Analysis of the *Rhizobium meliloti* Genes *exo*U, *exo*V, *exo*W, *exo*T, and *exo*I Involved in Exopolysaccharide Biosynthesis and Nodule Invasion: *exo*U and *exo*W Probably Encode Glucosyltransferases

Anke Becker, Annette Kleickmann, Helge Küster, Mathias Keller, Walter Arnold, and Alfred Pühler

Lehrstuhl für Genetik, Fakultät für Biologie, Universität Bielefeld, Postfach 100131, D-33501 Bielefeld 1, Federal Republic of Germany

Received 19 May 1993. Accepted 9 July 1993.

Sequence analysis of a 5.780-kb DNA fragment originating from megaplasmid 2 of *Rhizobium meliloti* 2011 involved in biosynthesis of exopolysaccharide I (EPS I) and invasion of alfalfa nodules revealed the presence of five *exo* genes designated *exo*U, *exo*V, *exo*W, *exo*T, and *exo*I. ExoT resembled transmembrane proteins, whereas ExoI displayed a characteristic signal peptide. Sequence comparisons with several polysaccharide-polymerizing enzymes of both prokaryotic and eukaryotic origin indicated that *exo*W and *exo*U encode glucosyltransferases. Moreover, ExoV displayed weak homologies to the ExoO, ExoA, ExoL, and ExoM proteins of *R. meliloti*, which are also discussed as glucosyltransferases. Using *exo-lac*Z transcription fusions in connection with plasmid integration mutagenesis, promoters were identified in front of *exo*I, *exo*T, *exo*W, *exo*V, and *exo*U. *R. meliloti* 2011 strains with mutations in *exo*T, *exo*W, *exo*V, and *exo*U produced no detectable EPS I and were unable to infect alfalfa nodules, whereas *exo*I mutants synthesized a reduced amount of EPS I and did infect alfalfa nodules.

A complex interaction between *Rhizobium meliloti* and alfalfa plants results in the formation of nitrogen-fixing root nodules (Long 1989). The establishment of an effective symbiosis requires the coordinated exchange of signals between both symbiotic partners. Considerable evidence indicates that the acidic exopolysaccharide (EPS I) produced by *R. meliloti* is essential for nodule invasion. *R. meliloti* mutants deficient in EPS I biosynthesis are unable to invade and colonize the central nodule tissue (Long *et al.* 1988; Müller *et al.* 1988). EPS I constitutes a polymer of octasaccharide subunits composed of one galactose and seven glucose residues joined by β-1,3-, β-1,4-, and β-1,6-glucosidic bonds. The exopolysaccharide is modified with one pyruvate, one acetate, and one succinate group per subunit (Aman *et al.* 1981).

Recently, Battisti *et al.* (1992) reported that the exogenous application of the most charged tetrameric EPS I repeating unit to alfalfa roots corrected the nodule invasion defect of *R. meliloti* EPS I^- mutants. This finding was supported by an *exo*P mutant able to infect nodules, although only oligosaccharides and no EPS I polymer could be detected in the culture supernatant of this mutant (Becker *et al.* 1993b). This indicated that oligosaccharide precursors from the EPS I biosynthesis pathway are essential for the recognition of *R. meliloti* and its host plant during infection.

Genes involved in EPS I biosynthesis (*exo* genes) were found to be clustered on a 24-kb DNA region of the *R. meliloti* megaplasmid 2 (Long *et al.* 1988; Müller *et al.* 1988). Nucleotide sequence analysis of 17.7 kb of the *exo* gene cluster resulted in the identification of 14 *exo* genes designated *exo*B, *exo*Z (Buendia *et al.* 1991), *exo*Q, *exo*F, *exo*Y, *exo*X (Müller *et al.* 1993), *exo*H, *exo*K, *exo*L (Becker *et al.* 1993a), *exo*A, *exo*M, *exo*O, *exo*N, and *exo*P (Becker *et al.* 1993b). Potential functions could be proposed for several *exo* gene products. ExoB and most probably ExoN are involved in the synthesis of the nucleotide sugar precursors UDP-galactose and UDP-glucose (Buendia *et al.* 1991, Becker *et al.* 1993b). In addition, it was demonstrated that ExoX negatively influences the production of EPS I (Müller *et al.* 1993). Homologies of ExoY to the *Salmonella typhimurium* RfbP protein indicated that ExoY encodes a galactosyltransferase (Müller *et al.* 1993). Moreover, homologies of ExoO to the *R. meliloti* NodC protein suggested that *exo*O might code for a glucosyltransferase (Becker *et al.* 1993b). The homology of ExoK to the *Bacillus subtilis* BgsA showed that *exo*K might specify an endo-β-1,3-1,4-glucanase (Becker *et al.* 1993a). Since *exo*H mutants produced a nonsuccinylated EPS I, Leigh *et al.* (1987) proposed that ExoH is involved in succinylation. Additionally, ExoF (probably a periplasmic protein) as well as ExoQ and ExoP (probably transmembrane proteins) may participate in EPS I export (Becker *et al.* 1993b, Müller *et al.* 1993).

In this study we present the nucleotide sequence of the 5.780-kb DNA region located between *exo*X and *exo*H, which completes the DNA sequence of the 24-kb *exo*-gene cluster of *R. meliloti.* In addition, we report on the phenotypes of defined *exo* mutants and present data on the transcriptional organization of the gene region analyzed.

Corresponding author: Alfred Pühler.

MPMI Vol. 6, No. 6, 1993, pp. 735-744
© 1993 The American Phytopatholgical Society

RESULTS

Sequence analysis of a 5.780-kb DNA fragment of the *R. meliloti exo* gene cluster revealed the presence of five open reading frames designated *exo*U, *exo*V, *exo*W, *exo*T, and *exo*I.

To complete the DNA sequence of the 24-kb *R. meliloti exo*-gene cluster, we sequenced a 5.780-kb DNA fragment located between *exo*X and *exo*H (Becker *et al.* 1993a, Müller *et al.* 1993) (Fig. 1).

Five open reading frames (ORFs) preceded by potential ribosome binding sites (Table 1) could be identified on this DNA fragment (Fig. 2). Two ORFs located upstream from *exo*H and one ORF located downstream from *exo*X were in the same orientation as these genes and were designated *exo*I, *exo*T, and *exo*U, respectively. The location of *exo*T corresponded to the location of the complementation group *exo*T reported by Reuber *et al.* (1991). Although the *exo*I sequence contained four potential start codons, plasmid integration mutagenesis of this coding region (see below) indicated that the ATG at nucleotide position 5171 (Fig. 2) represented the start codon used *in vivo*.

Two additional ORFs designated *exo*W and *exo*V were oriented in the opposite direction. The length and the molecular weight of the deduced amino acid sequences are listed in Table 1.

ExoT resembled transmembrane proteins and ExoI contained a signal peptide sequence.

According to the hydrophobic profiles (Eisenberg *et al.* 1984) of the deduced amino acid sequences, ExoW, ExoV, and ExoU are most probably located in the cytoplasm. In contrast, ExoT displayed 13 hydrophobic helices (Fig. 3). According to the hydrophobic moment plot (Eisenberg *et al.* 1984; data not shown), six of these were tentatively classified as being transmembrane helices; four were characterized as being part of globular domains; and three were predicted to be membrane-surface associated (Fig. 3). We therefore presume that ExoT represents a transmembrane protein.

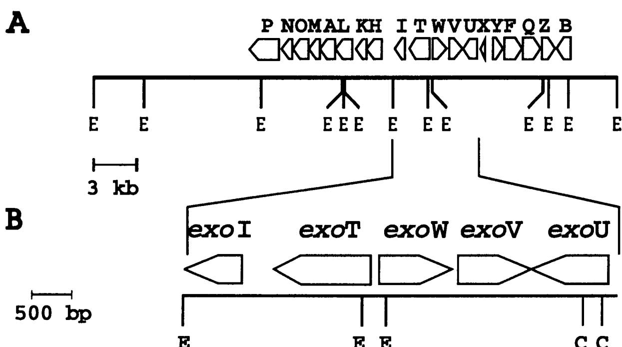


Fig. 1. Physical and genetic map of the *Rhizobium meliloti* 2011 DNA region on megaplasmid 2 involved in exopolyssacharide I biosynthesis and nodule infection. **A,** *Eco*RI restriction map of cosmid pRmAB839 (Becker *et al.* 1993a). Locations and designations of the *exo* genes already sequenced (Buendia *et al.* 1991; Becker *et al.* 1993a,b; Müller *et. al.* 1993) are indicated above the restriction map. **B,** Restriction map of a 5.780-kb DNA region carrying the genes *exo*I, *exo*T, *exo*W, *exo*V, and *exo*U. Arrows denote the location of coding regions. C, *Cla*I; E, *Eco*RI.

The N-terminal region of ExoI displayed significant similarities to prokaryotic signal peptides. According to the algorithm devised by von Heijne (1986), the −13/+2 region comprising amino acids 32–46 was found to have a score of 14.85. In addition, the potential cleavage site located between positions 44 and 45 conformed to the −3/−1 rule. Therefore, ExoI could be located in the periplasmic space.

ExoW and ExoU shared amino acid motifs with other polysaccharide-polymerizing enzymes

Sequence comparisons of ExoU and ExoW to other *R. meliloti exo* genes revealed homologies of 32 and 24% to the deduced amino acid sequence of the *exo*O gene (Becker *et al.* 1993b). In addition, searches of sequence databases indicated that both ExoU and ExoW belong to a family of polysaccharide synthases from several prokaryotic and eukaryotic organisms. Figure 4 shows an alignment of sequence parts from several of these proteins that are significantly homologous to the ExoO, ExoU, and ExoW sequences. Whereas the cellulose synthase function of the BesA from *Acetobacter xylinum* (Saxena *et al.* 1990) is based on experimental evidence, functions of other proteins have been concluded from genetic data and sequence homologies, as in the case of the NodC protein from *R. meliloti* and *R. loti* (Török *et al.* 1984; Atkinson and Long 1992; Debellé *et al.* 1992). The DG42 protein from *Xenopus laevis* (Rosa *et al.* 1988) is possibly involved in the synthesis of matrix polysaccharides by the developing embryo during gastrulation (Atkinson and Long 1992), whereas the Orf2 from *Anabaena* sp. strain PCC7120 is located in the vicinity of the *het*A gene, which takes part in the biosynthesis of the envelope polysaccharide (Holland and Wolk 1990).

The sequence alignment in Figure 4 revealed the presence of a number of highly conserved residues as well as the occurrence of several invariant residues. Since the EPS I subunit of *R. meliloti* is built up from one galactose and seven glucose residues and since ExoY has been shown to be a galactosyl transferase (Müller *et al.* 1993), we propose that ExoU and ExoW are likely to represent glucosyltransferases.

Additionally, ExoV displayed weak similarities to ExoW and ExoU as well as to ExoL, ExoA, and ExoM (Becker *et al.* 1993b) over its entire length (about 12–15%). Since these proteins have been thought to represent glucosyltransferases (Becker *et al.* 1993b), we propose that *exo*V could have a similar enzymatic function.

Exopolysaccharide production and nodule formation of defined *R. meliloti exo* mutants.

Mutants in *exo*I, *exo*T, and *exo*W have been constructed by insertion of a promoterless *lac*Z-Gm interposon (Becker *et al.*

Table 1. Putative ribosome binding sites of the *exo* genes and relevant features of the *exo* gene products[a]

Gene	Potential rbs	Distance to start codon (bp)	Gene product length (aa)	Molecular weight
*exo*U	GGAG	7	342	37,018
*exo*V	GGAG	6	316	35,322
*exo*W	AGGG	8	319	36,217
*exo*T	GGAA	9	494	53,316
*exo*I	GGGA	6	191	20,321

[a] aa, amino acids; rbs, ribosome binding site.

```
 S  *
 TCCTGAGCGGCTGGCGCCGCTCTCCTTCGCACTTGCAAAAAATTTTTCGCCTGACGCGTTTGCGACATGATCGCACGGCATGGTTCCAGC    90
                                                       M  T  A  A  A  P  T  D  V  C  I  I  I
 AACTTCGCGCCCTTGAGGCGCGACAGCCGATAACTCGCTGGAGGCCTCACATGACCGCAGCCGCGCCGACGGACGTCTGCATCATCATTT   180

 S  A  K  N  A  A  D  T  I  A  R  A  V  A  S  A  L  A  E  P  E  A  A  E  V  V  V  I  D  D
 CCGCGAAGAATGCCGCCGACACCATAGCGCGCGCGGTCGCTTCGGCGCTTGCCGAACCGGAGGCGGCAGAAGTCGTCGTCATCGATGACG   270
 G  S  T  D  D  S  A  S  V  A  R  A  A  D  D  G  T  G  R  L  N  V  V  R  F  E  E  N  R  G
 GCTCGACCGATGACAGCGCTTCGGTGGCACGCGCTGCGGATGACGGAACGGGGCGGCTGAATGTCGTTCGCTTCGAAGAAAACCGCGGCC   360
 P  A  A  A  R  N  H  A  I  A  I  S  H  S  P  L  I  G  V  L  D  A  D  D  F  F  F  P  G  R
 CCGCGGCCGCGCGCAATCATGCGATCGCGATCTCGCATTCTCCGCTTATCGGCGTGCTCGACGCAGATGATTTCTTCTTTCCGGGACGCC   450   e
154 L  G><Q  L  L  S  Q  D  G  W  D  F  I  A  D  N  I  A  F  I  D  A  A  Q  A  A  T  A  H  G
 TGGGCCAATTGCTTTCGCAGGACGGCTGGGACTTCATTGCGGACAACATCGCGTTCATCGATGCCGCGCAGGCGGCAACCGCGCATGGCA   540   x
 R  I  D  R  F  A  P  T  P  R  L  I  D  L  V  G  F  V  E  G  N  I  S  R  R  G  V  R  R  G
 GGATCGACCGATTCGCCCCTACCCCCCGGCTCATCGATCTCGTCGGCTTCGTCGAAGGAAACATATCGCGACGCGGCGTGCGGCGCGGCG   630   o
 E  I  G  F  L  K  P  L  M  R  R  A  F  L  D  Q  H  G  L  R  Y  N  E  T  L  R  L  G  E  D
 AAATAGGCTTTCTGAAGCCGCTGATGCGGCGCGCCTTCCTCGACCAGCATGGCCTGCGCTACAACGAGACGCTGCGCCTCGGCGAAGATT   720
 Y  D  L  Y  A  R  A  L  A  N  G  A  R  Y  K  I  I  H  S  C  G  Y  A  A  V  V  R  G  N  S                U
 ACGATCTCTACGCTCGCGCGCTCGCCAATGGCGCACGCTACAAGATCATCCACAGTTGCGGCTATGCCGCCGTCGTCCGCGGCAACTCGC   810
 L  S  G  S  H  R  T  I  D  L  K  R  L  Y  E  A  D  R  A  I  L  A  G  S  R  L  S  S  D  A
 TGAGCGGCAGTCACCGGACGATCGACCTCAAGCGTCTCTATGAGGCAGATCGTGCAATTCTCGCAGGGAGCCGACTGAGCAGCGACGCCG   900
 E  A  A  V  R  R  H  E  R  H  I  R  D  R  Y  E  L  R  H  F  L  D  L  K  N  Q  Q  G  F  G
 AAGCGGCGGTCCGCCGGCACGAGCGCCACATTCGCGACCGCTACGAACTACGGCATTTTCTCGATCTCAAGAACCAGCAGGGCTTTGGTC   990
 R  A  F  G  Y  A  L  T  H  P  A  A  L  P  A  I  I  G  G  I  L  A  D  K  T  E  R  F  R  P
 GTGCATTCGGCTACGCCCTGACCCACCCGGCGGCACTGCCGGCGATCATCGGCGGAATCCTTGCGGACAAGACCGAACGTTTTCGTCCGT  1080
 S  G  S  P  A  P  V  A  L  G  G  K  G  D  V  R  Y  L  L  E  T  L  A  V  D  Q  P  Q  K  *
 CCGGTTCGCCGGCGCCCGTCGCTCTCGGTGGCAAGGGCGACGTGCGCTATCTGCTCGAGACTCTTGCCGTGGATCAGCCTCAAAAATAGT  1170
2505      ><                                                                          *  F  Y  D
 CCCGGCGTATGCCGTCGAGCACCGTGCGGAAACGCTCCTTGCGCTCTGCCAGCGCACTGTCCTTGCTCAATTGCGGCGCCGCCCGGCTTG  1260
  R  R  I  G  D  L  V  T  R  F  R  E  K  R  E  A  L  A  S  D  K  S  L  Q  P  A  A  R  S  A
 CCTGCCACAGCGCCAGCGTCGCGGGCGCGGCGAGCGCTCGCTTGGCGGCAGCGCGCAATGAAGTCTGGCGTGGTTCACGCTCGGCGACCG  1350
  Q  W  L  A  L  T  A  P  A  A  L  A  R  K  A  A  A  R  L  S  T  Q  R  P  E  R  E  A  V  A
 CCAAGTCTCCGTCGATCTGTCGGCGGTTCGGGTCCTCCGGCTGCGGTTCCTTCGCCTGGAAATCCATGCCCCAGAAACGGGCACCCTTGA  1440
  L  D  G  D  I  Q  R  R  N  P  D  E  P  Q  P  E  K  A  Q  F  D  M  G  W  F  R  A  G  K  I
 TCATGGCCTCCGCCCGCGTCGACACCGGCACCCGCCGGGGGCGGTAGGTGACGCCGAGGGTCTGGGCCCAGTCGTTCCACTTGAAGCTGT  1530   e
  M  A  E  A  R  T  S  V  P  V  R  R  P  R  Y  T  V  G  L  T  Q  A  W  D  N  W  K  F  S  N
 TGATGCTGTGCGACGTGCTGACCGCGGTCCAGGGAACGCGGAACGCGTCGGCCAGGATGGCGCCGTGCATGGATTCAGCGACGATAAGCT  1620   x
  I  S  H  S  T  S  V  A  T  W  P  V  R  F  A  D  A  L  I  A  G  H  M  S  E  A  V  I  L  E
 CCGACTGGCCTATCTTTCGGATGACGTCCTTGGCTTCGCCACGCGGGTCGATATAGTTGAGCCCGACCGCGTCGCAGATCGCCGGCCACA  1710   o
  S  Q  G  I  K  R  I  V  D  K  A  E  G  R  P  D  I  Y  N  L  G  V  A  D  C  I  A  P  W  L
 ATCCGGCGATCGCCGATTCCCAGTGTGGTACGAAGCTCCGCTTATATATCTTTCGAAGGCCCTGAAACTCCGGAAGGTCGGCGACCATCA  1800
  G  A  I  A  S  E  W  H  P  V  F  S  R  K  Y  I  K  R  L  G  Q  F  E  P  L  D  A  V  M  V                V
 CCGCAGGATCGATGATACCGAGCTCGGGAGCCACGCCGACCTTCGCAGCCGTCAGCGGACCGCGGACCGAGCGGATATCCCATTCCTTAG  1890
  A  P  D  I  I  G  L  E  P  A  V  G  V  K  A  A  T  L  P  G  R  V  S  R  I  D  W  E  K  P
 GGTCGCTCATGTCGGGAAGCGTGCCGTAGCCGAAACCGCTGCCGATGACGAGCTTATGCGTAGCCTTGGGCAGCAGAGCGCGGTTCAAGA  1980
  D  S  M  D  P  L  T  G  Y  G  F  G  S  G  I  V  L  K  H  T  A  K  P  L  L  A  R  N  L  V
 CCGTTCCAACCCCCACGAGCAGCGTCTCGGGATAGACCTCGCGAAAACCGGGCAGCAGGAAATCCCAAAGCCACAGGTTGAGGTCGTCAC  2070
  T  G  V  G  V  L  L  T  E  P  Y  V  E  R  F  G  P  L  L  F  D  W  L  W  L  N  L  D  D  G
 CGAAATTTCCGTGCTGGGATTCCCAATAATACGGCTTCATGACCGTCTCCGTTTGCAATTGTGAGGCGACCTGACACGCGCGGTGGCGAA  2160
  F  N  G  H  Q  S  E  W  Y  Y  P  K  M
130                                                                  ><
 TGTCGTCGTTGATTTCGGGTGTTGTCAGGGCAAGCAGCGCGTGCCCGATCGGGCCGAAGGAGGTTAGCGGGATAGTTTTCCAACCGCCAA  2250
                                                                        *  R  S  L  K  G  V  A  L
 CTCGGCGGCGCTGCCGATCAATCGCGGGTCTCGCCAGAGCCAGCGGGCAAGAAGATCGAACTGCGGCAGTTTCCGGCGCTTTATCCGGCG  2340
  E  A  A  S  G  I  L  R  P  D  R  W  L  W  R  A  L  L  D  F  Q  P  L  K  R  R  K  I  R  R
 TGCCTGGCTCCATAGCGCCTGCCGCCGCGCGGTATGCTTGTAGGATCGGATCGCCTCCATCGCCTTCGGCCGGTTACGATATCGGCCTTC  2430
  A  Q  S  W  L  A  Q  R  R  A  T  H  K  Y  S  R  I  A  E  M  A  K  P  R  N  R  Y  R  G  E
 GAGCGTGTCGAGTGCAACCCAGGTATTGAACTGCTGCTTCAGGAACTGAGGCGAGTCGTTATCGATCGAATGAAAGATGTTGAGCCCTTC  2520
  L  T  D  L  A  V  W  T  N  F  Q  Q  K  L  F  Q  P  S  D  N  D  I  S  H  F  I  N  L  G  E            e
 GCCACGCACCGCGCCCGCTGCATCGCAAAGAACCACGCGCTTGGATGCGAGCACGCAATCGCAGAAGAAGAGCACGTCCTCGGCCGCTAG  2610
  G  R  V  A  G  A  A  D  C  L  V  V  R  K  S  A  L  V  C  D  C  F  F  L  V  D  E  A  A  L            x
 CTTGAGCGTGGCCTCGAAACGCACCTTTTCGAAGAGCTTGCGCCCGATCACCATGCAGGACATGTGGAGAAAGCTCCAGTTCTTCAGCAT  2700
  K  L  T  A  E  F  R  V  K  E  F  L  K  R  G  I  V  M  C  S  M  H  L  F  S  W  N  K  L  M            o
 GACGTCCTGAAGTTCGGGGAGCTCGACCACGAGGGGACTTTCCGAGAGGCGCGTGACGGTCTCGCTCTTTTCGAGATCCGCGACCCCGAA  2790
  V  D  Q  L  E  P  L  E  V  V  L  P  S  E  S  L  R  T  V  T  E  S  K  E  L  D  A  V  G  F
 ATGATAGTAGAAAGCGTCGCCGCCCGTGATCGAGGCCCAATAGCAATCGGCGTCGAAACGCGTCATGCTTTGATAGGCGTTCAGGAGATG  2880   W
  H  Y  Y  F  A  D  G  G  T  I  S  A  W  Y  C  D  A  D  F  R  T  M  S  Q  Y  A  N  L  L  H
 ATCCGGCGTCCAGACGTCGTCGGAATCGAGGAAAGCGACGAAGTCGCTGTCGGCAGGAACATTATCGAGACCGGTATTGCGCGCGCCACC  2970
  D  P  T  W  V  D  D  S  D  L  F  A  V  F  D  S  D  A  P  V  N  D  L  G  T  N  R  A  G  G
 CGGCCCGCCGTTAGGCTGGCGGATGACGGTGATGCGCTCCCTTTCTTCCTGGGCGAGCCCGGCGAGCTCGTCCGCTATCGGATAGGGGGA  3060
  P  G  G  N  P  Q  R  I  V  T  I  R  E  R  E  E  Q  A  L  G  A  L  E  D  A  I  P  Y  P  S
 TTCGTCATCGATCACGAGAACGTGGAAATCCTCGAGCGTCTGAGCAAAAACGGATGCCAGCGCGCGCCGGAGAATTCCGGGTTCCTTCTG  3150
  E  D  D  I  V  L  V  H  F  D  E  L  T  Q  A  F  V  S  A  L  A  R  R  L  I  G  P  E  K  Q
 GTAGTACGGAATGACAACTGTGAGCTTCGCCATCTTTTCGCCCCTGCGGTTTCTCTAAAAAGAGTCTGCGTGGTCTCCTCCCACCGGGTC  3240
  Y  Y  P  I  V  V  T  L  K  A  M
```

(Continued on next page)

Fig. 2. Nucleotide sequence of a 5.780-kb DNA fragment of megaplasmid 2 of *Rhizobium meliloti* 2011 carrying the genes *exo*I, *exo*T, *exo*W, *exo*V, and *exo*U. The nucleotide sequence of one DNA strand is presented in the 5′ to 3′ direction. The five open reading frames identified are marked by arrows. The deduced amino acid sequence is given in the one-letter code above the nucleotide sequence for genes encoded by the upper strand and below the nucleotide sequence for genes encoded by the lower strand. To connect the new sequence to the sequence reported by Müller *et al.* (1993), the last amino acid of ExoX is shown. Putative ribosome binding sites and the potential signal peptide of ExoI are underlined. Stop codons are marked by asterisks. Insertion sites of transposon Tn*5* or Tn*5*-B20 (Simon *et al.* 1989) are indicated by two arrow heads (><) above the nucleotide sequence. Designations of the respective *R. meliloti* transposon mutants are shown at the left side (154, Tn*5*-B20 mutant Rm154; 2505, Tn*5* mutant Rm2505; 130, Tn*5*-B20 mutant Rm130). The sequence has been submitted to the EMBL/GenBank/DDBJ databases under the accession number Z22646.

(Continued from preceding page)

```
                                                                                   M  T  P  T  V  N  A  K
GCGGCCGTATATCTAGAGTTTGTCCTTTGCATTCCTCCTTCAAAGGACTTTCACGGAATCCTCGCCAATGACCCCAACCGTTAACGCAAA  3330

 T  V  T  R  N  V  G  W  S  V  L  S  K  T  G  T  F  G  L  K  F  V  T  V  P  I  L  A  R  I
AACCGTAACGCGCAACGTCGGCTGGAGCGTTCTTTCCAAGACAGGGACATTCGGGCTTAAATTTGTCACGGTGCCGATTCTGGCCCGCAT  3420
 L  S  P  E  E  F  G  A  V  A  V  A  L  T  V  V  Q  F  L  A  M  I  G  G  A  G  L  T  S  A
TCTGTCTCCCGAGGAATTCGGCGCCGTCGCGGTTGCGCTCACCGTGGTGCAGTTCCTCGCCATGATCGGCGGCGCCGGCCTCACCTCCGC  3510
 L  V  I  Q  Q  H  E  E  M  E  T  V  H  S  V  F  W  A  N  L  A  I  A  L  M  M  A  L  G  L
ACTCGTCATCCAGCAGCATGAGGAGATGGAAACCGTGCACTCGGTTTTCTGGGCAAACCTCGCCATCGCGCTCATGATGGCGCTCGGACT  3600
 F  V  F  A  E  P  L  A  T  L  L  G  A  P  E  A  A  Y  L  L  R  I  M  S  L  L  I  P  L  Q
GTTCGTTTTTGCCGAGCCCCTGGCCACGCTGCTCGGCGCGCCGGAGGCTGCCTACCTGCTAAGGATCATGAGTCTGCTGATCCCGCTGCA  3690
 L  G  G  D  V  A  Y  S  L  L  V  R  R  M  N  F  R  K  D  A  V  W  S  M  I  S  E  S  L  G
GCTCGGCGGCGACGTTGCCTATTCGCTGCTTGTCCGGAGGATGAATTTCCGCAAGGATGCCGTCTGGAGCATGATCTCCGAATCACTCGG  3780
 A  V  I  A  V  L  L  A  L  L  G  F  G  I  W  S  L  L  A  Q  L  F  V  S  A  L  V  R  L  S
TGCCGTTATTGCCGTTCTTCTGGCGCTGCTCGGTTTCGGCATATGGTCGCTGCTCGCTCAGCTTTTCGTCTCCGCGCTGGTGCGCCTGTC  3870
 G  L  Y  A  V  S  R  Y  A  P  R  F  V  F  S  L  Q  R  V  L  A  L  S  R  F  S  F  G  M  M
CGGCCTCTATGCCGTTTCCCGTTACGCACCGCGCTTCGTATTCTCGCTGCAACGCGTTCTGGCGCTCAGCCGCTTCAGTTTCGGCATGAT  3960
 G  S  E  I  A  N  F  I  T  F  Q  S  P  M  V  V  I  S  R  Y  L  G  L  S  D  A  G  A  Y  S
GGGCTCCGAGATCGCGAACTTCATCACCTTCCAGTCCCCGATGGTGGTGATATCCCGCTATCTCGGGCTGTCCGACGCAGGCGCCTATTC  4050
 A  A  N  R  F  A  S  I  P  N  Q  V  V  L  S  A  V  M  G  V  L  F  P  T  F  G  Q  M  M  H
GGCGGCGAACCGGTTCGCGAGCATCCCGAACCAGGTCGTCCTCTCCGCCGTCATGGGGGTGCTGTTCCCAACCTTCGGCCAGATGATGCA  4140
 D  R  E  R  R  S  Q  A  L  M  L  S  T  Q  V  T  T  V  L  L  A  P  M  M  F  G  L  W  A  L
TGATCGCGAGCGGCGTTCACAGGCGCTGATGCTCAGCACTCAGGTGACCACCGTCCTGCTGGCGCCGATGATGTTCGGCCTCTGGGCACT  4230
 A  E  P  A  M  L  V  L  F  G  S  Q  W  A  Y  A  W  P  V  L  G  L  L  A  L  S  K  G  I  L
TGCCGAGCCGGCGATGCTCGTACTTTTCGGGAGCCAATGGGCCTATGCCTGGCCGGTCCTCGGACTTCTCGCACTATCGAAAGGCATTCT  4320
 T  P  C  S  T  F  I  P  Y  L  K  G  V  G  Q  G  A  V  L  F  W  W  A  L  I  R  A  V  A  T
CACCCCCTGCAGCACGTTCATCCCCTATCTCAAGGGGGTCGGCCAGGGAGCCGTCCTGTTCTGGTGGGCGCTGATTCGCGCGGTTGCGAC  4410
 T  G  A  V  A  Y  G  A  I  D  G  S  L  V  E  A  M  I  W  L  C  I  V  N  A  V  T  L  V  G
AACAGGCGCGGTCGCCTACGGTGCCATCGACGGCTCGCTGGTCGAGGCGATGATCTGGCTGTGCATCGTCAATGCCGTAACGCTTGTCGG  4500
 Y  S  W  V  V  F  R  A  D  S  T  P  F  L  K  G  L  F  I  S  S  R  P  M  I  A  A  L  L  M
ATACTCCTGGGTGGTGTTCCGTGCCGACAGCACGCCCTTCCTCAAGGGTTTGTTCATATCCAGCCGGCCGATGATCGCGGCGCTGCTGAT  4590
 A  L  V  V  R  F  L  L  E  H  F  G  A  H  V  P  N  A  V  L  Q  L  I  A  G  T  A  I  G  S
GGCGCTCGTCGTCCGCTTCCTGCTCGAGCATTTCGGCGCACATGTGCCGAACGCCGTCTTGCAATTGATCGCCGGAACGGCCATCGGCAG  4680
 V  I  Y  T  V  L  I  L  L  T  E  R  S  L  L  R  R  L  L  D  M  A  R  A  R  K  P  R  A  A
CGTGATCTACACGGTTCTGATCCTTCTGACGGAGCGATCGCTGTTGCGGAGGCTCCTGGACATGGCGCGAGCACGCAAGCCGAGAGCCGC  4770
 P  A  G  A  A  E  *
TCCCGCGGGAGCCGCCGAGTAGCCAGACGAGCCCGACTGCGGAAGAAAGAGTGACGGGTCAACCCGCCCTTTTGTTGCCGCCAGGGCGCG  4860
GCCGCGGGAACGGTTTGCGGCACGCCGCGTTACCGTTCGACGGCAACAACATGGATGTTCGAGATGAAAGGCATTATCCTTTGGTTGATG  4950
GGCGTTCCCCTCATCGTCATTATCCTGCTCTATATGTTCGTCTTTTGACCCCGCGGCATCGCGAATCATTCCGCCTGACGAGCAACAGCC  5040
GCGCGAAATACACGTGATCAGTCACGCCGATATGCACCCCGCGGTGGCATTCACCCATAGTTTTGAAGGATGTGAATGGCGCCCACCTGG  5130

                                                       M  T  R  I  K  S  A  V  A  A  G  G  R  R  A  P  H

GTTAGCCTGTTCGGGTGTTTCGGCTGAACGGGGATGGGGAATGACCCGCATCAAAAGCGCTGTAGCTGCCGGCGGCAGGCGAGCGCCGCA  5220

 S  A  R  L  G  S  A  S  T  R  T  I  G  A  V  L  A  A  L  L  M  T  H  D  A  G  A  A  E  P

CTCGGCACGACTGGGCTCTGCCAGCACGCGGACCATCGGCGCCGTTCTGGCGGCGCTTCTCATGACCCATGACGCCGGGGCCGCAGAGCC  5310
 I  I  G  Q  A  S  V  I  D  G  D  T  I  E  I  A  G  E  R  V  Q  L  N  S  V  D  A  P  E  E
GATCATCGGTCAAGCCTCGGTGATCGACGGCGATACTATTGAAATCGCGGGCGAACGCGTCCAACTTAACAGTGTGGACGCGCCGGAAGA  5400
 W  Q  V  C  L  D  E  R  G  A  D  Y  R  C  G  K  E  S  A  S  A  L  D  A  F  L  S  A  S  R
GTGGCAGGTCTGCCTCGATGAAAGAGGGGCAGATTATCGGTGCGGCAAGGAATCGGCCTCGGCCCTGGATGCATTCTTGTCGGCATCCCG  5490
 P  T  R  C  E  F  A  G  R  D  R  Y  G  R  F  V  G  T  C  F  R  A  D  G  K  D  V  N  R  W
TCCGACTCGCTGCGAATTTGCCGGCCGCGATCGTTATGGCCGCTTCGTCGGAACGTGTTTCCGGGCCGATGGCAAAGACGTGAACCGCTG  5580
 L  I  E  S  G  N  A  V  D  R  D  T  D  N  K  G  L  Y  A  S  A  Q  Q  T  A  K  S  N  G  A
GCTGATCGAGTCCGGCAATGCAGTCGATCGCGACACGGACAACAAAGGCCTCTATGCATCGGCGCAGCAAACTGCCAAATCGAACGGTGC  5670
 G  I  W  R  A  Q  P  E  H  A  C  A  A  R  V  G  R  V  N  R  K  P  S  C  *
TGGAATCTGGCGCGCACAGCCCGAGCATGCCTGCGCGGCGCGCGTCGGCCGGGTGAACCGAAAGCCGAGTTGTTGAAGGCTGCAAACGAG  5760
AGCCAATGATCTTAGAATTC  5780
```

exoT

exoI

1993b) into the respective coding regions. Restriction fragments of the *exo*-gene region were subcloned into the mobilizable vector pK18mob, which is unable to replicate in *R. meliloti*. After *in vitro* insertion of the *lac*Z-Gm interposon into the *exo* genes, the resulting plasmids were transferred to *R. meliloti* 2011 by *Escherichia coli* S17-1-mediated mobilization (Simon *et al.* 1983). The wild-type *exo* genes were replaced by the *exo* genes disrupted by a *lac*Z-Gm interposon due to two recombination events (Fig. 5).

R. meliloti strains carrying the *lac*Z-Gm interposon inserted into *exo*T (RmH11a) and *exo*W (RmH19a, RmH19b) induced white nodules resembling the pseudonodules induced by the *exo*Y mutant Rm0540 (Müller *et al.* 1988). These mutants were nonfluorescent on medium containing Calcofluor white. No EPS could be detected by cetyl pyridinium chloride (CPC) or ethanol precipitation (Fig. 6). In contrast, the *exo*I-*lac*Z-Gm interposon mutants (RmH10a, RmH10b) were able to infect alfalfa nodules, indicating that *exo*I is not essential for an effective symbiosis with alfalfa. About 50% of the EPS amount produced by the wild-type *R. meliloti* 2011 could be recovered from the culture supernatants of the *exo*I mutants by CPC precipitation, whereas about 20% were obtained by ethanol precipitation (Fig. 6). The phenotype of the *exo*I mutants therefore resembled the phenotype of *exo*O mutants described previously (Becker *et al.* 1993b). The difference in the EPS amount obtained by precipitation with CPC and ethanol could reflect the presence of an altered EPS I produced by these mutants.

We have previously isolated several transposon Tn*5*- and Tn*5*-B20-induced *R. meliloti* mutants located on the 7.8-kb

*Eco*RI DNA fragment of the *exo*-gene region (Keller *et al.* 1988, Müller *et al.* 1988) (Fig. 1A). Sequence analysis showed that the Tn*5*-B20 insertion of mutant Rm154 was located in the *exo*U coding region (Figs. 2 and 6), whereas the Tn*5* insertion site of mutant Rm2505 was located in *exo*V (Fig. 2). Since these *exo*U and *exo*V mutants failed to produce EPS and induced pseudonodules on alfalfa (Keller *et al.* 1988; Müller *et al.* 1988), both genes are necessary for EPS I biosynthesis and the establishment of an effective symbiosis with alfalfa.

The genes *exo*U, *exo*V, *exo*W, *exo*T, and *exo*I form monocistronic transcription units

To confirm the direction of *exo* gene transcription, *R. meliloti* strains carrying the promoterless *lac*Z-Gm interposon inserted into *exo*I, *exo*T, *exo*W, and *exo*V were assayed for β-galactosidase activity. The *lac*Z transcription fusions oriented opposite to the direction of *exo* gene transcription were expressed at the background level of the wild-type *R. meliloti* 2011 (2 ± 1 relative β-galactosidase units), whereas the *lac*Z transcription fusions oriented in the direction of *exo* gene transcription displayed β-galactosidase activities ranging from 53 to 69 relative units (Fig. 5). The promoterless *lac*Z gene encoded by transposon Tn*5*-B20 of *exo*U mutant Rm154 (Keller *et al.* 1988) was expressed at a lower level (31 ± 1 relative units) (Fig. 5).

Since the phenotype of the EPS I overproducing polar *exo*X::Tn*5*-B20 mutant Rm124 (Müller *et al.* 1993) differed from the phenotype of the EPS I⁻ *exo*U::Tn*5*-B20 mutant Rm154, we conclude that *exo*X and *exo*U represent monocistronic transcription units.

To investigate whether *exo*T and *exo*I or *exo*W and *exo*V are organized as operons, the *lac*Z-Gm interposon mutants described above were subjected to plasmid integration mutagenesis. Plasmids containing various subfragments of the 5.780-kb DNA fragment were integrated into the genome of the interposon mutants carrying the *lac*Z gene oriented in the direction of *exo*-gene transcription. The β-galactosidase activities of *lac*Z-transcription fusions located downstream from the vector insertion sites were determined (Fig. 5).

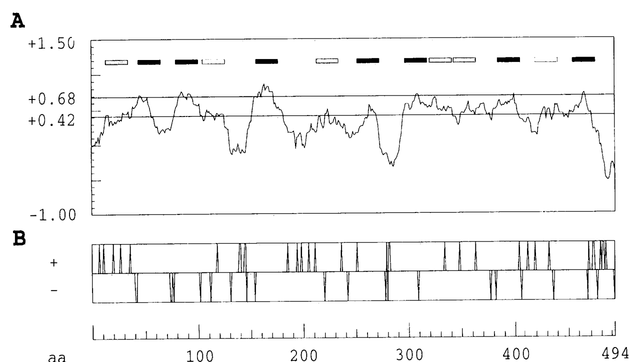


Fig. 3. Analysis of the deduced amino acid (aa) sequence of the *Rhizobium meliloti* 2011 *exo*T gene. **A,** Hydrophobicity plot according to Eisenberg *et al.* (1984), computed using a window of 21 amino acids. Potential transmembrane helices identified by the hydrophobic moment plot (data not shown) are marked by black boxes, helices belonging to globular domains by shaded boxes, and helices predicted to be membrane-surface-associated by open boxes. **B,** Distribution of charged amino acids.

Integration of plasmid pT65 (carrying the 5′ part of *exo*T and 176 bp upstream from the potential *exo*T start codon) into the genome of the *exo*T-*lac*Z-Gm and *exo*I-*lac*Z-Gm interposon mutants resulted in strains that displayed β-galactosidase activities similar to the activities of the *exo*T-*lac*Z-Gm and *exo*I-*lac*Z-Gm interposon mutants without integrated plasmid. This result indicated that a promoter is located on the 176-bp DNA region upstream from the *exo*T start codon. Since integration of plasmid pT85 carrying an internal *exo*T fragment prevented the expression of the *exo*T-*lac*Z transcription fusion but did not influence the expression of the *exo*I-*lac*Z transcription fusion, it can be concluded that another promoter is located in front of *exo*I (Fig. 5). This result was in accordance with the different phenotypes of *exo*T (EPS I⁻) and *exo*I (EPS I⁺) mutants. By integration of plasmids pI79, pI61, and pI81 into the genome of the *exo*I-*lac*Z-Gm interposon mutant, a 125-bp DNA region that was located upstream from *exo*I and displayed promoter activity could be identified (Fig. 5). The fragments cloned into these plasmids were bordered by the internal *Stu*I site of *exo*I at their 3′ ends. Plasmid pI81 contained 844 bp upstream from the potential *exo*I start codon, whereas plasmid pI61 contained only 557 bp upstream from the *exo*I coding region. The *exo*I-*lac*Z-Gm interposon mutant carrying one of these plasmids displayed β-galactosidase activities similar to those of the *exo*I-*lac*Z interposon mutant carrying no plasmid. Based on the insertion of plasmid pI79, we concluded that the GTG at nucleotide position 5054 (Fig. 2) could not represent the start codon of *exo*I. Although only 8 bp upstream from this codon were present in plasmid pI79, integration of this plasmid did not reduce the expression of the *exo*I-*lac*Z tran-

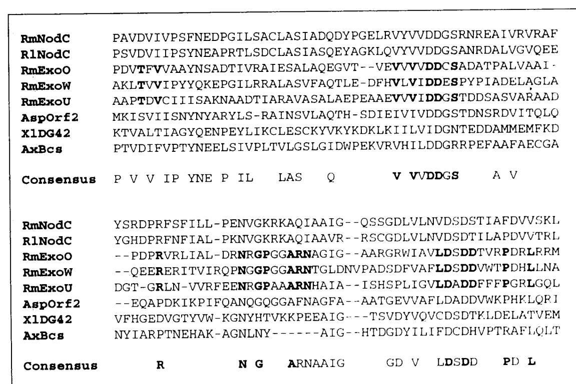


Fig. 4. Alignment of partial sequences of proteins supposed to be involved in polysaccharide polymerization. Partial sequences of *Rhizobium meliloti* 2011 ExoO (Becker *et al.* 1993b), ExoW, and ExoU are compared to the following proteins: *R. meliloti* NodC (Török *et al.*, 1984), *R. loti* NodC (EMBL/GenBank/DDBJ accession number L06241, 1993), ORF2 protein from *Anabaena* sp. strain PCC7120 (Holland and Wolk 1990), BcsA protein from *Acetobacter xylinum* (Saxena *et al.* 1990); and DG42 protein from *Xenopus laevis* (Rosa *et al.* 1988). Residues identical in ExoO, ExoW, and ExoU are indicated by bold letters. If four out of eight amino acids were identical, they are included in the consensus sequence. Consensus sequence residues being highly conserved between all three Exo proteins and at least two other sequences are printed in bold type. RmNodC, NodC of *R. meliloti*; RlNodC, NodC of *R. loti*; RmExoO, *R. meliloti* ExoO; RmExoW, *R. meliloti* ExoW; RmExoU, *R. meliloti* ExoU; AspOrf2, ORF2 protein from *Anabaena* sp. strain PCC7120; XlDG42, DG42 protein from *X. laevis*; Bcs, AxBcsA protein from *Acetobacter xylinum*.

scription fusion. We therefore propose that the next start codon preceded by a ribosome binding site at a reasonable distance (ATG at nucleotide position 5171, Fig. 2) represents the actual start codon of *exo*I.

To investigate whether the promoter located in front of *exo*T influences the transcription of *exo*H, which is located 855 bp downstream from *exo*I (Becker *et al.* 1993a), plasmids pI79, pI61, and pI81 were integrated into the genome of the *exo*H-*lac*Z-Gm interposon mutant RmH1a (Becker *et al.* 1993b). Compared to the β-galactosidase activity of the *exo*H-*lac*Z-Gm interposon mutant without integrated plasmids, no alteration of the β-galactosidase activities was found. This indicated that the transcription of *exo*H is independent of the promoter preceding *exo*I. Therefore, *exo*T and *exo*T form independent transcription units.

To investigate whether *exo*V and *exo*W represent monocistronic transcription units as well, we constructed the plasmids pW66, pW67, and pW64. Plasmid pW66 carried a 308-bp *Eco*RI fragment comprising 132 bp of the *exo*T 5′ part, the 124-bp intergenic region between *exo*T and *exo*W, as well as 52 bp of the *exo*W 5′ part, whereas the plasmids pW67 and pW64 contained longer parts of the *exo*T and *exo*W coding regions (Fig. 5).

Integration of these three plasmids into the genome of the *exo*W-*lac*Z-Gm interposon mutant resulted in the identification of a 256-bp DNA region located upstream from *exo*W, which was sufficient for expression of the *exo*W-*lac*Z transcription fusion at the level of the *exo*W-*lac*Z-Gm interposon mutant without integrated plasmids (Fig. 5).

To analyze the transcription of *exo*V, plasmids pVlacZ and pVlacZr were constructed. These plasmids contained the 2.631-kb *Eco*RI-*Cla*I fragment comprising the 3′ part of *exo*W, the complete *exo*V coding region disrupted by a *lac*Z-Gm interposon at the *Eco*RV site, and the 3′ part of *exo*U. The *lac*Z gene of plasmid pVlacZ was oriented in the direction of *exo*V transcription, whereas the *lac*Z gene of plasmid pVlacZr was oriented in the opposite direction. Since we failed to homogenotize these mutations, merodiploid *R. meliloti* strains carrying the *lac*Z-Gm interposon inserted into *exo*V in both orientations were constructed by integration of plasmids pVlacZ and pVlacZr into the *R. meliloti* genome (Fig. 5). Depending on the crossing over site, integration of plasmids pVlacZ and pVlacZr resulted in two different *R. meliloti* strains (Fig. 5). If the crossing over occurred upstream from the *lac*Z-Gm interposon, the region upstream from the interposon in the resulting strain represented the original situation of the wild-type genome. If the crossing over took place downstream from the interposon, only 1,261 bp comprising 911 bp of the *exo*W 3′ part and the intergenic region between *exo*W and *exo*V were located upstream from the *exo*V start codon. Compared to the expression of the *exo*V-*lac*Z transcription fusion preceded by the wild-type genomic structure, the β-galactosidase activity of the *exo*V-*lac*Z transcription fusion disrupted from the promoter located in front of *exo*W

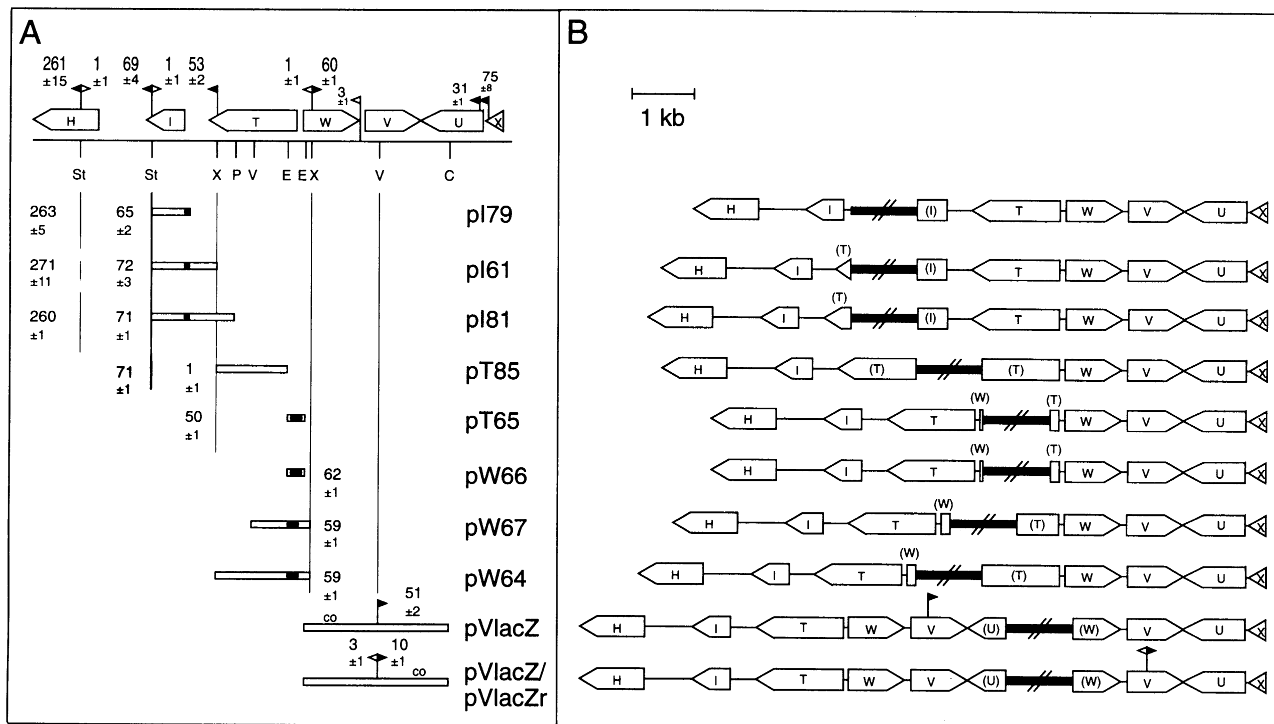
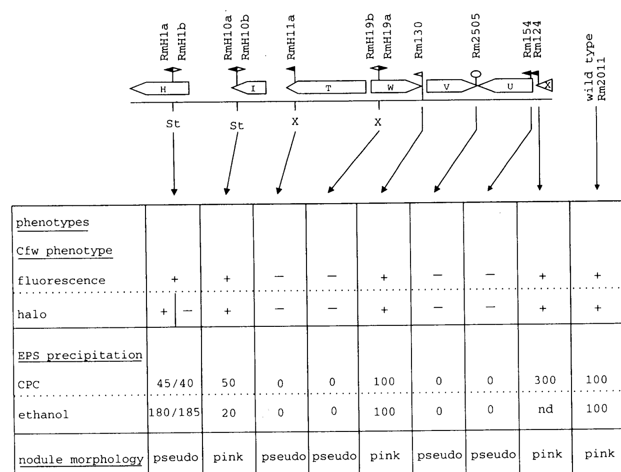


Fig. 5. Mapping of promoters on the *Rhizobium meliloti exo*H/*exo*I/*exo*T/*exo*W/*exo*V/*exo*U fragment. **A,** Relevant restriction sites and gene structure of the *exo*-gene region analyzed. The relative β-galactosidase units of *exo-lac*Z transcription fusions are shown above the genes. Below the genes, the position of restriction fragments and the designation of the plasmids used for integration mutagenesis are presented. Potential promoters located on the fragments used for plasmid integration mutagenesis are indicated by black dots. The relative β-galactosidase units of *exo-lac*Z transcription fusions located downstream from the vector integration sites are listed. Activities were calculated from at least five independent measurements. **B,** Genomic structures of the *exo* gene region resulting from plasmid integration experiments. Insertion sites of interposons or transposons in the genome of the recipient strains are not indicated, whereas the position of *lac*Z-Gm interposons inserted into plasmids used for integration mutagenesis are marked. Heavy lines indicate the vector part of the integrated plasmids. Incomplete *exo* genes are denoted in brackets. C, *Cla*I; E, *Eco*RI; P, *Pst*I; St, *Stu*I; V, *Eco*RV; X, *Xho*I; co, crossing over site.

by an integrated vector was reduced to 20% (Fig. 5). Since the low expression of the *exo*V-*lac*Z transcription fusion was significantly higher than the background activity of wild-type *R. meliloti* 2011, we propose that a weak promoter is located in front of *exo*V. This conclusion could be confirmed by the phenotype of strain Rm130 carrying a polar Tn*5*-B20 insertion 3 bp downstream from the *exo*W stop codon. This *R. meliloti* strain infected alfalfa and produced EPS I.



phenotypes									
Cfw phenotype									
fluorescence	+	+	−	−	+	−	−	+	+
halo	+ | −	+	−	−	+	−	−	+	+
EPS precipitation									
CPC	45/40	50	0	0	100	0	0	300	100
ethanol	180/185	20	0	0	100	0	0	nd	100
nodule morphology	pseudo	pink	pseudo	pseudo	pink	pseudo	pseudo	pink	pink

Fig. 6. Phenotypes of *Rhizobium meliloti* 2011 strains mutated in defined *exo* genes by transposon or interposon insertions. Map is of relevant restriction sites and gene structure of the 9.75-kb *exo*-gene region encoding *exo*H, *exo*I, *exo*T, *exo*W, *exo*V, *exo*U, and *exo*X. Long flags indicate the position and orientation of *lac*Z-Gm interposons, whereas Tn*5*-B20 insertions are marked by short flags. The position of a Tn*5* insertion is denoted by a hairpin. Cfw-phenotype (fluorescence on medium containing Calcofluor white): +, fluorescent; –, nonfluorescent; halo (appearence of a fluorescent halo around the bacterial colonies on medium containing Calcofluor white): +, halo, –, haloless; exopolysaccharide (EPS) precipitation: EPS amount (percent of the EPS amount produced by the wild type) as determined by the HCl/L-cysteine method after precipitation with cetyl pyridinium chloride (CPC) or ethanol (see Materials and Methods); EPS production of *R. meliloti* 2011: 1.42 mg per 10^9 cells per milliliter culture as detected after CPC precipitation; 0.95 mg per 10^9 cells per milliliter culture as detected after ethanol precipitation; nodule morphology: pink, effective pink nodules; pseudo, pseudonodules; nd, not determined. Restriction sites: St, *Stu*I; X, *Xho*I.

DISCUSSION

The 5.780-kb DNA sequence presented here completes the sequence analysis of a 24-kb exopolysaccharide biosynthesis gene cluster from *R. meliloti* 2011. Five new *exo* genes designated *exo*U, *exo*V, *exo*W, *exo*T, and *exo*I could be identified by coding region and mutational analysis. The *exo*T gene corresponded to a complementation group previously identified by Reuber *et al.* (1991). According to the map reported by Long *et al.* (1988), most of their Tn*5* mutations were located within *exo* gene coding regions, but the phenotypes of these mutants differed from the phenotypes of the interposon and transposon mutants reported here.

Analysis of the hydropathic profile of ExoT indicated that this protein contains several transmembrane as well as several membrane-associated helices. This finding was confirmed by several Tn*pho*A insertions in *exo*T, reported by Reuber *et al.* (1991), that displayed alkaline phosphatase activity.

A characteristic signal peptide with a striking score of 14.85 (according to von Heijne 1986) could be identified for the N-terminal region of *exo*I. Together with the results of plasmid integration mutagenesis, this supports that the ATG at nucleotide position 5171 represents the *exo*I start codon used *in vivo*.

Sequence comparisons of ExoW and ExoU indicated the presence of significant homologies to the ExoO protein, which was discussed as functioning as a glucosyltransferase (Becker *et al.* 1993b). In addition, these three proteins shared common amino acid motifs with several polysaccharide-polymerizing enzymes. Since several glucosyltransferases have to be involved in the synthesis of EPS I, it is likely that these genes encode enzymes with glucosyltransferase activity. Moreover, ExoV showed weak similarities to the deduced amino acid sequences of the *exo*L, *exo*A, *exo*M, and *exo*O (Becker *et al.* 1993b) as well as to the *exo*W genes from the *R. meliloti* 2011 *exo*-gene region, which were proposed as encoders of glucosyltransferases. Together these data indicate the presence of seven genes encoding glucosyltransferases in the *exo*-gene region. This corresponds to biochemical evidence that up to seven different glucosyltransferases are necessary to synthesize the EPS I subunit (Aman *et al.* 1981). The fact that the EPS I subunit consists of one galactose and seven glucose molecules joined by β-1,3-, β-1,4-, and β-1,6-glucosidic bonds could explain the weak and substantially different

Table 2. Strains used and constructed in this work

Strain	Relevant characteristics	Reference
E. coli		
JM83	Δ(*lac-pro*AB),Φ80dlacZΔM15 Host for pUC and pSVB plasmids	Vieira and Messing (1982)
DH5α	*rec*A1,Δ*lac*U169,Φ80d*lac*ZΔM15 Host for pUC and pSVB plasmids	Bethesda Research Laboratories
S17-1	*E. coli* 294 RP4-2-Tc::Mu-Km::Tn*7* chromosomally integrated	Simon *et al.* (1983)
R. meliloti		
Rm2011	Wild type, Nod^+, Fix^+, Inf^+, EPS^+, Cfw^+, Sm^r, Nx^R	Casse *et al.* (1979)
Rm124	*exo*X::Tn*5*-B20 mutant of *R. meliloti* 2011	Keller *et al.* (1988)
Rm130	*R. meliloti* 2011 carrying Tn*5*-B20 inserted between *exo*W and *exo*V	Keller *et al.* (1988)
Rm154	*exo*U::Tn*5*-B20 mutant of *R. meliloti* 2011	Müller *et al.* (1988)
Rm2505	*exo*V::Tn*5* mutant of *R. meliloti* 2011	Müller *et al.* (1988)
RmH1a,b	Rm2011 *exo*H-*lac*Z-Gm[a], insertion site: *Stu*I	Becker *et al.* (1993b)
RmH10a,b	Rm2011 *exo*I-*lac*Z-Gm[a], insertion site: *Stu*I	This work
RmH11a	Rm2011 *exo*T-*lac*Z-Gm[a], insertion site: *Xho*I	This work
RmH19a,b	Rm2011 *exo*W-*lac*Z-Gm[a], insertion site: *Xho*I	This work

[a] Strain a carried the *lac*Z gene oriented in the direction of *exo* gene transcription and strain b carried the *lac*Z gene oriented opposite to the direction of *exo* gene transcription.

homologies (12–32%) between the Exo proteins discussed as glucosyltransferases.

To analyze the genetic organization of the DNA fragment sequenced, *exo-lac*Z transcription fusions were constructed. Subsequently, these strains were subjected to plasmid integration mutagenesis using various subfragments of the *exo*-gene region. The results demonstrated that *exo*U, *exo*V, *exo*W, *exo*T, and *exo*I form monocistronic transcription units. Additionally, it was shown that the transcription of *exo*H (Becker *et al.* 1993a,b) is independent of the promoter located in front of *exo*T. It could be demonstrated that two promoters were involved in the transcription of *exo*V. One of these was located in front of *exo*W, whereas the other, much weaker promoter was identified in front of *exo*V. The promoter activity of this weak promoter was sufficient for EPS I biosynthesis at the wild-type level.

Since all fragments used for integration mutagenesis were subcloned into the same site of the same vector and since integration of plasmid pT85 prevented the expression of the *exo*T-*lac*Z transcription fusion located downstream from the vector integration site, it is rather unlikely that a vector-specific promoter was responsible for the expression of the *exo-lac*Z transcription fusions. This is further confirmed by our previous analysis of the *exo*HKLAMONP region, in which the same vector system was used to elucidate the transcriptional organization (Becker *et al.* 1993b).

This report completes the sequence analysis of the 24-kb *exo*-gene region of *R. meliloti* 2011. In further experiments we are going to confirm several *exo*-gene functions deduced from sequence comparisons by biochemical methods. The elucidation of the complete genetic organization and the biosynthetic pathway of EPS I should contribute to the understanding of the molecular basis of the recognition between alfalfa and its microsymbiont during the infection process.

MATERIALS AND METHODS

Bacterial strains and plasmids.

Strains and plasmids used in this study are listed in Table 2.

Media and growth conditions.

E. coli strains were grown in Penassay broth (Difco Laboratories) or in LB medium (Maniatis *et al.* 1982) at 37° C. *R. meliloti* strains were grown in TY medium (Beringer 1974), Vincent minimal medium (Vincent 1970), M9 medium (Miller 1972), or LB medium (Maniatis *et al.* 1982) at 30° C.

Antibiotics were supplemented as required at the following concentrations (micrograms per milliliter): for *R. meliloti*: streptomycin, 600; nalidixic acid, 8; tetracycline, 8; gentamicin, 40; and neomycin, 120; for *E. coli*: tetracycline, 10; chloramphenicol, 25; ampicillin, 100; gentamicin, 10; and kanamycin, 50.

DNA biochemistry.

Plasmid DNA from *E. coli* was prepared as described by Priefer (1984). DNA restriction, agarose gel electrophoresis, cloning procedures, and Southern hybridization followed established protocols (Maniatis *et al.* 1982). Total DNA from rhizobia was isolated according to Meade et al. (1982). Transformation of *E. coli* cells was performed according to the method of Morrison (1977).

DNA sequencing.

DNA sequences were obtained either from defined restriction fragments cloned into the sequencing vectors pSVB30 or pK18 (Arnold and Pühler 1988; Pridmore 1987) or by using overlapping nested deletion clones generated by exonuclease III digestion according to Henikoff (1984). Sequencing reactions were performed using the Auto Read Sequencing Kit (Pharmacia LKB) according to a protocol devised by Zimmermann *et al.* (1990). Sequence data were obtained and processed using the A.L.F. DNA Sequencer (Pharmacia LKB) according to the manufacturer's instructions.

The Tn*5*-B20 insertion junctions were sequenced using an oligodeoxynucleotide primer complementary to the first nucleotides of the *lac*Z gene integrated in IS50L (Simon *et al.* 1989).

Analysis of nucleotide and protein sequences.

The nucleotide and amino acid sequences were analyzed using the computer programs of Staden (1986). The coding probability was calculated according to the codon usage method (Staden and McLachlan 1982), employing an *R. meliloti* codon usage table as described by Buendia *et al.* (1991). The amino acid sequences deduced from the nucleotide sequence were compared to the EMBL/GenBank/DDBJ database using the FASTA programs (Pearson and Lipman 1988). The deduced amino acid sequences were analyzed for their hydrophobicity, following the procedure described by Eisenberg *et al.* (1984). Using the weight matrix proposed by von Heijne (1986), the N-terminal regions were compared to prokaryotic signal peptides.

Interposon mutagenesis.

Using appropiate restriction sites, a promoterless *lac*Z-Gm interposon (Becker *et al.* 1993b) was inserted in both orientations into fragments of the *exo*-gene region, which were subcloned into the mobilizable suicide vector pK18mob (Table 3). Resulting hybrid plasmids were transferred from the broad host range mobilizing strain *E. coli* S17-1 (Simon *et al.* 1983) to *R. meliloti* 2011 as described by Simon (1984). Homogenotization of the *lac*Z-Gm insertions was performed as described by Masepohl *et al.* (1988). All homogenotes were verified by Southern hybridization.

Exopolysaccharide production.

EPS production by *R. meliloti* strains was detected by the Calcofluor white staining method as described by Hynes *et al.* (1986). Quantitative estimation of the EPS production was performed according to Müller *et al.* (1988). The total hexose content was quantified by the HCl/L-cysteine method (Chaplin and Kennedy 1986).

Plant nodulation assays.

R. meliloti strains were assayed for their symbiotic phenotypes on *Medicago sativa* 'Du Puits' (Saatgutveredlung Lippstadt, FRG). Seeds were surface-sterilized and germinated as described by Müller *et al.* (1988). The seedlings were inoculated using late log phase bacterial cultures. The plantlets were grown on nitrogen-free medium as described by Rolfe *et al.* (1980) and analyzed after 3–4 wk.

Table 3. Plasmids used and constructed in this work

Plasmid	Relevant characteristics	Source or reference
pSVB30	Ap^r, *lacZ′*	Arnold and Pühler (1988)
pK18	Km^r, *lacZ′*	Pridmore (1987)
pK19	Km^r, *lacZ′*	Pridmore (1987)
pK18mob	Km^r, pK18 derivative, *mob*-site	A. Schäfer, Bielefeld
pK19mob	Km^r, pK19 derivative, *mob*-site	A. Schäfer, Bielefeld
pAB2001	Promoterless *lac*Z-Gm interposon	Becker *et al.* (1993b)
pRmAB839	Cosmid based on pSUP205 carrying a 36-kb contiguous region of *R. meliloti* 2011 megaplasmid 2	Becker *et al.* (1993a)
pI79	583-bp fragment[a] (125 bp-*exo*I)[b] cloned in pK18mob	This work
pI61	1,015-bp *Xho*I-*Stu*I fragment (557 bp-*exo*I)[b] cloned in pK18mob	This work
pI81	1,302-bp *Pst*I-*Stu*I fragment (844 bp-*exo*I)[b] cloned in pK18mob	This work
pT85	1,184-bp *Xho*I-*Eco*I fragment (1,737 bp-*exo*I)[b] cloned in pK19mob	This work
pT65	308-bp *Eco*RI fragment (176 bp-*exo*T)[b] cloned in pK18mob	This work
pW66	308-bp *Eco*RI fragment cloned (256 bp-*exo*W)[b] in pK19mob	This work
pW67	920-bp *Xho*I-*Eco*RV fragment (828 bp-*exo*W)[b] cloned in pK18 mob	This work
pW64	1,526-bp *Xho*I fragment (1,434 bp-*exo*W)[b] cloned in pK18mob	This work
pVlacZ	2,631-bp *Eco*RI-*Cla*I fragment cloned in pK18mob carrying a *lac*Z-Gm interposon inserted into the internal *Eco*RV site of *exo*V oriented in the direction of *exo*V transcription	This work
pVlacZr	2,631-bp *Eco*RI-*Cla*I fragment cloned in pK18mob carrying a *lac*Z-Gm interposon inserted into the internal *Eco*RV site of *exo*V oriented opposite to the direction of the *exo*V transcription	This work

[a] Fragments generated by exonuclease III digestion.
[b] The distance of the 5′ end of the fragment cloned to the start codon of the next *exo* gene located downstream is given in base pairs (bp).

β-Galactosidase assay of *R. meliloti* strains carrying *lac*Z-Gm insertions.

R. meliloti strains were grown to an optical density (600 nm) of 0.6–0.8 in LB medium. The assay for β-galactosidase activity followed the protocol devised by Aguilar *et al.* (1985). The reaction was stopped after 20–50 min. Relative β-galactosidase units were calculated per cell number according to Miller (1972).

Analysis of the transcriptional organization.

Subfragments of the *exo*-gene cluster were cloned in several steps into the mobilizable suicide vectors pK18mob and pK19mob (Table 3). The *exo* genes analyzed, which were encoded by the cloned fragments, were transcribed opposite to the orientation of the vector *lac*Zα promoter. Hybrid plasmids were transferred to *R. meliloti* 2011 *exo-lac*Z-Gm homogenotes. Integration of hybrid plasmids into the *R. meliloti* genome by single crossing over was selected by the vector-encoded antibiotic resistance. The occurrence of the recombination events within the homologous *exo* sequences of the plasmids and the *exo*-gene region of the *R. meliloti* genome was verified by Southern hybridization. Transconjugants were assayed for their β-galactosidase activities.

ACKNOWLEDGMENTS

We are very grateful to A. Schäfer for providing some vectors. We also acknowledge the excellent technical assistance of M. Bellanco-Garcia. A. Becker was supported by a scholarship from the Studienstiftung des Deutschen Volkes. This work was financially supported by Deutsche Forschungsgemeinschaft Pu28/13.

LITERATURE CITED

Aguilar, O. M., Kapp, D., and Pühler, A. 1985. Characterisation of a *Rhizobium meliloti* fixation gene (*fix*F) located near the common nodulation region. J. Bacteriol. 164:245-254.

Aman, P., McNeil, M., Franzen, L., Darvill, A. G., and Albersheim, P. 1981. Structural elucidation using HPLC-MS and GLC-MS of the acidic polysaccharide secreted by *Rhizobium meliloti* strain 1021. Carbohydr. Res. 95:263-282.

Arnold, W., and Pühler, A. 1988. A family of high-copy-number plasmid vectors with single end-label sites for rapid nucleotide sequencing. Gene 70:171-179.

Atkinson, E. M., and Long, S. R. 1992. Homology of *Rhizobium meliloti* NodC to polysaccharide polymerizing enzymes. Mol. Plant-Microbe Interact. 5:439-442.

Battisti, L., Lara, J. C., and Leigh, J. A. 1992. Specific oligosaccharide form of the *Rhizobium meliloti* exopolysaccharide promotes nodule invasion in alfalfa. Proc. Natl. Acad. Sci. USA 89:5625-5629.

Becker, A., Kleickmann, A., Arnold, W., and Pühler, A. 1993a. Analysis of the *Rhizobium meliloti exo*H/*exo*K/*exo*L fragment: *exo*K shows homology to excreted β-1,3-1,4-glucanases and ExoH resembles membrane proteins. Mol. Gen. Genet. 238:145-154.

Becker, A., Kleickmann, A., Arnold, W., Keller, M., and Pühler, A. 1993b. Identification and analysis of the *Rhizobium meliloti exo*AMONP genes involved in exopolysaccharide biosythesis and mapping of promoters located on the *exo*HKLAMONP fragment. Mol. Gen. Genet. 241:367-379.

Beringer, J. E. 1974. R factor transfer in *Rhizobium leguminosarum*. J. Gen. Microbiol. 84:188-198.

Buendia, A. M., Enenkel, B., Köplin, R., Niehaus, K., Arnold, W., and Pühler, A. 1991. The *Rhizobium meliloti exo*Z/*exo*B fragment of megaplasmid 2: ExoB functions as a UDP-glucose-4-epimerase and ExoZ shows homology to NodX of *Rhizobium leguminosarum* biovar *viciae* strain TOM. Mol. Microbiol. 5:1519-1530.

Casse, F., Boucher, C., Julliot, J. S., Michel, M., and Dénarié, J. 1979. Identification and characterisation of large plasmids in *Rhizobium meliloti* using agarose gel electrophoresis. J. Bacteriol. 113:229-242.

Chaplin, M. F., and Kennedy, S. F. 1986. Carbohydrate Analysis. A Practical Approach. IRL Press, Oxford.

Debellé, F., Rosenberg, C., and Dénarié, J. 1992. The *Rhizobium*, *Bradyrhizobium*, and *Azorhizobium* NodC proteins are homologous to yeast chitin synthases. Mol. Plant-Microbe Interact. 5:443-446.

Eisenberg, D., Schwarz, E., Komaromy, M., and Wall, R. 1984. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J. Mol. Biol. 179:125-142.

Henikoff, S. 1984. Unidirectional digestions with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28:351-359.

Holland, D., and Wolk, P. 1990. Identification and characterization of *het*A, a gene that acts early in the process of morphological differentiation of heterocysts. J. Bacteriol. 172:3131-3137.

Hynes, M. F., Simon, R., Müller, P., Niehaus, K., Labes, M., and Pühler, A. 1986. The two megaplasmids of *Rhizobium meliloti* are involved in the effective nodulation of alfalfa. Mol. Gen. Genet. 202:356-362.

Keller, M., Müller, P., Simon, R., and Pühler, A. 1988. *Rhizobium meliloti* genes for exopolysaccharide synthesis and nodule infection located on megaplasmid 2 are actively trancribed during symbiosis. Mol. Plant-Microbe Interact. 1:267-274.

Leigh, J. A., Reed, J. W., Hanks, J. F., Hirsch, A. M., and Walker, G. C. 1987. *Rhizobium meliloti* mutants that fail to succinylate their Calcofluor-binding exopolysaccharide are deficient in nodule invasion. Cell 51:579-587.

Long, S., Reed, J. W., Himawan, J., and Walker, G. C. 1988. Genetic analysis of a cluster of genes required for synthesis of the Calcofluor-binding exopolysaccharide of *Rhizobium meliloti*. J. Bacteriol. 170:4239-4248.

Long, S. R. 1989. *Rhizobium*-legume nodulation: Life together in the underground. Cell 56:203-214.

Maniatis, T., Fritsch, E. F., and Sambrook, J. 1982. Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Habor, NY.

Masepohl, B., Klipp, W., and Pühler, A. 1988. Genetic characterisation and sequence analysis of the duplicated *nif*A/*nif*B gene region of *Rhodobacter capsulatus*. Mol. Gen. Genet. 212:27-37.

Meade, H. M., Long, S. R., Ruvkun, G. B., Brown, S. E., and Ausubel, F. M. 1982. Physical and genetic characterisation of symbiotic and auxotrophic mutants of *Rhizobium meliloti* induced by transposon Tn*5* mutagenesis. J. Bacteriol. 149:114-122.

Miller, J. H. 1972. Experiments in Molecular Genetics. Cold Spring Habor Laboratory Press, Cold Spring Harbor, NY.

Morrison, D. A. 1977. Transformation in *Escherichia coli*: Cryogenic preservation of competent cells. J. Bacteriol. 132:349-351.

Müller, P., Hynes, M., Kapp, D., Niehaus, K., and Pühler, A. 1988. Two classes of *Rhizobium meliloti* infection mutants differ in exopolysaccharide production and in coinoculation properties with nodulation mutants. Mol. Gen. Genet. 211:17-26.

Müller, P., Keller, M., Weng, W. M., Quandt, J., Arnold, W., and Pühler, A. 1993. Genetic analysis of the *Rhizobium meliloti exo*YFQ operon: ExoY is homologous to sugar transferases and ExoQ represents a transmembrane protein. Mol. Plant-Microbe Interact. 6:55-65.

Pearson, W. R., and Lipman, D. J. 1988. Improved tools for biological sequence analysis. Proc. Natl. Acad. Sci. USA 85:2444-2448.

Pridmore, R. D. 1987. New and versatile cloning vectors with kanamycin-resistance marker. Gene 56:309-312.

Priefer, U. 1984. Isolation of plasmid DNA. Pages 14-25 in: Advanced Molecular Genetics. A. Pühler and K. N. Timmis, eds. Springer-Verlag, Berlin.

Reuber, T. L., Long, S., and Walker, G. C. 1991. Regulation of *Rhizobium meliloti exo* genes in free-living cells and *in planta* examined by using Tn*pho*A fusions. J. Bacteriol. 173:426-434.

Rolfe, B. G., Gresshoff, P. M., and Shine, J. 1980. Rapid screening for symbiotic mutants of *Rhizobium* and white clover. Plant Sci. Lett. 19:277-284.

Rosa, F., Sargent, T. D., Rebbert, M. L., Michaels, G. S., Jamrich, M., Grunz, H., Jonas, E., Winkles, J. A., and Dawid, I. B. 1988. Accumulation and decay of DG42 gene products follow a gradient pattern during *Xenopus* embryogenesis. Dev. Biol. 129:114-123.

Saxena, I. M., Lin, F. C., and Brown, R M., Jr. 1990. Cloning and sequencing of the cellulose synthase catalytic subunit gene of *Acetobacter xylinum*. Plant Mol. Biol. 15:673-683.

Simon, R. 1984. High frequency mobilisation of gram-negative bacterial replicons by the *in vitro* constructed Tn*5-Mob* transposon. Mol. Gen. Genet. 196:413-420.

Simon, R., Priefer, U., and Pühler, A. 1983. A broad host range mobilisation system for *in vivo* genetic engineering: transposon mutagenesis in gram-negative bacteria. Biotechnology 1:784-791.

Simon, R., Quandt, J., and Klipp, W. 1989. New derivatives of transposon Tn*5* suitable for mobilization of replicons, generation of operon fusions and induction of genes in Gram-negative bacteria. Gene 80:161-169.

Staden, R. 1986. The current status and portability of our sequence handling software. Nucleic Acids Res. 14:217-232.

Staden, R., and McLachlan, A. D. 1982. Codon preference and its use in identifying protein coding regions in large DNA sequences. Nucleic Acids Res. 10:141-156.

Török, J., Kondorosi, E., Stepkowski, I., Posfa, J., and Kondorosi, A. 1984. Nucleotide sequence of *Rhizobium meliloti* nodulation genes. Nucleic Acids Res. 12:9509-9524.

Vieira, J., and Messing, J. 1982. The pUC plasmids, an M13mpL7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19:259-268.

Vincent, J. M. 1970. A Manual for the Practical Study of Root Nodule Bacteria. Handb. No. 15. IBP, Oxford.

von Heijne, G. 1986. A new method for predicting sequence cleavage sites. Nucleic Acids Res. 14:4683-4690.

Zimmermann, J., Voss, H., Schwager, C., Stegemann, J., Erfle, H., Stucky, K., Kristensen, T., and Ansorge, W. 1990. A simplified method protocol for fast plasmid DNA sequencing. Nucleic Acids Res. 18:1067.