goes to 6), (1 goes to 3)-alpha-D-glucan. *Carbohydr. Res.* **101**, 57–74 (1982).
- van Leeuwen, S. S. *et al.* Structural analysis of the alpha-D-glucan (EPS35-5) produced by the *Lactobacillus reuteri* strain 35-5 glucansucrase GTFA enzyme. *Carbohydr. Res.* **343**, 1251–1265 (2008).
- Brisson, Y. *et al.* Functional and structural characterization of alpha-(1->2) branching sucrose derived from DSR-E glucansucrase. *J. Biol. Chem.* **287**, 7915–7924 (2012).
- Vuillemin, M. *et al.* Characterization of the first alpha-(1->3) branching sucrases of the GH70 family. *J. Biol. Chem.* **291**, 7687–7702 (2016).
- MacGregor, E. A., Jespersen, H. M. & Svensson, B. A circularly permuted alpha-amylase-type alpha/beta-barrel structure in glucan-synthesizing glucosyltransferases. *FEBS Lett.* **378**, 263–266 (1996).
- Janecek, S., Svensson, B. & MacGregor, E. A. alpha-Amylase: an enzyme specificity found in various families of glycoside hydrolases. *Cell Mol. Life Sci.* **71**, 1149–1170 (2014).
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–5 (2014).
- MacGregor, E. A., Janecek, S. & Svensson, B. Relationship of sequence and structure to specificity in the alpha-amylase family of enzymes. *Biochim. Biophys. Acta* **1546**, 1–20 (2001).
- Meng, X. *et al.* Structure-function relationships of family GH70 glucansucrase and 4,6-alpha-glucanotransferase enzymes, and their evolutionary relationships with family GH13 enzymes. *Cell Mol. Life Sci.* **73**, 2681–2706 (2016).
- Janecek, S. & Gabrisko, M. Remarkable evolutionary relatedness among the enzymes and proteins from the alpha-amylase family. *Cell Mol. Life Sci.* **73**, 2707–2725 (2016).
- Vujičić-Zagar, A. *et al.* Crystal structure of a 117 kDa glucansucrase fragment provides insight into evolution and product specificity of GH70 enzymes. *Proc. Natl. Acad. Sci. USA* **107**, 21406–21411 (2010).
- Pijning, T., Vujicic-Zagar, A., Kralj, S., Dijkhuizen, L. & Dijkstra, B. W. Structure of the alpha-1,6/alpha-1,4-specific glucansucrase GTFA from *Lactobacillus reuteri* 121. *Acta Crystallogr. Sect. F. Struct. Biol. Commun.* **68**, 1448–1454 (2012).

22. Ito, K. *et al.* Crystal structure of glucansucrase from the dental caries pathogen *Streptococcus mutans*. *J. Mol. Biol.* **408**, 177–186 (2011).
23. Gangoit, J., Pijning, T. & Dijkhuizen, L. The *Exiguobacterium sibiricum* 255-15 GtfC enzyme represents a novel glycoside hydrolase 70 subfamily of 4,6- $\alpha$ -glucanotransferase enzymes. *Appl. Environ. Microbiol.* **82**, 756–766 (2015).
24. Kralj, S. *et al.* 4,6- $\alpha$ -Glucanotransferase, a novel enzyme that structurally and functionally provides an evolutionary link between glycoside hydrolase enzyme families 13 and 70. *Appl. Environ. Microbiol.* **77**, 8154–8163 (2011).
25. Leemhuis, H. *et al.* 4,6- $\alpha$ -Glucanotransferase activity occurs more widespread in *Lactobacillus* strains and constitutes a separate GH70 subfamily. *Appl. Microbiol. Biotechnol.* **97**, 181–193 (2013).
26. Gangoit, J., van Leeuwen, S. S., Vafiadi, C. & Dijkhuizen, L. The gram-negative bacterium *Azotobacter chroococcum* NCIMB 8003 employs a new glycoside hydrolase family 70 4,6- $\alpha$ -glucanotransferase enzyme (GtfD) to synthesize a reuteran like polymer from maltodextrins and starch. *Biochim Biophys Acta* **1860**, 1224–1236 (2016).
27. Bai, Y. *et al.* Biochemical characterization of the *Lactobacillus reuteri* glycoside hydrolase family 70 GTFB type of 4,6- $\alpha$ -glucanotransferase enzymes that synthesize soluble dietary starch fibers. *Appl. Environ. Microbiol.* **81**, 7223–7232 (2015).
28. Bai, Y. *et al.* *Lactobacillus reuteri* strains convert starch and maltodextrins into homoexopolysaccharides using an extracellular and cell-associated 4,6- $\alpha$ -glucanotransferase. *J. Agric. Food Chem.* **64**, 2941–2952 (2016).
29. Irague, R. *et al.* Combinatorial engineering of dextranase specificity. *PLoS One* **8**, e77837 (2013).
30. Kralj, S., Eeuwema, W., Eckhardt, T. H. & Dijkhuizen, L. Role of asparagine 1134 in glucosidic bond and transglycosylation specificity of reuteransucrase from *Lactobacillus reuteri* 121. *FEBS J.* **273**, 3735–3742 (2006).
31. Meng, X., Pijning, T., Dobruchowska, J. M., Gerwig, G. J. & Dijkhuizen, L. Characterization of the functional roles of amino acid residues in acceptor-binding subsite +1 in the active site of the glucansucrase GTF180 from *Lactobacillus reuteri* 180. *J. Biol. Chem.* **290**, 30131–30141 (2015).
32. Meng, X., Dobruchowska, J. M., Pijning, T., Gerwig, G. J. & Dijkhuizen, L. Synthesis of new hyperbranched  $\alpha$ -glucans from sucrose by *Lactobacillus reuteri* 180 glucansucrase mutants. *J. Agric. Food Chem.* **64**, 433–442 (2016).
33. Yin, Y. *et al.* dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **40**, W445–51 (2012).
34. Lechat, P., Souche, E. & Moszer, I. SynTVView - an interactive multi-view genome browser for next-generation comparative microorganism genomics. *BMC Bioinformatics* **14**, 277-2105-14-277 (2013).
35. Tamura, K., Stecher, G., Peterson, D., Filipiński, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
36. van Leeuwen, S. S. *et al.* Structural analysis of the  $\alpha$ -D-glucan (EPS180) produced by the *Lactobacillus reuteri* strain 180 glucansucrase GTF180 enzyme. *Carbohydr. Res.* **343**, 1237–1250 (2008).
37. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Pyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
38. Ganzle, M. G. & Follador, R. Metabolism of oligosaccharides and starch in lactobacilli: a review. *Front. Microbiol.* **3**, 340 (2012).
39. Moller, M. S. *et al.* Enzymology and structure of the GH13\_31 glucan 1,6- $\alpha$ -glucosidase that confers isomaltooligosaccharide utilization in the probiotic *Lactobacillus acidophilus* NCFM. *J. Bacteriol.* **194**, 4249–4259 (2012).
40. Macfarlane, G. T., Steed, H. & Macfarlane, S. Bacterial metabolism and health-related effects of galacto-oligosaccharides and other prebiotics. *J. Appl. Microbiol.* **104**, 305–344 (2008).
41. Nguyen, H. A. *et al.* Heterologous expression and characterization of an N-acetyl-beta-D-hexosaminidase from *Lactococcus lactis* ssp. *lactis* IL1403. *J. Agric. Food Chem.* **60**, 3275–3281 (2012).
42. Beekwilder, J. *et al.* Characterization of Rhamnosidases from *Lactobacillus plantarum* and *Lactobacillus acidophilus*. *Appl. Environ. Microbiol.* **75**, 3447–3454 (2009).
43. Goddard-Borger, E. D. *et al.* Mechanistic insights into the 1,3-xylanases: useful enzymes for manipulation of algal biomass. *J. Am. Chem. Soc.* **134**, 3895–3902 (2012).
44. B ath, 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.* **253**, 75–82 (2005).
45. van Hijum, S. A., Kralj, S., Ozimek, L. K., Dijkhuizen, L. & van Geel-Schutten, I. G. Structure-function relationships of glucansucrase and fructansucrase enzymes from lactic acid bacteria. *Microbiol. Mol. Biol. Rev.* **70**, 157–176 (2006).
46. van Leeuwen, S. S. *et al.* Structural characterization of bioengineered  $\alpha$ -D-glucans produced by mutant glucansucrase GTF180 enzymes of *Lactobacillus reuteri* strain 180. *Biomacromolecules* **10**, 580–588 (2009).
47. Kralj, S., van Geel-Schutten, I. G., Faber, E. J., van der Maarel, M. J. & Dijkhuizen, L. Rational transformation of *Lactobacillus reuteri* 121 reuteransucrase into a dextranase. *Biochemistry* **44**, 9206–9216 (2005).
48. van Leeuwen, S. S., Kralj, S., Gerwig, G. J., Dijkhuizen, L. & Kamerling, J. P. Structural analysis of bioengineered  $\alpha$ -D-glucan produced by a triple mutant of the Glucansucrase GTF180 enzyme from *Lactobacillus reuteri* strain 180: generation of ( $\alpha$ 1  $\rightarrow$  4) linkages in a native (1  $\rightarrow$  3)(1  $\rightarrow$  6)- $\alpha$ -D-glucan. *Biomacromolecules* **9**, 2251–2258 (2008).
49. Uitdehaag, J. C., van Alebeek, G. J., van Der Veen, B. A., Dijkhuizen, L. & Dijkstra, B. W. Structures of maltohexaose and maltoheptaose bound at the donor sites of cyclodextrin glycosyltransferase give insight into the mechanisms of transglycosylation activity and cyclodextrin size specificity. *Biochemistry* **39**, 7772–7780 (2000).
50. Barends, T. R. *et al.* Three-way stabilization of the covalent intermediate in amylomaltase, an  $\alpha$ -amylase-like transglycosylase. *J. Biol. Chem.* **282**, 17242–17249 (2007).
51. Synnysya, A. & Novak, M. Structural diversity of fungal glucans. *Carbohydr. Polym.* **92**, 792–809 (2013).
52. Olafsdottir, E. S. & Ingolfsdottir, K. Polysaccharides from lichens: structural characteristics and biological activity. *Planta Med.* **67**, 199–208 (2001).
53. Tsumuraya, Y., Misaki, A., Takaya, S. & Torii, M. A new fungal  $\alpha$ -D-glucan, elsinan, elaborated by *Elsinoe leucospila*. *Carbohydr. Res.* **66**, 53–65 (1978).
54. Olafsdottir, E. S. *et al.* Immunologically active (1  $\rightarrow$  3)-(1  $\rightarrow$  4)- $\alpha$ -D-glucan from *Cetraria islandica*. *Phytomedicine* **6**, 33–39 (1999).
55. Bock, K., Gagnaire, D., Vignon, M. & Vincendon, M. High resolution nuclear magnetic resonance studies of nigeran. *Carbohydr. Polym.* **3**, 13–22 (1983).
56. Cordeiro, L. M., Stocker-Worgotter, E., Gorin, P. A. & Iacomini, M. Elucidation of polysaccharide origin in *Ramalina peruviana* symbiosis. *FEMS Microbiol. Lett.* **238**, 79–84 (2004).
57. Takeda, T. *et al.* Further studies on the structure of polysaccharides from the lichen *Flavoparmelia caperata* (L.) HALE. *Chem. Pharm. Bull. (Tokyo)* **51**, 1436–1438 (2003).
58. Wolski, E. A. *et al.* An  $\alpha$ -glucan elicitor from the cell wall of a biocontrol binucleate *Rhizoctonia* isolate. *Carbohydr. Res.* **340**, 619–627 (2005).
59. Gorin, P. A. J. & Iacomini, M. Polysaccharides of the lichens *Cetraria islandica* and *Ramalina usnea*. *Carbohydr. Res.* **128**, 119–132 (1984).

## Acknowledgements

This work was financially supported by Nestec Ltd and by the University of Groningen. We are very grateful to Albert Woortman (Macromolecular Chemistry and New Polysaccharide Materials, University of Groningen) for the HPSEC-MALLS-RI analyses.

## Author Contributions

J.G., C.V. and L.D. designed the study. J.G. designed, performed and analyzed most of experiments. S.S.v.L. conducted the NMR analysis of polysaccharide shown in Table S3 and Fig. S3 and the Smith degradation analysis shown in Fig. S4. S.D. annotated the carbohydrate-active enzymes present in the *L. fermentum* NCC 2970 genome. T.P. analyzed the protein sequence in combination with the homology model of the *L. fermentum* GtfB in Fig. 2. G.J.G. contributed to the methylation analysis in Table 1. J.G., S.S.v.L. and L.D. wrote the paper. All the authors reviewed the results, revised the manuscript and approved the final version of the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Gangoiti, J. *et al.* 4,3- $\alpha$ -Glucanotransferase, a novel reaction specificity in glycoside hydrolase family 70 and clan GH-H. *Sci. Rep.* 7, 39761; doi: 10.1038/srep39761 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017