JBC Papers in Press. Published on June 3, 2020 as Manuscript RA120.014029
The latest version is at https://www.jbc.org/cgi/doi/10.1074/jbc.RA120.014029

Cross-utilization of β-galactosides and cellobiose in *Geobacillus stearothermophilus*

Smadar Shulami[1#], Arie Zehavi[1#], Valery Belakhov[2], Rachel Salama[1], Shifra Lansky[3], Timor Baasov[2], Gil Shoham[3]* and Yuval Shoham[1]*

[1] Department of Biotechnology and Food Engineering, Technion-Israel Institute of Technology, Haifa, Israel 32000
[2] Schulich Faculty of Chemistry, Technion - Israel Institute of Technology, Haifa 32000, Israel.
[3] Institute of Chemistry and the Laboratory for Structural Chemistry and Biology, The Hebrew University of Jerusalem, Jerusalem 91904, Israel.

# These authors contributed equally to this work.

*Corresponding authors: Yuval Shoham, E-mail: yshoham@technion.ac.il; Gil Shoham, gil2@mail.huji.ac.il

Running title: *Growth of G. stearothermophilus on galactosides & cellobiose*

Keywords: glycoside hydrolases; 6-phospho-β-galactosidase; 6-phospho-β-glucosidase; PTS system; *G. stearothermophilus*; galactose; galactosyl-glycerol; lactose; ABC transport system; bacterial metabolism; cellobiose

Abstract

Strains of the Gram-positive, thermophilic bacterium *Geobacillus stearothermophilus* possess elaborate systems for the utilization of hemicellulolytic polysaccharides, including xylan, arabinan, and galactan. These systems have been studied extensively in strains T-1 and T-6, representing microbial models for the utilization of soil polysaccharides, and many of their components have been characterized both biochemically and structurally. Here, we characterized routes by which *G. stearothermophilus* utilizes mono- and disaccharides such as galactose, cellobiose, lactose, and galactosyl-glycerol. The *G. stearothermophilus* genome encodes a phosphoenolpyruvate carbohydrate phospho-transferase system (PTS) for cellobiose. We found that the cellobiose-PTS system is induced by cellobiose and characterized the corresponding GH1 6-phospho-β-glucosidase, Cel1A. The bacterium also possesses two transport systems for galactose, a galactose-PTS system and an ABC galactose transporter. The ABC galactose transport system is regulated by a three-component sensing system. We observed that both systems, the sensor and the transporter, utilize galactose-binding proteins that also bind glucose with the same affinity. We hypothesize that this allows the cell to control the flux of galactose into the cell in the presence of glucose. Unexpectedly, we discovered that *G. stearothermophilus* T-1 can also utilize lactose and galactosyl-glycerol via the cellobiose-PTS system together with a bifunctional 6-phospho-β-galactosidase/glucosidase, Gan1D. Growth curves of strain T-1 growing in the presence of cellobiose, with either lactose or galactosyl-glycerol, revealed initially logarithmic growth on cellobiose and then linear growth supported by the additional sugars. We conclude that Gan1D allows the cell to utilize residual galactose-containing disaccharides, taking advantage of the promiscuity of the cellobiose-PTS system.

Introduction

Geobacillus stearothermophilus T-1 is a thermophilic, Gram-positive, soil bacterium, which is capable of utilizing plant cell-wall derived polysaccharides, including xylan, arabinan and galactan (1-3). Utilization of these polysaccharides include extracellular and intracellular hemicellulolytic enzymes, ATP-binding cassette (ABC) sugar-transport systems, carbohydrate sensing systems, sugar metabolism enzymes and regulatory proteins (4-8). A major challenge for soil bacteria in the natural environment is to sense the scarce carbon sources and compete on these sources with nearby microorganisms. Indeed, we have demonstrated that *G. stearothermophilus* employs a unique strategy for the efficient utilization of polysaccharides in its immediate environment. First, it utilizes two- or three-component systems to sense minute amounts of mono- or di-saccharides in the surrounding, which signal the presence of the corresponding polysaccharides (2,9). The two- or three-component sensing systems then activate dedicated ABC sugar transporters that transfer the sugars into the cell and induce the corresponding systems for expressing extracellular endo-type glycoside hydrolases that partially degrade the high molecular weight polysaccharides into short (usually decorated) oligosaccharides. Additional ABC transporters for oligosaccharides transporters transfer the large oligosaccharides into the cell, and those are further hydrolyzed into sugar monomers by a battery of specific intracellular glycoside hydrolases (10-16). This utilization strategy allows the bacterium to a) react rapidly to presence of potential polysaccharides such as xylan, arabinan and galactan in the immediate environment; b) transfer efficiently the degradation products into the bacterium cell; c) almost exclusively utilize the degraded decorated oligosaccharides, since these are rarely imported by other organisms. Essentially the same strategy was recently demonstrated for the utilization of yeast mannan by the human gut *Bacteroidetes* and was coined a “selfish mechanism” (17).

In addition to the complete utilization machinery for xylan, arabinan and galactan, *G. stearothermophilus* also possesses scavenging mechanisms for the utilization of mono- or di-saccharides that are often found in the surroundings, resulting from the extensive degradation of the corresponding polysaccharides by other soil microorganisms. In this case, the sugars are imported by the bacterium into the cell *via* a different type of transporters, the phosphoenolpyruvate-dependent phospho-

transferase systems (PTS). The PTS systems usually use phosphoenolpyruvate (PEP) as the phosphoryl donor for sugar phosphorylation, together with three essential catalytic entities, termed enzyme I, enzyme II, and HPr (heat-stable, histidine-phosphorylatable protein) (18). During their import *via* the PTS systems, sugars are simultaneously phosphorylated at the C6 hydroxyl group of the terminal sugar unit at the non-reducing end, and are further cleaved (inside the cell) by dedicated 6-phospho-β-glycosidases/galactosidases (19).

In the present study, we identified in *G. stearothermophilus* strain T-1 such PTS systems for cellobiose and galactose, and biochemically characterized the corresponding enzymes involved. These are the 6-phospho-β-galactosidase Gan1D, belonging to glycoside hydrolase (GH) family GH1, and the 6-phospho-β-glycosidase Cel1A belonging also to family GH1. In addition, we demonstrate that the bacterium can utilize the β-galactosides lactose and galactosyl-glycerol, using the cellobiose PTS system together with the bifunctional 6-phospho-β-galactosidase, Gan1D.

Results

Identifying of the cellobiose utilization gene cluster

As part of a systematic search for hemicellulolytic utilization systems in *G. stearothermophilus*, we identified in strain T-1 a 4.3 kb gene cluster, which appears to be involved in the utilization of cellobiose (Figure 1A). The cluster was identified *via* bioinformatics analysis of the genome sequence of strains T-1 and includes an operon of four genes and a transcriptional regulator gene. Based on sequence homology, the *celBCD* genes encode for a complete PTS system. The CelD protein is homologous to proteins from the lactose/cellobiose EIIC family that bind cellobiose or lactose. CelC and CelB are the EIIA and EIIB domains, respectively, which transfer the phosphoryl group from HPr to the transported sugar. The forth gene in this operon, *celA*, encodes for a 6-phospho-β-glucosidase, Cel1A, which shows sequence homology to a number of GH1 6-phospho-glucosidases, including those from *Streptococcus pyogenes* (51% sequence identity), *Lactobacillus plantarum* (32%) and *Streptococcus mutans* (32%). Adjacent to the operon lays the *celR* gene encoding for a transcriptional regulator belonging to the GntR family. The genes for the EI and HPr proteins are located elsewhere on the chromosome.

mRNA expression of cellobiose utilizing genes

G. stearothermophilus T-1 is unable to grow on cellulose but grows very well on cellobiose. To test whether the *celBCD*, *celA* and *celR* genes are involved in cellobiose utilization, we measured the corresponding mRNA levels in cultures grown on cellobiose, xylose or glucose. Total mRNA was extracted from mid-exponential phase cultures and the cDNA was amplified with primers specific to *celA, celB, celC, celD and celR* as well as for the isocitrate dehydrogenase gene, *ict*, that was used for normalization. Relative expression was measured by real-time RT PCR, as presented in Figure 1B. The expression levels of the corresponding genes were 7-15-fold higher in the cellobiose-grown culture, as compared to cultures grown on xylose or glucose. The relatively high levels of *celA* mRNA compared to the *celBCD* genes, may reflect differences in mRNA stability. These results suggest that the *celBCDA* operon is induced by cellobiose and thus highly likely to be involved in cellobiose utilization.

Biochemical characterization of Cel1A

The biochemical activity of Cel1A was determined using different chromogenic and natural substrates (Table 1). Cel1A exhibited a significant catalytic activity on substrates with glucose-6-phosphate at the glycon moiety, and had no detectable activity toward unphosphorylated substrates. These results suggest that Cel1A is a 6-phospho-β-glucosidase, with high specificity toward glucose-6-phosphate as the glycon moiety. The effect of temperature on Cel1A activity was determined at pH 6.5, using the chromogenic substrate oNPβGlc6P. The optimal temperature in a 20 min reaction was 70°C (Figure 2A). The thermal stability was determined after incubating Cel1A at temperatures between 30°C to 85°C for 10 minutes. The enzyme was stable at temperatures below 65°C and lost over 90% of its activity at 75°C (Figure 2B). The activity of Cel1A was determined at different pH values in the range of

3.5 - 9.5 (Figure 2C). The pH profile is a typical bell-shaped curve, which may reflect the ionization of the two catalytic carboxylates. The enzyme was most active at the pH range of 6.5 to 8, and lost about 90% of its activity at pH 3.5 and 9.5. Based on sequence alignment of Cel1A with other retaining GH1 enzymes, the putative acid/base and nucleophile catalytic residues are Glu174 and Glu373, respectively. These residues were substituted with alanine and the effect of the mutations on the activity of the enzyme was determined using *o*NPβGlc6P as the substrate. The Michaelis-Menten catalytic constants of the Cel1A-E174A catalytic mutant towards *o*NP-Glc6P were 0.01 mM, 2.1 s^{-1}, and 2.1×10^5 s^{-1} M^{-1} for K_m, k_{cat} and k_{cat}/K_m, respectively. The k_{cat} value was about 50-fold lower, compared to the wild type, and the K_m was 10-fold lower, compared to the wild type, suggesting the accumulation of a glycosyl-enzyme intermediate. Such results are expected for an acid-base mutant since the relatively high reactivity of the *o*-nitrophenol leaving group (pK_a 7.22) elevates the rate of the first glycosylation step even without a proton assistant. The second deglycosylation step remains considerably slow due to the loss of the base required for the activation of the catalytic water molecule (20). The alanine replacement of the Glu373 nucleophile, resulted in a non-detectable activity toward *o*NP-Glc6P, reflecting the inability of this mutant to bind the substrate and stabilize the first transition state. In similar retaining glycoside hydrolases, the nucleophile is usually involved in a strong interaction with the C2 hydroxyl group of the glycone sugar (21).

Identification of the sensing and transport systems for galactose

We have demonstrated previously that *G. stearothermophilus* T-1 consumes galactan efficiently, utilizing the galactan utilization gene cluster, *ganREFGBA* (3). Following bioinformatics analysis of the bacterium genome sequence we identified a new cluster that appears to be involved in galactose utilization. This new 12.5-kb cluster is composed of genes encoding for a putative three-component regulatory sensing system (GalPST2), an ABC sugar transport system (GalE2F2G2), a regulatory protein (GanR2) and two 6-phospho glycosidases (Gan4C and Gan1D) (Figure 3).

GalP exhibits significant sequence similarity to periplasmic sugar-binding proteins with a 27 amino acid signal peptide at its N-terminus. GalS exhibits characteristic features of bacterial histidine kinase proteins, including two transmembrane (TM) helices (TM1, residues 7-25 and TM2, residues 177-197) flanking an extracellular domain (residues 26-176) and a conserved C-terminal cytoplasmic region containing the ATP-binding kinase domain. Downstream to the *galS* gene lays the *galT2* gene, which encodes for a protein with high sequence similarity to response regulators. As in the case for many response regulators, GalT2 has a predicted two-domain architecture, with an N-terminus signal receiver domain (REC), linked to a C-terminus effector domain (22). The N-terminal signal receiver domain (positioned at residues 5-118) shares a significant homology with the Rec superfamily of the REC domains, whereas the C-terminal domain (residues 157-225) contains a putative helix-turn-helix motif, resembling the AraC-type DNA-binding domain (23). Taken together, these observations suggest that the *galPST2* and *galE2F2G2* clusters encode for a three-component sensing system and an ABC galactose transport system, respectively.

mRNA expression levels of the galactose related genes

Based on genome sequence analysis, strain T-1 encodes for four sugars-specific PTS systems, two for the disaccharides, cellobiose and trehalose, and two for the mono-sugars, mannitol and galactose (Figure 4A). To determine whether the *galPST2* and *galE2F2G2* gene clusters are involved in galactose utilization and to identify the PTS system for galactose, the mRNA levels of the *galPST2*, *galE2F2G2* and the four sugar-PTS systems were measured in cultures grown in the presence of galactose and related sugars (Figure 4). The expression *of galE2* was 5-fold higher on galactose-grown cultures, compared to cultures grown on either glucose, xylose or galactan (Figure 4B). These results support our original suggestion that the *galE2F2G2* operon constitutes an ABC galactose transporter. The expression of the *galP* gene, which is part of a three-component sensing system, appears to be relatively low and constant, regardless of the carbon source used (Figure 4B).

Although these results cannot correlate the system to galactose, it is expected that the sensing systems will be expressed constitutively. The mRNA levels of the four PTS systems were also measured in cultures grown on either galactose or glucose (Figure 4C). The expression level of the *ptsA* gene in galactose-grown culture was about 9-fold higher than for a culture grown on glucose, suggesting that *ptsA* is part of the galactose PTS transporter, dedicated for galactose import. Taken together, these results suggest that *G. stearothermophilus* T-1 has two transport systems for galactose.

GalP and GalE2 bind galactose

The putative three-component system for galactose, GalPST2, and the galactose ABC transport system, GalE2F2G2, both have dedicated sugar-binding proteins (GalP and GalE2, respectively) that are tethered to the local membrane. The ability of these proteins to bind galactose was confirmed by isothermal titration calorimetry (ITC). The calorimetric titration curves are shown in Figure 5 and the thermodynamic binding parameters are summarized in Table 2. These results demonstrate that GalE2 and GalP are able to bind galactose quite tightly, with dissociation constants in the micromolar range, K_D = 1.4 μM and K_D = 6.1 μM, respectively. Similar titrations of GalP with cellobiose and lactose did not result in a significant enthalpy change. Surprisingly, however, both proteins bind glucose in similar affinities to galactose. Nevertheless, although glucose seems to bind these proteins quite tightly, it is rather unlikely that it can activate the corresponding sensing systems, since the expression of the operons of these systems are practically unaffected by glucose (Figure 4B).

GalT2 binds to the *galE2F2G2* promoter region

The fact that the potential three-component sensing system for galactose (*galPST2*) is located adjacent to a putative ABC transport system (*galE2F2G2*), suggests that the sensing system is functionally linked to this transport system, and that GalT2 is therefore a response regulator that regulates the expression of the transporter. This type of adjacent arrangement of gene clusters was found in *G. stearothermophilus* also for the arabinose and xylotriose utilization systems (2,9). Sequence analysis of the promoter region of *galE*2 revealed a putative -35 sequence (TTGATA), which is a relatively close match to the σ^A consensus sequence, TTGACA. This -35 sequence is separated by 18-bp from the potential -10 region (CAACAT), which differs from the *B. subtilis* consensus, TATAAT, by three nucleotides (24). The upstream region of the -35 of the *galE2* promoter contains two direct repeats, CAAAAAAGT, separated by 11-bp, which may function as the recognition sequences for the response regulator GanlT2 (Figure 6A). This putative binding site, upstream of the -35 region, can allow direct interaction of the activator with the carboxy-terminal domain of the α subunit of RNA polymerase (25,26). To test whether GalT2 can bind the *galE2* promoter region we utilized gel mobility shift assays. These studies indicated that the GalT2 protein (using N-His_6-GalT2) can bind significantly to a 113-bp DNA fragment containing the two direct repeats discussed above, and a nearly complete shift was obtained in the presence of 0.3 μM GalT2 (Figure 6B). This relatively high concentration of GalT2 (reflecting a relatively weak protein-DNA binding) may originate from the fact that the protein was not fully phosphorylated, as often observed for other response regulators (27). To test whether the phosphorylation of GalT2 increase significantly its binding to DNA, we phosphorylated GalT2 *in vitro*, using Mg^{+2} and acetyl phosphate as the phosphate donor. As shown by size-exclusion chromatography, such phosphorylation changed the oligomeric state of GalT2 in solution, transforming it from a monomer (not phosphorylated) to a dimer (phosphorylated) (Figure 6C), as usually observed for related response regulators. Indeed, the fully-phosphorylated GalT2 protein (in its dimeric form) gave a complete shift at 0.07 μM (Figure 6D), demonstrating a significant increase in its DNA binding capabilities. These results further support the identification of GalT2 as a response regulator, as such regulators usually change their conformation upon phosphorylation, allowing them to enhance their binding to their target DNA segments (28,29).

Biochemical characterization of Gan1D

The 12.5-kb galactose utilization cluster also contains the *gan1D* gene, which is expressed in cultures grown on galactose or galactan (Figure 4B). We have previously described the 3D crystal structure of Gan1D, as well as its catalytic mutants complexed with substrates and products (30). These Gan1D structures revealed the structural features that allow the enzyme to accommodate both 6-phospho-galactose and 6-phospho-glucose at the glycone binding site (-1) (30). To further characterize the specificity of Gan1D, its catalytic activity was studied using different chromogenic and natural substrates (Table 3). Gan1D readily hydrolyzed substrates with a glycoside-6-phosphate as the glycone moiety such as lactose-6-phosphate. The enzyme had no detectable activity toward non-phosphorylated substrates, in a good correlation with similar studies conducted on other GH1 6-phospho-β-glycosidases (20,31). Interestingly, however, the enzyme exhibited bifunctional specificity, showing similar catalytic activities on substrates containing glucose-6-phosphate or galactose-6-phosphate as their glycone moiety (Table 3).

Growth of strain T-1 on β-galactosides substrates in the presence of cellobiose

The physiological role of Gan1D in *G. stearothermophilus* T-1 was initially puzzling since we failed to identify a dedicated PTS system for galacto-disaccharides. Indeed, neither lactose nor galactosyl-glycerol were found to support the growth of *G. stearothermophilus* T-1. Lactose is not expected to be abundant in soil, the natural niche of the bacterium, however, galactosyl-glycerol, the degradation product of galacto-lipids, should be available there. The most abundant galacto-lipids in plants are those of mono-galactosyl-diacyl-glycerol (MGDG) and digalactosyl-diacyl-glycerol (DGDG), mainly found in plant chloroplast membranes (32,33). The degradation of both of these molecules by lipases releases galactosyl-glycerol, a β-galactoside bound to glycerol (1-0-β-galactosyl-glycerol). Interestingly, unlike the genetic context of many phospho-glycosidases, the *gan1D* gene in strain T-1 as well as its homologous enzymes in many other bacteria, do not lay adjacent to the genes of the corresponding PTS system for disaccharides. Considering these observations, we hypothesized that the utilization of β-galactosides (lactose or galactosyl-glycerol) in strain T-1 is linked to an alternative PTS system. A somewhat supporting evidence for that is the observation that *Lactococcus lactis* IL1403, can utilize lactose only in the presence of cellobiose (34). In analogy, it was tempting to speculate that the utilization of β-galactosides (lactose or galactosyl-glycerol) in *G. stearothermophilus* T-1 is actually depended on its cellobiose PTS system (*celBCD*). To test this hypothesis, strain T-1 was grown in a defined mBSM medium, containing either 0.4% lactose or galactosyl-glycerol in the presence of various concentrations of cellobiose. The resulting growth curves were characterized with an initial exponential growth on cellobiose, followed by a linear growth on lactose or galalctosyl-glycerol (Figure 7AB). The extent of the logarithmic growth was proportional to the initial cellobiose concentration, suggesting that in this phase, it is cellobiose that is mostly consumed. Following the logarithmic phase, the cultures exhibited linear growth, this time likely consuming lactose or galactosyl-glycerol. The slopes of the linear growth were proportional to the turbidity (cell mass) at end of the logarithmic phase (see inserts in Figure 7AB). These unexpected results can be explained by the following behavior of the bacterium. At the end of logarithmic phase, it seems that all of the cellobiose is already consumed, and at that stage, the cells do not express the now non-induced cellobiose PTS genes (*celBCD*). Hence, at the end of the exponential growth, the number of PTS systems per cell seem to remain constant, neglecting obvious natural turnover. At that point, the limiting factor of the growth rate is likely to be the transport of lactose or galactosyl-glycerol into the cell, probably *via* the already fixed population of the cellobiose PTS systems. This hypothesis seems to be in good correlation with the linear cell growth observed from this point on, which is now proportional to the (now unchanged) overall number of cellobiose PTS systems per cell.

Taken these observations together, it appears as if *G stearothermophilus* T-1 is capable of utilizing galacto-disaccharides using the alternative cellobiose PTS system for import and the Gan1D enzyme for processing. In support of this

hypothesis, it is noted that *G. stearothermophilus* T-6, a close relative of strain T-1, possesses a similar cellobiose PTS system, but lacks a 1825-bp segment containing the *gan1D* homologous gene. In turn, this strain grows very well on cellobiose, yet fails to grow on lactose or galactosyl-glycerol in the presence of cellobiose (Figure 7C).

Discussion

G. stearothermophilus possesses a cellobiose- and galactose- PTS systems. *G. stearothermophilus* encodes for four putative PTS systems, dedicated for the import of cellobiose, galactose, mannitol and trehalose. The cellobiose-PTS system was identified based on the observation that the *celBCDA* operon is expressed in the presence of cellobiose (Figure 1B) and that Cel1A rapidly hydrolyzes 6-phospho-cellobiose (Table 1). The cellobiose-PTS system is most likely regulated by CelR. A sequence resembling an operator site for CelR includes an inverted repeat of 8-bp (5'-TTTTTATT -N_{13}N – AATAAAAA-3') and it is located within the promoter region of the *celBCDA* operon. *G. stearothermophilus* can readily utilize cellobiose, although it does not encode for cellulolytic genes. Thus, this bacterium seems to possess the ability to scavenge from the surroundings residual cellobiose, which is the main product of cellulose degradation by cellulolytic microorganisms.

The PTS system for galactose was identified in *G. stearothermophilus* based on the measurements of mRNA levels of the *ptsA* gene, in the presence of galactose. The system is encoded by a four-gene operon, consisting of a multi-domain EII protein, with the corresponding domains of EIIA, EIIB and EIIC fused together (*ptsA*), a histidine-containing phospho-carrier (HPr)-like protein (*ptsB*), and a phosphoenolpyruvate-protein kinase (the EI component) (*ptsC*). The first gene in the operon is *glcT*, encoding for an anti-terminator protein of the BlgG family. Both genes (*glcT* and *ptsA*) are separated by a palindromic sequence that functions as a transcriptional terminator. GlcT possesses an N-terminal RNA binding domain, followed by two PTS regulation domains, which are likely to be phosphorylated and thereby modulate the RNA binding activity (35,36). Based on these observations, the galactose PTS system of *G. stearothermophilus* seems to be regulated by the anti-terminator GlcT.

In *G. stearothermophilus* strain T-1, galactose can also be imported by a non-PTS ABC galactose transporter (*galE2G2F2*) (Figure 8). The operon is upregulated in the presence of galactose and the corresponding substrate binding protein (SBP) GalE2, was shown to have a significant galactose binding (Table 2). In *Lactococcus lactis*, the import of galactose occurs *via* high affinity non-PTS system (K_d in the μM range), whereas the galactose PTS system is a low-affinity transporter (K_d in the mM range). The specific binding affinity of the strain T-1 galactose PTS system is yet to be determined, however our current data indicates that GalE2 binds galactose with a relatively high-affinity, with K_d values in the μM range (Table 2).

A three-component sensing system regulates the expression of the galactose ABC transporter. The *galP*, *galS* and *galT2* gene products constitute a three-component sensing system, in which GalP is a galactose-binding lipoprotein, GalS is a class-I histidine kinase and GalT2 is the response regulator, which seems to be phosphorylated by GalS (Figure 8). In a good correlation with similar sensing systems, GalP binds extracellular galactose and is likely to activate the sensor histidine kinase, either by direct protein-protein interaction, or by presenting the sugar to the histidine kinase sensor domain. In turn, the auto-phosphorylated GalS histidine kinase phosphorylates the response regulator, GalT2, thereby passing the sensing signal into the cell. Thus, GalP seems to be a typical SBP, functioning here as an auxiliary component that assists this three-component system in sensing small amounts of extracellular galactose.

GalT2 seems to function as a typical response regulator, and accordingly possesses a phospho-acceptor receiver domain (REC), a common module in a variety of response regulators (27). Interestingly, many response regulators are able to catalyze their own phosphorylation *in vitro* in the presence of Mg^{+2} and a suitable phosphoryl donor, and in the absence of any auxiliary proteins (37). Phosphorylation of the REC domains usually induce dimerization of the response regulator, commonly associated with conformational changes, thereby facilitating its binding to the target DNA segment (usually palindromic or direct repeat sequences) (38,39). Indeed, we were able

to show that *in vitro* auto-phosphorylation of GalT2 promoted its dimer formation in solution, and facilitated its ability to bind DNA. The potential binding sites for GalT2 are located upstream of the -35 region of *galE2* (the first gene in the galactose ABC transporter operon) which is made of two direct repeat sequences, CAAAAAAGT, separated by 11 nucleotides. A binding upstream of the -35 region usually enables a direct interaction of the bound activator with the alpha carboxy-terminal domain of the corresponding RNA polymerase (26). Direct repeats are common binding sites for activators (38,39), and in *G. stearothermophilus* such sequences are found in the upstream region of *xynE* (a part of the xylotriose ABC transporter) and *araE* (a part of the arabinose ABC transporter) (2,9).

Three-component sensing systems for mono-sugars is likely to provide the bacterium with a highly sensitive and rapidly responding mechanism for detecting and utilizing potential high-molecular-weight polysaccharides in the extracellular environment. Such sensing system was recently characterized in *Clostridium Beijerinckii*, dedicated for the utilization of xylose (40). A somewhat different mechanism for sensing was found in *Clostridium thermocellum* for the regulation of cellulosomal genes. This mechanism is based on a transmembranal RsgI-like protein, which is made of an extracellular carbohydrate binding module (CBM) and an intracellular anti-sigma peptide. In such systems, the binding of the target saccharide to the CBM sensing domain leads to the release of the corresponding alternative sigma factor, thereby enabling the expression of the corresponding genes (41,42).

The 6-Phospho-galactose metabolism in G. stearothermophilus T-1. In principle, 6-phospho-galactose and 6-phospo-lactose can be metabolized by the tagatose-6-phosphate (Tag6P) pathway, while galactose can be also metabolized *via* the Leloir pathway (43). In *G. stearothermophilus* T-1 D-galactose-6-phosphate is formed after the transport of lactose or galactose *via* the corresponding PTS systems (Figure 8). D-galactose-6-phosphate is catabolized by three enzymes of the Tag6P pathway, resulting in dihydroxyacetone phosphate (DHAP) and D-glyceraldehyde 3-phosphate (G3P) (Figure 8). Using sequence alignment, we identified in *G. stearothermophilus* T-1, homologous genes for these three enzymes, tagatose isomerase, tagatose 6-P kinase and D-tagatose-1,6-bisphosphate aldolase. Interestingly, these three genes are not clustered together, as common in other Gram-positive bacteria (44,45). The putative tagatose aldolase gene from strain T-1 shares 40% identity with the tagatose-bisphosphate aldolase gene (*gatY*), which in *Bacillus licheniformis* is part of a five-gene cluster involved in the D-tagatose pathway (46). Although such bioinformatics analysis suggests a similar tagatose pathway in *Geobacillus*, we cannot rule out the involvement of a phosphatase, which, in principle, could alternatively catalyze the hydrolysis of Gal6P into galactose and phosphate (47).

Regulation of the galactose related genes. In *G. stearothermophilus*, regulation of the galactose utilization related genes appears to involve several regulatory mechanisms. As noted above, the galactose-PTS system is regulated by the anti-terminator transcriptional protein GlcT, and the expression of the ABC galactose transporter (*galE2G2F2*) is positively regulated by the response regulator GalT2. Surprisingly, the sugar binding protein of the galactose ABC transport system (GalE2) binds with similar affinities to both galactose and glucose, as was also observed for GalP, the SBP component of the galactose sensing system. These observations could suggest that glucose competes with galactose over the binding to both GalP and GalE2, thereby preventing the expression of galactose-related genes when the glucose concentration is relatively high. Such glucose-galactose competition will ensure that the activation of the galactose utilization system, occurs only when the glucose levels are relatively low, following a mechanism similar to “inducer exclusion” (48). In *E. coli*, the non-phosphorylated form of the EIIA protein, a part of the glucose PTS system, is known to inhibit sugar transporters, exerting the known “glucose effect” (48). In support for this hypothesis is the absence of catabolic responsive elements in the *ganE2* promoter, which usually involve the binding of the catabolite control protein A (CcpA). CcpA is known to be a major regulator of carbon metabolism in Gram-positive bacteria (49,50).

The regulation of the *gan1D* gene probably involves the negative regulator GanR. We have previously characterized the galactan utilization

cluster (*ganEFGBA*). This operon is regulated by GanR, which binds to an invert repeat upstream the -35 site (3). Galacto-oligosaccharides, although not galactose, function as the molecular inducers. Very similar binding sequence is located in the *gan1D* promoter region, suggesting that *gan1D* may also be regulated by GanR. Interestingly, almost the same sequence appears in the promoter region of the galactose metabolism operon, *galKETR* (the Leloir pathway), which is most likely regulated by a different negative regulator, GalR.

Gan1D is a bifunctional 6-phospho-β-galactosidase/glucosidase. In the GH1 family, two catalytic specificities towards phosphorylated substrates have been reported so far, those of 6-phospho-glucosidase and 6-phospho-galactosidase. Glucose-6-phosphate (Glc-6P) and Galactose-6-phosphate (Gal-6P) differ only by the exact configuration of their C4 atom, where the O4 hydroxyl group adopts an axial position in Gal-6P, and an equatorial position in Glc6P. Surprisingly, Gan1D exhibited relatively similar catalytic activities towards sugar substrates with Glc-6P or Gal-6P in their glycone moiety. From the genomic context, however, it was expected that Gan1D would function only as a 6-phospho-β-galactosidase. These results appear to be somewhat less surprising in view of the catalytic specificity data reported for homologous GH1 6-phospho-β-galactosidases, many of which were also shown to be less specific in regards to the glycone sugar of their substrate, and specifically with regards to the configuration of the OH group at the C4 position of this sugar (e.g. glucose *vs*. galactose) [21, 23].

The natural biological role of Gan1D. 6-Phospho-β-galactosidases are relatively common in lactic bacteria, used for the utilization of lactose, which is imported into the cell *via* a lactose-PTS system (51). In the case of *G. stearothermophilus*, however, the fact that strain T-1 encodes for an enzyme **(**Gan1D**)** with 6-phospho-β-galactosidase activity, but lacks a dedicated lactose-PTS system seemed to be enigmatic. Nevertheless, we were surprised to find out that strain T-1 can utilize lactose and galactosyl-glycerol in the presence of cellobiose. These unexpected observations suggested that the cellobiose-PTS system could also transfer lactose and galactosyl-glycerol. Growth curves of strain T-1, growing in presence of cellobiose together with either lactose or galactosyl-glycerol showed an initial logarithmic growth followed by a linear growth (Figure 7). The slope curves of the linear growth were proportional to the turbidity (cell mass) at end of the logarithmic phase. These results suggest that cellobiose is utilized first, supporting logarithmic growth, and when depleted, either lactose or galactosyl-glycerol can enter the cell *via* the same cellobiose-PTS system, although at a somewhat lower affinity. The linear growth is likely to reflect the fact that the transfer of lactose (or galactosyl-glycerol) is a rate limiting step, and that in the absence of cellobiose, the cellobiose-PTS operon is not expressed. With this in mind, it seems that Gan1D could be involved in two important functions in *G. stearothermophilus* T-1. First, it allows the cell to utilize residual galactose containing disaccharides, especially the more abundant galactosyl-glycerol, taking advantage of the relative promiscuity of the cellobiose-PTS system. Considering the competition on the scarce carbon sources in the soil, this could represent a significant growth advantage. Second, it is also possible that the promiscuous cellobiose-PTS system allows the entrance of other non-metabolizable galactose-based compounds, which may accumulate in the cell and inhibit growth. In such cases, Gan1D can potentially hydrolyze these compounds to facilitate their removal. In this respect, the *E. coli lacA* gene product, thiogalactoside transacetylase, could have a somewhat similar function (52) as suggested here for Gan1D.

Experimental procedures

Bacterial growth conditions

Growth media for *G. stearothermophilus* was modified Basic Salt Medium (mBSM) (53) supplemented with 1% of carbon source. Liquid mBSM contained the following per liter: KH_2PO_4, 0.4 g; $MgSO_4.7H_2O$, 0.1 g; $(NH_4)_2SO_4$, 2 g; [MOPS 3-(N-morpholino) propanesulfonic acid]-Sodium salt, adjusted to pH 7.0, 10 g. After autoclave, mBSM was supplemented with (per liter); nitrilotriacetic acid, 0.2 g, NaOH, 0.125 g, 1 ml of amino acids stock solution (0.05 g/lit of 19 amino acids except tyrosine), and 1 ml of three different trace elements solutions. Solution I contained per liter: $MgSO_4{\cdot}7H_2O$, 0.145 g; Solution II contained per liter: $CaCl_2{\cdot}2H_2O$, 0.132g; and

Solution III contained in g per liter: $FeSO_4 \cdot 7H_2O$, 0.998 g; $MnSO_4 \cdot 4H_2O$, 0.592 g; $ZnCl_2$, 0.42 g; and $CuSO_4 \cdot 5H_2O$, 0.624g. The pH of the latter solution was adjusted to 2.0 with sulfuric acid.

DNA and RNA isolation and manipulation

G. stearothermophilus T-1 genomic DNA was isolated according to the procedure published by Marmur (54) as outlined by Johnson (55). Plasmid DNA was purified using the DNA Clean-Up System (Promega). DNA was manipulated by standard procedures (56) Total RNA was isolated with the RNeasy kit (Qiagen GmbH, Hilden, Germany) according to the manufacturer's protocol.

Real time RT-PCR analysis

Total RNA was isolated from mid-exponential culture of *G. stearothermophilus* T-1 grown on modified basic salt medium (mBSM) supplemented with 0.5% (wt/vol) of different sugars as the sole carbon source (53). Cell pellets of about 10^9 cells were suspended in 1 ml of Tri reagent (Sigma), sonicated (Sonicator model W-375; Heat System-Ultrasonics Inc., Plainview, NY), and total RNA was extracted according to the Tri reagent protocol. Contaminating genomic DNA was removed by DNase I treatment (Qiagen, Hilden, Germany) followed by a clean-up protocol with the RNeasy-Kit (Qiagen, Hilden, Germany). Reverse transcription (RT) of RNA (1 µg of total RNA) was performed with the qScript cDNA Synthesis kit (Quanta BioSciences, Beverly, MD, USA) following the manufacturer's instructions. The primers were designed to generate amplicons ranging in size from 90 to 200 bp and control reactions were performed in the absence of reverse transcriptase. The Reactions were performed using the Applied Biosystems 7300 Real Time PCR system and contained (20 µl volume) template cDNA, reverse and forward primers (10 µM each) and PerfeCTa SYBR green Fastmix (Quanta BioSciences, Beverly, MD, USA). All measurements were performed in triplicate and Data analysis was performed with the Applied Biosystems 7300 system software by using the housekeeping gene *ict* (encoding for isocitrate dehydrogenase) for normalization.

Synthesis of 6-Phospho-β-glycosides

The synthetic substrate *o*NP-β-D-galactopyranoside-6-phosphate (*o*NPβGal-6P) was purchased from Carbosynth (Berkshire, UK). The phosphorylated forms of *p*NP-β-D-glucopyranoside, *o*NP-β-D-glucopyranoside, cellobiose, lactose and gentiobiose were made enzymatically using β-glucosidase kinase (BglK) from *Klebsiella pneumonia* as described previously with minor modifications (57). The phosphorylation reaction contained 12.5 mM HEPES buffer pH 7.5, 1 mM $MgSO_4$, 50 mM ATP (adjusted to pH 7.5), 50 mM substrate and 25 µg/ml BglK. During the 2 hr incubation at room temperature, the pH of the reaction was maintained at 7.5 by periodic addition of 40 µl 3 M NH_4OH solution. Following the incubation, the pH of the solution was adjusted to 8.2 and barium acetate was gradually added to a final concentration of 0.6 M. The sample was centrifuged and the soluble fraction (containing the phosphorylated sugar) was filtrated, chilled on ice, and incubated at 4 ºC overnight with 4 volumes of absolute ethanol. Following centrifugation, the pellet was dried at 37 ºC, resuspended in 4 ml aqueous solution of AG 50W-X2 Resin (Bio-Rad, Hercules, CA, USA), filtrated *via* 0.45 µm-pore-size filters (Merck Millipore, Darmstadt, Germany) and lyophilized. Typically, a total yield of 100-150 mg phosphorylated substrate was obtained in these reactions. The structure and purity of *o*NPβGlu-6-phosphate (*o*NPβGlu-6P), *p*NPβGlu-6-phosphate (*p*NPβGlu-6P), cellobiose-6-phosphate (Cel-6P), lactose-6-phosphate (Lac-6P) and Gentiobiose-6-phosphate (Gen-6P) were confirmed by NMR and mass spectrometry. ^{1}H-NMR and ^{13}C-NMR spectroscopy was performed in D_2O on Bruker Avance AV-III 400 and AV-III 500 spectrometers. Signal assignments were confirmed by correlation spectroscopy, total correlation spectroscopy and heteronuclear correlation experiments. The mass spectrum was obtained on a Bruker Daltonix Apex 3 mass spectrometer under electrospray ionization (ESI).

Synthesis of galactosyl-glycerol

The target glycerol derivative of galactose (compound **5**) was synthesized from the commercial penta-*O*-acetyl-*β*-D-galactopyranoside **1** and D,L-*α,β*-isopropylideneglycerol **3** in four chemical steps as illustrated in Scheme 1 (58-60). First, treatment of **1** with HBr (33% solution in acetic acid) resulted the α-bromide **2** (60), which then reacted with the compound **3** in the presence of Ag_2CO_3 to afford selectively the β-glycoside **4**. Sequential two-steps deprotection of **4**, removal of acetate esters by treatment with sodium methoxide followed by acid treatment to remove the isopropylidene ketal, resulted the target β-galactoside **5**. The product was characterized by NMR and MS spectral analysis and the data was identical to the previously reported (58,59).

The ^{1}H NMR spectra (including COSY, DEPT, HSQC and HMBC) were recorded on a Bruker Avance™ 400 spectrometer or 600 spectrometer, and chemical shifts reported (in ppm) are relative to internal Me_4Si (d = 0.0) with $CDCl_3$ as the solvent, and to HOD (d = 4.63) with D_2O as the solvent. ^{13}C NMR spectra were recorded on a Bruker Avance™ 500 spectrometer or at 125.8 MHz, and the chemical shifts reported (in ppm) relative to the residual solvent signal for $CDCl_3$ (d = 77.00), or to external sodium 2,2-dimethyl-2-silapentane sulfonate (d = 0.0) for D_2O as the solvent. Mass spectra were obtained either on a MALDI Micromass spectrometer under electron spray ionization (ESI). Reactions were monitored by TLC on Silica Gel 60 F254 (0.25 mm, Merck), and spots were visualized by charring with a yellow solution containing $(NH_4)Mo_7O_{24}{\cdot}4H_2O$ (120 g) and $(NH_4)_2Ce(NO_3)_6$ (5 g) in 10% H_2SO_4 (800 mL). Column chromatography was performed on a Silica Gel 60 (70-230 mesh). All reactions were carried out under an argon atmosphere with anhydrous solvents, unless otherwise noted. Penta-*O*-acetyl-*β*-D-galactopyranoside **1** was purchased from Tokyo Chemical Industry Co., Ltd (TCI, Japan) and D,L-*α,β*-isopropylideneglycerol **3** was from Sigma–Aldrich (USA).

Penta-O-acetyl-β-D-galactopyranoside bromide **(2).** To a solution of compound **1** (15.0 g, 0.038 mol) in 80 mL anhydrous dichloromethane was added at 0 °C a solution of HBr (9.33 g, 0.12 mol, 33% solution in acetic acid). The reaction was stirred at 0 °C for 1 h and then an additional 3 h at 20 °C. Reaction progress was monitored by TLC (ethyl acetate/hexanes, 1:1). When the reaction was complete, the mixture was cooled to 0 °C and diluted with cold dichloromethane, followed by addition of cold solution of $NaHCO_3$ (5% in water). The aqueous layer was washed with 3 portions of cold dichloromethane; the combined extracts were dried with anhydrous $MgSO_4$ and concentrated at reduced pressure. Compound **2** was rapidly decomposed upon the storage and/or on column chromatography on silica gel. Therefore, it was used in the following step without additional purification. R_f 0.67 (ethyl acetate/hexanes, 1:1). MALDI-TOF MS: calculated for $C_{14}H_{19}BrO_9$ *m/e* 412.20; measured 413.0.

1,2-Isopropylidene-3-O-(β-2',3',4',6'-tetra-O-acetyl-galactopyranosyl)-rac-glycerol ***4.*** The mixture of D,L-*α,β*-isopropylideneglycerol **3** (2.25 g, 0.017 mol) in anhydrous dichloromethane (40 mL), Ag_2CO_3 (4.69 g, 0.017 mol), and fine Drierite (8.5 g) was stirred in the dark under argon. After being stirred during 40 min, iodine (0.43 g, 0.0017 mol) was added, followed by dropwise addition of the bromide **2** (7.0 g, 0.017 mol) in anhydrous dichloromethane (30 mL) during ~1 h at room temperature. The reaction mixture was stirred overnight at room temperature, and then filtered. The filtrate was evaporated and the residue (orange oil) was washed twice with anhydrous dichloromethane and the combined fractions were evaporated to dryness. The crude was purified by column chromatography on silica gel (EtOAc/hexanes 3:2) to afford **4** (3.75 g, 47%) as a mixture (1:1) of two diastereomers. R_f 0.60 (ethyl acetate/hexanes, 7:3).

^{1}H NMR (400 MHz, $CDCl_3$): δ_H 1.31 (s, 3H, isopropylidene-CH_3), 1.37 (s, 3H, isopropylidene-CH_3), 1.92 (s, 3H, acetate-CH_3), 1.98 (s, 3H, acetate-CH_3), 2.00 (s, 3H, acetate-CH_3), 2.09 (s, 3H, acetate-CH_3), 3.55-4.25 m (2 x 8H), 4.48 (d, 1H, *J* = 7.8 Hz), 4.51 (d, 1H, *J*=7.8 Hz), 4.95 (d, 2 x 1H, *J* = 10.2 Hz), 5.13 (t, 2 x 1H, *J* = 8.2 Hz), 5.32 (s, 2 x 1H). MALDI-TOF MS calculated for $C_{20}H_{30}O_{12}Na$ ($[M+Na]^+$) *m/e* 485.45; measured 485.02.

3-O-(β-galactopyranosyl)-rac-glycerol ***5.*** To a solution of compound **4** (4.06 g, 0.009 mol) in anhydrous methanol (50 mL) at 0 °C was added sodium methylate (1.9 g, 0.035 mol) and the reaction progress was monitored by TLC (methanol/chloroform, 1:4). After about 1 hour at 0

°C, the reaction was warmed to room temperature and stirring was continued until its completion. The reaction mixture was loaded on silica gel column and eluted with methanol. Yield: 1.83 g, 71 %, R_f 0.52 (methanol/chloroform, 1:4). ^{1}H NMR (400 MHz, $CDCl_3$): δ_H 1.33 (s, 3H, isopropylidene-CH_3), 1.38 (s, 3H, isopropylidene-CH_3), 3.56 (m, 1H, H-5'), 3.59 (m, 1H, H-3'), 3.62 (m, 1H, H-2'), 3.82 (br s, 1H, H-6'-methylene), 4.02 (br s, 1H, H-4'), 4.24 (d, 1H, H-1', *J*=7.8 Hz), 4.12-4.25 (m, 4H, 2 methylene groups). MALDI-TOF MS calculated for $C_{12}H_{22}O_8$ *m/e* 294.30; measured 295.10.

The observed material from the previous step (1.75 g, 0.006 mol) was stirred in a mixture of acetic acid-water (60 mL, 4:1) at 60 °C during 4.5 h. The reaction progress was monitored by TLC (methanol/chloroform, 1:4). After completion, the reaction mixture was evaporated; the residue was washed 4 times with methanol and the crude was purified by column chromatography on silica gel (chloroform/methanol, 1:1) to afford compound **5** as colorless oil. Yield: 1.16 g, 77 %, R_f 0.53 (chloroform-methanol, 1:1). ^{1}H NMR (600 MHz, MeOD-d_4): δ_H 3.52 (dd, 1H, H-1a, $J_{1a,2}$ = 3.6 Hz, $J_{1a,1b}$= 12.0 Hz), 3.77 (dd, 1H, H-1b, $J_{1b,2}$ = 3.5 Hz, $J_{1b,1a}$ = 12.0 Hz), 3.58 (m, 1H, H-2), 3.47 (dd, 1H, H-3a, $J_{3a,2}$ = 5.2 Hz, $J_{3a,3b}$ = 11.0 Hz), 3.69 (dd, 1H, H-3b, $J_{3b,2}$ = 5.1 Hz, $J_{3b,3a}$ =11.0 Hz), 4.23 (d, 1H, H-1', $J_{1',2'}$ = 7.6 Hz), 3.76 (t, 1H, H-2', $J_{2',3'}$ = 6.4 Hz), 3.55 (dd, H-3', $J_{3',2'}$ = 6.4 Hz, $J_{3',4'}$ = 5.5 Hz), 3.68 (dd, 1H, H-4', $J_{4',3'}$ = 5.5 Hz, $J_{4',5'}$ = 4.2 Hz), 3.72 (ddd, 1H, H-5', $J_{5',4'}$ = 4.2 Hz, $J_{5',6a}$ = 4.7 Hz, $J_{5',6b}$ = 5.5 Hz), 3.81 (m, 2H, H-6'). ^{13}C NMR (600 MHz, MeOD-d_4): δ_C 62.51 (C-6'), 64.07 (C-1), 70.31 (C-3), 72.19 (C-2), 72.42 (C-2'), 72.58 (C-4'), 72.64 (C-3'), 76.75 (C-5'), 105.21 (C-1'). MALDI-TOF MS calculated for $C_9H_{18}O_8Na$ ($[M+Na]^+$) *m/e* 277.20; measured 277.10.

Production and purification of His_6-tagged GalP, GanE2, Gan1D, Cel1A and GalT2

The *galP, ganE2, gan1D, celA and galT2* open reading frame were all cloned into pET9d vector (Novagen). The lipoproteins GalP and and GalE2 were cloned without their 27 and 30 N-terminal lipoprotein coding sequences, respectively. All the primers were designed to allow in-frame cloning of the genes into the T7 polymerase expression vector using restriction sites at the 5' terminus and at the 3' terminus (Table 4). The primers contained six histidine codons to provide His_6-fused products (Table 4). Mutagenesis was performed using the QuikChange site directed mutagenesis kit (Stratagene, La Jolla, CA, USA). The mutated genes were sequenced to confirm that only the desired mutations were inserted. For protein production of Gan1D, Cel1A GalP and GalE2, *E. coli* BL21(DE3) cultures containing the appropriate vector (pET9d- Gan1D, - Cel1A, GalP or - GalE2) were grown overnight in terrific broth (56) with kanamycin (25 g/ml) 0.5 liter in 2-liter baffled shake flasks shaken at 230 rpm at 37°C) to a final turbidity at 600 nm of 15 units. The cultures were harvested, resuspended in 30 ml of buffer (20 mM imidazole, 20 mM phosphate buffer, 500 mM NaCl, pH 7.0), disrupted by two passages through homogenizer (Avestin, Emulsiflex), and centrifuged (14,000 × g for 15 min) to obtain soluble extracts. The His-tagged, fused proteins were purified using a 5-ml His-trap column mounted on a ÄKTA-avant FPLC system (GE healthcare). In the case of the purified response regulator GalT2 *E. coli* BL21(DE3) (pET9d-*galT2*) culture was grown in Terrific Broth (TB) with kanamycin (25 μg ml^{-1}) (1×500 ml in 2-L shake flasks) at 37°C without induction, until turbidity reached 3-4 OD_{600}. To improve solubility, growth was then carried out at 18°C in the presence of 0.4 mM IPTG for 16 h to a final turbidity at 600 nm of about 8 to 11 units. The cultures were harvested, resuspended in 30 ml of buffer (20 mM imidazole, 20 mM phosphate buffer, 500 mM NaCl, pH 7.0), disrupted by two passages through homogenizer (Avestin, Emulsiflex), and centrifuged (14,000 × g for 15 min) to obtain soluble extracts. The His-tagged GalT2 was mounted on a ÄKTA-avant FPLC system (GE healthcare), according to the manufacturer's instructions, and isolated using a 5-ml His-trap column (GE healthcare). The purified response regulator, GalT2, was dialyzed overnight against 2 liters of buffer containing 50 mM Tris-HCl (pH 7.0) and 100 mM KCl, followed by the addition of EDTA and glycerol to final concentrations of 1 mM and 10 %, respectively.

Biochemical characterization of Gan1D

The activity of Gan1D towards aryl-phospho-β-D-glycosides was determined spectroscopically at 405

nm, following the release of the *p*-nitrophenyl or *o*-nitrophenyl. The reactions were performed at 40°C in 100 mM citric acid-Na_2HPO_4 buffer pH 6.5, 1 mg/ml BSA (Bovine Serum Albumin) and 0.13 µg/ml Gan1D. Substrate concentrations were in the range of 0-5 mM. The reactions were initialized by adding 20µl of appropriately diluted pre-warmed substrate to 180 µl pre-warmed enzyme solution in a 96-well-plate. The absorbance at 405 nm was measured at 30 sec intervals for 20 min using a Synergy HT Multi-Mode Microplate Reader (Bio-Tek Instruments, Winooski, VT, USA). The molar extinction coefficients of *o*-nitrophenol and *p*-nitrophenol as determined under the described conditions were $\Delta\varepsilon$=0.63 mM^{-1} cm^{-1} and $\Delta\varepsilon$=3.03 mM^{-1} cm^{-1}, respectively. All measurements were performed in duplicates and the relative error was less than 5%. The kinetic studies with natural substrates were performed by following the release of glucose. The reactions were performed at 40°C in 100 mM citric acid-Na_2HPO_4 buffer pH 6.5, 1 mg/ml BSA and 0.5 µg/ml Gan1D. Substrate concentrations were in the range of 0-24 mM. The reactions were initialized by adding 20 µl of appropriately diluted pre-warmed substrate to 80 µl pre-warmed enzyme solution in 1.7 ml tubes. After 10 min the reactions were terminated by the addition of 25 µl of 0.2 M Na_2CO_3. The amount of glucose released was measured using a high-performance-anion-exchange-chromatography (HPAEC) system equipped with a PA1 column (Dionex, Sunnyvale, CA, USA). The elution was carried out using two buffer eluents, eluent A (150 mM NaOH) and eluent B (150 mM NaOH and 500 mM sodium acetate). The gradient was programmed as follows: 0–1 min, isocratic elution using 100% A and 0% B; 1–50 min, linear gradient to 0% A and 100% B. The elution rate was 1 ml/min, where glucose and galactose (of known concentrations) were used as standards.

Isothermal titration calorimetry (ITC*)*

ITC measurements were performed at 40°C with a MicroCal iTC200 titration calorimeter (Malvern Inst., Worcestershire, UK). Protein solutions were dialyzed overnight against McIlvaine buffer (100 mM citric acid-Na_2HPO_4, pH 6.5). Ligand solutions of galactose, glucose, cellobiose, and lactose were prepared by dilution with the protein dialysis buffer. Ligand solution aliquots (2-4 µl), at least 10 times the molar concentration of the protein, were added by a 40 µl rotating stirrer-syringe to a reaction cell containing 200 µl of 0.04 mM protein solution. Separate titrations of the ligand into the buffer solution determined that the heat of dilution was negligible. Calorimetric data analysis was carried out with the Origin 7.0 software (OriginLab).

Mobility-shift DNA-binding assays

To map the DNA region to which GalT2 is bind, fluorescently labeled double-stranded DNA probes were generated via PCR, using Cy5-labeled primers (Table 1). Binding was performed in binding reaction mixture (30-µl total volume) containing 20 µl of solution comprised of 50 mM Tris-Cl [pH 7.5], 100 mM KCl, 10% glycerol, 1 mM EDTA, 2 µg of salmon sperm DNA, 0.66 mM dithiothreitol, 33 µg of bovine serum albumin, 0.1 pmol of labeled probe and N-terminus His6-GalT2. The binding mixture was incubated for 30 min at 45°C and then separated on a 6.6% nondenaturing polyacrylamide gel prepared in Tris-borate-EDTA buffer that run for 1 h. Fluorescence was detected directly on gel using a Fusion FX system (Vilber).

Phosphorylation of His_6-GalT2

Phosphorylation of His_6-GalT2 was accomplished by following the protocol of Lukat et al. (37). The GalT2 protein (250 µg) was incubated for 30 min at RT in 0.2 ml of a buffer containing 100 mM Tris-HCl (pH 7.0), 8.5 mM $MgCl_2$ and 50 mM acetyl phosphate (Sigma). The phosphorylated form of GalT2 (dimer in solution) was separated by size exclusion using an ÄKTA-avant system (GE healthcare) equipped with a Superose 12 HR gel-filtration column (GE Healthcare Life Sciences) of 24 ml total column volume. Protein samples (0.2 µl) were applied onto the column and eluted at room temperature with a solution consisting of 50 mM Tris-HCl buffer pH 7.0, 100 mM NaCl and 0.02% sodium azide, at a flow rate of 0.5 ml min^{-1}. Molecular weights were determined from regression analysis of the log relative molecular weight (Mr) of protein standards. The protein standards used were Thyroglobulin (Mr, 669,000), Ferritin (Mr, 440,000), Aldolase (Mr, 158,000), Conalbumin

(Mr, 75,000) and Ovalbumin (Mr, 44,000) (GH Healthcare).

Growth of *G. stearothermophilus* T-1 on lactose, and galactosyl-glycerol

G. stearothermophilus T-1 was grown on modified basic salt medium (mBSM) (53) supplemented with 0.4% (wt/vol) of cellobiose as the sole carbon source. The cellobiose-induced logarithmic culture that reached 0.4 OD_{600} was washed twice with mBSM and then diluted 10-fold in 96-well plate containing mBSM supplemented, for example with 0.05% cellobiose and 0.4% lactose. The plate (200 µl total volume per well) was incubated with shaking at 60˚C, and monitored for cell growth (OD_{600}) in a microplate spectrophotometer (Epoch, BioTek).

Data Availability

All data presented are contained in the manuscript. The 14,067-bp sequence containing the galactose utilization region, the cellobiose PTS system, and the galactose PTS system from *G. stearothermophilus* T-1 have been deposited in the GenBank™ under accession numbers KF840174.1, MT424749 and MT441723, respectively.

Acknowledgments
We thank Steve Withers for the *Klebsiella pneumonia* β-glucosidase kinase (BglK) clone. SL is grateful to the Azrieli Foundation for the award of an Azrieli Fellowship. YS holds the Erwin and Rosl Pollak Chair in Biotechnology at the Technion.

Funding and additional information

This work was supported by the United States-Israel Binational Science Foundation (BSF) Jerusalem, Israel (grant No. 96-178 to YS); the Israel Science Foundation Grants 500/10 and 152/11 (to YS) and 1905/15 (to GS); the Israel Ministry of Science and Technology Grant 3-12484 (to YS and GS); the I-CORE Program of the Planning and Budgeting Committee; the Ministry of Environmental Protection and the Grand Technion Energy Program (GTEP), and comprises part of The Leona M. and Harry B. Helmsley Charitable Trust Reports on Alternative Energy series of the Technion, Israel Institute of Technology and the Weizmann Institute of Science. YS acknowledges partial support by the Russell Berrie Nanotechnology Institute and The Lorry I. Lokey Interdisciplinary Center for Life Science and Engineering, Technion.

Conflict of Interest: The authors declare no conflicts of interest in regards to this manuscript.

References

1. Shulami, S., Gat, O., Sonenshein, A. L., and Shoham, Y. (1999) The glucuronic acid utilization gene cluster from *Bacillus stearothermophilus* T-6. *J Bacteriol* **181**, 3695-3704
2. Shulami, S., Raz-Pasteur, A., Tabachnikov, O., Gilead-Gropper, S., Shner, I., and Shoham, Y. (2011) The L-Arabinan utilization system of *Geobacillus stearothermophilus*. *J Bacteriol* **193**, 2838-2850
3. Tabachnikov, O., and Shoham, Y. (2013) Functional characterization of the galactan utilization system of *Geobacillus stearothermophilus*. *FEBS J* **280**, 950-964
4. Bravman, T., Mechaly, A., Shulami, S., Belakhov, V., Baasov, T., Shoham, G., and Shoham, Y. (2001) Glutamic acid 160 is the acid-base catalyst of beta-xylosidase from *Bacillus stearothermophilus* T-6: a family 39 glycoside hydrolase. *FEBS Lett* **495**, 115-119
5. Zaide, G., Shallom, D., Shulami, S., Zolotnitsky, G., Golan, G., Baasov, T., Shoham, G., and Shoham, Y. (2001) Biochemical characterization and identification of catalytic residues in alpha-glucuronidase from *Bacillus stearothermophilus* T-6. *Eur J Biochem* **268**, 3006-3016
6. Shallom, D., Belakhov, V., Solomon, D., Shoham, G., Baasov, T., and Shoham, Y. (2002) Detailed kinetic analysis and identification of the nucleophile in alpha-L-arabinofuranosidase from *Geobacillus stearothermophilus* T-6, a family 51 glycoside hydrolase. *J Biol Chem* **277**, 43667-43673
7. Hovel, K., Shallom, D., Niefind, K., Belakhov, V., Shoham, G., Baasov, T., Shoham, Y., and Schomburg, D. (2003) Crystal structure and snapshots along the reaction pathway of a family 51 alpha-L-arabinofuranosidase. *EMBO J* **22**, 4922-4932
8. Golan, G., Shallom, D., Teplitsky, A., Zaide, G., Shulami, S., Baasov, T., Stojanoff, V., Thompson, A., Shoham, Y., and Shoham, G. (2004) Crystal structures of

Geobacillus stearothermophilus alpha-glucuronidase complexed with its substrate and products: mechanistic implications. *J Biol Chem* **279**, 3014-3024

9. Shulami, S., Zaide, G., Zolotnitsky, G., Langut, Y., Feld, G., Sonenshein, A. L., and Shoham, Y. (2007) A two-component system regulates the expression of an ABC transporter for xylo-oligosaccharides in *Geobacillus stearothermophilus*. *Appl Environ Microbiol* **73**, 874-884
10. Teplitsky, A., Mechaly, A., Stojanoff, V., Sainz, G., Golan, G., Feinberg, H., Gilboa, R., Reiland, V., Zolotnitsky, G., Shallom, D., Thompson, A., Shoham, Y., and Shoham, G. (2004) Structure determination of the extracellular xylanase from *Geobacillus stearothermophilus* by selenomethionyl MAD phasing. *Acta Crystallogr D Biol Crystallogr* **60**, 836-848
11. Czjzek, M., Ben David, A., Bravman, T., Shoham, G., Henrissat, B., and Shoham, Y. (2005) Enzyme-substrate complex structures of a GH39 beta-xylosidase from *Geobacillus stearothermophilus*. *J Mol Biol* **353**, 838-846
12. Solomon, V., Teplitsky, A., Shulami, S., Zolotnitsky, G., Shoham, Y., and Shoham, G. (2007) Structure-specificity relationships of an intracellular xylanase from *Geobacillus stearothermophilus*. *Acta Crystallogr D Biol Crystallogr* **63**, 845-859
13. Alalouf, O., Balazs, Y., Volkinshtein, M., Grimpel, Y., Shoham, G., and Shoham, Y. (2011) A new family of carbohydrate esterases is represented by a GDSL hydrolase/acetylxylan esterase from *Geobacillus stearothermophilus*. *J Biol Chem* **286**, 41993-42001
14. Salama, R., Alalouf, O., Tabachnikov, O., Zolotnitsky, G., Shoham, G., and Shoham, Y. (2012) The *abp* gene in *Geobacillus stearothermophilus* T-6 encodes a GH27 beta-L-arabinopyranosidase. *FEBS Lett* **586**, 2436-2442
15. Lansky, S., Salama, R., Solomon, H. V., Feinberg, H., Belrhali, H., Shoham, Y., and Shoham, G. (2014) Structure-specificity relationships in Abp, a GH27 beta-L-arabinopyranosidase from *Geobacillus stearothermophilus* T6. *Acta Crystallogr D Biol Crystallogr* **70**, 2994-3012
16. Solomon, H. V., Tabachnikov, O., Lansky, S., Salama, R., Feinberg, H., Shoham, Y., and Shoham, G. (2015) Structure-function relationships in Gan42B, an intracellular GH42 beta-galactosidase from *Geobacillus stearothermophilus*. *Acta Crystallogr D Biol Crystallogr* **71**, 2433-2448
17. Cuskin, F., Lowe, E. C., Temple, M. J., Zhu, Y., Cameron, E., Pudlo, N. A., Porter, N. T., Urs, K., Thompson, A. J., Cartmell, A., Rogowski, A., Hamilton, B. S., Chen, R., Tolbert, T. J., Piens, K., Bracke, D., Vervecken, W., Hakki, Z., Speciale, G., Munoz-Munoz, J. L., Day, A., Pena, M. J., McLean, R., Suits, M. D., Boraston, A. B., Atherly, T., Ziemer, C. J., Williams, S. J., Davies, G. J., Abbott, D. W., Martens, E. C., and Gilbert, H. J. (2015) Human gut *Bacteroidetes* can utilize yeast mannan through a selfish mechanism. *Nature* **517**, 165-169
18. Saier, M. H., Jr. (2015) The Bacterial Phosphotransferase System: New Frontiers 50 Years after Its Discovery. *J Mol Microbiol Biotechnol* **25**, 73-78
19. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**, D490-495
20. Yu, W. L., Jiang, Y. L., Pikis, A., Cheng, W., Bai, X. H., Ren, Y. M., Thompson, J., Zhou, C. Z., and Chen, Y. (2013) Structural insights into the substrate specificity of a

6-phospho-beta-glucosidase BglA-2 from *Streptococcus pneumoniae* TIGR4. *J Biol Chem* **288**, 14949-14958
21. Shallom, D., Leon, M., Bravman, T., Ben-David, A., Zaide, G., Belakhov, V., Shoham, G., Schomburg, D., Baasov, T., and Shoham, Y. (2005) Biochemical characterization and identification of the catalytic residues of a family 43 beta-D-xylosidase from *Geobacillus stearothermophilus* T-6. *Biochemistry* **44**, 387-397
22. West, A. H., and Stock, A. M. (2001) Histidine kinases and response regulator proteins in two-component signaling systems. *Trends Biochem Sci* **26**, 369-376
23. Gallegos, M. T., Schleif, R., Bairoch, A., Hofmann, K., and Ramos, J. L. (1997) Arac/XylS family of transcriptional regulators. *Microbiol Mol Biol Rev* **61**, 393-410
24. Jarmer, H., Larsen, T. S., Krogh, A., Saxild, H. H., Brunak, S., and Knudsen, S. (2001) Sigma A recognition sites in the *Bacillus subtilis* genome. *Microbiology* **147**, 2417-2424
25. Rhodius, V. A., and Busby, S. J. (1998) Positive activation of gene expression. *Curr Opin Microbiol* **1**, 152-159
26. Barnard, A., Wolfe, A., and Busby, S. (2004) Regulation at complex bacterial promoters: how bacteria use different promoter organizations to produce different regulatory outcomes. *Curr Opin Microbiol* **7**, 102-108
27. Buschiazzo, A., and Trajtenberg, F. (2019) Two-Component Sensing and Regulation: How Do Histidine Kinases Talk with Response Regulators at the Molecular Level? *Annu Rev Microbiol* **73**, 507-528
28. Kenney, L. J. (2000) Response-regulator phosphorylation and activation: a two-way street? response. *Trends Microbiol* **8**, 155-156
29. Bogel, G., Schrempf, H., and Ortiz de Orue Lucana, D. (2007) DNA-binding characteristics of the regulator SenR in response to phosphorylation by the sensor histidine autokinase SenS from *Streptomyces reticuli*. *FEBS J* **274**, 3900-3913
30. Lansky, S., Zehavi, A., Belrhali, H., Shoham, Y., and Shoham, G. (2017) Structural basis for enzyme bifunctionality - the case of Gan1D from *Geobacillus stearothermophilus*. *FEBS J* **284**, 3931-3953
31. Schulte, D., and Hengstenberg, W. (2000) Engineering the active center of the 6-phospho-beta-galactosidase from *Lactococcus lactis*. *Protein Eng* **13**, 515-518
32. Sastry, P. S. (1974) Glycosyl glycerides. *Adv Lipid Res* **12**, 251-310
33. Gounaris, K., and Barber, J. (1983) Monogalactosyldiacylglycerol: The most abundant polar lipid in nature. *Trends in Biochemical Sciences* **8**, 378-381
34. Aleksandrzak-Piekarczyk, T., Kok, J., Renault, P., and Bardowski, J. (2005) Alternative lactose catabolic pathway in *Lactococcus lactis* IL1403. *Appl Environ Microbiol* **71**, 6060-6069
35. Knezevic, I., Bachem, S., Sickmann, A., Meyer, H. E., Stulke, J., and Hengstenberg, W. (2000) Regulation of the glucose-specific phosphotransferase system (PTS) of *Staphylococcus carnosus* by the antiterminator protein GlcT. *Microbiology* **146 (Pt 9)**, 2333-2342
36. Himmel, S., Grosse, C., Wolff, S., Schwiegk, C., and Becker, S. (2012) Structure of the RBD-PRDI fragment of the antiterminator protein GlcT. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **68**, 751-756

37. Lukat, G. S., McCleary, W. R., Stock, A. M., and Stock, J. B. (1992) Phosphorylation of bacterial response regulator proteins by low molecular weight phospho-donors. *Proc Natl Acad Sci U S A* **89**, 718-722
38. Blanco, A. G., Sola, M., Gomis-Ruth, F. X., and Coll, M. (2002) Tandem DNA recognition by PhoB, a two-component signal transduction transcriptional activator. *Structure* **10**, 701-713
39. Bordi, C., Ansaldi, M., Gon, S., Jourlin-Castelli, C., Iobbi-Nivol, C., and Méjean, V. (2004) Genes Regulated by TorR, the Trimethylamine Oxide Response Regulator of *Shewanella oneidensis*. *Journal of Bacteriology* **186**, 4502-4509
40. Sun, Z., Chen, Y., Yang, C., Yang, S., Gu, Y., and Jiang, W. (2015) A novel three-component system-based regulatory model for D-xylose sensing and transport in *Clostridium beijerinckii*. *Mol Microbiol* **95**, 576-589
41. Sand, A., Holwerda, E. K., Ruppertsberger, N. M., Maloney, M., Olson, D. G., Nataf, Y., Borovok, I., Sonenshein, A. L., Bayer, E. A., Lamed, R., Lynd, L. R., and Shoham, Y. (2015) Three cellulosomal xylanase genes in *Clostridium thermocellum* are regulated by both vegetative SigA (sigma(A)) and alternative SigI6 (sigma(I6)) factors. *FEBS Lett* **589**, 3133-3140
42. Nataf, Y., Bahari, L., Kahel-Raifer, H., Borovok, I., Lamed, R., Bayer, E. A., Sonenshein, A. L., and Shoham, Y. (2010) *Clostridium thermocellum* cellulosomal genes are regulated by extracytoplasmic polysaccharides via alternative sigma factors. *Proc Natl Acad Sci U S A* **107**, 18646-18651
43. Frey, P. A. (1996) The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose. *FASEB J* **10**, 461-470
44. Rosey, E. L., Oskouian, B., and Stewart, G. C. (1991) Lactose metabolism by *Staphylococcus aureus*: characterization of *lacABC*D, the structural genes of the tagatose 6-phosphate pathway. *J Bacteriol* **173**, 5992-5998
45. Jagusztyn-Krynicka, E. K., Hansen, J. B., Crow, V. L., Thomas, T. D., Honeyman, A. L., and Curtiss, R., 3rd. (1992) *Streptococcus mutans serotype* c tagatose 6-phosphate pathway gene cluster. *J Bacteriol* **174**, 6152-6158
46. Van der Heiden, E., Delmarcelle, M., Lebrun, S., Freichels, R., Brans, A., Vastenavond, C. M., Galleni, M., and Joris, B. (2013) A pathway closely related to the (D)-tagatose pathway of gram-negative enterobacteria identified in the gram-positive bacterium *Bacillus licheniformis*. *Appl Environ Microbiol* **79**, 3511-3515
47. Mokhtari, A., Blancato, V. S., Repizo, G. D., Henry, C., Pikis, A., Bourand, A., de Fatima Alvarez, M., Immel, S., Mechakra-Maza, A., Hartke, A., Thompson, J., Magni, C., and Deutscher, J. (2013) *Enterococcus faecalis* utilizes maltose by connecting two incompatible metabolic routes via a novel maltose 6'-phosphate phosphatase (MapP). *Mol Microbiol* **88**, 234-253
48. Saier, M. H., Jr., and Crasnier, M. (1996) Inducer exclusion and the regulation of sugar transport. *Res Microbiol* **147**, 482-489
49. Deutscher, J. (2008) The mechanisms of carbon catabolite repression in bacteria. *Curr Opin Microbiol* **11**, 87-93
50. Gorke, B., and Stulke, J. (2008) Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol* **6**, 613-624

51. Honeyman, A. L., and Curtiss, R., 3rd. (1993) Isolation, characterization and nucleotide sequence of the *Streptococcus mutans* lactose-specific enzyme II (lacE) gene of the PTS and the phospho-beta-galactosidase (lacG) gene. *J Gen Microbiol* **139**, 2685-2694
52. Danchin, A. (2009) Cells need safety valves. *Bioessays* **31**, 769-773
53. Shulami, S., Shenker, O., Langut, Y., Lavid, N., Gat, O., Zaide, G., Zehavi, A., Sonenshein, A. L., and Shoham, Y. (2014) Multiple regulatory mechanisms control the expression of the *Geobacillus stearothermophilus* gene for extracellular xylanase. *J Biol Chem* **289**, 25957-25975
54. Marmur, J. (1961) A procedure for the isolation of deoxyribonucleic acid from micro-organisms. *J Mol Biol* **3**, 208–218
55. Johnson, J. L. (1981) Manual of Methods for General Bacteriology in (Gerhardt, P., CostilowR.G.E., E.W., N., Wood W.A., N.R., Krieg, R., and Philips, G. B. eds.), AmericanSocietyforMicrobiology,Washington,D.C. . pp 450–472
56. Sambrook, J., and Russell, D. W. (2001) *Molecular cloning: a laboratory manual*, 3rd ed. ed., Cold Spring Harbor,, NY.
57. Thompson, J., Lichtenthaler, F. W., Peters, S., and Pikis, A. (2002) Beta-glucoside kinase (BglK) from *Klebsiella pneumoniae*. Purification, properties, and preparative synthesis of 6-phospho-beta-D-glucosides. *J Biol Chem* **277**, 34310-34321
58. Cateni, F., Bonivento, P., Procida, G., Zacchigna, M., Scialino, G., and Banfi, E. (2007) Chemoenzymatic synthesis and in vitro studies on the hydrolysis of antimicrobial monoglycosyl diglycerides by pancreatic lipase. *Bioorg Med Chem Lett* **17**, 1971-1978
59. Janwitayanuchit, W., Suwanborirux, K., Patarapanich, C., Pummangura, S., Lipipun, V., and Vilaivan, T. (2003) Synthesis and anti-herpes simplex viral activity of monoglycosyl diglycerides. *Phytochemistry* **64**, 1253-1264
60. Mechaly A., Belakhov V., Shoham Y., and Baasov T. (1997) An Efficient Chemical-Enzymatic Synthesis of 4-Nitrophenyl β-Xylobioside - a Chromogenic Substrate for Xylanases. *Carbohydrate Research* **304**, 111-115

Table 1. Michaelis–Menten catalytic constants for the hydrolysis of synthetic and natural substrates by Cel1A.

Substrate	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($s^{-1}\cdot M^{-1}$)
*o*NP-β-D-glucopyranoside-6-p (*o*NPβGlc6P)	106	0.22	4.8×10^5
*o*NP-β-D-galactopyranoside-6-p (*o*NPβGal6P)	ND	ND	ND
Cellobiose-6-phosphate	60	5.1	1.2×10^4
Lactose-6-phosphate	ND	ND	ND
oNP-β-D-glucopyranoside (*o*NPβGlc)	ND	ND	ND

The reactions were performed at 40°C, in 100 mM citric acid-Na_2HPO_4 buffer pH 7.
ND= Non-detectable.

Table 2. Thermodynamic parameters for the binding of potential substrates to GalE2 and GalP.

Protein	Ligand	*n*	K_D (μM)	ΔH_B (kcal·mol^{-1})	$T\Delta S_B$ (kcal·mol^{-1})	ΔG_B (kcal·mol^{-1})
GalE2	Galactose	1.12	1.4	-5.8± 0.3	0.3	-6.1
	Glucose	1.04	3.1	-7.0 ± 0.4	0.8	-7.8
GalP	Galactose	0.65	6.1	-10± 0.6	-2.6	-12.6
	Glucose	0.71	4.2	-9.0 ± 0.2	-1.3	-10.2

Table 3. Michaelis-Menten catalytic constants for the hydrolysis of synthetic and natural substrates by Gan1D[a].

Substrate	k_{cat} **(s^{-1})**	K_m**(mM)**	k_{cat}/K_m**($s^{-1}\cdot M^{-1}$)**
*o*NPβGal-6P	148	0.32	4.9×10^5
*o*NPβGlc-6P	143	0.17	8.4×10^5
*p*NPβGlc-6P	114	0.12	9.3×10^5
Cel-6P	74	0.86	8.6×10^4
Lac-6P	25	1.1	2.3×10^4
Gen-6P	12	2.9	4.1×10^3
*o*NPβGal	ND[b]	ND	ND
*o*NPβGlc	ND	ND	ND

[a] The reactions were performed at 40°C, in 100 mM citric acid-Na_2HPO_4 buffer pH 6.5. [b] ND- Non-detectable. All measurements were performed in duplicate and the relative error was less than 5%.

Table 4. Oligonucleotides used in this study

Primer	Sequence (5'-3') a,b	Application
gan1D N-ter	5'-TATAG**TCATGA**TACATCACCACCACCACCATGAGCATCGTCATCTTAAAC-3'	Cloning into pET9d
gan1D C-ter	5'-TATAC**GGATCC**CTACAGCTCGGCACCGTTC-3'	
celA N-ter	5'-TATAC**CCATGG**GGCATCATCATCACCACCACAAACGATTGAAAATG-3'	
celA C-ter	5'-ATGCG**AGATCT**TTAGTTTCCAGCTACTGTTTTG-3'	
galP N-ter	5'-TATAC**CCATGG**GGCACCATCATCATCATCATGCTTGGAACGTGTATGCGTATCC -3'	
galP C-ter	5'-TATCA**GGATCC** TCATGGCGTGTTGACCTCATAATTGC-3'	
galT2 N-ter	5'-ATATAT**CCATGG**TGCATCATCATCATCATCATAAAGTGGCGATCGTCGAT-3'	
galT2 C-ter	5'-TATATA**GGATCC**TCAGCGCGCCTGTACTTTTTT-3'	
galE2 N-ter	5'-TATCG**CCATGG**GGCATCATCATCATCATCAT GATGACGCCAAACAA-3'	
galE2 C-ter	5'-ATGCG**AGATCT**TACTTCTTCAGTTCTG-3'	
Fwr *gan1D*	5'-GTTCCCGCCTGAGTTTTT-3'	Real Time RT PCR
Rev *gan1D*	5'-CCTTTTCCATCTTCGTTCCA-3'	
Fwr *galE2*	5'- CTCTTCTTCGTCCGATGA-3'	
Rev *galE2*	5'- GTCCACCAGCTGAAAATCT-3'	
Fwr *galP*	5'-CGGCAGGCGAATATTGATG-3'	
Rev *galP*	5'-CAAATTCCGCTTCCGTCAAC-3'	
Fwr *ptsA*	5'-GTTTGGCAACGCCTTGAAA-3'	
Rev *ptsA*	5'-GCCTGCTTGCTCCATCACAT-3'	
Fwr *ptsD*	5'-CGCTCAAAGAGGAAGGAAA-3'	
Rev *ptsD*	5'-CCGATGACGACTTGAAACTG-3'	
Fwr *mltE*	5'-ACGAGACGTTTGCGAAGCT-3'	
Fwr *mltE*	5'-ATTTTGCCGCCCGTATAG-3'	
Fwr *citC*	5'-GACTTGCGGCCGTGTTC-3	
Rev *citC*	5'- GAAGACATCTACGCTGGCATT-3	
Fwr *celB*	5'-ACAAGCTTGCTAGTGACGAA-3'	
Rev *celB*	5'-GTGGGCCAATAAGAATGAC-3'	
Fwr *celC*	5'-GATTCAAGAAGCAGAGCAAG-3'	
Rev *celC*	5'-TGACATGACATGGTCTTCC-3'	
Fwr *celD*	5'-CTAAGCAAAGTTCTCGTTCC-3'	
Rev *celD*	5'-TTTCACTCATCACCTTGTCC-3'	
Fwr *celA*	5'-GGTTTACCCACAATGAACC-3'	
Rev *celA*	5'-CATCGTATGGTAAGCCACTT-3'	
Fwr *celR*	5'-CACCGGTGTACCACTTAATC-3'	
Rev *celR*	5'-GAGTTCGTGAGTGATGTAACG-3'	
Fwr *galE2*	5'-CATAAAAGCGAAATGGCAGTC-3'	DNA-binding assays
Cy5 Rev *galE2*	5'-CTTATCAATCGGCGGCTTG-3'	

[a] Boldface bases indicate engineered restriction sites. [b] Cy5 - Cyanine 5'end-labeled primers

Figures

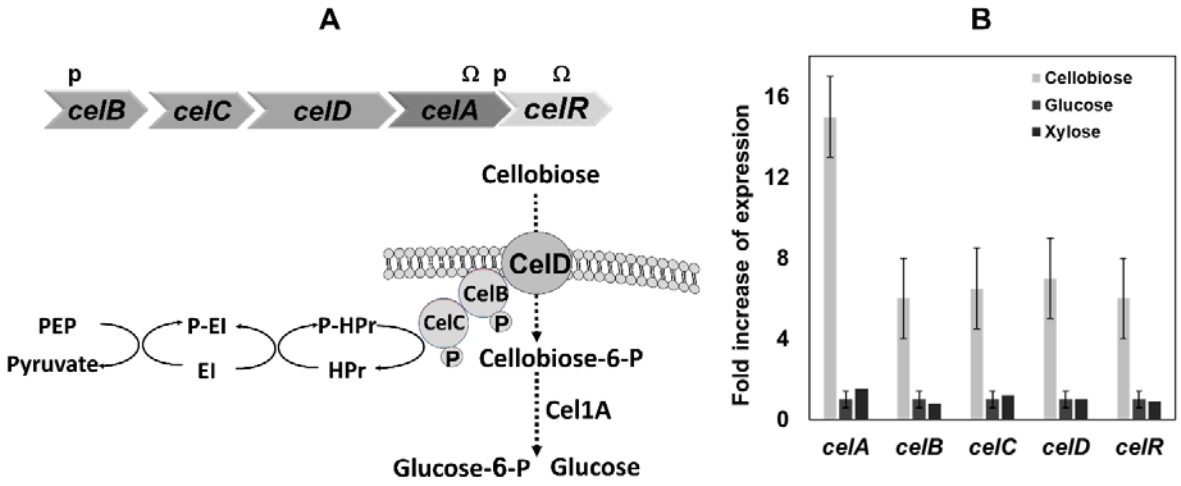


Figure 1. **Schematic representation of the cellobiose PTS system and mRNA expression levels of the related genes. (A)** The gene cluster encodes for the cellobiose PTS enzyme system, *celBCD*, a GH1 6-phospho-glucosidase, *celA*, and a GntR family transcriptional regulator, *celR*. The letter P indicates the proposed promoter site, and Ω indicates the position of the putative rho-independent transcription termination site. **(B)** Real-time RT PCR analysis of the cellobiose utilization gene cluster. Total RNA was extracted from mid-logarithmic phase cultures of *G. stearothermophilus* grown in minimal medium supplemented with either 0.5% cellobiose, xylose or glucose as the sole carbon source. Normalization was performed using the housekeeping gene *ict*, which encodes for isocitrate dehydrogenase. The error bars (in bar charts) are SD values.

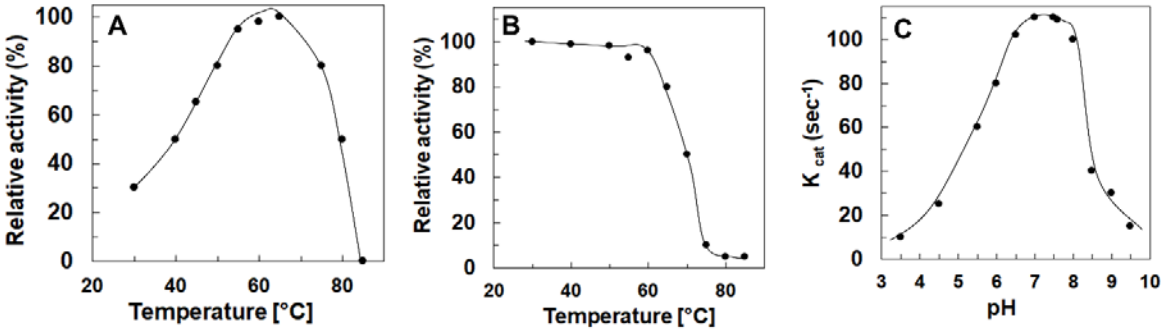


Figure 2. Effect of temperature and pH in a 20 minutes reaction on Cel1A activity toward *o*NPβGlc6P. **(A)** Relative activity at pH 7 at different temperatures. **(B)** Thermostability of Cel1A at pH 7. Residual activity was measured after 10 min incubation at the indicated temperatures. **(C)** Initial rate activity at different pHs. The reactions were performed at 40°C in either sodium phosphate-citric acid buffer (pH3-7.8) or Atkins&Pantin buffer (pH 7.5-9.5).

Figure 3. Schematic representation of the 12.5 kb cluster containing galactose utilization genes. The cluster contains a Tn31 transposon between *gan4C* and *galP*. The letter P indicates the proposed promoter site, and Ω indicates the position of the putative rho-independent transcription termination site.

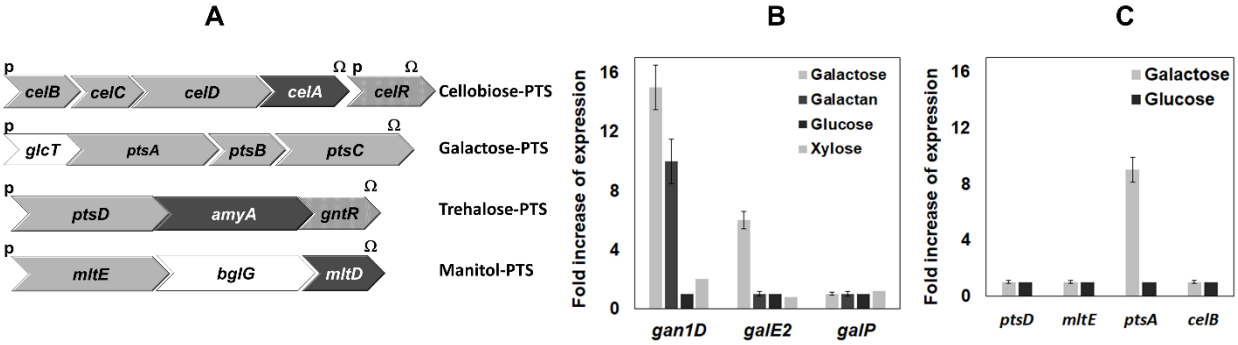


Figure 4. Presentation of four putative PTS systems in strain T-1 and the expression of galactose related genes in the presence of galactose. **(A)** Schematic gene organization of the PTS systems for cellobiose, galactose, trehalose and mannitol. Cellobiose PTS enzyme II complex is encoded by *celB*, *celC*, and *celD*; *ptsABC* ORFs encode for a PTS transporter for galactose; *ptsD* encodes for trehalose PTS EIIBC; and *mltE* encodes for manitol PTS EIIBC. The *celR* and *gntR* genes encode for transcriptional regulators; *glcT* and *blgG* encode for anti-terminator proteins; *celA* encodes for 6-phospho-β-glucosidase; *amyA* encodes for trehalose-6-phosphate hydrolase and *mltD* encodes for mannitol-1-phosphate 5-dehydrogenase **(B)** Real-time RT PCR analysis of the galactose related gene cluster in *G. stearothermophilus* T-1. **(C)** Real-time RT PCR analysis of the four putative PTS systems in strain T-1. Total RNA was extracted from mid-logarithmic phase cultures of *G. stearothermophilus* grown in minimal medium supplemented with different sugars (0.5%) as the sole carbon source. Normalization was performed using the housekeeping gene for isocitrate dehydrogenase, *ict*. The error bars (in bar charts) are SD values

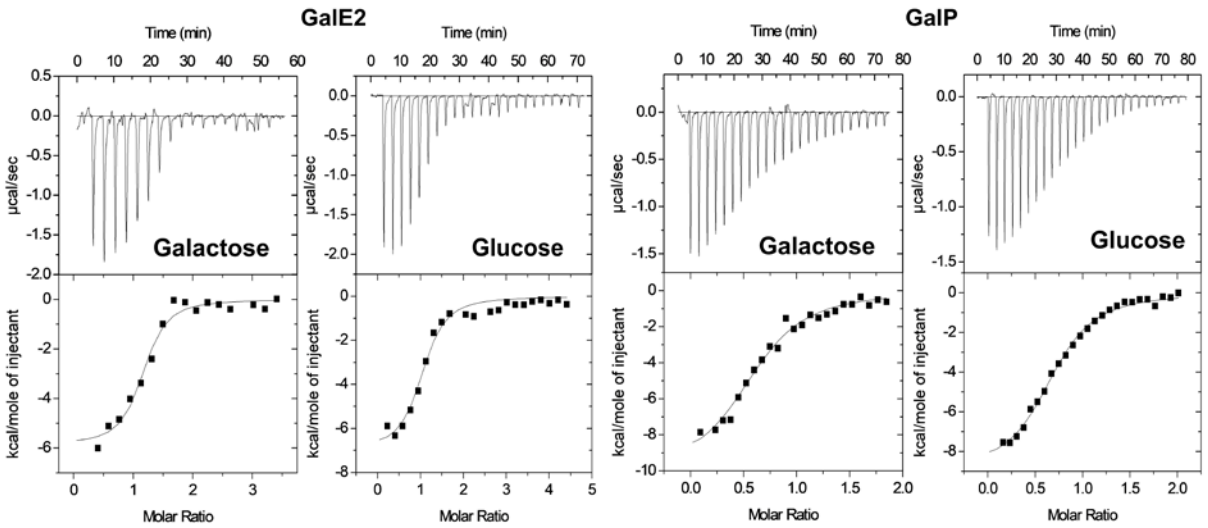


Figure 5. Calorimetric titration of GalE2 and GalP galactose and glucose. The top half of each experiment shows the raw data for calorimetric titration of the protein with the ligand, and the lower half displays the integrated injection heats from the upper panel. The solid line is the curve of best fit to a single binding site (n=1) model that was used to derive the binding parameters. Experiments were performed at 40°C.

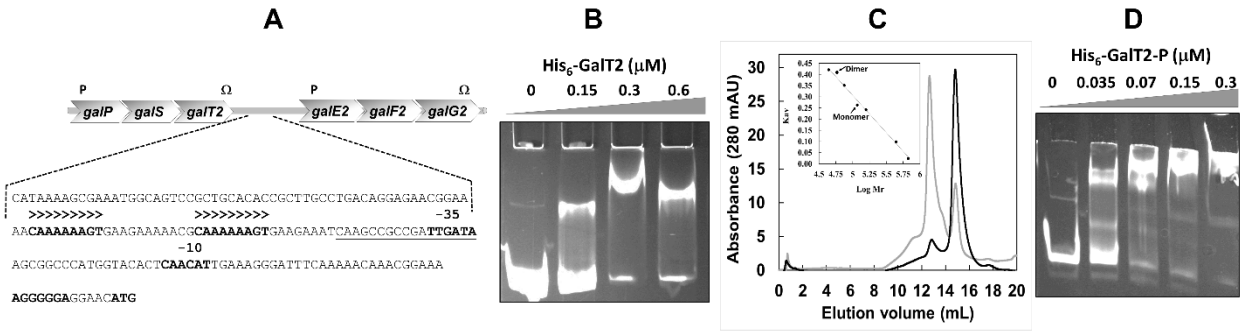


Figure 6. Gel shift assays demonstrating binding of unphosphorylated and phosphorylated His_6-GalT2 to the *galE2* promoter. **(A)** The promoter region of *galE2*. The horizontal arrows indicate repeat binding sites for the response regulator GalT2. The -35 and -10 regions, the ribosome-binding site (RBS), and the ATG initiating codon are in bold. The sequences of the primers used to synthesize a PCR product of 113-bp DNA fragment the for the gel shift assay are underlined. **(B)** Fluorescently labeled 113-bp DNA fragment containing the *galE2* promoter was incubated with increasing amounts of His_6-GalT2. **(C)** Size-exclusion chromatograms (overlaid) of unphosphorylated His_6-GalT2 (solid black line) and phosphorylated His_6-GalT2 (solid gray line). The inset shows the linear regression of the protein standards used for *Mr* determination (filled circles) and the position of the dimeric and monomeric forms of GalT2 on this curve are marked with arrows. **(D)** The fluorescently labeled 113-bp DNA fragment was incubated with increasing amounts of phosphorylated His_6-GalT2.

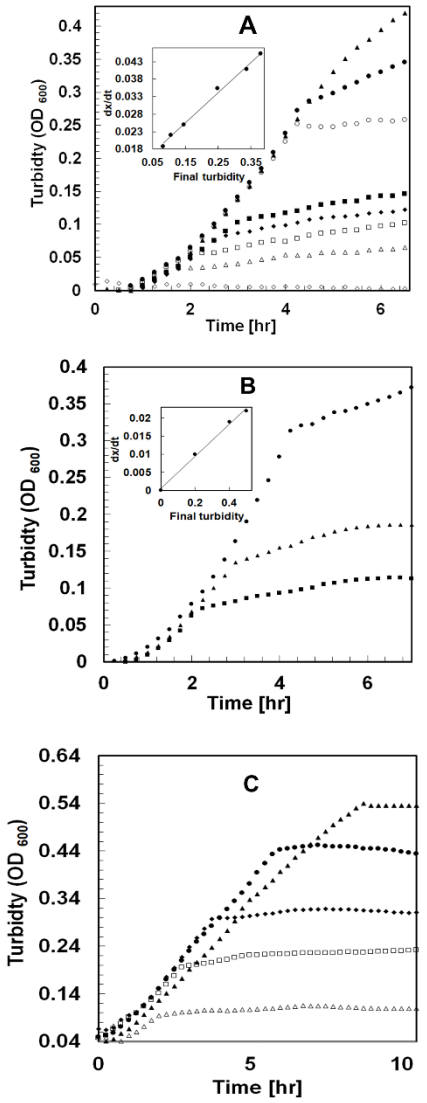


Figure 7. Growth of *G. stearothermophilus strain* T-1 and T-6 on lactose or galactosyl-glycerol in the presence of cellobiose. Cellobiose-induced logarithmic culture at 0.4 OD_{600} was washed twice with mBSM and then diluted 10-fold in 96-well plate containing **(A)** mBSM supplemented with 0.4 % lactose and varoius concentrations of cellobiose. Cellobiose concentrations were: (△) 0.01%, (□) 0.02%, (◆) 0.04%, (■) 0.05%, (●) 0.1 %, (▲) 0.2% cellobiose. mBSM supplemented only with 0.1% cellobiose or 0.4% lactose as a sole carbon source is shown in (○) and (◊), respectively. **(B)** mBSM supplemented with 0.4 % galactosyl-glycerol and varoius concentrations of cellobiose. Cellobiose concentrations were: (■) 0.05%, (▲) 0.1 %, (●) 0.2%. (C) Growth of *G. stearothermophilus* T-6 (lacks the *gan1D* gene) on mBSM supplemented with 0.4 % lactose and varoius concentrations of cellobiose Concentrations of cellobiose were as follow: (△) 0.01%, (□) 0.02%, (◆) 0.04%, (●) 0.1 %, (▲) 0.2%. The path length in the wells was 0.6 cm. The inserts show the growth rate (dx/dt) on lactose or galactosyl-glycerol as function of final cell mass grown on cellobiose.

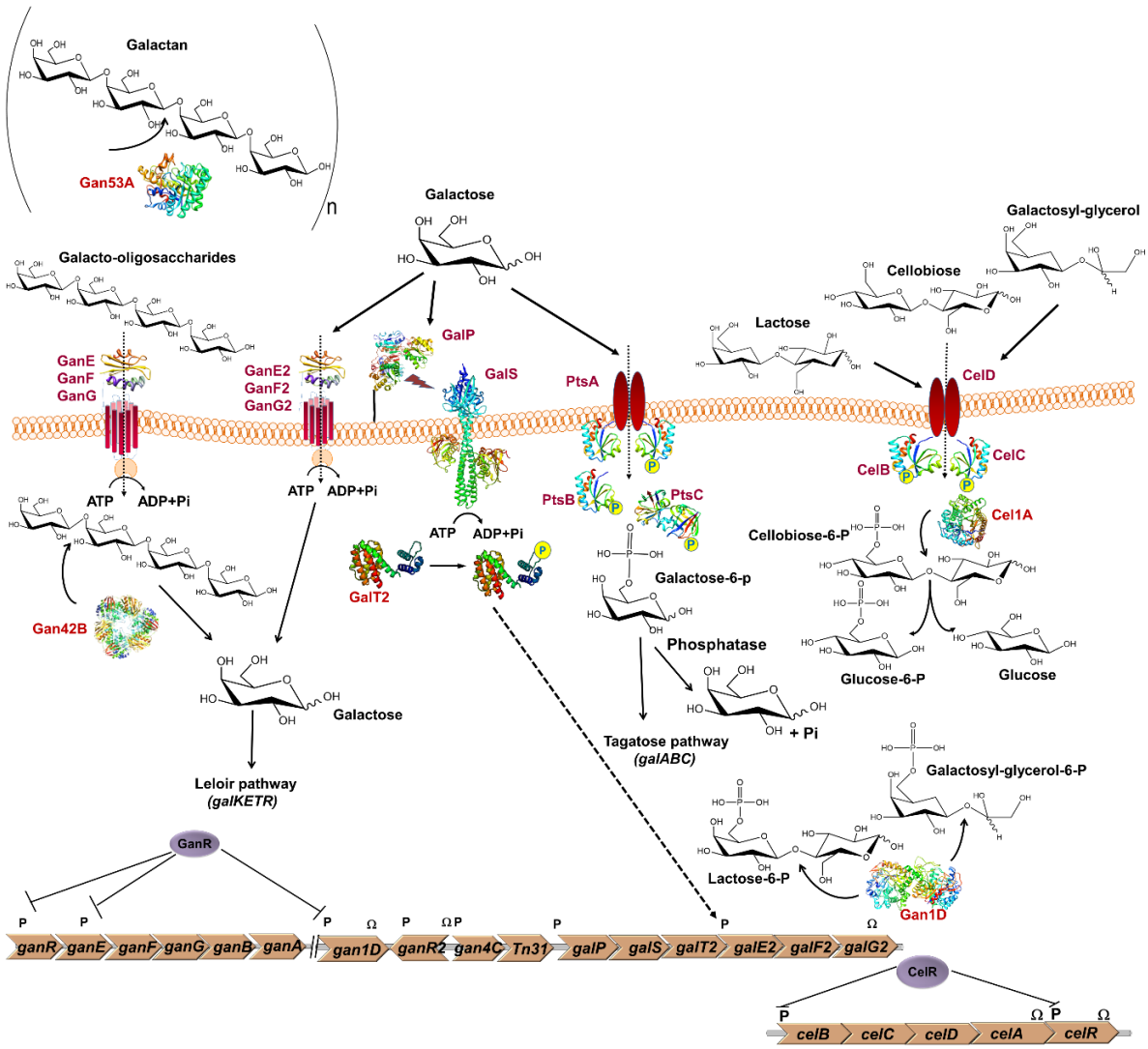
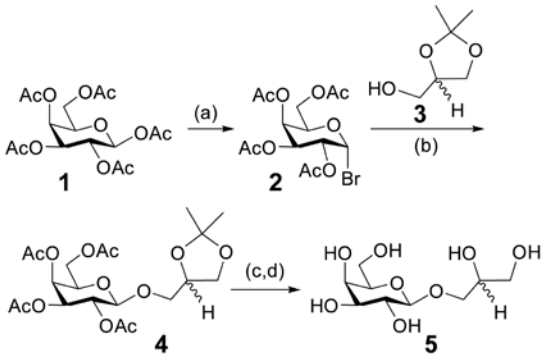


Figure 8. Proposed pathways for the utilization of galactan, galactose, cellobiose, and β-galactosides in *G. stearothermophilus* T-1. The extracellular β-1,4-galactanase Gan53A cleaves galactan and generates short galacto-oligosaccharides which enter the cell *via* a specific ABC transport system (GanEFG), and further degraded by the intracellular β-galactosidase Gan42B into galactose (3). Galactose is metabolized *via* the Leloir pathway involving the *galKETR* operon encoding for galactokinase (*galK*), UDP-glucose 4-epimerase (*galE*), galactose-1-phosphate uridyltransferase (*galT*), and a transcriptional regulator (*galR*). Extracellular galactose binds the high affinity sugar-binding protein, GalP, which presents the sugar to the sensor histidine kinase (GalS). GalS phosphorylates the response regulator (GalT2), which in turn binds to the *galE*2 promoter and activates the expression of the ABC transport system for galactose (GalE2F2G2). In addition, galactose can be imported by a specific PTS for galactose (PtsABC). Inside the cell, the phosphorylated galactose (galactose-6-phosphate) is further metabolized *via* the tagatose-6-phosphate pathway, to dihydroxyacetone phosphate (DHAP) and glyceraldehyde 3-phosphate (GAP) or

alternatively, hydrolyzed by a phosphatase to galactose and Pi. Strain T-1 encodes for homologous tagatose pathway genes including, galactose-6-phosphate isomerase (*galA)*, tagatose-6P kinase (*galB)* and tagatose 1,6-bisphosphate aldolase (*galC)*. Cellobiose is imported to the cell by a dedicated PTS (*celBCD*) and is further hydrolyzed inside the cell by 6-phospho-β-glucosidase (Cel1A) to glucose-6-phoshate and glucose. Other galacto-moiety di-saccharides including lactose, and galactosyl-glycerol can be also imported by the PTS for cellobiose. Lactose and galactosyl-glycerol are further hydrolyzed by Gan1D to galactose-6-phosphate and glucose or glycerol, respectively. The letter P on the genetic map indicate proposed promoter sites, and Ω indicates the position of a hypothetical rho-independent transcription termination sites.



Scheme 1. Reagents and conditions: (a) HBr (33% AcOH), CH_2Cl_2, 0 – 20 °C; (b) **3**, Ag_2CO_3, Drierite, iodine, CH_2Cl_2, rt; (c) NaOMe, MeOH, 0 – 20 °C; (d) AcOH, H_2O, 60 °C.

Cross-utilization of β-galactosides and cellobiose in *Geobacillus stearothermophilus*

Smadar Shulami, Arie Zehavi, Valery Belakhov, Rachel Salama, Shifra Lansky, Timor Baasov, Gil Shoham and Yuval Shoham

J. Biol. Chem. published online June 3, 2020
