Identification of genes in *Rhizobium leguminosarum* bv. *trifolii* whose products are homologues to a family of ATP-binding proteins

Jarosław Król and Anna Skorupska

Author for correspondence: Anna Skorupska. Tel: +48 81 537 59 72. Fax: +48 81 537 51 02. e-mail: genet@biotop.umcs.lublin.pl

The specific interaction between rhizobia and their hosts requires many genes that influence both early and late steps in symbiosis. Three new genes, designated *prsD*, *prsE* (protein secretion) and *orf3*, were identified adjacent to the *exo133* mutation in a cosmid carrying the genomic DNA of *Rhizobium leguminosarum* bv. *trifolii* TA1. The *prsDE* genes share significant homology to the genes encoding ABC transporter proteins PrtDE from *Erwinia chrysanthemi* and AprDE from *Pseudomonas aeruginosa* which export the proteases in these bacteria. *PrsD* shows at least five potential transmembrane hydrophobic regions and a large hydrophilic domain containing an ATP/GTP binding cassette. *PrsE* has only one potential transmembrane hydrophobic domain in the N-terminal part and is proposed to function as an accessory factor in the transport system. ORF3, like PrtF and AprF, has a typical N-terminal signal sequence but has no homology to these proteins. The insertion of a kanamycin resistance cassette into the *prsD* gene of the *R. leguminosarum* bv. *trifolii* TA1 wild-type strain created a mutant which produced a normal amount of exopolysaccharide but was not effective in the nodulation of clover plants.

**Keywords**: *Rhizobium leguminosarum* bv. *trifolii*, transport system genes, ATP-binding proteins, *exo* region

**INTRODUCTION**

*Rhizobium*, *Bradyrhizobium* and *Azorhizobium* are bacterial genera that form nitrogen-fixing nodules on legume roots (Schulze et al., 1994). *Rhizobium leguminosarum* bv. *trifolii* induces nitrogen-fixing nodules on the roots of its host plant, clover (*Trifolium* spp.). Regulatory (nod*DE*), common (nod*A*B*C*I*) and host-specific (nod*FERL*, nod*MN*) nodulation (nod) genes are located on symbiotic plasmid pSym (Djordjevic et al., 1985; Schofield et al., 1986). Besides the nodulation and nitrogen-fixation genes, genes encoding extracellular polysaccharides (*exo*) and lipo-polysaccharides (*lps*) are needed for successful nodulation (Leigh & Walker, 1994). Mutants of *R. l. trifolii* deficient in exopolysaccharide synthesis form nodules which do not fix nitrogen (Chen et al., 1985; Derylo et al., 1986; Bialek et al., 1995; Skorupska et al., 1995). In view of the close interactions that occur between *Rhizobium* and its host legumes it seems likely that products of several genes may be exported proteins. Transport systems are known which translocate substances across the two membranes of the Gram-negative bacterial cell envelope in a single step (Higgins, 1992; Fath & Kolter, 1993). Among the best characterized of these transporters are ABC-type (ATP-binding cassette) systems which export proteins such as haemolysin of *Escherichia coli*, proteases of *Erwinia chrysanthemi* and *Pseudomonas aeruginosa*, peptides such as colicin V and non-protein substrates such as *Klebsiella pneumoniae* capsular polysaccharide and *Agrobacterium tumefaciens* and *Rhizobium meliloti* β-1,2-glucans (for reviews see Fath & Kolter, 1993; Higgins, 1992). The secretion apparatus consists of a transmembrane protein with an ATP-binding domain and additional proteins located in the periplasm and in the outer membrane which function to facilitate the export of products through both membranes (Fath & Kolter, 1993). Here we present the nucleotide sequence of three genes
from *R. l. trifolii* TA1: *prsD*, *prsE* (protein secretion) and *orf3*. *PrsD* is significantly related to the bacterial ABC exporter of proteases *PrsD* from *Erw. chrysanthemi* (Letoiffé et al., 1990) and *AprD* from *P. aeruginosa* (Duong et al., 1992). The *prsE* structural gene was identified as the next gene linked to *prsD*, with homology to *prtF* and *aprF* genes encoding accessory factors in related export systems. The third gene, *orf3*, located immediately downstream of *prsE* is not homologous to *PrtF* or *AprF* outer-membrane factors.

**METHODS**

**Bacterial strains, plasmids and growth conditions.** The strains and plasmids used in this study are listed in Table 1. *E. coli* strains were grown in LB medium at 37 °C (Sambrook et al., 1989). *R. l. trifolii* was grown in 79CA medium at 28 °C (Skorupska et al., 1991). Antibiotics were added at the following concentrations (μg mL⁻¹) for *E. coli* and *R. l. trifolii*: tetracycline, 10; rifampicin, 40; kanamycin, 40. Exopolysaccharide production was detected as described by Becker et al. (1995). To assay protease activity, 79CA agar plates containing skimmed milk (1.5%) were used (Guzzo et al., 1991). For testing cellulase activity, 79CA plates with carboxymethylcellulose (2%) were used and stained with Congo red (0.1%). To measure polygalacturonate lyase activity, 79CA plates containing polygalacturonic acid (0.25%) were used and stained with cetyltrimethylammonium bromide (1%) (Mateos et al., 1992).

**DNA manipulations and bacterial matings.** Recombinant DNA techniques such as restriction analysis, cloning procedures, transformation, labelling of DNA and hybridization were carried out according to Sambrook et al. (1989). Plasmids were transferred from *E. coli* to *R. l. trifolii* TA1 using triparental mating (Skorupska et al., 1991). Homologous recombination in *R. l. trifolii* TA1 was carried out with plasmid pPH1J1 which is incompatible with the cosmid vector (Ruvkun & Ausubel, 1981). Homogenization was verified by Southern hybridization with pARF136 as a probe.

**DNA sequencing.** DNA sequencing reactions were performed by the dideoxy-termination method using M13mp18/mp19 vectors with [γ-32P]dATP (Amersham) as the radioactive label. Sequenase version 2.0 (Amersham) was used according to the manufacturer’s (USB) guidelines. Both strands of DNA were

---

**Table 1. Strains, plasmids and phages used in this study**

<table>
<thead>
<tr>
<th>Strain, plasmid or phage</th>
<th>Relevant characteristics*</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>R. l. trifolii</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TA1</td>
<td>Wild-type <em>bv. trifolii</em></td>
<td>M. Djordjevic</td>
</tr>
<tr>
<td>24.1</td>
<td>Wild-type <em>bv. trifolii</em>, Sm*</td>
<td>Skorupska et al. (1991)</td>
</tr>
<tr>
<td><strong>E. coli</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JM101</td>
<td>Δlac-proAB supE thi (F' traD36 proAB lacZΔM1)</td>
<td>Sambrook et al. (1989)</td>
</tr>
<tr>
<td>DH55</td>
<td>supE44 ΔlacU169 (Δ80 lacZΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1</td>
<td>Sambrook et al. (1989)</td>
</tr>
<tr>
<td>HB101</td>
<td>F' supE44 thi recA13 hsdS20 lacY1 rpsL20 (Sm*)</td>
<td>Sambrook et al. (1989)</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pRK7813</td>
<td>IncP, mob Tra cosmid vector, Tc*</td>
<td>Stanley et al. (1987)</td>
</tr>
<tr>
<td>pRK2013</td>
<td>Helper plasmid, Km*</td>
<td>Ditta et al. (1980)</td>
</tr>
<tr>
<td>pARF136</td>
<td>pLAFR3 clone with 19 kb insert of TA1</td>
<td>Skorupska et al. (1991)</td>
</tr>
<tr>
<td>pARF1368</td>
<td>5·6 kb <em>BamHI</em> fragment of pARF136 cloned into pRK7813</td>
<td>Skorupska et al. (1991)</td>
</tr>
<tr>
<td>pJ6</td>
<td>2·19 kb EcoRI–HindIII fragment of pARF136 cloned into pRK7813</td>
<td>This study</td>
</tr>
<tr>
<td><strong>Phages</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M13mp18, mp19</td>
<td>Sequencing vectors</td>
<td>Sambrook et al. (1989)</td>
</tr>
<tr>
<td>pJE1</td>
<td>4·0 kb <em>BglII–EcoRI</em> fragment of pARF136 cloned into M13</td>
<td>This study</td>
</tr>
<tr>
<td>pJ61</td>
<td>1·15 kb <em>EcoRI–BamHI</em> fragment of pARF136 cloned into M13mp18</td>
<td>This study</td>
</tr>
<tr>
<td>p216</td>
<td>1·04 kb <em>BamHI–HindIII</em> fragment of pARF1368 cloned into M13</td>
<td>This study</td>
</tr>
<tr>
<td>pM1</td>
<td>1·02 kb <em>HindIII–PstI</em> fragment of pARF1368 cloned into M13</td>
<td>This study</td>
</tr>
</tbody>
</table>

* Km, kanamycin; Se, streptomycin; Tc tetracyclim.
**R. leguminosarum** genes encoding ATP-binding proteins

**Plant tests.** *R. l. trifolii* strains were assayed for their symbiotic phenotypes on *Trifolium pratense* (cv. Ulka). Seeds were surface-sterilized, germinated and grown as described previously (Bialek et al., 1995). After 5 weeks the plants were checked for symbiotic parameters. Nitrogen fixation activity was determined by the acetylene reduction assay (Hardy et al., 1982).

**RESULTS**

**Subcloning of prsD, prsE and orf3**

The recombinant cosmid clone pARF136 was identified by complementation of the Exo- mutation (exo93) in *R. l. trifolii* 24.1 with the cosmid genome bank of *R. l. trifolii* TA1 (Skorupska et al., 1991). By random mutagenesis of pARF136 with Tn5, we identified a Tn5 insertion, designated pKm133, which mapped in a 5.6 kb *BamHI* fragment of pARF136 (Fig. 1). Recombination of this mutation into *R. l. trifolii* TA1 resulted in an Exo- mutant phenotype. Clover infected with *R. l. trifolii* TA1 exo233 formed non-nitrogen-fixing nodules.

**DNA sequence analysis**

The nucleotide sequence analysis of the 5.263 kb *PstI*- *BglII* fragment adjacent to the exo133 mutation (Fig. 1) revealed at least three ORFs; one, designated prsD, extends from an ATG initiation codon at nt 1256 to a TAG termination codon at nt 2968 and encodes a polypeptide of 570 aa with a deduced *M* of 60887. The sequence GAGG, which was found 2 bp further upstream, resembles a typical ribosome binding site. No obvious promoter sequence upstream of prsD was found.

A second ORF, prsE, 15 nt downstream of prsD starts from a GTG initiation codon at nt 2984 and encodes a polypeptide of 433 aa with a deduced *M* of 47146. No N-terminal signal sequence was identified in this protein. A potential ribosome binding sequence (AGAAGGA-AG) 2 bp upstream of GTG codon was identified.

A third ORF, orf3, 150 nt downstream of prsE initiates probably at the ATG codon at nt 4433 and encodes a polypeptide of 433 aa with a deduced *M* of 47146. No N-terminal signal sequence was identified in this protein. A potential ribosome binding sequence (AGAAGGA-AG) 2 bp upstream of the ATG codon was identified.

A second ORF, prsE, 15 nt downstream of prsD starts from a GTG initiation codon at nt 2984 and encodes a polypeptide of 433 aa with a deduced *M* of 47146. No N-terminal signal sequence was identified in this protein. A potential ribosome binding sequence (AGAAGGA-AG) 2 bp upstream of the ATG codon was identified.

A third ORF, orf3, 150 nt downstream of prsE initiates probably at the ATG codon at nt 4433 and encodes a polypeptide of 149 aa with a predicted *M* of 16502. A putative ribosome binding site was identified (AGGGT) 6 bp upstream of the initiation codon.

**Fig. 2.** Sequence alignments of (a) PrsD (*R. l. trifolii*), PrtD (*Erw. chrysanthemi*) and AprD (*P. aeruginosa*) and (b) PrsE (*R. l. trifolii*), PrtE (*Erw. chrysanthemi*) and AprE (*P. aeruginosa*). Conservative or identical amino acids in all three proteins are indicated by a black box. Amino acid residues are numbered on the right. The conserved Walker motifs are double underlined. The extents of putative transmembrane domains in PrsD and PrsE are overlined; the membrane-associated domain in PrsE is indicated by a dotted line.

---

**DNA sequence**

The nucleotide sequence analysis of the 5-263 kb *PstI*- *BglII* fragment adjacent to the exo133 mutation (Fig. 1) revealed at least three ORFs; one, designated prsD, extends from an ATG initiation codon at nt 1256 to a TAG termination codon at nt 2968 and encodes a polypeptide of 370 aa with a deduced *M* of 60887. The sequence GAGG, which was found 2 bp further upstream, resembles a typical ribosome binding site. No obvious promoter sequence upstream of prsD was found.

A second ORF, prsE, 15 nt downstream of prsD starts from a GTG initiation codon at nt 2984 and encodes a polypeptide of 433 aa with a deduced *M* of 47146. No N-terminal signal sequence was identified in this protein. A potential ribosome binding sequence (AGAAGGA-AG) 2 bp upstream of the ATG codon was identified.

A third ORF, orf3, 150 nt downstream of prsE initiates probably at the ATG codon at nt 4433 and encodes a polypeptide of 149 aa with a predicted *M* of 16502. A putative ribosome binding site was identified (AGGGT) 6 bp upstream of the initiation codon.

---

**Fig. 2.** Sequence alignments of (a) PrsD (*R. l. trifolii*), PrtD (*Erw. chrysanthemi*) and AprD (*P. aeruginosa*) and (b) PrsE (*R. l. trifolii*), PrtE (*Erw. chrysanthemi*) and AprE (*P. aeruginosa*). Conservative or identical amino acids in all three proteins are indicated by a black box. Amino acid residues are numbered on the right. The conserved Walker motifs are double underlined. The extents of putative transmembrane domains in PrsD and PrsE are overlined; the membrane-associated domain in PrsE is indicated by a dotted line.
Comparison of PrsD, PrsE and ORF3 with other bacterial secretion systems

A comparison of the deduced amino acid sequence of PrsD to proteins in the database revealed significant homologies to prokaryotic cytoplasmic membrane transporter proteins. Common structural features of these transport systems include an integral membrane domain and a highly conserved hydrophilic, ATP-binding domain with ATPase activity.

The hydropathy profile (Eisenberg et al., 1984) of the PrsD protein revealed at least five hydrophobic regions, probably corresponding to transmembrane or membrane-associated regions (Fig. 2a). A putative ATP-binding site was found in the central region of the hydrophilic domain. The latter domain contains two Walker motifs required for the binding of ATP (Fig. 2a) (Walker et al., 1982).

The most significant homology was to the P. aeruginosa AprD protein involved in the secretion of alkaline protease (Duong et al., 1992) and to the Erw. chrysanthemi PrtD protein required for the export of protease (Letoffé et al., 1990). The alignments of these proteins indicated that PrsD is 44.4% identical and 82.4% similar to AprD and 40.6% identical and 78.6% similar to PrtD (Fig. 2a).

The hydropathy profile of PrsE shows one N-terminal hydrophobic, transmembrane domain, a C-terminal membrane-associated globular domain and the lack of a signal peptide. This protein is homologous to the second component of this type of secretion system in Erw. chrysanthemi (PrsE, 27.8% identity, 71.4% similarity; Letoffé et al., 1990) and in P. aeruginosa (AprE, 26.7% identity, 72.1% similarity; Duong et al., 1992) (Fig. 2b). These proteins, referred to as accessory factors (AF; Fath & Kolter, 1993), are anchored in the inner membrane and span the periplasm. They probably act to connect the inner and outer membranes and facilitate the export of substrates through both membranes. PrtE is localized to both the inner and outer membrane (Delepelaire & Wandersman, 1991).

The predicted ORF3 product showed no sequence similarity to the outer-membrane proteins PrtF of Erw. chrysanthemi and AprF of P. aeruginosa, nor to other prokaryotic proteins. However, analysis of the deduced amino acid sequence of ORF3 revealed the presence at its N terminus of a typical prokaryotic signal peptide containing four positively charged residues followed by a hydrophobic domain (residues 17–34) and ending at a bacterial signal peptide cleavage site (between residues 37 and 38). Analogous signal peptides were found in the AprF, PrtF and TolC outer-membrane factors in related export systems (Letoffé et al., 1990; Duong et al., 1992).

Mutagenesis of prsD

To construct a prsD mutant, a BamHI kanamycin resistance cassette isolated from pSUP2021 (Simon et al., 1983) was inserted in vitro into the BamHI site of prsD on pJ6 (Fig. 1). Transformants were tested for the correct constructs and inserts were introduced into the genome of the wild-type strain R. l. trifolii TA1 by marker exchange, resulting in the prsD mutant TA1.34. The TA1.34 prsD mutant formed normal mucoid colonies resulting from the production of exopolysaccharide in the same amount as the wild-type strain. The number of nodules per clover plant was similar as in the inoculation with strain TA1. However, the nodules formed on Trifolium pratense did not fix nitrogen, as determined by an acetylene reduction test for nitrogenase activity.

DISCUSSION

In this study we report the identification of three new R. l. trifolii TA1 genes, designated prsD, prsE and orf3. prsD and prsE are homologous to a superfamily of ATP-binding proteins and are probably involved in transport. In addition to homology with P. aeruginosa aprDE genes (Duong et al., 1992) and Erw. chrysanthemi prtDE genes (Letoffé et al., 1990), we also found significant homology to the secretion guard for the Serratia marcescens lipase lipBC genes (Akatsuka et al., 1995) and metalloprotease hasDE genes (Letoffé et al., 1993). These data suggest that the exported substrate could be an extracellular protease. However, we could not detect extracellular proteolytic activity in R. l. trifolii TA1. Hydrolytic enzymes that cleave glycosidic bonds in plant cell walls, such as pectinases, cellulases and hemicellulases, have been found in other wild-type strains of R. l. trifolii but only at low levels of activity (Mateos et al., 1992; Jimenez-Zurdo et al., 1996). We did not find such extracellular enzymes in TA1.

Although the genes involved in the synthesis and secretion of the extracellular proteins are usually clustered, the lipase and metalloprotease genes are not linked to lipBCD in S. marcescens but are linked to genes encoding products homologous to GDP-mannose pyrophosphorylase, an enzyme involved in the biosynthesis of lipopolysaccharide in Salmonella typhimurium and E. coli (Akatsuka et al., 1995). The structure of LPS plays an important role in the localization of secretory components such as TolC, PrtF and even PrtE in the outer membrane (Wandersman & Letoffé, 1993). We have also considered the involvement of an ABC transporter protein, PrsD, and accessory factor, PrsE, in the export of exopolysaccharides because of the location of the prsDE and orf3 genes in the vicinity of the exo133 mutation and the partial homology of PrsE to ExoF (22.6% identity over 199 aa). ExoF is a membrane protein involved in the first step of the biosynthesis of EPSI in R. meliloti (Reuber & Walker, 1993). The identification of the homology of PrsE to PrtE and of ORF3 as a putative outer-membrane protein with an N-terminal signal sequence argues against this hypothesis because at present none of the systems involved in the secretion of non-protein substrates have accessory factors like HlyD or PrtE and outer membrane factors such as TolC or PrtF (Fath & Kolter, 1993). Recently,
Becker et al. (1995) identified the exsA gene in R. meliloti, linked to exoB, encoding a protein homologous to ABC exporters that secrete non-protein substrates. ExsA is homologous to NvdA involved in the export of cyclic β-(1,2)-glucans (Stanfield et al., 1988).

The R. leguminosarum bv. viciae nodulation protein NodO has been described as partially homologous to haemolysin and is secreted from E. coli into the growth medium by HlyBD/ToIC and PrtDEF protein secretion systems. This is strong evidence that an ABC exporter is required for NodO secretion (Scheu et al., 1992; Sutton et al., 1994). The Rhizobium genes encoding the proteins required for NodO secretion are unlinked to the nodO gene which is located on the symbiotic plasmid. Several rhizobia, including R. l. trifolii, containing the cloned nodO gene secrete NodO into the medium (Scheu et al., 1992). Our results provide further evidence of the existence of a specialized secretion system widespread among Gram-negative bacteria.

ACKNOWLEDGEMENTS

This research was supported by the Polish Committee for Scientific Research grant no 6 PO4B 014 10. We thank Dr Jean-Marc Neefs (Oxford Molecular Group, Oxford, UK) for computer assistance. The authors are indebted to Mrs Maria Malek for excellent technical work.

REFERENCES


Received 18 July 1996; revised 2 December 1996; accepted 11 December 1996.