

Identification of a novel pyruvyltransferase using ^{13}C solid-state NMR to analyze rhizobial exopolysaccharides

Derek H. Wells,^{a*†} Nicolette F. Goularte,^{b*} Melanie J. Barnett,^a Lynette Cegelski,^{c#} and Sharon R. Long^{b#}

^aDepartment of Biology, Stanford University, Stanford, California, USA

^bDepartment of Structural Biology, Stanford University, Stanford, California, USA

^cDepartment of Chemistry, Stanford University, Stanford, California, USA

*These authors contributed equally to the project. Order was determined by order of participation.

[†]Present address: Derek H. Wells, ExoPolymer, Inc., 733 Industrial Rd, San Carlos, California, USA.

#Address correspondence to Lynette Cegelski (cegelski@stanford.edu) and Sharon R. Long (srl@stanford.edu).

ABSTRACT

The alphaproteobacterium *Sinorhizobium meliloti* secretes two acidic exopolysaccharides (EPS), succinoglycan (EPSI) and galactoglucan (EPSII), which differentially enable it to adapt to a changing environment. Succinoglycan is essential for invasion of plant hosts, and thus for formation of nitrogen-fixing root nodules. Galactoglucan is critical for population-based behaviors such as swarming and biofilm formation, and can facilitate invasion in the absence of succinoglycan on some host plants. Biosynthesis of galactoglucan is not as completely understood as that of succinoglycan. We devised a pipeline to: identify putative pyruvyltransferase and acetyltransferase genes; construct genomic deletions in strains engineered to produce either succinoglycan or galactoglucan; and analyze EPS from mutant bacterial strains. EPS samples were examined by ^{13}C cross-polarization magic-angle spinning (CPMAS) solid-state nuclear magnetic resonance (NMR). CPMAS NMR is uniquely suited to defining chemical composition in complex samples and enable detection and quantification of distinct EPS functional groups. Galactoglucan was isolated from mutant strains, with deletions in five candidate acyl/acetyltransferase genes (*exoZ*, *exoH*, *SMb20810*, *SMb21188*, *SMa1016*) and a putative pyruvyltransferase (*wgaE* or *SMb21322*). Most samples were similar in composition to wild-type EPSII by CPMAS NMR analysis. However, galactoglucan produced from a strain lacking *wgaE* exhibited a significant reduction in pyruvylation. Pyruvylation was restored through ectopic expression of plasmid-encoded *wgaE*. Our work has thus identified WgaE as a galactoglucan pyruvyltransferase. This exemplifies how the systematic combination of genetic analyses and solid-state NMR detection is a rapid means to identify genes responsible for modification of rhizobial exopolysaccharides.

IMPORTANCE

Nitrogen-fixing bacteria are crucial for geochemical cycles and global nitrogen nutrition. Symbioses between legumes and rhizobial bacteria establish root nodules, where bacteria convert dinitrogen to ammonia for plant utilization. Secreted exopolysaccharides (EPS) produced by *Sinorhizobium meliloti* (succinoglycan and galactoglucan) play important roles in soil and plant environments. Biosynthesis of galactoglucan is not as well characterized as succinoglycan. We employed solid-state nuclear magnetic resonance (NMR) to examine intact EPS from wild type and mutant *S. meliloti* strains. NMR analysis of EPS isolated from a *wgaE* gene mutant revealed a novel pyruvyltransferase that modifies galactoglucan. Few EPS pyruvyltransferases have been characterized. Our work provides insight into biosynthesis of an important *S. meliloti* EPS and expands knowledge of enzymes that modify polysaccharides.

INTRODUCTION

The alpha-proteobacterium *Sinorhizobium meliloti* is an important soil saprophyte that forms an intracellular, nitrogen-fixing symbiosis with legumes such as alfalfa (1). During life in the soil and symbiotic development, *S. meliloti* must respond appropriately to a changing environment and plant signals. Symbiosis begins with a molecular signal exchange between plant root and bacteria that establishes bacteria-host recognition (2). Tight association with plant roots and initiation of host invasion require live bacteria and involve bacterial envelope polysaccharides (3). Concurrent with invasion, plant cells grow and divide to form a root nodule (4). Bacteria are released into root nodule cells, where they differentiate into nitrogen-fixing bacteroids by genome endoreduplication, cell envelope remodeling, and novel gene expression (5-7).

The *S. meliloti* cell envelope plays an important role in soil survival and symbiosis (8). As with other Gram-negative bacteria, the *S. meliloti* envelope consists of inner and outer membranes, with a peptidoglycan layer in the periplasmic space between the two membranes (9). In addition to lipids and proteins, the *S. meliloti* envelope is associated with at least five symbiotically relevant polysaccharides: lipopolysaccharide, capsular polysaccharide, cyclic β -D-glucan, two acidic exopolysaccharides (EPSI and EPSII) (10-12), as well as others whose roles remain to be defined (13-17).

Numerous studies report that exopolysaccharides (EPS) help cultured *S. meliloti* tolerate stresses such as detergent, salt, acidic pH, heat, antimicrobial peptides, and reactive oxygen species (18-28). It is likely that EPS protect the bacteria from a variety of stresses encountered in the soil, rhizosphere and during symbiosis (29-31). EPS also suppress host plant defense responses (32, 33) and may act as a specific signaling molecule to the plant, interacting with host plant receptors (34).

Although it has the potential to produce two different EPS molecules, *S. meliloti* strain Rm1021 (a commonly used lab strain) depends only upon EPSI (succinoglycan) for symbiosis with alfalfa. It does not express genes for EPSII (galactoglucan) biosynthesis because the

genome carries a native insertion element in *expR*, encoding the required transcription activator (35). When functional ExpR is restored to *S. meliloti* Rm1021, galactoglucan can substitute for succinoglycan for successful invasion of alfalfa, but not of other host plants tested (35). Further experiments determined that an *S. meliloti* 1021 mutant strain unable to synthesize any EPS was able to form nodules on alfalfa when purified low molecular weight (LMW) forms of galactoglucan, but not high molecular weight (HMW) forms, were applied to plant roots along with the bacteria (36). ExpR and galactoglucan are important for *S. meliloti* population behaviors such as biofilm formation, root colonization, cell population movement across a surface, and perhaps even protection from predation (37-43).

The enzymes for succinoglycan biosynthesis are encoded by the *exo* genes (10, 44, 45), while genes in the *wga*, *wgb*, *wgc*, *wgd*, and *wge* gene cluster (formerly *exp* genes) are required for galactoglucan production (35, 46). In contrast to what is known about succinoglycan biosynthesis, these collective *wgx* genes have not been assigned to specific enzymatic steps, and the galactoglucan biosynthetic pathway has not been delineated.

Structurally, succinoglycan consists of repeating units containing one galactose and seven glucose molecules, decorated with acetyl, 1-carboxyethylidene (i.e., pyruvyl), and succinyl substituents, and joined by β -1,4-, β -1,3-, and β -1,6-glycosidic linkages (47, 48). Galactoglucan consists of a disaccharide repeating unit: β -D-glucose-(1 $\rightarrow$ 3)- α -D-galactose-(1 $\rightarrow$ 3) that is 6-O-acetylated on most D-glucose residues and pyruvate-ketal-modified [i.e., 4,6-O-(1-carboxyethylidene) groups] on D-galactose residues (49, 50).

Early structural studies of *S. meliloti* LMW EPS samples employed ^1H and ^{13}C solution-state NMR analysis, chromatography, and mass spectrometry (51, 52). HMW EPS, like other large polymers and biomolecules, poses a challenge to quantitative analysis by solution-state NMR due to reduced molecular mobility (i.e., high viscosity) of the samples. For this reason, analysis of HMW EPS composition usually integrates chemical and/or enzymatic digestion to produce smaller soluble species, including purified octasaccharides from succinoglycan and oligosaccharide fragments of galactoglucan (49-55). The use of chemically digested HMW EPS samples enables reasonable ^1H and ^{13}C NMR spectra to be obtained but can be accompanied by undesired modification to the polysaccharide. For instance, early ^{13}C solution-state NMR analysis required succinoglycan depolymerase and acid hydrolysis to detect ^{13}C peaks, which resulted in deacetylation of EPS (55). In principle, chemical digestion can be avoided by using a suitable depolymerase; for example, Her et al. successfully used a crude *S. meliloti* enzyme preparation to convert HMW galactoglucan to LMW forms (49). However, digestion and solubilization may be incomplete, providing only partial analysis of HMW EPS samples. Solid-state NMR is a unique direct detection method that overcomes the challenges faced by solution-state NMR for large biomolecules and enables quantification of carbons in heterogeneous and insoluble systems at natural abundance levels, even in intact cells and tissues (56-61). ^{13}C cross-polarization magic-angle spinning (CPMAS) solid-state NMR (62), in particular, has a rich history in defining and quantifying composition and phenomena such as crosslink formation in polysaccharides and protein-polysaccharide composite materials, including cellulose (62-64),

chitin (65), pectin (66), bacterial and plant cell walls (60, 67-69), and insect exoskeletons (70, 71). For example, ^{13}C CPMAS solid-state NMR was employed to investigate the extracellular matrix polysaccharide of *E. coli* biofilms and revealed the unanticipated presence of phosphoethanolamine cellulose, the first example of a naturally chemically modified cellulose (72).

Although galactoglucan biosynthesis appears to be important for a wide range of *S. meliloti* behaviors, its biosynthesis remains largely uncharacterized, especially in comparison to succinoglycan. By combining genetic and phenotypic tests with ^{13}C CPMAS solid-state NMR, we examined genes that may be involved in galactoglucan modification, quantified the extent of modifications in intact succinoglycan and galactoglucan, and assigned a ketal-pyruvylation function to one of two putative pyruvyltransferases encoded in the *wgx* gene cluster.

RESULTS AND DISCUSSION

Identification of candidate EPS modification genes and construction of mutants.

Succinoglycan produced by *S. meliloti* strain Rm1021 has succinyl, acetyl and pyruvyl modifications that are added by the ExoH succinyltransferase, ExoZ acetyltransferase, and ExoV pyruvyltransferase, respectively (45). Galactoglucan produced by Rm1021 has acetyl and pyruvyl modifications; however, enzymes that perform these modifications have not been identified (49, 50).

To identify candidate galactoglucan pyruvyltransferases and acetyltransferases, we mined the *S. meliloti* genome in the Pfam protein families database (73). Our intent was to obtain an overview of the types and numbers of predicted pyruvyltransferase and acetyl/acyltransferase domains, and then predict which domain-containing proteins were most likely to be involved in EPS modification. *S. meliloti* appears to encode enzymes with only one type of Pfam pyruvyltransferase domain (PF04230, PS_pyruv_trans), which matched five *S. meliloti* proteins (Figure 1A), including the succinoglycan pyruvyltransferase ExoV, and two predicted pyruvyltransferases (WgaD and WgaE) in the galactoglucan biosynthesis gene cluster (Figure S1; Table S1). Although WgaD and WgaE are encoded by adjacent genes, which are predicted to be cotranscribed (74), they are only 45% identical (Figure S2), suggesting they could have different functions or specificities.

Among dozens of predicted acetyl/acyltransferases in the *S. meliloti* genome, we focused on a group of proteins with Pfam domain PF01757 (Acyl_transf_3). This group includes membrane-bound proteins known to add acyl groups to polysaccharides (75), including the *S. meliloti* succinyltransferase ExoH and the acetyltransferase ExoZ (Figure 1B). Other members of this family in *S. meliloti* are SMb21188 and SMb20810, whose genes are located in previously identified polysaccharide biosynthesis clusters (14, 15), and Sma1016 (Figure 1B).

Of the predicted pyruvyltransferases in Figure 1A, we made strains (Table 1) with non-polar deletions of *wgaE* (CH23), *wgaD* (MB1235) or *wgaD* and *wgaE* (CH25). We pursued these particular genes first, because we believed it to be more likely that galactoglucan would be

modified by pyruvyltransferase(s) encoded in the galactoglucan *wgx* gene cluster. The *SMB20253* gene is not located in a polysaccharide biosynthesis gene cluster and the *SMB20812* protein is predicted to modify LPS core or lipid A (14), making them less attractive candidates for galactoglucan pyruvylation. Another likely possibility is that galactoglucan would be modified by the pyruvyltransferase ExoV, which is known to act on succinoglycan (45).

Since the *wgx* gene cluster does not appear to encode any acetyl/acyltransferases (Figure 1, Table S1), we made single deletion mutants for three genes (*SMA1016*, *SMB20810*, *SMB21188*) encoding proteins with PF01757 acyltransferase domains (Figure 1B; Table 1) to evaluate whether these influence galactoglucan composition. Galactoglucan could be acetylated by either ExoZ or ExoH, which act on succinoglycan, making them strong candidates. We therefore evaluated the effect of eliminating the enzymes ExoZ and ExoH on galactoglucan, using a strain with the entire *exo* cluster deleted (DW28). We also constructed and evaluated a strain (MB1279) lacking all five predicted Pfam domain PF01757 acyltransferases.

Lastly, since *S. meliloti* Rm1021 with restored ExpR produces substantial amounts of both succinoglycan and galactoglucan, we designed strains in which production of each exopolysaccharide was deliberately blocked, using targeted deletions of known biosynthesis genes. For the analysis of succinoglycan in the absence of galactoglucan, we employed a strain with a *wgeB* deletion (DW37), which abolishes galactoglucan biosynthesis. For the analysis of galactoglucan, the primary focus of this study, we used a strain with an *exoY* deletion (DW11), which abolishes succinoglycan biosynthesis. Strains lacking ExoY, which catalyzes the first committed step in EPSI biosynthesis (45), and WgeB, a bifunctional glycosyltransferase (38), have been previously used to eliminate EPS biosynthesis.

Succinoglycan analysis in wild-type and mutant strains by solid-state NMR. As proof of principle, we implemented ^{13}C CPMAS solid-state NMR spectroscopy to detect succinoglycan substituents on intact polysaccharide samples at natural abundance ^{13}C levels. No isotopic labeling was necessary (Materials and Methods). ^{13}C CPMAS is a powerful method of accounting for carbon spectral contributions from distinct chemical environments, such as those of each EPS substituent, and for monitoring changes in composition. The analysis of well-studied succinoglycan produced by *S. meliloti* allowed us to confirm the feasibility of this approach to identify key chemical modifications on polysaccharides.

S. meliloti succinoglycan (EPSI) contains pyruvyl, O-acetyl, and O-succinyl substituents on a repeating backbone containing galactose and glucose (Figure 2A). We analyzed succinoglycan from a $\Delta wgeB$ strain that is unable to produce galactoglucan. We compared the *expR*⁺ $\Delta wgeB$ strain (DW37) to the *expR*⁺ $\Delta wgeB\Delta exoH$ strain (DW50), which produces EPSI lacking the succinyl modification (76). The ^{13}C NMR spectrum of wild-type succinoglycan from the *expR*⁺ $\Delta wgeB$ strain (Figure 2B) clearly showed carbon contributions associated with the polysaccharide: anomeric and sugar ring carbons for the polysaccharide backbone (55-110 ppm); carbonyl carbons associated with acetyl, pyruvyl and succinyl groups (~175 ppm); and three aliphatic carbon populations associated with succinyl methylene carbons (31 ppm), acetyl methyl

carbons (25 ppm) and pyruvyl methyl carbons (21 ppm). The ^{13}C NMR spectrum of succinoglycan prepared from the *expR⁺ Δ wgeB Δ exoH* strain (Figure 2B) revealed a complete loss of succinyl methylene carbons at 31 ppm and a concomitant reduction in carbonyl carbon intensity, consistent with the absence of succinylation (Figure 2B). More specifically, the higher frequency succinyl carbonyl centerband peak at 180 ppm is completely abolished in the *expR⁺ Δ wgeB Δ exoH* (DW50) spectrum, along with a comparable decrease in intensity for the other succinyl carbonyl carbon that overlaps with the acetyl and pyruvyl carbonyls centered at 175 ppm (Figure 2B). Thus, we demonstrated that ^{13}C CPMAS NMR provides a simple and clear readout of carbon composition, validating our approach.

Quantitative ^{13}C CPMAS analysis of succinoglycan modification. To interrogate the extent of chemical modifications in succinoglycan (acetylation, pyruvylation, and succinylation), quantitative cross polarization (CP) array experiments were performed. In ^{13}C CPMAS NMR experiments, a single spectrum is insufficient to permit quantitative accounting of carbon types through peak integration due to possible differences in CP efficiency from ^1H nuclei to ^{13}C nuclei and in relaxation. For quantification of carbon contributions in proton-rich samples, peak areas are determined for each peak through a series of CPMAS NMR spectra acquired with different CP contact times, and the resulting data are extrapolated to define the theoretical magnetization at time zero (77).

The extent of modification for succinoglycan isolated from *expR⁺ Δ wgeB* strain DW37 was determined through analysis of the acetyl (25 ppm) and pyruvyl (21 ppm) methyl carbon and succinyl methylene carbon (centered at 31 ppm) peaks that are unique to the modifications in comparison to the anomeric carbon magnetization. Although there is spectral overlap of the three sets of aliphatic modification carbon contributions, they were sufficiently separated and exhibited standard lineshapes to be fit with modeled peaks (Materials and Methods). The quantitative CPMAS spectral analysis revealed that 27% of the sugar units were modified (Figure 2C and Table S2). More specifically, approximately 10% of the sugar units were modified with acetylation, 10% with pyruvylation and 7% by succinylation (Table S2). The carbonyl intensity also supported this determination, wherein each pyruvyl group and acetyl group contains one carbonyl carbon for each methyl carbon. Each succinyl group contains two carbonyl carbons with two methylene carbons. The ratio of carbonyl carbons to aliphatic carbons was 1:1, as expected. Full data analysis parameters are provided in Methods and Tables S2 and S3.

The overall extent of modification and relative ratio of acetyl:pyruvyl:succinyl modifications is similar to what has been estimated previously, with earlier studies estimating a 1:1:1 ratio of acetyl:pyruvyl:succinyl modifications (35, 45, 47, 48). Previous estimates (47-49, 55, 78) enabled elucidation of the succinoglycan biosynthesis pathway (45), yet there are limitations to using solution-based analytical tools to interrogate insoluble polymers, namely selective analysis of small soluble components that result from sample preparation. For example, solution-state NMR of highly viscous and partially insoluble polymers is useful for

assigning peaks of more soluble carbons and protons, if detectable above signals attributed to high molecular weight/low mobility components, but is unable to quantitatively determine their relative contributions compared to less mobile components. Solution-state ^{13}C NMR spectra have been reported for two intact EPS samples from *S. meliloti* 104A14 and a comparative mutant (53). However, the HMW EPS spectra exhibited broad ^{13}C peaks and much lower intensity carbonyl carbon peaks than expected, attributed to reduced mobility of the polymer, preventing quantitative analysis (53). Similarly, Glazebrook and Walker's 1989 publication identifying *S. meliloti* EPSII presented ^1H NMR spectra, where broader peak contributions were observed under strong sharper peaks (35). In addition, pioneering efforts implemented a variety of chemical digestion and secondary modifications to permit separations by HPLC and detection by mass spectrometry for structure sequencing and characterization (47-49, 55), including structure determination of an octasaccharide (78). However, undigested material was not analyzed in such experiments. The use of CPMAS solid-state NMR provides a complementary and straightforward approach to quantify carbons in intact polysaccharides.

Galactoglucan analysis in wild-type and mutant strains by solid-state NMR. We analyzed galactoglucan prepared from strains with deletions in candidate genes with possible pyruvyltransferase or acetyltransferase activities as described above. Galactoglucan consists of a backbone containing galactose and glucose with pyruvyl and acetyl substituents, respectively (Figure 3A). The ^{13}C CPMAS NMR spectrum of galactoglucan prepared from the *expR⁺ΔexoY* strain (DW11) revealed the obvious anomeric and sugar ring carbons, with chemical shifts reflective of the distinct electronic environment around the galactoglucan carbons (Figure 3B). As expected, the acetyl and pyruvyl carbonyl carbons appear together centered at 175 ppm and acetyl methyl carbons are observed at 25 ppm and pyruvyl methyl carbons at 21 ppm.

The three ^{13}C NMR spectra of galactoglucan prepared from strains lacking putative acyltransferases SMb21188, SMb20810 or SMa1016, appeared similar to wild-type EPSII, indicating that none of these genes was solely responsible for acetylation (Figure S3). Further, the ^{13}C NMR spectra of galactoglucan produced by a strain with the entire *exo* gene cluster deleted (DW28) was similar to wild type (Figure S4A), indicating that the enzymes responsible for succinoglycan acetylation, pyruvylation, and succinylation were not required for galactoglucan modification, although it does not rule out possible redundancy between these and other acyltransferases. To test for functional redundancy among the five putative acyltransferases, we introduced deletions of *SMb21188*, *SMb20810*, and *SMa1016* into *Δexo* strain DW28. The EPS spectrum of the resulting strain (Figure S4B; MB1279) was indistinguishable from wild type and DW28, indicating that none of the five acyltransferases are sufficient for galactoglucan O-acetylation under our growth conditions.

Although we did not identify an acetyltransferase required for galactoglucan modification, we observed major changes in the ^{13}C NMR spectrum associated with galactoglucan produced by a *ΔwgaE* strain (CH23), which lacks one of two putative pyruvyltransferases encoded in the galactoglucan biosynthesis gene cluster. The ^{13}C NMR

spectrum revealed a significant reduction in the pyruvyl methyl carbon peak and concomitant reduction in the carbonyl carbon intensity (Figure 3B), indicative of a significant reduction of pyruvylation of the polysaccharide. Residual intensity at 21 ppm suggests that some limited pyruvylation still remains in the absence of *wgaE*. It is possible that this residual pyruvylation is due to the activity of WgaD. We were unable to examine galactoglucan from the $\Delta wgaD$ (MB1235) and $\Delta wgaDE$ (CH25) strains because, unlike the $\Delta wgaE$ strain, they failed to produce enough EPS for analysis. In addition, the $\Delta wgaD$ strain was severely growth impaired, a phenotype similar to that of strains with mutations in *exoV* encoding the succinoglycan pyruvyltransferase (44), and which could be overcome by ectopic expression of *wgaD* on plasmid pMB1035 but not by ectopic expression of *wgaE* on pDW24. The disparate mutant phenotypes of $\Delta wgaE$ and $\Delta wgaD$ mutant strains suggest that WgaD has some role other than pyruvylation of galactoglucan and/or that its function is somehow intrinsically linked to synthesis, polymerization, or export. Finally, in order to validate the role of *wgaE* as a novel pyruvyltransferase gene, WgaE production was restored through complementation by ectopically expressing *wgaE* from the *lac* promoter on pSRKKm (pDW24; Materials and Methods). As shown in Figure S5, the complemented strain (DW128) fully restores pyruvylation and its spectrum matches that of native galactoglucan.

Quantitative analysis of galactoglucan modification by solid-state NMR. We quantified the extent of modification for galactoglucan as performed above for succinoglycan. From inspection of the CPMAS spectra, galactoglucan is clearly modified to a greater extent (Figure 3). Pyruvyl and acetyl groups each have a single carbonyl carbon and a single methyl carbon. The extent of modification for galactoglucan isolated from *expR*⁺ Δ *exoY* strain DW11 was determined by comparing magnetization of carbonyl and methyl carbons unique to the pyruvyl and acetyl groups to the anomeric carbon magnetization. This analysis revealed that 68% of sugars on the glucose galactose backbone are modified with either an acetyl or a pyruvyl group. (Figure 3C and Table S3). We next examined the prevalence of acetylation versus pyruvylation modifications. Although there is spectral overlap of the methyl carbon contributions, they were separated enough to be deconvoluted and modeled with Lorentzian peak fits (Figure S6). Analysis of these peak fits revealed that approximately 43% of the sugars were acetylated and 25% were pyruvylation (Table S3; Figure S6). Therefore, EPSII examined in this study exhibits more than twice the extent of overall modification as that of succinoglycan (EPSI), yet not all sugar residues are modified. The first reports of EPSII modification suggested EPSII has alternating glucose and galactose residues with all galactose residues pyruvylation (35, 49, 50). However, this estimate was based on solution-state NMR analyses of "viscous" samples and spectra were not shown. It is likely that parts of the polysaccharide analyzed in those studies were not sufficiently mobile and may have had broadened peaks that were indistinguishable from the baseline and unable to be observed, while sharper peaks for mobile components were analyzed. Indeed, solution-state ¹H NMR spectra reported in 1989 show broadened peaks consistent with less mobile components (35). These previous experiments also included a size-

exclusion chromatography step that may have resulted in inadvertent sample enrichment of soluble and LMW components that were more highly modified (35, 49, 50).

Conclusions and outlook. The complete *S. meliloti* galactoglucan biosynthesis pathway remains to be elucidated. In this study, we combined genetic techniques, including a rapid method for creating precise genomic deletions, with state-of-the-art CPMAS solid-state NMR to identify a new enzyme involved in polysaccharide modification. This enzyme, WgaE, is the primary pyruvyltransferase of galactoglucan. WgaE adds a 4,6-O-(1-carboxyethylidene) modification to galactose, the most common pyruvylated monosaccharide in the Carbohydrate Structure Database (79).

Compared to enol-pyruvylation of N-acetylglucosamine, a well-characterized step in bacterial peptidoglycan precursor synthesis, little is known about ketal-pyruvylation of polysaccharides, including the role such modifications play in bacterial physiology (79). To date, only a handful of bacterial pyruvyltransferases has been functionally characterized: ExoV (45), PssM (80), WcfO (81), CsaB (82) and GumL (83, 84). Our work now adds an additional member to the cohort of enzymes that performs this unique function.

How pyruvyl modifications contribute to EPS function in *S. meliloti* is unknown, but studies of other bacteria suggest that pyruvylation may alter the ability of EPS to chelate calcium ions, which in turn may affect adhesion, biofilm formation, and virulence (85). The degree of pyruvylation may also affect secondary structure and viscosity of exopolysaccharides (86). An *S. meliloti* mutant lacking the WgaE pyruvyltransferase still produces abundant galactoglucan, unlike mutants lacking the ExoV pyruvyltransferase, which fail to produce detectible succinoglycan (45). The former observation, taken together with the presence of two pyruvyltransferase genes in the ESPII gene cluster, suggests that galactoglucan biosynthesis is more complex than assumed. Indeed, it is unclear why there are at least five putative glycosyltransferases in the galactoglucan cluster (Figure S1) yet there are only two sugar residues in the repeating unit of the exopolysaccharide. The small peak of residual pyruvylation in the $\Delta wgaE$ mutant strain, according to CPMAS NMR, and the impaired growth and inability of the $\Delta wgaD$ mutant strain to produce galactoglucan, indicate that although both genes have pyruvyltransferase domains, they have non-overlapping activities in biosynthesis. One of several interpretations of these data would be that synthesis of the known galactoglucan structure first requires the synthesis of an unidentified molecular precursor of different sugar composition, which itself is a substrate for pyruvylation. Another interpretation is that not one, but two or more repeating units of galactoglucan, requiring multiple glycosyltransferases, are assembled in the cytoplasm prior to polymerization and export. For example, if galactoglucan were a multimer, there would be galactose residues in different structural contexts, and these may be substrates for independent pyruvyltransferases. If one of these modified residues were required for the subsequent steps of polymerization or export, similar to the proposed role of ExoV in succinoglycan biosynthesis (45), it could result in a cell growth defect due to “backing up” of galactoglucan production. In support of this hypothesis, normal growth was restored for a *wgaD*

deletion strain lacking either ExpR or WgeB, which block galactoglucan biosynthesis early in the pathway (Table 1; MB1228 and MB1245). Our work illuminates one crucial step in the biosynthesis of galactoglucan. Ahead lie further questions about the overall mechanism of galactoglucan assembly and export.

MATERIALS AND METHODS

Strains and plasmids. Table 1 lists strains and plasmids used in this study. *S. meliloti* strains were routinely grown in LB (5 g/l NaCl) (87) or TY medium (88) at 30°C. *E. coli* strains were grown in LB medium at 37°C. Antibiotics were used at the following concentrations: chloramphenicol (Cm), 25-50 µg ml⁻¹; gentamicin (Gm), 5 µg ml⁻¹ for *E. coli* and 25-50 µg ml⁻¹ for *S. meliloti*; kanamycin (Km), 25-40 µg ml⁻¹ for *E. coli*; neomycin (Nm), 50-100 µg ml⁻¹ for *S. meliloti*; and streptomycin (Sm), 500 µg ml⁻¹ for *S. meliloti*. Triparental conjugations were used (89) to transfer both replicative and nonreplicative plasmids to *S. meliloti*. Oligonucleotide primers used in this study are listed in the Supplemental Data Set. We used standard techniques for cloning and PCR amplification. *E. coli* NEB5α (New England Biolabs; Ipswich, MA) was used as a plasmid host.

For production of EPS, *S. meliloti* was propagated in a modified M9 sucrose growth medium: 2% sucrose; 0.02% yeast extract; 1 mM MgSO₄; 1X M9 salts (10X M9 salts stock = 58 g l⁻¹ NaHPO₄, 30 g l⁻¹ K₂PO₄, 5 g l⁻¹ NaCl, 10 g l⁻¹ NH₄Cl); 1X vitamins (1000X vitamin stock = 10 mg ml⁻¹ calcium pantothenate, 10 mg ml⁻¹ thiamine, and 0.2 mg ml⁻¹ biotin); and 1X trace elements (1000X trace elements stock = 40 g l⁻¹ citric acid•H₂O, 30 g l⁻¹ MnSO₄•H₂O, 10 g l⁻¹ NaCl, 1 g l⁻¹ FeSO₄•7H₂O, 1 g l⁻¹ CoCl₂•6H₂O, 1 g l⁻¹ ZnSO₄•7H₂O, 0.1 g l⁻¹ CuSO₄•5H₂O, 0.1 g l⁻¹ H₃BO₃, 0.1 g l⁻¹ Na₂MoO₄•2H₂O and pH adjusted to 3.0 with NaOH). Typically, 100 ml of EPS production medium in a 250 ml flask was inoculated with 1 ml of a saturated *S. meliloti* culture (grown in LB or TY). Cultures were incubated at 30°C with shaking at 250 RPM for 72 hours prior to EPS harvest.

Deletion strains. To generate non-polar deletion mutants of candidate genes identified in this study, as well as those whose function has already been identified, we used pJQ200SK, which allows for integration and then sucrose counter-selection for targeted genome modification in *S. meliloti* (90). Deletion cassettes were designed to retain the translation initiation codon of a given gene, directly followed by the native stop codon. In cases where genes were within an operon, and where deletion of the targeted open reading frame (ORF) would alter the coding sequence(s) of adjacent ORFs, deletion plasmids were designed to preserve the ribosome binding sites and correct reading frame of those adjacent genes. To generate a single cassette for insertion into pJQ200SK, DNA fragments ~500 bp flanking the ORF to be deleted were overlap-extension PCR amplified (91), using *S. meliloti* 1021 genomic DNA as a template. The resulting DNA was cloned via Circular Polymerase Extension Cloning (CPEC) (92) in a reaction with pJQ200SK, which had been amplified using primers oDW11 and oDW12 (Supplemental Data Set). CPEC reactions were incubated overnight at 37°C with DpnI (New England Biolabs) to remove template DNA. Plasmids were verified using pJQ200SK-specific primers (oDW18 and oDW44) for sequencing then introduced into *S. meliloti* 1021 by triparental conjugation (89). Plasmid integration events from homologous recombination were identified by colony PCR with a pJQ200SK-specific primer (oDW44 or oDW18) and a primer external to either of the homologous genomic target regions. Genomic insertions were

transferred from *S. meliloti* 1021 to other strains using N3 phage transduction as described (93). Sucrose counterselection was performed on LB plates containing 2% sucrose. Sucrose-resistant, Gm-sensitive colonies were checked by PCR for the deletion, using primers external to the *S. meliloti* DNA cloned in pJQ200SK (Supplemental Data Set). Additionally, for the mutant strain lacking all five predicted Pfam domain PF01757 acyltransferases (MB1279), we eliminated the unlikely possibility that one or more of the deleted ORFs was retained elsewhere in the genome by performing PCR amplification using ORF-specific primers (Supplemental Data Set).

Restoration of ExpR function to S. meliloti 1021. In *S. meliloti* 1021, *expR* is interrupted by insertion of the native insertion element IS_{Rm1} (35). We used pJQ200SK and sucrose counter-selection, as described above, to replace *expR* DNA in strain 1021 with WT *expR* DNA amplified (primers ExpR-SpeI-2 and ExpR-XhoI-2, Supplemental Data Set) from the *S. meliloti* 1021-derived strain, Rm8530 that carries a functional *expR* (35). Double homologous recombination of WT *expR* DNA into the 1021 genome was verified by phenotype and PCR (primers ExpR-5-CK and ExpR-3).

Ectopic-expression plasmids. pSRKKm is a pBBR-family, broad-host-range vector that encodes the LacI repressor, allowing for regulatable expression of cloned genes from *Plac*. To ectopically express *wgaE* (pDW24) or *wgaD* (pMB1035), each ORF was PCR-amplified from *S. meliloti* 1021 genomic DNA as a NdeI-XbaI DNA and introduced into NdeI-XbaI-digested pSRKKm (94). Cloned DNA was verified by sequencing. To induce expression of ORFs cloned into pSRKKm, we added 0.5 mM IPTG to the growth medium.

EPS harvest and purification. For EPS production, strains were propagated in modified M9 medium as described above. After approximately 72 hours of growth, cultures were decanted into polypropylene centrifuge bottles and centrifuged at room temperature at 10,000 g for 10 minutes in an SLA-1500 rotor in an RC5B centrifuge. For EPSI-producing strains, the culture volume was doubled with water prior to centrifugation to reduce viscosity and improve cell separation. Cleared supernatant was decanted from the cell pellet, and 1 mM CaCl₂ (according to starting culture volume) was added. EPS was precipitated with two volumes of cold isopropanol and storage at 4°C for at least four hours. Precipitated EPS was isolated by centrifugation at 6000 g for five minutes and washed 1X with 70% ethanol. After washing, samples were centrifuged at 6000 g for five minutes, and ethanol was decanted. Recovered EPS was incubated in a drying oven overnight at 60°C or until sample weight stabilized. Prior to solid-state NMR, samples were ground to a fine powder using a mortar and pestle.

Solid-state NMR measurements. All NMR experiments were performed in an 89-mm-bore 11.7T magnet using either an Agilent triple resonance BioMAS probe with a DD2 console (Agilent Technologies) or a home-built four-frequency ¹H/¹⁹F/¹³C/¹⁵N all transmission-line probe with a Varian VNMRs console. Samples were spun at 7143 Hz in either 36-μl-capacity 3.2-mm zirconia rotors or thin-walled 5-mm-outer-diameter zirconia rotors. The field strength for cross-polarization was 50 kHz with ramped CP at 57 kHz on ¹H. ¹H decoupling was performed at

83kHz. Recycle time was 2 s for all experiments. Spectrometer ^{13}C chemical shift referencing was performed by setting the high-frequency adamantane peak to 38.5 ppm (95). All spectra were the result of 32,768 scans, except for CP array experiments with 4096 scans, 8192 scans, or 65,536 scans as indicated for each CP contact time (Tables S2 and S3). NMR spectra were processed with 80-Hz line broadening.

CPMAS NMR peak fitting and quantitative CPMAS. Selected carbons from the EPSI and EPSII spectra were quantified to determine the extents of chemical modifications. Carbon peaks corresponding to the modification aliphatic carbons (methyl and methylene carbons) and anomeric C1 carbons were deconvoluted and modeled using MATLAB with Lorentzian peak fits (shown in Figure S6) as used successfully in other biomaterials including cellulose (96). In quantitative CPMAS experiments using CPMAS array experiments, the magnetization corresponding to each peak as a function of CP time depends on the buildup of magnetization through cross polarization (T_{IS}) and the relaxation dictated by $T_{1\rho}(\text{H})$, the proton spin-lattice relaxation time in the rotating frame (77). Quantitative carbon intensities were obtained by using data from long CP time acquisitions after magnetization buildup was complete, for $T_{IS} \ll T_{1\rho}(\text{H})$, to calculate the magnetization according to $I_t = I_0 [\exp\{-t/T_{1\rho}(\text{H})\}]$ where I_0 is the maximum signal intensity (77, 97). Carbonyl carbons are characterized by a large chemical shift anisotropy and can be difficult to quantify with peak contributions spread out among a carbonyl centerband peak (~175 ppm) and two spinning sidebands (118 ppm and 232 ppm) resulting from magic-angle spinning. The analysis on succinoglycan, which is less highly modified, was possible because over 64,000 scans were collected for each spectrum in the CP array analysis to limit error from spectral noise with full parameters provided in Tables S2 and S3.

ACKNOWLEDGMENTS

We are grateful to Cynthia Hao for assistance with plasmid and strain construction. We thank Aaron Duthoy and Dr. Sabrina Werby for critically reading our manuscript. N.F.G was supported by a Stanford Graduate Fellowship and an NSF Graduate Research Fellowship. This work was funded by NSF Award 2001189 and CAREER/PECASE Award 1453247 to L.C., and NIH Grant R01 GM093628 and NSF Award 2015870 to S.R.L. The authors declare the following competing interest: D.H.W. and S.R.L. are co-founders of ExoPolymer, Inc.

REFERENCES

1. Poole P, Ramachandran V, Terpolilli J. 2018. Rhizobia: from saprophytes to endosymbionts. *Nat Rev Microbiol* 19:291-303.
2. Long SR. 2016. SnapShot: signaling in symbiosis. *Cell* 167:582-582 e1.
3. Jones KM, Kobayashi H, Davies BW, Taga ME, Walker GC. 2007. How rhizobial symbionts invade plants: the *Sinorhizobium-Medicago* model. *Nat Rev Microbiol* 5:619-633.
4. Oldroyd GE, Downie JA. 2008. Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annu Rev Plant Biol* 59:519-546.
5. Gibson KE, Kobayashi H, Walker GC. 2008. Molecular determinants of a symbiotic chronic infection. *Annu Rev Genet* 42:413-441.
6. Kondorosi E, Mergaert P, Kereszt A. 2013. A paradigm for endosymbiotic life: cell differentiation of *Rhizobium* bacteria provoked by host plant factors. *Annu Rev Microbiol* 67:611-628.
7. Udvardi M, Poole PS. 2013. Transport and metabolism in legume-rhizobia symbioses. *Annu Rev Plant Biol* 64:781-805.
8. Haag AF, Arnold MFF, Myka KK, Kerscher B, Dall'Angelo S, Zanda M, Mergaert P, Ferguson GP. 2013. Molecular insights into bacteroid development during *Rhizobium*-legume symbiosis. *FEMS Microbiol Rev* 37:364-383.
9. Silhavy TJ, Kahne D, Walker S. 2010. The bacterial cell envelope. *Cold Spring Harb Perspect Biol* 2:a000414.
10. González JE, York GM, Walker GC. 1996. *Rhizobium meliloti* exopolysaccharides: synthesis and symbiotic function. *Gene* 179:141-146.
11. Downie JA. 2010. The roles of extracellular proteins, polysaccharides and signals in the interactions of rhizobia with legume roots. *FEMS Microbiol Rev* 34:150-170.
12. Pellock BJ, Cheng HP, Walker GC. 2000. Alfalfa root nodule invasion efficiency is dependent on *Sinorhizobium meliloti* polysaccharides. *J Bacteriol* 182:4310-4318.
13. Beck S, Marlow VL, Woodall K, Doerrler WT, James EK, Ferguson GP. 2008. The *Sinorhizobium meliloti* MsbA2 protein is essential for the legume symbiosis. *Microbiology* 154:1258-1270.

14. Finan TM, Weidner S, Wong K, Buhrmester J, Chain P, Vörholter FJ, Hernandez-Lucas I, Becker A, Cowie A, Gouzy J, Golding B, Pühler A. 2001. The complete sequence of the 1,683-kb pSymB megaplasmid from the N₂-fixing endosymbiont *Sinorhizobium meliloti*. *Proc Natl Acad Sci U S A* 98:9889-9894.
15. Griffiths JS, Long SR. 2008. A symbiotic mutant of *Sinorhizobium meliloti* reveals a novel genetic pathway involving succinoglycan biosynthetic functions. *Mol Microbiol* 67:1292-1306.
16. Schäper S, Steinchen W, Krol E, Altegoer F, Skotnicka D, Søgaard-Andersen L, Bange G, Becker A. 2017. AraC-like transcriptional activator CuxR binds c-di-GMP by a PilZ-like mechanism to regulate extracellular polysaccharide production. *Proc Natl Acad Sci U S A* 114:E4822-E4831.
17. Pérez-Mendoza D, Rodríguez-Carvajal MA, Romero-Jiménez L, Farias Gde A, Lloret J, Gallegos MT, Sanjuán J. 2015. Novel mixed-linkage beta-glucan activated by c-di-GMP in *Sinorhizobium meliloti*. *Proc Natl Acad Sci U S A* 112:E757-E765.
18. Lehman AP, Long SR. 2013. Exopolysaccharides from *Sinorhizobium meliloti* can protect against H₂O₂-dependent damage. *J Bacteriol* 195:5362-5369.
19. Miller-Williams M, Loewen PC, Oresnik IJ. 2006. Isolation of salt-sensitive mutants of *Sinorhizobium meliloti* strain Rm1021. *Microbiology* 152:2049-2059.
20. Morris J, González JE. 2009. The novel genes *emmABC* are associated with exopolysaccharide production, motility, stress adaptation, and symbiosis in *Sinorhizobium meliloti*. *J Bacteriol* 191:5890-5900.
21. Vriezen JA, de Bruijn FJ, Nusslein K. 2007. Responses of rhizobia to desiccation in relation to osmotic stress, oxygen, and temperature. *Appl Environ Microbiol* 73:3451-3459.
22. Davies BW, Walker GC. 2007. Identification of novel *Sinorhizobium meliloti* mutants compromised for oxidative stress protection and symbiosis. *J Bacteriol* 189:2110-2113.
23. Hellweg C, Pühler A, Weidner S. 2009. The time course of the transcriptomic response of *Sinorhizobium meliloti* 1021 following a shift to acidic pH. *BMC Microbiol* 9:37.
24. Rüberg S, Tian ZX, Krol E, Linke B, Meyer F, Wang Y, Pühler A, Weidner S, Becker A. 2003. Construction and validation of a *Sinorhizobium meliloti* whole genome DNA microarray: genome-wide profiling of osmoadaptive gene expression. *J Biotechnol* 106:255-268.

25. Barnett MJ, Bittner AN, Toman CJ, Oke V, Long SR. 2012. Dual RpoH sigma factors and transcriptional plasticity in a symbiotic bacterium. *J Bacteriol* 194:4983-94.
26. Penterman J, Abo RP, De Nisco NJ, Arnold MF, Longhi R, Zanda M, Walker GC. 2014. Host plant peptides elicit a transcriptional response to control the *Sinorhizobium meliloti* cell cycle during symbiosis. *Proc Natl Acad Sci U S A* 111:3561-3566.
27. Tiricz H, Szücs A, Farkas A, Pap B, Lima RM, Maróti G, Kondorosi E, Kereszt A. 2013. Antimicrobial nodule-specific cysteine-rich peptides induce membrane depolarization-associated changes in the transcriptome of *Sinorhizobium meliloti*. *Appl Environ Microbiol* 79:6737-6746.
28. Primo E, Bogino P, Cossoyich S, Foresto E, Nievas F, Giordano W. 2020. Exopolysaccharide II is relevant for the survival of *Sinorhizobium meliloti* under water deficiency and salinity stress. *Molecules* 25:4876.
29. Santos R, Hérouart D, Sigaud S, Touati D, Puppo A. 2001. Oxidative burst in alfalfa-*Sinorhizobium meliloti* symbiotic interaction. *Mol Plant Microbe Interact* 14:86-89.
30. Geddes BA, González JE, Oresnik IJ. 2014. Exopolysaccharide production in response to medium acidification is correlated with an increase in competition for nodule occupancy. *Mol Plant Microbe Interact* 27:1307-1317.
31. Maróti G, Downie JA, Kondorosi E. 2015. Plant cysteine-rich peptides that inhibit pathogen growth and control rhizobial differentiation in legume nodules. *Curr Opin Plant Biol* 26:57-63.
32. Aslam SN, Newman MA, Erbs G, Morrissey KL, Chinchilla D, Boller T, Jensen TT, De Castro C, Ierano T, Molinaro A, Jackson RW, Knight MR, Cooper RM. 2008. Bacterial polysaccharides suppress induced innate immunity by calcium chelation. *Curr Biol* 18:1078-1083.
33. Jones KM, Sharopova N, Lohar DP, Zhang JQ, VandenBosch KA, Walker GC. 2008. Differential response of the plant *Medicago truncatula* to its symbiont *Sinorhizobium meliloti* or an exopolysaccharide-deficient mutant. *Proc Natl Acad Sci U S A* 105:704-709.
34. Maillet F, Fournier J, Mendis HC, Tadege M, Wen J, Ratet P, Mysore KS, Gough C, Jones KM. 2020. *Sinorhizobium meliloti* succinylated high-molecular-weight succinoglycan and the *Medicago truncatula* LysM receptor-like kinase MtLYK10 participate independently in symbiotic infection. *Plant J* 102:311-326.

35. Glazebrook J, Walker GC. 1989. A novel exopolysaccharide can function in place of the calcofluor-binding exopolysaccharide in nodulation of alfalfa by *Rhizobium meliloti*. *Cell* 56:661-672.
36. González JE, Reuhs BL, Walker GC. 1996. Low molecular weight EPS II of *Rhizobium meliloti* allows nodule invasion in *Medicago sativa*. *Proc Natl Acad Sci U S A* 93:8636-8641.
37. Fujishige NA, Kapadia NN, De Hoff PL, Hirsch AM. 2006. Investigations of *Rhizobium* biofilm formation. *FEMS Microbiol Ecol* 56:195-206.
38. Bahlawane C, McIntosh M, Krol E, Becker A. 2008. *Sinorhizobium meliloti* regulator MucR couples exopolysaccharide synthesis and motility. *Mol Plant Microbe Interact* 21:1498-1509.
39. Rinaudi LV, Giordano W. 2010. An integrated view of biofilm formation in rhizobia. *FEMS Microbiol Lett* 304:1-11.
40. Gao M, Coggin A, Yagnik K, Teplitski M. 2012. Role of specific quorum-sensing signals in the regulation of exopolysaccharide II production within *Sinorhizobium meliloti* spreading colonies. *PLoS One* 7:e42611.
41. Rinaudi LV, González JE. 2009. The low-molecular-weight fraction of exopolysaccharide II from *Sinorhizobium meliloti* is a crucial determinant of biofilm formation. *J Bacteriol* 191:7216-7224.
42. Nogales J, Bernabeu-Roda L, Cuéllar V, Soto MJ. 2012. ExpR is not required for swarming but promotes sliding in *Sinorhizobium meliloti*. *J Bacteriol* 194:2027-2035.
43. Pérez J, Jiménez-Zurdo JJ, Martínez-Abarca F, Millan V, Shimkets LJ, Muñoz-Dorado J. 2014. Rhizobial galactoglucan determines the predatory pattern of *Myxococcus xanthus* and protects *Sinorhizobium meliloti* from predation. *Environ Microbiol* 16:2341-2350.
44. Glucksmann MA, Reuber TL, Walker GC. 1993. Genes needed for the modification, polymerization, export, and processing of succinoglycan by *Rhizobium meliloti*: a model for succinoglycan biosynthesis. *J Bacteriol* 175:7045-7055.
45. Reuber TL, Walker GC. 1993. Biosynthesis of succinoglycan, a symbiotically important exopolysaccharide of *Rhizobium meliloti*. *Cell* 74:269-280.
46. Becker A, Rüberg S, Küster H, Roxlau AA, Keller M, Ivashina T, Cheng HP, Walker GC, Pühler A. 1997. The 32-kilobase *exp* gene cluster of *Rhizobium meliloti* directing the

- biosynthesis of galactoglucan: genetic organization and properties of the encoded gene products. *J Bacteriol* 179:1375-1384.
47. Åman P, McNeil M, Franzén L-E, Darvill AG, Albersheim P. 1981. Structural elucidation, using H.P.L.C.-M.S. and G.L.C.-M.S., of the acidic polysaccharide secreted by *Rhizobium meliloti* strain 1021. *Carbohydr Res* 95:263-282.
 48. Reinhold BB, Chan SY, Reuber TL, Marra A, Walker GC, Reinhold VN. 1994. Detailed structural characterization of succinoglycan, the major exopolysaccharide of *Rhizobium meliloti* Rm1021. *J Bacteriol* 176:1997-2002.
 49. Her GR, Glazebrook J, Walker GC, Reinhold VN. 1990. Structural studies of a novel exopolysaccharide produced by a mutant of *Rhizobium meliloti* strain Rm1021. *Carbohydr Res* 198:305-312.
 50. Levery SB, Zhan H, Lee CC, Leigh JA, Hakomori S. 1991. Structural analysis of a second acidic exopolysaccharide of *Rhizobium meliloti* that can function in alfalfa root nodule invasion. *Carbohydr Res* 210:339-347.
 51. Chouly C, Colquhoun IJ, Jodelet A, York G, Walker GC. 1995. NMR studies of succinoglycan repeating-unit octasaccharides from *Rhizobium meliloti* and *Agrobacterium radiobacter*. *Int J Biol Macromol* 17:357-363.
 52. Wang LX, Wang Y, Pellock B, Walker GC. 1999. Structural characterization of the symbiotically important low-molecular-weight succinoglycan of *Sinorhizobium meliloti*. *J Bacteriol* 181:6788-6796.
 53. Summers ML, Botero LM, Busse SC, McDermott TR. 2000. The *Sinorhizobium meliloti* Lon protease is involved in regulating exopolysaccharide synthesis and is required for nodulation of alfalfa. *J Bacteriol* 182:2551-2558.
 54. Wei GJ, Song SC, Lin LP, Her GR. 1996. Structural studies of extracellular polysaccharide produced by *Rhizobium fredii* Tu6, a polysaccharide with nonasaccharide repeating units. *Bot Bull Acad Sin* 37:127-131.
 55. Harada T, Amemura A, Jansson PE, Lindberg B. 1979. Comparative studies of polysaccharides elaborated by *Rhizobium*, *Alcaligenes*, and *Agrobacterium*. *Carbohydr Res* 77:285-288.
 56. Cegelski L. 2015. Bottom-up and top-down solid-state NMR approaches for bacterial biofilm matrix composition. *J Magn Reson* 253:91-97.

57. Cegelski L, Schaefer J. 2005. Glycine metabolism in intact leaves by *in vivo* ^{13}C and ^{15}N labeling. *J Biol Chem* 280:39238-39245.
58. McCrate OA, Zhou X, Reichhardt C, Cegelski L. 2013. Sum of the parts: composition and architecture of the bacterial extracellular matrix. *J Mol Biol* 425:4286-4294.
59. Rabiah NI, Romaniuk JAH, Fuller GG, Scales CW, Cegelski L. 2019. Carbon compositional analysis of hydrogel contact lenses by solid-state NMR spectroscopy. *Solid State Nucl Magn Reson* 102:47-52.
60. Romaniuk JA, Cegelski L. 2015. Bacterial cell wall composition and the influence of antibiotics by cell-wall and whole-cell NMR. *Philos Trans R Soc Lond B Biol Sci* 370:20150024.
61. Werby SH, Cegelski L. 2019. Spectral comparisons of mammalian cells and intact organelles by solid-state NMR. *J Struct Biol* 206:49-54.
62. Schaefer J, Stejskal EO. 1976. Carbon-13 nuclear magnetic resonance of polymers spinning at the magic angle. *J Am Chem Soc* 98:1031-1032.
63. Atalla RH, Gast JC, Sindorf DW, Bartuska VJ, Maciel GE. 1980. ^{13}C NMR spectra of cellulose polymorphs. *J Am Chem Soc* 102:3249-3251.
64. Earl W, Vanderhart D. 1980. High resolution, magic angle sampling spinning ^{13}C NMR of solid cellulose I. *J Am Chem Soc* 102:3251-3252.
65. Peter MG, Grün L, Förster H. 1984. CP/MAS- ^{13}C -NMR spectra of sclerotized insect cuticle and of chitin. *Angew Chem, Int Ed Engl* 23:638-639.
66. Sinitsya A, Copiková J, Pavliková H. 1998. ^{13}C CP/MAS NMR spectroscopy in the analysis of pectins. *J Carbohydr Chem* 17:279-292.
67. Jacob GS, Schaefer J, Wilson GE, Jr. 1983. Direct measurement of peptidoglycan cross-linking in bacteria by ^{15}N nuclear magnetic resonance. *J Biol Chem* 258:10824-10826.
68. Kelly JE, Chrissian C, Stark RE. 2020. Tailoring NMR experiments for structural characterization of amorphous biological solids: A practical guide. *Solid State Nucl Magn Reson* 109:101686.
69. Cegelski L, O'Connor RD, Stueber D, Singh M, Poliks B, Schaefer J. 2010. Plant cell-wall cross-links by REDOR NMR spectroscopy. *J Am Chem Soc* 132:16052-16057.

70. Schaefer J, Kramer KJ, Garbow JR, Jacob GS, Stejskal EO, Hopkins TL, Speirs RD. 1987. Aromatic cross-links in insect cuticle: detection by solid-state ^{13}C and ^{15}N NMR. *Science* 235:1200-1204.
71. Kramer KJ, Hopkins TL, Schaefer J. 1995. Applications of solids NMR to the analysis of insect sclerotized structures. *Insect Biochem Mol Biol* 25:1067-1080.
72. Thongsomboon W, Serra DO, Possling A, Hadjineophytou C, Hengge R, Cegelski L. 2018. Phosphoethanolamine cellulose: A naturally produced chemically modified cellulose. *Science* 359:334-338.
73. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. *Nucleic Acids Res* 47:D427-D432.
74. Schlüter JP, Reinkensmeier J, Barnett MJ, Lang C, Krol E, Giegerich R, Long SR, Becker A. 2013. Global mapping of transcription start sites and promoter motifs in the symbiotic alpha-proteobacterium *Sinorhizobium meliloti* 1021. *BMC Genomics* 14:156.
75. Pearson CR, Tindall SN, Herman R, Jenkins HT, Bateman A, Thomas GH, Potts JR, Van der Woude MW. 2020. Acetylation of surface carbohydrates in bacterial pathogens requires coordinated action of a two-domain membrane-bound acyltransferase. *mBio* 11:e01364-20.
76. Leigh JA, Reed JW, Hanks JF, Hirsch AM, Walker GC. 1987. *Rhizobium meliloti* mutants that fail to succinylate their calcofluor-binding exopolysaccharide are defective in nodule invasion. *Cell* 51:579-587.
77. Stejskal EO, Schaefer J, Steger TR. 1978. High-resolution ^{13}C nuclear magnetic resonance in solids. *Faraday Symp Chem Soc* 13:56-62.
78. Jansson PE, Kenne L, Lindberg B, Ljunggren H, Lönngren J, Rudén U, Svensson S. 1977. Demonstration of an octasaccharide repeating unit in the extracellular polysaccharide of *Rhizobium meliloti* by sequential degradation. *J Am Chem Soc* 99:3812-3815.
79. Hager FF, Sutzl L, Stefanovic C, Blaukopf M, Schaffer C. 2019. Pyruvate substitutions on glycoconjugates. *Int J Mol Sci* 20:4929.
80. Ivashina TV, Fedorova EE, Ashina NP, Kalinchuk NA, Druzhinina TN, Shashkov AS, Shibaev VN, Ksenzenko VN. 2010. Mutation in the *pssM* gene encoding ketal pyruvate

- transferase leads to disruption of *Rhizobium leguminosarum* bv. *viciae*-*Pisum sativum* symbiosis. J Appl Microbiol 109:731-742.
81. Sharma S, Erickson KM, Troutman JM. 2017. Complete tetrasaccharide repeat unit biosynthesis of the immunomodulatory *Bacteroides fragilis* capsular polysaccharide A. ACS Chem Biol 12:92-101.
 82. Hager FF, López-Guzmán A, Krauter S, Blaukopf M, Polter M, Brockhausen I, Kosma P, Schäffer C. 2018. Functional characterization of enzymatic steps involved in pyruvylation of bacterial secondary cell wall polymer fragments. Front Microbiol 9:1356.
 83. Katzen F, Ferreiro DU, Oddo CG, Ielmini MV, Becker A, Pühler A, Ielpi L. 1998. *Xanthomonas campestris* pv. *campestris* gum mutants: effects on xanthan biosynthesis and plant virulence. J Bacteriol 180:1607-1617.
 84. Marzocca MP, Harding NE, Petroni EA, Cleary JM, Ielpi L. 1991. Location and cloning of the ketal pyruvate transferase gene of *Xanthomonas campestris*. J Bacteriol 173:7519-7524.
 85. Bianco MI, Toum L, Yaryura PM, Mielnichuk N, Gudesblat GE, Roeschlin R, Marano MR, Ielpi L, Vojnov AA. 2016. Xanthan pyruvilation is essential for the virulence of *Xanthomonas campestris* pv. *campestris*. Mol Plant Microbe Interact 29:688-699.
 86. Schmid J, Sieber V, Rehm B. 2015. Bacterial exopolysaccharides: biosynthesis pathways and engineering strategies. Front Microbiol 6:496.
 87. Sambrook J, Fritsch EF, Maniatis TA. 1989. Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
 88. Beringer JE. 1974. R factor transfer in *Rhizobium leguminosarum*. J Gen Microbiol 84:188-198.
 89. Finan TM, Kunkel B, De Vos GF, Signer ER. 1986. Second symbiotic megaplasmid in *Rhizobium meliloti* carrying exopolysaccharide and thiamine synthesis genes. J Bacteriol 167:66-72.
 90. Quandt J, Hynes MF. 1993. Versatile suicide vectors which allow direct selection for gene replacement in Gram-negative bacteria. Gene 127:15-21.
 91. Pogulis RJ, Vallejo AN, Pease LR. 1996. In vitro recombination and mutagenesis by overlap extension PCR. Methods Mol Biol 57:167-176.

92. Quan J, Tian J. 2009. Circular polymerase extension cloning of complex gene libraries and pathways. PLoS One 4:e6441.
93. Martin MO, Long SR. 1984. Generalized transduction in *Rhizobium meliloti*. J Bacteriol 159:125-129.
94. Khan SR, Gaines J, Roop II RM, Farrand SK. 2008. Broad-host-range expression vectors with tightly regulated promoters and their use to examine the influence of TraR and TraM expression on Ti plasmid quorum sensing. Appl Environ Microbiol 74:5053-5062.
95. Morcombe CR, Zilm KW. 2003. Chemical shift referencing in MAS solid state NMR. J Magn Reson 162:479-486.
96. Atalla RH, VanderHart DL. 1999. The role of solid state ^{13}C NMR spectroscopy in studies of the nature of native celluloses. Solid State Nucl Magn Reson 15:1-19.
97. Kolodziejewski W, Klinowski J. 2002. Kinetics of cross-polarization in solid-state NMR: A guide for chemists. Chem Rev 102:613-628.
98. Reuber TL, Walker GC. 1993. The acetyl substituent of succinoglycan is not necessary for alfalfa nodule invasion by *Rhizobium meliloti* Rm1021. J Bacteriol 175:3653-3655.
99. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785-786.
100. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 37:420-423.
101. Meade HM, Long SR, Ruvkun GB, Brown SE, Ausubel FM. 1982. Physical and genetic characterization of symbiotic and auxotrophic mutants of *Rhizobium meliloti* induced by transposon Tn5 mutagenesis. J Bacteriol 149:114-122.

FIGURES

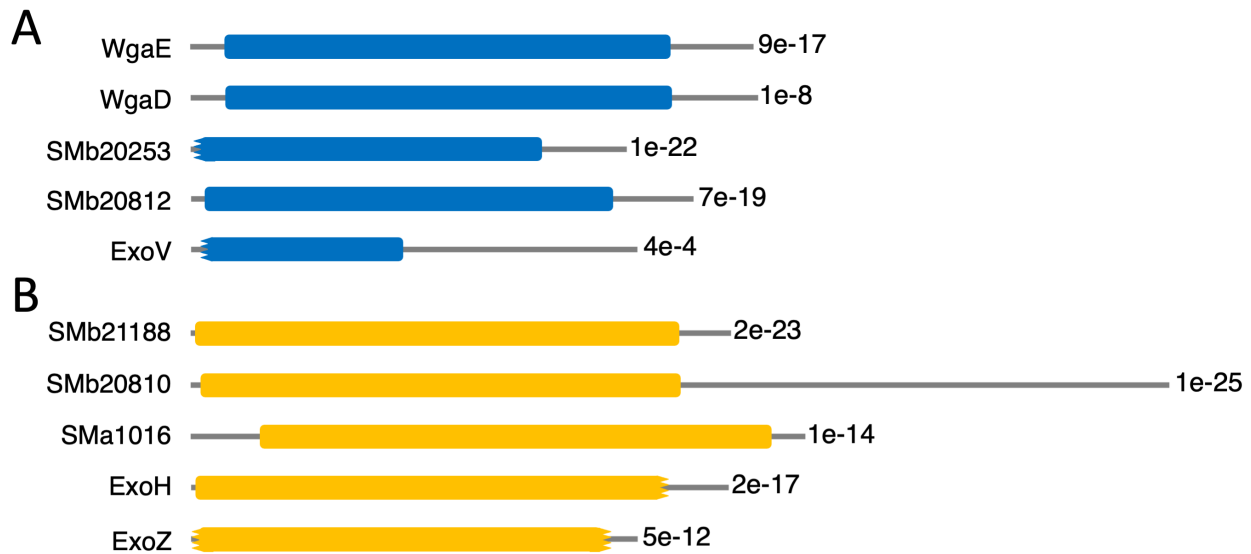


Figure 1. Predicted pyruvyltransferases and acyltransferases in *S. meliloti*. (A) Regions with similarity to polysaccharide pyruvyltransferase domains (Pfam PF04320) are shown in blue. (B) Regions with similarity to cytoplasmic membrane-located acyltransferase domains (Pfam PF01757), including *S. meliloti* ExoZ, the acyltransferase that acetylates succinoglycan (98), are shown in gold. Pfam release 32.0 was used to identify proteins and predict domains (73). Rectangles with smooth ends indicate a full-length match to the Pfam profile hidden Markov model (HMM). Jagged ends indicate a match to a portion of the Pfam HMM. At the end of each protein schematic, an E-value for the Pfam domain match is shown. The E-value is the number of hits expected to have a score equal or better than the value shown by chance alone. SignalP version 4.1 (99) and version 5.0 (100) predicted that none of the proteins contain an N-terminal signal sequence and thus are unlikely to be secreted.

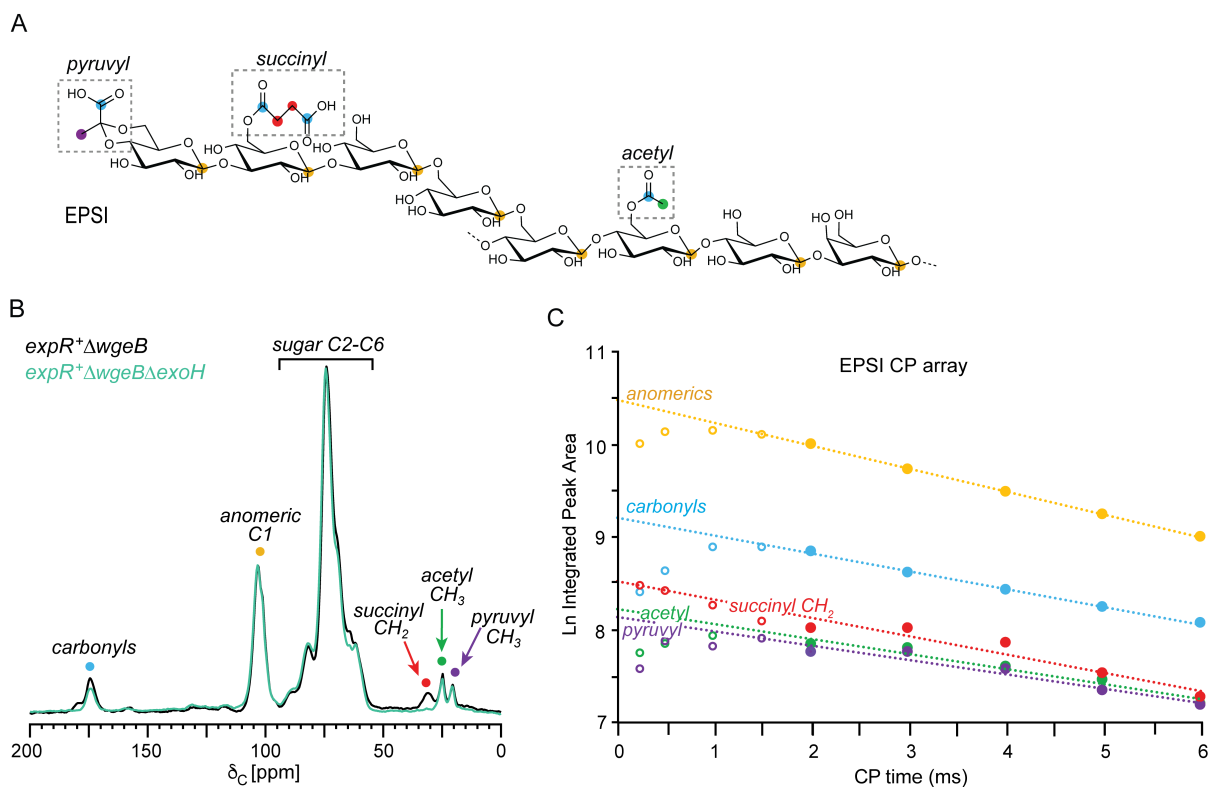


Figure 2. Analysis of *S. meliloti* succinoglycan (EPSI) (A) Structure of succinoglycan (EPSI) containing pyruvyl, acetyl, and O-succinyl substituents on a backbone of octamers composed of glucose and galactose (47, 48). Colored dots indicate distinct carbons with carbonyls (blue), anomeric C1 (yellow), succinyl CH₂ (red), acetyl CH₃ (green), and pyruvyl CH₃ (purple) groups. (B) ¹³C CPMAS NMR spectrum of EPSI purified from the *expR*⁺ Δ *wgeB* strain (DW37, black) compared to EPSI from the *expR*⁺ Δ *wgeB* Δ *exoH* strain (DW50, green). Number of scans: 32,768. (C) Quantitative CP array plot of integrated intensities for fitted peaks of the spectrum for EPSI purified from the *expR*⁺ Δ *wgeB* strain (DW37). Tabulated data and scan counts for each experiment are provided in Table S2.

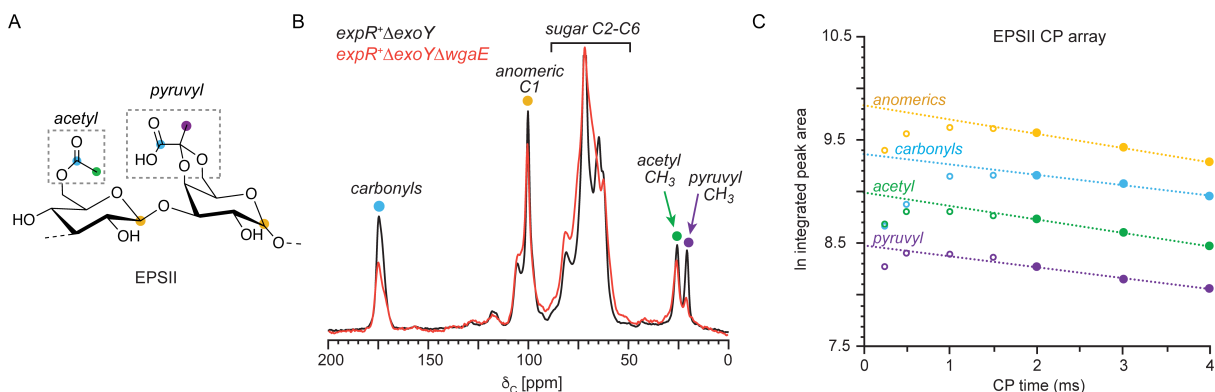


Figure 3. Analysis of *S. meliloti* galactoglucan (EPSII) and validation of novel pyruvyltransferase WgaE. (A) Structure of *S. meliloti* galactoglucan (EPSII) containing the pyruvyl and acetyl substituents on a backbone of glucose-galactose repeating units (49, 50). Colored dots indicate distinct carbons with carbonyls (blue), anomeric C1 (yellow), acetyl CH₃ (green), and pyruvyl CH₃ (purple) groups. (B) Solid state NMR spectrum of EPSII purified from the *expR*⁺Δ*exoY* strain (DW11, black) compared to that from the *expR*⁺Δ*exoY*Δ*wgaE* strain (CH23, red). Number of scans: 32,768. (C) Quantitative CP array plot of integrated intensities for fitted peaks of the spectrum for EPSII purified from the *expR*⁺Δ*exoY* (DW11) strain. Tabulated data and scan counts for each experiment are provided in Table S3.

TABLE 1. Strains and plasmids used in this study.

Strain or plasmid	Description	Reference
<i>S. meliloti</i> strains		
Rm1021	Wild type SU47; Sm ^r	(101)
MB1187	Rm1021, <i>expR</i> corrected; Sm ^r	This study
DW11	MB1187, Δ <i>exoY</i> ; Sm ^r	This study
DW28	MB1187, 22-kb <i>exo</i> deletion (Δ <i>exoZ-exoP</i>); Sm ^r	This study
DW37	MB1187, Δ <i>wgeB</i> (<i>SMb21313</i>); Sm ^r	This study
DW50	MB1187, Δ <i>wgeB</i> Δ <i>exoH</i> ; Sm ^r	This study
DW128	DW11, Δ <i>wgaE</i> (<i>SMb21322</i>); pDW24; Sm ^r Nm ^r	This study
CH20	DW11, Δ <i>SMb21188</i> ; Sm ^r	This study
CH21	DW11, Δ <i>SMb20810</i> ; Sm ^r	This study
CH22	DW11, Δ <i>SMa1016</i> ; Sm ^r	This study
CH23	DW11, Δ <i>wgaE</i> (<i>SMb21322</i>); Sm ^r	This study
CH25	DW11, Δ <i>wgaD</i> Δ <i>wgaE</i> ; Sm ^r	This study
MB1228	Rm1021, Δ <i>wgaD</i> (<i>SMb21321</i>); Sm ^r	This study
MB1235	DW11, Δ <i>wgaD</i> (<i>SMb21321</i>); Sm ^r	This study
MB1245	DW37, Δ <i>wgaD</i> ; Sm ^r	This study
MB1279	DW28, Δ <i>SMb20810</i> Δ <i>SMb21188</i> Δ <i>SMa1016</i> ; Sm ^r	This study
Plasmids		
pRK600	ColE1; provides RK2 transfer functions; Cm ^r	(89)
pJQ200SK	<i>sacB</i> vector; P15a <i>ori</i> , nonreplicative in <i>S. meliloti</i> ; Gm ^r	(90)
pMB971	pJQ200SK, <i>expR</i> ⁺ DNA, used to correct Rm1021 <i>expR</i> allele; Gm ^r	This study
pDW1	pJQ200SK, for making Δ <i>exoH</i> ; Gm ^r	This study
pDW3	3' <i>exoP</i> and 3' <i>exoZ</i> for <i>exo</i> region deletion; Gm ^r	This study
pDW5	pJQ200SK, for making Δ <i>wgeB</i> ; Gm ^r	This study
pCH1	pJQ200SK, for making Δ <i>SMb21188</i> ; Gm ^r	This study
pCH2	pJQ200SK, for making Δ <i>SMb20810</i> ; Gm ^r	This study
pCH3	pJQ200SK, for making Δ <i>SMa1016</i> ; Gm ^r	This study
pCH4	pJQ200SK, for making Δ <i>wgaE</i> ; Gm ^r	This study
pCH5	pJQ200SK, for making Δ <i>wgaD</i> ; Gm ^r	This study
pCH6	pJQ200SK, for making Δ <i>wgaD</i> Δ <i>wgaE</i> ; Gm ^r	This study
pSRKKm	Broad-host-range <i>Plac</i> expression vector; Km ^r	(94)
pDW24	pSRKKm, <i>wgaE</i> ; Km ^r	This study
pMB1035	pSRKKm, <i>wgaD</i> ; Km ^r	This study

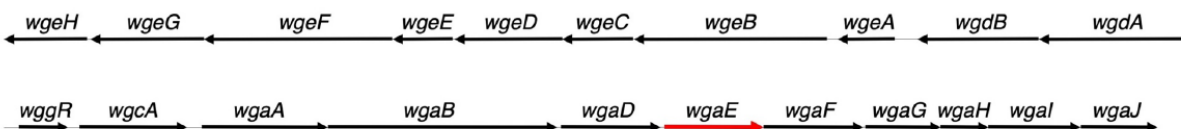


Figure S1. The 27.4-kb *S. meliloti* region containing EPSII (galactoglucan) biosynthesis genes. The region shown corresponds to nucleotides 969820 to 997228 on the pSymB megaplasmid. *wgaE* encoding a ketal pyruvyltransferase is shown in red.

		20		40			
WgaE	MR--IL	LTGIPSYLQRTVS	CVSCAEVRHRPYFDEVR	TKKE	38		
WgaD	MRPRIL	VTGIPGHYTRLAN	CAQCLSVSYSERQKQPE	TKEE	40		
		60		80			
WgaE	LIEQVKK	IANTCNYLIGEGAAH	ALRGHDVTYLPFWHLVN	S	78		
WgaD	FLQELRN	ISNTCNYLIGEGALR	AIAPH-AKQVPFWHLN	C	79		
		100		120			
WgaE	RNSDEAYES	LNNEEFDICVFAS	ANLLRPCYSADLEAEV	FEK	118		
WgaD	SQNGVGLDEF	NANFDICVFETC	ANLLRKCLSDADAEAEV	LGK	119		
		140		160			
WgaE	LKMPVV	IMGIGIQRR	REGLENLPAGTLKF	LEVLRNKE	SFF	158	
WgaD	LKMPV	VMLGIGLQNR	RDLENSLPECTKRL	LDVLKER	EHYF	159	
		180		200			
WgaE	LTRGYFTA	EFLREQGMKF	VKPTGCCPSLFFAP	AEMKRSLAA	198		
WgaD	LTRGFETAG	FLKDQGF	SYVQPTGCCPSIYLM	PHNMRTSLKK	199		
		220		240			
WgaE	LANPELASSQK	IAFGGYLC	SVADTIVDAHAL	LKPD	SVASY	238	
WgaD	LPNV	PVGKARTI	-FSCGYLC	ANHDCIVDA	AALAPEGSRPQY	238	
		260		280			
WgaE	VIQDEV	VVAYNLS	LPVDDDL	PVYDRASGRIT	QGTAYKHSE	EK	278
WgaD	VIQDEF	FLHFDMNVE	ANGDGRVYDS	ASGTMLGELS	YPGTER		278
		300		320			
WgaE	WQRKHELLV	FFDTNQWR	SWVSARD	LCFCRRFHGCI	IIGMQA	318	
WgaD	LKTPFDVRT	FFDTNQWR	AWASMD	FNFCRRFHGSI	IIAMQA	318	
		340		360			
WgaE	GVP	SLMI	AVDDRMREMLE	FIFPYME	AAVWNREQDR	KAYL	358
WgaD	AVP	SLMV	AVDDRMREML	GYTGLPAIDA	VEVDKAENR	AEFV	358
		380					
WgaE	ASF	LSKIDSQAVI	DRYSACETNFR	NALAHVGL	--	390	
WgaD	ADH	LAGL	NASELVDRYS	DRERTFRSAL	REIGIGQ	392	

Figure S2. Amino acid sequence alignment of *S. meliloti* pyruvyltransferases WgaE (SMb21322) and WgaD (SMb21321). Identical amino acid residues are indicated in red.

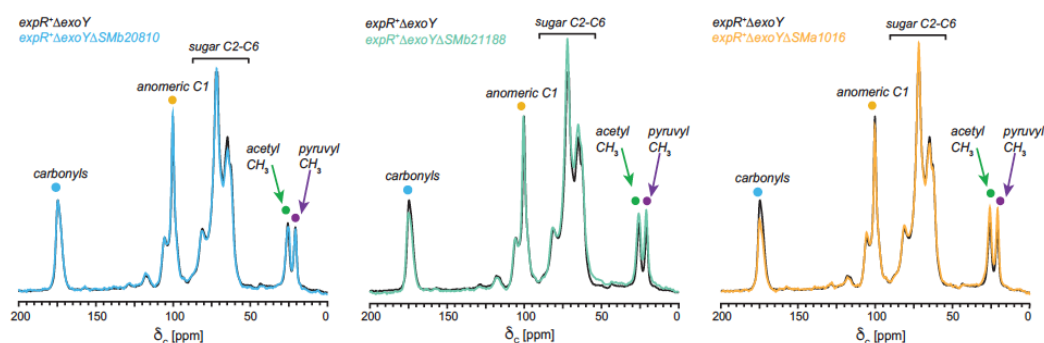


Figure S3. ^{13}C CPMAS solid-state NMR spectrum of galactoglucan (EPSII) purified from the *expR*⁺ Δ *exoY* strain (DW11, black) compared to spectra of EPSII from *expR*⁺ Δ *exoY* derivative strains each carrying a deletion of a putative acyltransferase gene. SMB20810 acyltransferase (blue, CH21); SMB21188 acyltransferase (green, CH20); and SMA1016 acyltransferase (yellow, CH22). Colored dots indicate distinct carbons with carbonyls (blue), anomeric C1 (yellow), acetyl CH_3 (green), and pyruvyl CH_3 (purple) groups. Each spectrum is the result of 32,768 scans.

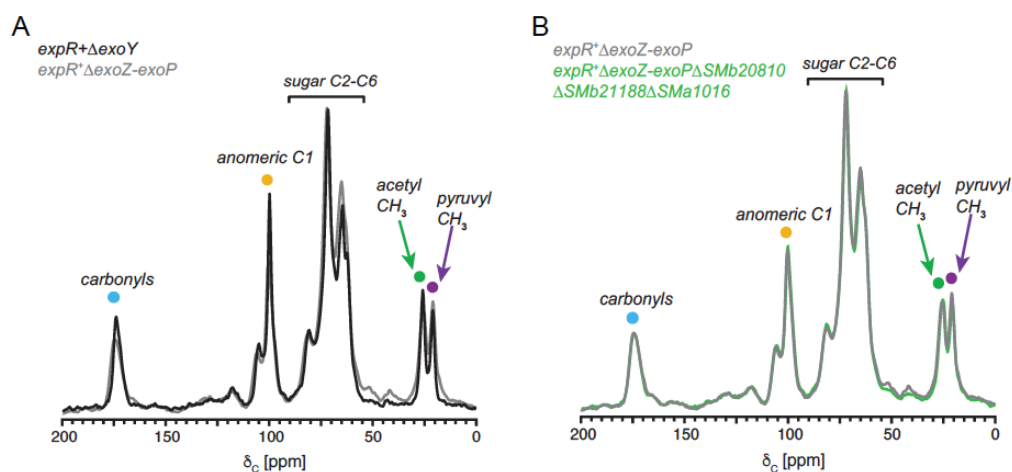


Figure S4. ^{13}C CPMAS solid-state NMR spectrum of galactoglucan (EPSII) purified from the *expR*⁺ 22-kb *exo* deletion (Δ *exoZ*-*exoP*) strain (DW28, grey) compared to (A) the wild type EPSII-producing strain *expR*⁺ Δ *exoY* (DW11, black) and (B) the spectrum of EPSII from strain MB1279 (green) lacking all five predicted Pfam PF01757 acetyl/acyltransferases (ExoH, ExoZ, SMA1016, SMB20810, SMB21188). Colored dots indicate distinct carbons with carbonyls (blue), anomeric C1 (yellow), acetyl CH_3 (green), and pyruvyl CH_3 (purple) groups. Each spectrum is the result of 32,768 scans.

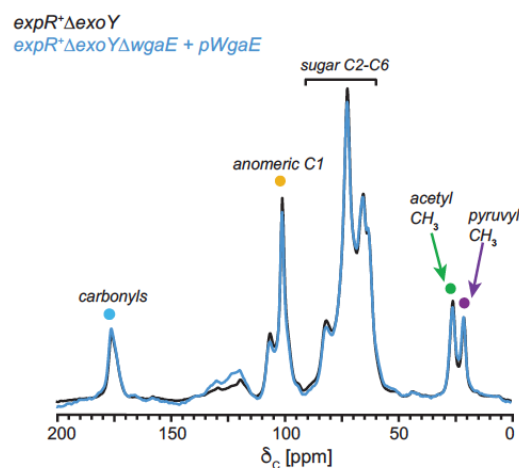


Figure S5. ^{13}C CPMAS solid-state NMR spectrum of galactoglucan (EPSII) purified from the *expR*⁺ Δ *exoY* strain (DW11, black) compared to the spectrum of EPSII from the *expR*⁺ Δ *exoY* Δ *wgaE* strain with *wgaE* ectopically expressed from plasmid pDW24 (DW128, blue) showing the pyruvyltransferase complementation. Colored dots indicate distinct carbons with carbonyls (blue), anomeric C1 (yellow), acetyl CH_3 (green), and pyruvyl CH_3 (purple) groups. Each spectrum is the result of 32,768 scans.

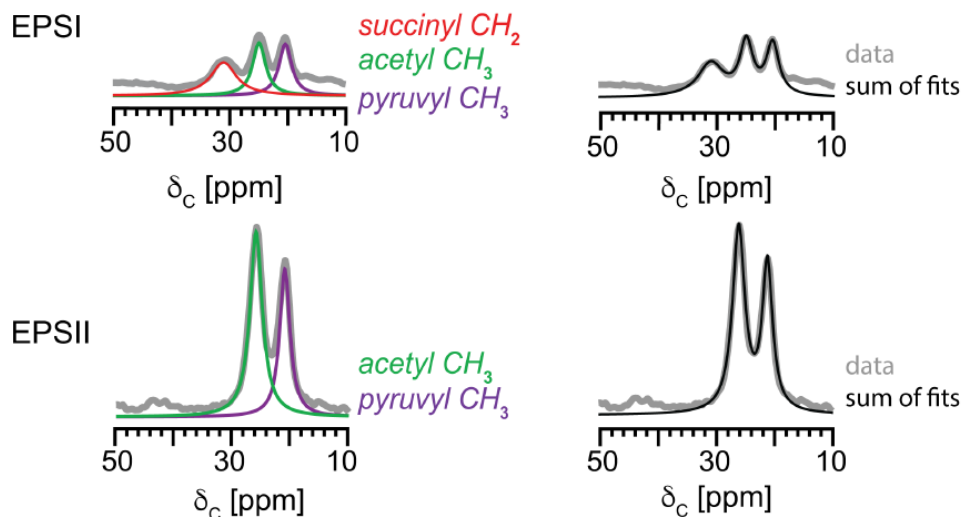


Figure S6. Comparison of experimental CPMAS spectra and peak fitting. ^{13}C CPMAS solid-state NMR spectra are shown in grey for spectra from the CP array experiment obtained with a CP time of 2ms as an example for EPSI (top) and EPSII (bottom). Modeled Lorentzian peak fits are shown as overlays: the modeled succinyl CH_2 (red), the modeled acetyl CH_3 (green), and the modeled pyruvyl CH_3 (purple). The sums of the modeled fits are in black.

Table S1. Genes in the EPSII biosynthesis cluster.

Gene	Former name(s)	Protein predicted function
<i>wgeH</i>	<i>expE8</i>	Unknown
<i>wgeG</i>	<i>expE7</i>	Glycosyltransferase
<i>wgeF</i>	<i>expE6</i>	Unknown
<i>wgeE</i>	<i>expE5</i>	Unknown, possible membrane-anchored protein
<i>wgeD</i>	<i>expE4</i>	Glycosyltransferase
<i>wgeC</i>	<i>expE3</i>	Methyltransferase
<i>wgeB</i>	<i>expE2</i>	Glycosyltransferase, bifunctional
<i>wgeA</i>	<i>expE1</i>	RTX-motif calcium-binding protein, secreted by WgdAB
<i>wgdB</i>	<i>expD2</i>	Type I secretion, membrane fusion protein
<i>wgdA</i>	<i>expD1</i>	Type I secretion, ATP-binding protein
<i>wggR</i>	<i>expG</i>	Transcription activator
<i>wgcA</i>	<i>expC</i>	Glycosyltransferase
<i>wgaA</i>	<i>expA1</i>	Unknown
<i>wgaB</i>	<i>expA2, expA3</i>	Glycosyltransferase, bifunctional
<i>wgaD</i>	<i>expA4</i>	Pyruvyltransferase
<i>wgaE</i>	<i>expA5</i>	Pyruvyltransferase
<i>wgaF</i>	<i>expA6</i>	Unknown, possible periplasmic location
<i>wgaG</i>	<i>expA7</i>	Glucose-1-phosphate thymidyltransferase
<i>wgaH</i>	<i>expA8</i>	dTDP-4-dehydrorhamnose 3,5-epimerase
<i>wgaI</i>	<i>expA9</i>	dTDP-glucose, 4, 6-dehydratase
<i>wgaJ</i>	<i>expA10</i>	dTDP-4-dehydrorhamnose reductase

Table S2. Numerical analysis of succinoglycan (EPSI) CPMAS NMR CP array peak integrals with spectral scan counts tabulated (A) and signal-to-noise analysis (B).

A

EPSI	CP Time (ms)									Fit at 0ms	e*x/1000
	0.25	0.5	1	1.5	2	3	4	5	6		
anomeric C1	9.991	10.130	10.140	10.090	9.996	9.723	9.483	9.244	9.004	10.475	35.419
	y = -0.2463x + 10.475 R ² = 0.9993										
carbonyls	8.406	8.631	8.879	8.890	8.848	8.612	8.431	8.249	8.067	9.211	10.011
	y = -0.1925x + 9.2114 R ² = 0.9969										
acetyl CH ₃	7.738	7.843	7.926	7.906	7.846	7.798	7.603	7.460	7.204	8.231	3.756
	y = -0.1622x + 8.231 R ² = 0.9582										
pyruval CH ₃	7.575	7.867	7.811	7.898	7.762	7.751	7.571	7.346	7.188	8.145	3.445
	y = -0.1553x + 8.1448 R ² = 0.946										
succinyl CH ₂	8.467	8.408	8.252	8.091	8.012	8.015	7.856	7.536	7.268	8.524	5.035
	y = -0.1967x + 8.5242 R ² = 0.9052										
scan count	8192	8192	8192	8192	8192	65536	65536	65536	65536		

B

EPSI	Std dev	Peak height	Std dev/Peak height
CP: 2ms			
acetyl CH ₃	0.19921	9.349	2.13%
pyruval CH ₃	0.19921	8.401	2.37%
succinyl CH ₂	0.19921	5.487	3.63%
CP: 3ms			
acetyl CH ₃	0.12166	8.043	1.51%
pyruval CH ₃	0.12166	7.751	1.57%
succinyl CH ₂	0.12166	4.771	2.55%
CP: 4ms			
acetyl CH ₃	0.11915	6.817	1.75%
pyruval CH ₃	0.11915	6.498	1.83%
succinyl CH ₂	0.11915	3.733	3.19%
CP: 5ms			
acetyl CH ₃	0.10652	5.312	2.01%
pyruval CH ₃	0.10652	5.345	1.99%
succinyl CH ₂	0.10652	2.985	3.57%
CP: 6ms			
acetyl CH ₃	0.10475	4.331	2.42%
pyruval CH ₃	0.10475	4.288	2.44%
succinyl CH ₂	0.10475	2.200	4.76%

Table S3. Numerical analysis of galactoglucan (EPSII) CPMAS NMR CP array peak integrals with spectral scan counts tabulated (A) and signal-to-noise analysis (B).

A

EPSII	CP Time (ms)							Fit at 0ms	e ^x /1000
	0.25	0.5	1	1.5	2	3	4		
anomeric C1	9.402	9.561	9.623	9.608	9.569	9.434	9.295	9.845	18.862
	y = -0.1374x + 9.8449 R ² = 0.9999								
carbonyls	8.662	8.875	9.153	9.158	9.162	9.079	8.962	9.368	11.703
	y = -0.0999x + 9.3676 R ² = 0.9909								
acetyl CH ₃	8.686	8.808	8.809	8.766	8.735	8.603	8.474	8.995	8.059
	y = -0.1302x + 8.9946 R ² = 0.9999								
pyruval CH ₃	8.273	8.403	8.395	8.361	8.278	8.156	8.063	8.488	4.858
	y = -0.1076x + 8.4884 R ² = 0.9943								
scan count	4096	4096	4096	4096	4096	4096	4096		

B

EPSII	Std dev	Peak height	Std dev/Peak height
CP: 2ms			
acetyl CH ₃	0.31939	38.92	0.82%
pyruval CH ₃	0.31939	31.15	1.03%
CP: 3ms			
acetyl CH ₃	0.33494	34.06	0.98%
pyruval CH ₃	0.33494	28.14	1.19%
CP: 4ms			
acetyl CH ₃	0.32465	30.24	1.07%
pyruval CH ₃	0.32465	25.18	1.29%

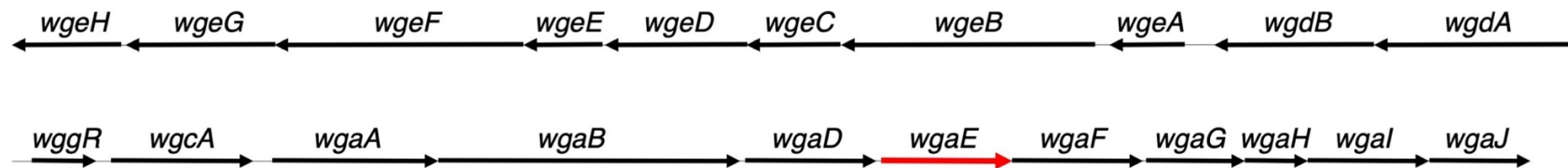


Figure S1. The 27.4-kb *S. meliloti* region containing EPSII (galactoglucan) biosynthesis genes. The region shown corresponds to nucleotides 969820 to 997228 on the pSymB megaplasmid. *wgaE* encoding a ketal pyruvyltransferase is shown in red.

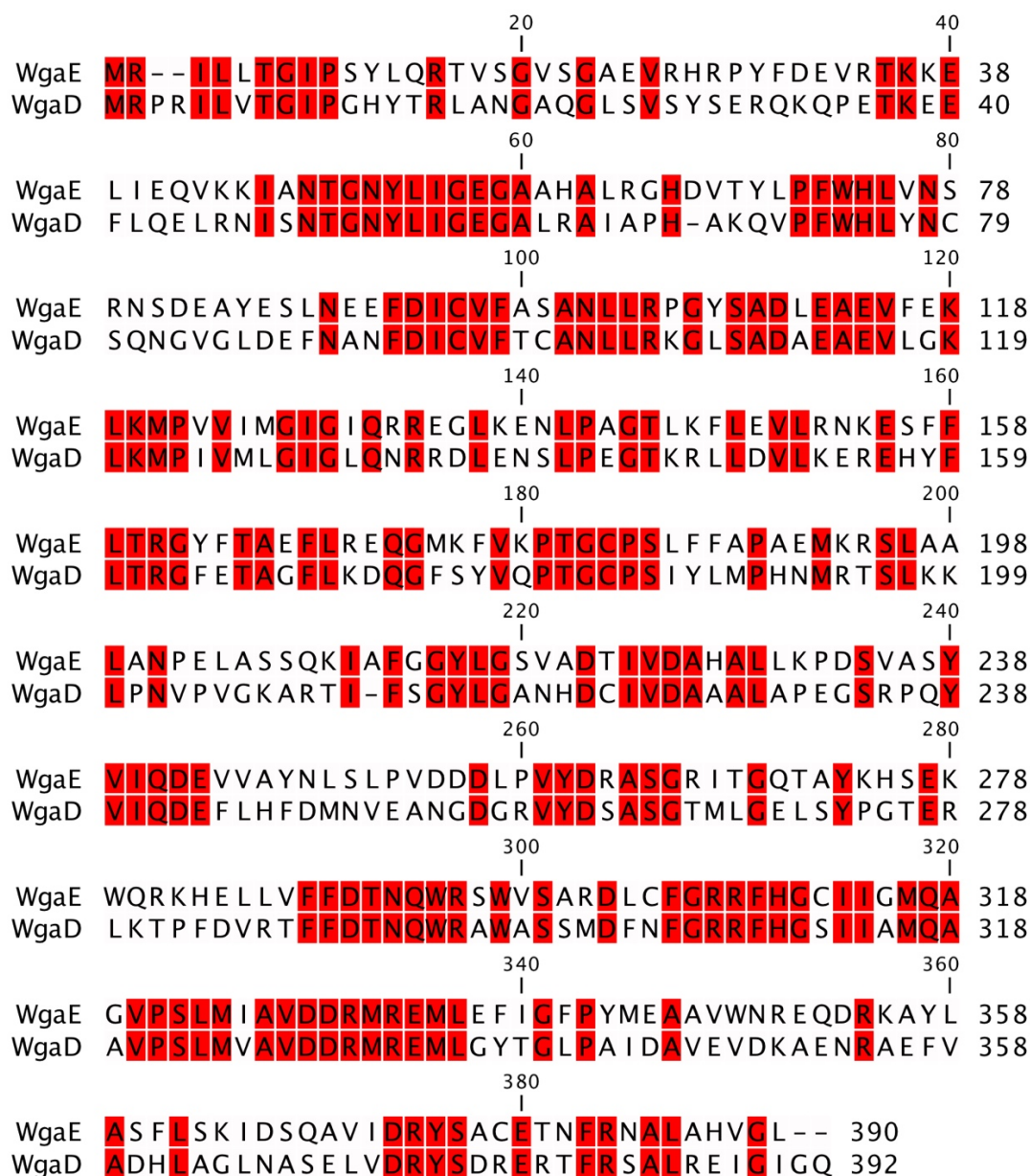


Figure S2. Amino acid sequence alignment of *S. meliloti* pyruvyltransferases WgaE (SMb21322) and WgaD (SMb21321). Identical amino acid residues are indicated in red.

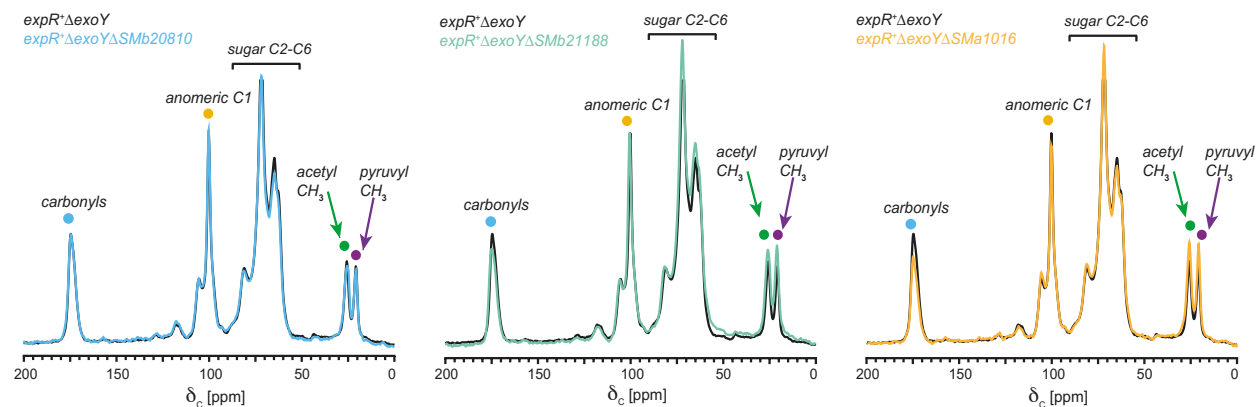


Figure S3. ^{13}C CPMAS solid-state NMR spectrum of galactoglucan (EPSII) purified from the $\text{expR}^+\Delta\text{exoY}$ strain (DW11, black) compared to spectra of EPSII from $\text{expR}^+\Delta\text{exoY}$ derivative strains each carrying a deletion of a putative acyltransferase gene. SMb20810 acyltransferase (blue, CH21); SMb21188 acyltransferase (green, CH20); and SMA1016 acyltransferase (yellow, CH22). Colored dots indicate distinct carbons with carbonyls (blue), anomeric C1 (yellow), acetyl CH_3 (green), and pyruvyl CH_3 (purple) groups. Each spectrum is the result of 32,768 scans.

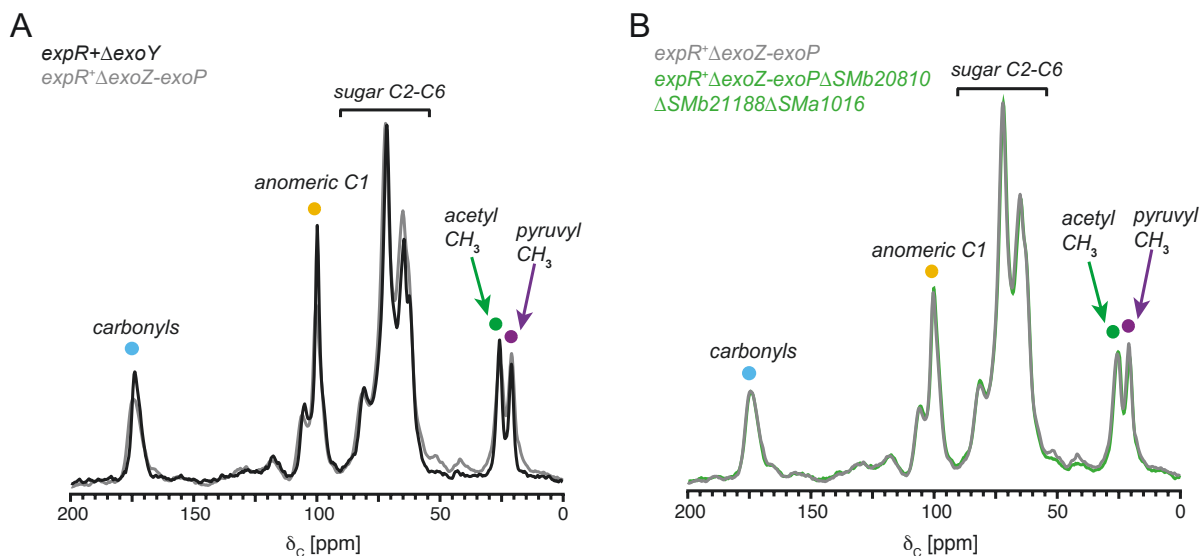


Figure S4. ^{13}C CPMAS solid-state NMR spectrum of galactoglucan (EPSII) purified from the $expR^+$ 22-kb *exo* deletion ($\Delta exoZ-exoP$) strain (DW28, grey) compared to (A) the wild type EPSII-producing strain $expR^+\Delta exoY$ (DW11, black) and (B) the spectrum of EPSII from strain MB1279 (green) lacking all five predicted Pfam PF01757 acetyl/acyltransferases (ExoH, ExoZ, Sma1016, SMb20810, SMb21188). Colored dots indicate distinct carbons with carbonyls (blue), anomeric C1 (yellow), acetyl CH_3 (green), and pyruvyl CH_3 (purple) groups. Each spectrum is the result of 32,768 scans.

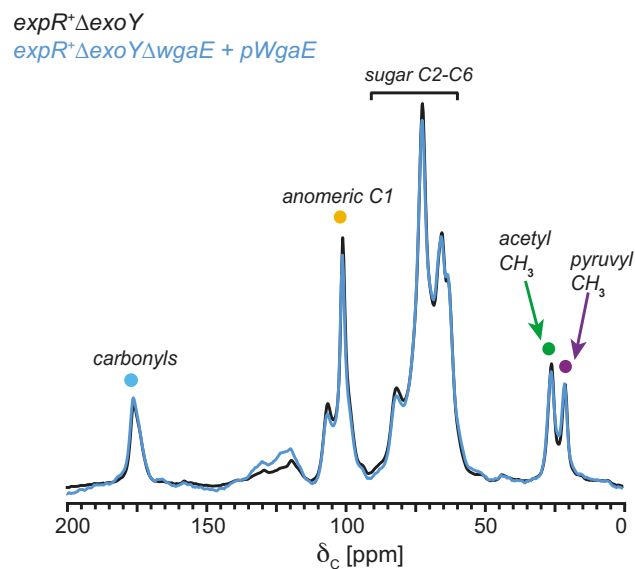


Figure S5. ^{13}C CPMAS solid-state NMR spectrum of galactoglucan (EPSII) purified from the *expR*⁺ Δ *exoY* strain (DW11, black) compared to the spectrum of EPSII from the *expR*⁺ Δ *exoY* Δ *wgaE* strain with *wgaE* ectopically expressed from plasmid pDW24 (DW128, blue) showing the pyruvyltransferase complementation. Colored dots indicate distinct carbons with carbonyls (blue), anomeric C1 (yellow), acetyl CH₃ (green), and pyruvyl CH₃ (purple) groups. Each spectrum is the result of 32,768 scans.

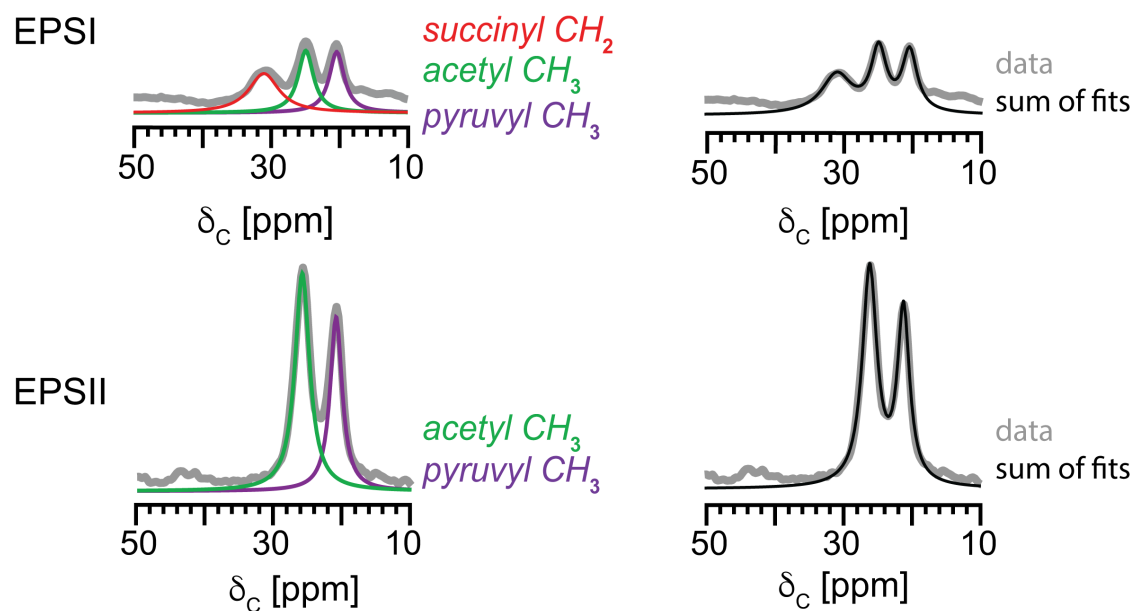


Figure S6. Comparison of experimental CPMAS spectra and peak fitting. ^{13}C CPMAS solid-state NMR spectra are shown in grey for spectra from the CP array experiment obtained with a CP time of 2ms as an example for EPSI (top) and EPSII (bottom). Modeled Lorentzian peak fits are shown as overlays: the modeled succinyl CH_2 (red), the modeled acetyl CH_3 (green), and the modeled pyruvyl CH_3 (purple). The sums of the modeled fits are in black.

Table S1. Genes in the EPSII biosynthesis cluster.

Gene	Former name(s)	Protein predicted function
<i>wgeH</i>	<i>expE8</i>	Unknown
<i>wgeG</i>	<i>expE7</i>	Glycosyltransferase
<i>wgeF</i>	<i>expE6</i>	Unknown
<i>wgeE</i>	<i>expE5</i>	Unknown, possible membrane-anchored protein
<i>wgeD</i>	<i>expE4</i>	Glycosyltransferase
<i>wgeC</i>	<i>expE3</i>	Methyltransferase
<i>wgeB</i>	<i>expE2</i>	Glycosyltransferase, bifunctional
<i>wgeA</i>	<i>expE1</i>	RTX-motif calcium-binding protein, secreted by WgdAB
<i>wgdB</i>	<i>expD2</i>	Type I secretion, membrane fusion protein
<i>wgdA</i>	<i>expD1</i>	Type I secretion, ATP-binding protein
<i>wggR</i>	<i>expG</i>	Transcription activator
<i>wgcA</i>	<i>expC</i>	Glycosyltransferase
<i>wgaA</i>	<i>expA1</i>	Unknown
<i>wgaB</i>	<i>expA2, expA3</i>	Glycosyltransferase, bifunctional
<i>wgaD</i>	<i>expA4</i>	Pyruvyltransferase
<i>wgaE</i>	<i>expA5</i>	Pyruvyltransferase
<i>wgaF</i>	<i>expA6</i>	Unknown, possible periplasmic location
<i>wgaG</i>	<i>expA7</i>	Glucose-1-phosphate thymidyltransferase
<i>wgaH</i>	<i>expA8</i>	dTDP-4-dehydrorhamnose 3,5-epimerase
<i>wgaI</i>	<i>expA9</i>	dTDP-glucose, 4, 6-dehydratase
<i>wgaJ</i>	<i>expA10</i>	dTDP-4-dehydrorhamnose reductase

Table S2. Numerical analysis of succinoglycan (EPSI) CPMAS NMR CP array peak integrals with spectral scan counts tabulated (A) and signal-to-noise analysis (B).

A

EPSI	CP Time (ms)									Fit at 0ms	e ^x /1000
	0.25	0.5	1	1.5	2	3	4	5	6		
anomeric C1	9.991	10.130	10.140	10.090	9.996	9.723	9.483	9.244	9.004	10.475	35.419
	y = -0.2463x + 10.475 R ² = 0.9993										
carbonyls	8.406	8.631	8.879	8.890	8.848	8.612	8.431	8.249	8.067	9.211	10.011
	y = -0.1925x + 9.2114 R ² = 0.9969										
acetyl CH ₃	7.738	7.843	7.926	7.906	7.846	7.798	7.603	7.460	7.204	8.231	3.756
	y = -0.1622x + 8.231 R ² = 0.9582										
pyruval CH ₃	7.575	7.867	7.811	7.898	7.762	7.751	7.571	7.346	7.188	8.145	3.445
	y = -0.1553x + 8.1448 R ² = 0.946										
succinyl CH ₂	8.467	8.408	8.252	8.091	8.012	8.015	7.856	7.536	7.268	8.524	5.035
	y = -0.1967x + 8.5242 R ² = 0.9052										
scan count	8192	8192	8192	8192	8192	65536	65536	65536	65536		

B

EPSI	Std dev	Peak height	Std dev/Peak height
CP: 2ms			
acetyl CH ₃	0.19921	9.349	2.13%
pyruval CH ₃	0.19921	8.401	2.37%
succinyl CH ₂	0.19921	5.487	3.63%
CP: 3ms			
acetyl CH ₃	0.12166	8.043	1.51%
pyruval CH ₃	0.12166	7.751	1.57%
succinyl CH ₂	0.12166	4.771	2.55%
CP: 4ms			
acetyl CH ₃	0.11915	6.817	1.75%
pyruval CH ₃	0.11915	6.498	1.83%
succinyl CH ₂	0.11915	3.733	3.19%
CP: 5ms			
acetyl CH ₃	0.10652	5.312	2.01%
pyruval CH ₃	0.10652	5.345	1.99%
succinyl CH ₂	0.10652	2.985	3.57%
CP: 6ms			
acetyl CH ₃	0.10475	4.331	2.42%
pyruval CH ₃	0.10475	4.288	2.44%
succinyl CH ₂	0.10475	2.200	4.76%

Table S3. Numerical analysis of galactoglucan (EPSII) CPMAS NMR CP array peak integrals with spectral scan counts tabulated (A) and signal-to-noise analysis (B).

A

EPSII	CP Time (ms)							Fit at 0ms	e^x/1000
	0.25	0.5	1	1.5	2	3	4		
anomeric C1	9.402	9.561	9.623	9.608	9.569	9.434	9.295	9.845	18.862
	y = -0.1374x + 9.8449 R² = 0.9999								
carbonyls	8.662	8.875	9.153	9.158	9.162	9.079	8.962	9.368	11.703
	y = -0.0999x + 9.3676 R² = 0.9909								
acetyl CH₃	8.686	8.808	8.809	8.766	8.735	8.603	8.474	8.995	8.059
	y = -0.1302x + 8.9946 R² = 0.9999								
pyruval CH₃	8.273	8.403	8.395	8.361	8.278	8.156	8.063	8.488	4.858
	y = -0.1076x + 8.4884 R² = 0.9943								
scan count	4096	4096	4096	4096	4096	4096	4096		

B

EPSII	Std dev	Peak height	Std dev/Peak height
CP: 2ms			
acetyl CH ₃	0.31939	38.92	0.82%
pyruval CH ₃	0.31939	31.15	1.03%
CP: 3ms			
acetyl CH ₃	0.33494	34.06	0.98%
pyruval CH ₃	0.33494	28.14	1.19%
CP: 4ms			
acetyl CH ₃	0.32465	30.24	1.07%
pyruval CH ₃	0.32465	25.18	1.29%