Epitope Tagging of Legume Root Nodule Extensin Modifies Protein Structure and Crosslinking in Cell Walls of Transformed Tobacco Leaves

Sébastien Gucciardo, Elizabeth A. Rathbun, Michael Shanks, Susan Jenkyns, Lora Mak, Marcus C. Durrant, and Nicholas J. Brewin

John Innes Centre, Norwich NR4 7UH, U.K.

Submitted 15 July 2004. Accepted 17 September 2004.

Root nodule extensins (RNEs) are highly glycosylated plant glycoproteins localized in the extracellular matrix of legume tissues and in the lumen of *Rhizobium*-induced infection threads. In pea and other legumes, a family of genes encode glycoproteins of different overall length but with the same basic composition. The predicted polypeptide sequence reveals repeating and alternating motifs characteristic of extensins and arabinogalactan proteins. In order to monitor the behavior of individual RNE gene products in the plant extracellular matrix, the coding sequence of PsRNE1 from *Pisum sativum* was expressed in insect cells and in tobacco leaves. RNE products extracted from tobacco tissues were of high molecular weight (in excess of 80 kDa), indicating extensive glycosylation similar to that in pea tissues. Epitope-tagged derivatives of PsRNE1 could be localized in cell walls. However, the introduction of epitope tags at the C-terminus of RNE altered the behavior of RNE in the extracellular matrix, apparently preventing intermolecular crosslinking of RNE molecules and their covalent association with other cell wall components. These observations are discussed in the light of a computational model for the RNE glycoprotein that is consistent with an extended rod-like structure. It is proposed that RNE can undergo three classes of tyrosine-based crosslinking. Intramolecular crosslinking of vicinal Tyr residues is rod stiffening, end-to-end linkage is rod lengthening, and side-to-side intermolecular crosslinking is rod bundling. The control of these interconversions could have important implications for the biomechanics of infection thread growth.

Legume crops are important for the development of sustainable agriculture and legume nodules provide an excellent model for studying fundamental aspects of plant-microbe signaling and cell morphogenesis (Limpens and Bisseling 2003). During the differentiation of root nodules, the development of specialized tissues from primordia and meristems is closely coupled to the subcellular process of tissue and cell invasion by *Rhizobium* spp. (Brewin 1991, 2004).

Symbiosis transforms *Rhizobium* spp. from soil saprophytes into nitrogen-fixing endosymbionts through the progressive remodeling of the plant-microbial cell interface. *Rhizobium* spp.-derived Nod factors cause root hair deformations by perturbing plant cell wall growth and activating cell division processes in the root cortex. Following these pre-infection events, a

matrix-filled tunnel, the infection thread, serves as a conduit for colonizing bacteria. Its transcellular orientation is controlled by the cytoskeleton, and its extensibility apparently is controlled by peroxide-driven protein crosslinking (Brewin 1998; 2004).

Root nodule extensin (matrix glycoprotein) first was identified as a major component of the infection thread matrix using a monoclonal antibody, MAC265 (VandenBosch et al. 1989). The corresponding set of glycoproteins was isolated by immunopurification from pea tissues and shown by Fourier transform infrared (FTIR) spectroscopy to be composed of approximately 70% carbohydrate. Following protein sequencing and polymerase chain reaction (PCR)-based cDNA cloning, the matrix glycoproteins (MAC265 antigens) were found to be encoded by a family of genes in *Pisum sativum*, *Vicia faba*, and *Medicago truncatula* (Rathbun et al. 2002). Different genes in a single plant species apparently encode root nodule extensins (RNEs) of different lengths, but with very similar molecular structures. All RNEs have repeating motifs that are characteristic of extensins and of arabinogalactan proteins (AGPs), and these are interspersed along the length of the macromolecule. For example, PsRNE1 from pea (Fig. 1) has 11 extensin motifs (Ser-Pro)_x that are predicted to carry relatively small arabinose glycosylations. There are also six motifs with noncontiguous (hydroxy)-proline residues that are predicted to be targets for large glycan substitutions, each comprising a 1-3, β -linked galactose backbone with one or two oligosaccharide side chains (Kieliszewski 2001). Extensin-AGP copolymers apparently are widespread in legumes but they are not common elsewhere in the plant kingdom. Gum arabic, the product of *Acacia senegal*, is a member of the same family of proteins as RNEs, although there still is not a complete sequence of the polypeptide component of this glycoprotein (Goodrum et al. 2000).

At present, there are no X-ray or nuclear magnetic resonance (NMR) structural data for complete extensin molecules. Electron micrographs of individual extensin molecules from carrot showed them to take the form of kinked rods of variable overall geometry (Stafstrom and Staehelin 1986a). Deglycosylation resulted in a collapse of the structure to a more globular form; therefore, the carbohydrate component of extensin seems to be required in order to keep the molecule in the elongated conformation required to fulfill its biological function (Stafstrom and Staehelin 1986b).

In order to monitor the behavior of individual RNE gene products in the plant extracellular matrix, the coding sequence of PsRNE1 was expressed in insect cells and in tobacco cells following transient transformation. The structure of RNE also has been investigated by protein modeling based on ab initio calculation of the geometry and energy of the peptide

Corresponding author: N. J. Brewin; Telephone: +44 1603 450273; Fax: +44 1603 450045; E-mail: nick.brewin@bbsrc.ac.uk

and glycan parts of the molecule. The results are consistent with a rod-shaped structure capable of intermolecular conjugation through crosslinking of tyrosine residues. Introduction of epitope tags at the C-terminus of the glycoprotein apparently modified its ability to form intermolecular bonds by end-to-end conjugation. We conclude that the properties of RNE and its state of glycosylation or deglycosylation may have important implications for the biomechanics of infection thread growth.

RESULTS

In the present study, the properties of three complete coding sequences derived from pea tissues were investigated by reverse-genetic approaches. Coding sequences of PsRNE1 and PsRNE3 are illustrated in Figure 1. The sequence for PsRNE2 is identical to that of PsRNE1 except that the run of seven consecutive prolines is replaced by a sequence of only five. Strategies for the introduction of epitope tags into PsRNE1 are illustrated in Figure 2. A similar set of constructs also was developed for PsRNE3 (*not shown*). Both of these sets of vectors were introduced by transformation into tobacco leaves. For comparison, the coding sequences for PsRNE1, -2, and -3 (including the leader peptides) also were introduced by transformation into insect cells in order to examine the effects of the plant cell glycosylation machinery on the nature of the protein product.

Expression of PsRNE1 in insect cells.

Following introduction of the coding sequences into an insect cell expression system, the expressed proteins from BV907 (PsRNE1) and BV908 (PsRNE2) were named RNEX1 and RNEX2, respectively. (Attempts to express PsRNE3 were unsuccessful.) The protein product was detectable by MAC265 antibody (Fig. 3). The same antigen band (approximately 28 to 30 kDa) also was detected by R82, an antiserum raised against a synthetic peptide derived from the predicted sequence of

PsRNE1 (data not shown). Following sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (Fig. 3), the observed mobility of the RNE protein, relative to the predicted molecular weight (20 kDa), suggested that a certain amount of glycosylation may have taken place in the insect cell system. The minor difference in mobility observed for PsRNE1 relative to PsRNE2 was consistent with the absence of two proline residues from the latter. For comparison, the mobility of RNE antigens derived from root nodules also is shown. The major mobility shift probably is the result of extensive glycosylation with arabinogalactan (AG) blocks in plant tissues.

Localization of RNE-yellow fluorescent protein in cell walls of tobacco leaves.

Transient transformation of tobacco leaves was used as a test system to investigate the expression of PsRNE1 in a higher plant. (A nonlegume system was chosen in order to reduce the risk of possible interference from or molecular interaction with other cell wall glycoproteins. Moreover, stable transformation of *M. truncatula* was unsuccessful using the constructs described in Figure 2.) Following transformation of tobacco

Fig. 2. Expression constructs and epitope-tagged derivatives of PsRNE1. Triple Myc epitope (Mycx3) was introduced close to the N-terminus (with reiteration of the NQ dipeptide), or close to the C-terminus (with reiteration of the YHH tripeptide). Yellow fluorescent protein (YFP) was introduced as a direct addition to the C-terminus. A similar range of constructs was developed for PsRNE2 and PsRNE3 (*not shown*). 35S, CaMV 35S promoter; SP, signal peptide sequence; RNE, coding sequence; CaMV term, termination sequence.

Fig. 1. Predicted polypeptide sequences for pea root nodule extensins (RNEs). Each of the two coding sequences, PsRNE1 and PsRNE3, is displayed to emphasize the higher order repeat motifs present (in variable number) in these two polypeptides (and in other members of the RNE family). PsRNE1, PsRNE2, and PsRNE3 are full-length coding sequences corresponding to PsExt5.5, PsExt.RT1, and PsExt.RT3, respectively (Rathbun et al. 2002). PsRNE2 (*not shown*) is similar to PsRNE1 but with SerPro₅ instead of SerPro₇. The end of the signal peptide is marked by an asterisk. Extensin motifs (SerPro_x) are underlined: contiguous proline residues usually are hydroxylated and then glycosylated with oligomers of arabinose and the serines usually are O-substituted with galactose. Motifs involving alternating prolines are marked in italics and probably are sites for addition of large arabinogalactan motifs (Kieliszewski 2001). The positions of epitope insertions are marked by vertical arrowheads.

leaves, the distribution of PsRNE1-yellow fluorescent protein (YFP) was investigated by immunofluorescence. As expected, the label was confined largely to plant cell walls and was somewhat punctate in distribution (Fig. 4A), perhaps reflecting the points of new growth or turnover in the plant cell wall. In a few instances (data not shown), some of the fluorescent label was observed in cytoplasmic vesicles adjacent to the wall. As a control, the distribution of unconjugated YFP also was observed to be cytoplasmic in vacuolated epidermal cells of transformed tobacco leaves (Fig. 4B).

Instability of the RNE-YFP conjugate.

The size distribution of extractable RNE-YFP conjugates was examined by electrophoresis and immunoblotting (Fig. 5). Using anti-green fluorescent protein (GFP) antibody, a single

Fig. 3. Expression of PsRNE1 and PsRNE2 in insect cells. Western blot, showing immunolabeling with MAC265 antibody for root nodule extensin (PsRNE1 and PsRNE2) expressed in insect cells, relative to staining with Coomassie blue. Attempts to express PsRNE3 were unsuccessful (data not shown). As a negative control, proteins from insect cells infected with vector alone are shown (left); as a positive control, RNE from pea root nodule extracts is shown (MGP).

antigen band was observed corresponding in gel mobility to what would be expected for the RNE-YFP conjugate. The same band also reacted with MAC265, which detects the RNE component of the protein conjugate. As expected, the YFP conjugate with PsRNE1 was of higher molecular weight than the conjugate with PsRNE3. Furthermore, the difference in mobility was greater than might be expected from the difference in polypeptide sequence alone (Fig. 1). This can be taken as further evidence that the AG glycosylation site of PsRNE1 probably is used in transformed tobacco tissues.

It was observed, both for RNE1-YFP and for RNE3-YFP, that a relatively low molecular weight product comprising YFP alone also was present in the tobacco leaf extracts (Fig. 5). This indicated that the YFP epitope could be cleaved from the RNE component in transformed leaf tissues. When extracts were stained with MAC265 antibody, additional bands were observed that could not be stained with the anti-GFP antiserum. One of these bands corresponded in mobility to the glycosylated PsRNE1 monomer, presumably after cleavage and release of the YFP epitope. Additional bands also were observed at lower electrophoretic mobility. These are presumed to be conjugates of RNE. The fact that these RNE bands do not react with anti-GFP antiserum could indicate that the GFP epitope is modified or masked in all positions in the multimeric state so that they are no longer antigenic. Alternatively, it could indicate that cleavage of the C-terminal YFP epitope is a precondition for end-to-end conjugation of RNE macromolecules. This second interpretation seems more probable and suggests that RNE molecules may join end to end through their C-terminal sequences. In other words, RNE without the YFP epitope may be capable of intermolecular crosslinking, but crosslinking is restricted by the presence of a C-terminal epitope.

Effects of epitope tags on extractability of RNE antigens.

The presence of a C-terminal epitope on RNE1 or RNE3 also was found to enhance the extractability of MAC265 antigen from transformed leaf tissues of tobacco (Fig. 6). Derivatives carrying YFP as a C-terminal epitope tag were strongly extractable, whereas derivatives carrying Myc as a C-terminal

Fig. 4. Confocal immunofluorescence microscopy of vacuolated epidermal cells of tobacco leaves that have been transiently transformed with **A**, PsRNE1-yellow fluorescent protein (YFP) conjugate or **B**, YFP. Bar represents 10 μ m.

epitope tag were weakly extractable. In contrast, derivatives carrying Myc as an N-terminal epitope were nonextractable, as was the native RNE1 sequence. This indicates that C-terminal epitope tagging of RNE apparently inhibits the irreversible incorporation of these molecules into plant cell walls.

Previous research (de Lorenzo et al. 1998; VandenBosch et al. 1989) has indicated that the monoclonal antibodies MAC265, MAC236, and MAC204 each recognize slightly different epitopes associated with RNE macromolecules. For example, following treatment of Western blots with periodate, reactivity of antigen to MAC265 was unaffected, reactivity to MAC 236 was enhanced, and reactivity to MAC 204 was reduced (E. A. Rathbun, *unpublished observations*). However, the exact biochemical nature of the epitope recognized by each of these antibodies is still unknown. In the present study, it was noted that, in extracts from tobacco leaves, MAC204 antibody

preferentially labeled bands of lower mobility than did MAC265 (Fig. 7). This suggests that, whereas MAC265 preferentially recognizes the monomeric state of RNE1, MAC204 and MAC236 preferentially label the dimer and multimeric forms. (MAC236 gave a similar staining pattern to that of MAC204; data not shown.) Moreover, the fact that the unconjugated form of PsRNE1 seems to accumulate as a dimer recognizable by MAC204 and MAC236 provides a further indication that C-terminal epitope tagging inhibits the intermolecular crosslinking of RNE.

The status of N-terminal epitope tagging could not be completely deduced from these gels because of the weak immunostaining signal obtained with all antibodies. Following immunostaining of Western blots, trace quantities of antigen corresponding to the monomer of RNE1-5'Myc conjugate could be detected with MAC265. (However, this is not obvious from the Western blots presented in Figures 6 and 7.) Expression of RNE1-5'Myc was emphasized in extracts from tobacco leaves that had been incubated for a longer time period following transformation (data not shown). Extractable quantities of RNE1-5'Myc were significantly greater than for extracts de-

Fig. 5. Western blot of tobacco leaf extracts transformed with PsRNE1-yellow fluorescent protein (YFP), unconjugated YFP, and PsRNE3-YFP. Blots were immunostained with MAC265 (left) or anti-green fluorescent protein antiserum (right). Note the apparent instability of PsRNE1-YFP and PsRNE3-YFP conjugates (closed arrowheads) and the higher order structure of MAC265 antigens (double arrowheads) apparently following cleavage of the YFP epitope from the C-terminus. The mobility of unconjugated YFP is indicated (open arrowheads).

Fig. 6. Effects of epitope tag on extractability of root nodule extensin (RNE) derivatives from transformed tobacco leaves, as revealed by immunostaining of Western blots with MAC265. Derivatives carrying yellow fluorescent protein (YFP) as a C-terminal epitope tag were strongly extractable (for both RNE1 and for RNE3). Derivatives carrying Myc as a C-terminal epitope tag were weakly extractable (for both RNE1 and for RNE3). Derivatives carrying Myc as an N-terminal epitope were nonextractable, as was the native RNE1 sequence.

Fig. 7. Effects of epitope tag on extractability of root nodule extensin (RNE) derivatives from transformed tobacco leaves, as revealed by immunostaining of Western blots with **A**, MAC265 and **B**, MAC204. Positions of RNE1 monomer are indicated by an arrowhead and of higher-order structures by a double arrowhead.

rived from leaves transformed with unsubstituted RNE1, but they always were much less than for either of the RNE derivatives carrying epitope substitutions at the C-terminal end.

Attempts to detect the Myc-epitope associated with RNE were largely unsuccessful, except in the case of PsRNE1-myc3', where an antigen band of the expected size was observed using the anti-Myc antibody (9E10). In the other cases, the lack of detectable antigenicity for the relatively small Myc epitope may have been the result of steric hindrance following heavy glycosylation of the polypeptide in the extracellular matrix.

Comparison of C-terminal region of RNE sequences from different legumes.

In view of the apparent importance of the C-terminal end of RNE molecules for extractability and covalent dimerization (Figs. 6 and 7), the RNE C-terminal sequence was used for homology searching of plant EST and peptide sequence databases. It was found that all closely matching C-terminal sequences were derived from legumes and a high degree of conservation of the C-terminus was observed across a wide range of taxonomic groups (Fig. 8). Although related sequences were found to be associated with extensins in other plant species, for example in *Brassica napus* and *Arabidopsis thaliana*, the C-terminal part differed from that of legume RNEs, being -SPPHH PYLYKSPPPPYHY for these two species. The penultimate legume-specific motif SPPPPHYLYK has been replaced in *Arabidopsis* spp. by SPPHHPYLYK (Fig. 8). Moreover, the extensin-type nonlegume sequences did not exhibit the interspersed extensin-AGP motifs that apparently are characteristic of the legume-specific RNE group.

The number of matching sequences from expressed sequence tag (EST) databases was rather variable: 76 of 187,763 for *M. truncatula*, 106 of 346,582 for *Glycine max*, but only 4 of 110,563 for *Lotus japonicus*. This may reflect differences in transcript abundance in the EST libraries used or it may reflect differences in the way that these libraries were generated. Interestingly, the single RNE sequence obtained from *Lotus* spp. ESTs differed from corresponding sequences at the C-terminus of RNEs from *Pisum*, *Vicia*, and *Medicago* spp. by the absence of PPHV from the motif VHTYPPHVPHPVYHS. The motif from *Lotus* spp., VHTYPPHVPYHS, is similar to the motif found three times in PsRNE1 and twice in PsRNE3 as a histidine-rich target for AG glycosylation (Fig. 1).

Computer modeling of RNE glycoprotein structure.

Given the lack of an experimental atomic resolution structure for RNE, we carried out molecular modeling ab initio

(Fig. 9), using the consensus sequence for pea RNE derived by Rathbun and associates (2002). We first attached tetraarabinose units to the hydroxyprolines of a polyproline II helix, using the NMR studies of Akiyama and associates (1980) to establish the connectivities of the monomers. As expected, systematic conformational searching of the C-O torsion angles in a Ser(Gal)(Hyp[Ara]₄)₅ block showed that the sugar chains enjoy a high degree of conformational freedom, with a wide range of energetically similar conformers arranged around the polyproline II core. We chose the minimum energy structure from these searches, and assembled individual Ser(Gal)(Hyp[Ara]₄)_{4/5} blocks using the appropriate peptide linkers. For the two-residue Ala-Tyr linkage, we used systematic ϕ/ψ backbone torsion angle searching to optimize the geometry; whereas, for the longer peptides, we used the most similar fragments identified by short sequence searches of the Protein Data Bank (Berman et al. 2000).

In order to identify the most favorable geometry for the intramolecular isodityrosine linkages, we carried out macrocyclic conformational searching on a YAY motif, using the method described by Davies and associates (1999). This procedure identified eight low-energy conformers for the macrocyclic portion of the isodityrosine-linked YAY peptide. To characterize the extent to which these structures induce kinks in the peptide backbone, we added alanines to both termini, and measured the angle between the α -carbons of these alanines and the central Ala of the YAY motif. In an ideal extended conformation, this angle should be 180°. For the lowest energy conformer, this angle was 161°; the highest energy conformer (+5.6 kcal mol⁻¹ with respect to the lowest energy conformer) had the most acute angle (88°), and the other six conformers had intermediate energies (+1.4 to +5.0 kcal mol⁻¹) and angles (107 to 162°). This suggests that the isodityrosine linkages are quite flexible but, nevertheless, tend to reinforce a fairly linear backbone.

We used the minimum energy isodityrosine conformation when constructing the complete molecular model for RNE. The resulting rod-like structure (Fig. 9) is very reminiscent of the electron micrographs described by Stafstrom and Staehelin (1986a). Furthermore, manipulation of ϕ/ψ angles within the allowed regions of the Ramachandran plot for individual amino acids readily gave more heavily kinked geometries for the molecule, consistent with a high degree of conformational flexibility, rather than a single well-defined fold, for RNE. Modeling of RNE with substituted AG blocks was not attempted at this stage because of the size and complexity of these polysaccharide side chains. We note, however, that the structural formula of a hydroxyproline-AG polysaccharide has very recently

UniProt		
Q94ES6	PYKYSPPPPPAHTYPPHVPTPVYHSPPPPAYSPPPPA YYK SPPPPYHH	<i>Pisum</i>
Q9S9A4	PYKYSPPPPPVHTYPPHVPHPVYHSPPPPVYSPPPPH YYK SPPPPYHH	<i>Vicia</i>
Q9FS16	KHYSPPPVYHSPPPPKEKYVKSPPPPVHHYSPPHHPYLYKSPPPPYHY	<i>Arabidopsis</i>
ESTs		
AL377095	KPVYHSPPPP-VHTYPPHVPHPVYHSPPPPVHSPPPPH YYK SPPPPYHN	<i>Medicago</i>
BI419010	PYKYASPPPPPVHTY----PFPVYHSPPPPVHSPPPPH YYK SPPPPYHH	<i>Lotus</i>
BI788181	TPVYHSPPPP-YEKP-----YKYKSPPPPVHSPPPPH YYK SPPPPYHY	<i>Glycine</i>
CA411436	PYKYSPPPPPVHTYPPHTPHSIYN-----SPPPPH YYK SPPPSHHY	<i>Lupinus</i>
AJ309076	-----GTPPH YYK SPPPPHHY	<i>Sesbania</i>

Fig. 8. Predicted C-terminal sequence of PsRNE2 with ClustalW multiple sequence alignment to compare the most closely matched C-terminal sequences from other legumes and *Arabidopsis* spp. The query was the last 18 residues of the Q94ES6 sequence. The universal occurrence of the YYK motif is highlighted (bold). Sequences were derived either from the UniProt database or from expressed sequence tag (EST) databases. Not illustrated are UniProt sequences for *Pisum* spp. (Q94ES2, Q94ES3, Q94ES4, Q94ES5, and Q94ES9) and *Medicago* spp. (O24099) which were very similar to those presented above. There are two more UniProt entries (O49225 for *Glycine* spp. and Q41400 for *Sesbania* spp.), but both are short fragments only. If more than one matching EST sequence was found, only one entry per legume is shown here.

been deduced by NMR, and used to construct a three-dimensional model of this species (Tan et al. 2004).

DISCUSSION

Expression of the complete open reading frames (ORFs) of PsRNE1 and PsRNE3 in insect cells (Fig. 3) was the first direct evidence that these cDNA clones actually corresponded to the matrix glycoprotein that first was isolated by immunopurification with MAC265 antibody (Rathbun et al. 2002). Furthermore, it demonstrated that recognition of RNE by MAC265 was not dependent on a plant-specific glycosylation system. The difference in mobility between RNE expressed in insect cells (Fig. 3) compared with RNE expressed in tobacco cells (Fig. 5) is presumed to be correlated with substitution of very large AG side chains in the plant but not in the insect system.

RNE represents a family of glycoproteins with interspersed extensin and AGP motifs (Fig. 1). This family of chimeric glycoproteins has been described only among members of the legume family (Fig. 8) and, consequently, it may have co-evolved with the *Rhizobium* spp.–legume symbiosis itself. Furthermore, the close topological involvement of RNE with bacterial cells in the infection thread lumen (Rae et al. 1992) suggests an important role in the *Rhizobium* spp.–legume symbiosis. Previously, the results of cDNA cloning (Rathbun et al. 2002) indicated that, in a single legume plant, a number of genes encode RNE (glyco)proteins of varying length but of broadly similar composition. Therefore, it was anticipated that a mutational approach might not yield a distinct phenotype that could point toward the functional role of RNE molecules in symbiosis. Another approach might have been to introduce complete ORFs from pea into *Medicago* spp. to investigate their functional role. However, it was found during the course of the present study that, when leaf explants of *M. truncatula* were transformed with RNE expressed behind a 35S promoter, following

the method of Trinh and associates (1998), the callus failed to regenerate as whole plants (S. Gucciardo, *unpublished observations*). This suggests that the introduction of RNE by transformation into host legumes may require the use of tissue-specific or inducible promoters, in order to minimize possible negative effects on cell growth and development.

The sequence composition of RNE polypeptides is intriguing and provides some interesting indicators for protein complex formation and possible functional roles in symbiotic development. For example, PsRNE1 has a coding sequence of 156 residues, following cleavage of the leader peptide (Fig. 1). There are 20 Tyr residues that offer scope for peroxide-based protein crosslinking, through either intramolecular isodityrosine residues or intermolecular crossbridges (Brady and Fry 1997). In the present article, our modeling studies indicate that near-adjacent Tyr residues can be conjugated and that the preferred conformation of the resulting macrocyclic units has a fairly linear backbone. Hence, these intramolecular bridges would help to strengthen the extended rod structure predicted for the RNE macromolecule (Fig. 9).

Both the N-terminus and the C-terminus of PsRNE1 have Tyr clusters that could be involved in intermolecular crosslinking, forming di-isodityrosine linkages (Brady and Fry 1997). In the present study (Fig. 5), we demonstrated that the introduction of a YFP epitope tag at the C-terminus of PsRNE1 apparently inhibits intermolecular crosslinking. Higher molecular weight forms of the RNE antigen observed on Western blots carried the MAC265 epitope, but the YFP epitope was not detectable, suggesting that cleavage of the C-terminal epitope tag is a precondition for protein crosslinking. It is not yet clear whether the C-terminal sequences are involved in end-to-end conjugation of RNE molecules (to make dimers and other concatemers) or whether there is conjugation to some other cell wall component. Preliminary indications suggest that both processes may occur in transformed tobacco leaf cells.

Fig. 9. Ab initio protein model of root nodule extensins (RNEs). **A**, Model of the entire RNE molecule (predicted length, approximately 0.056 μ m), as taken from the generic structure of pea-derived RNE (Rathbun et al. 2002). The N-terminus is on the left. Color coding: Ara, orange; Gal, magenta; isodityrosine YXY motifs, red; other amino acids, green. **B**, A section of the molecule at higher resolution, showing an isodityrosine (IDT) linkage on the left, followed by an extensin motif (SP₅) with tetraarabinose and galactose substituents, an arabinogalactan protein motif (PHPHP) which is the target for substitution with large blocks of arabinogalactan (not shown), and an SP₄ extensin motif. The peptide sequence for the whole fragment is PYKYSSPPPPVHTYHPHPVYHSPPPPV. The YKY isodityrosine moiety was constructed from the output of a conformational search on the YAY isodityrosine linkage (see text for details). Color coding: arabinose, orange; galactose, magenta; Tyr, red; Ser, light green; Thr, dark green; Val, yellow; Pro, purple; His, blue; Lys, cyan. Sugars and amino acids are rendered in stick-and-ball and stick representations, respectively. **C**, End-on view of an SP₅ extensin motif, showing the sugar residues wrapped around the central polyproline II helix. Color coding as in B.

Extractability of the C-terminal epitope-tagged derivatives of PsRNE1 (both YFP-tagged and Myc-tagged variants) is much stronger than with the N-terminal Myc-tagged variant or with the native RNE sequence (Fig. 6). Assuming that the RNE antigen has been expressed at an equivalent level in transformed leaf tissues and that the extract has been loaded at an equivalent level on the protein gels, this implies that a C-terminal epitope tag has inhibited the irreversible crosslinking of RNE into the cell wall.

Evidence that the C-terminus of RNE may be involved in dimerization of RNE molecules also can be deduced from Figure 7. In tobacco cells, MAC204 and MAC236 preferentially labeled a higher molecular weight form of RNE (probably a dimer), whereas MAC265 preferentially labeled the monomeric glycoprotein. A similar distinction was observed previously in extracts from nodules of white lupin (de Lorenzo et al. 1998). With C-terminal epitope substitution, the proportion of monomeric RNE extractable from transformed tobacco leaves was increased relative to the putative dimeric form, implying that substitutions at this end of the molecule inhibit intermolecular crosslinking. By contrast, in the case of N-terminal substitution, only very small quantities of monomeric or multimeric forms of RNE could be detected. Following N-terminal substitution, the absence of RNE dimers detectable by MAC204 implies that a modified N-terminus inhibits the accumulation of RNE dimers but indirectly promotes irreversible crosslinking into the plant cell wall. Thus, the data presented in Figures 5, 6, and 7 give some preliminary evidence that the C-terminus and the N-terminus of RNE may both be involved in intermolecular crosslinking, although the exact mechanisms remain unclear. Furthermore, it seems that the C-terminal region of RNE may play a key role in the crosslinking of RNE, and it may be significant that there is strong conservation of peptide sequence across a wide range of legumes (Fig. 8). A specific binding protein might be predicted to associate with this C-terminal motif.

The presence of interspersed extensin and AGP glycomotifs apparently is a distinguishing feature of the legume RNEs involved in infection thread formation (Brewin 2004). Clearly, the presence or absence of the AG oligosaccharide blocks as substituents on the RNE polypeptide backbone would have a profound effect on the physical properties of the macromolecule. It was shown previously by FTIR spectroscopy (Rathbun et al. 2002) that the RNE proteins isolated from pea roots comprised more than 70% carbohydrate. This is consistent with the observed difference in electrophoretic mobility for RNE expressed in insect cells compared with the product in tobacco or pea nodules (Fig. 3). Extensins normally are incorporated irreversibly into plant cell walls. Therefore, it is reasonable to assume that the presence of the AG blocks is the distinctive feature that allows the RNE of legumes to be maintained as a soluble component of the lumen of the infection thread.

We previously have proposed a biochemical model illustrating potential mechanisms of peroxide-driven tyrosine crosslinking for RNE in the extracellular matrix (Wisniewski et al. 2000). Possible sources of peroxide are from diamine oxidase (a copper-containing enzyme that is abundant in the extracellular matrix) and NADPH oxidase. The distribution of tyrosine residues in RNE sequences (Fig. 1) indicates the potential for three types of tyrosine crosslinking. Intramolecular crosslinking is rod strengthening, end-to-end conjugation is rod lengthening, and side-to-side intermolecular crosslinking is rod bundling and may depend on prior removal of the AG blocks. The results of the present study are consistent with a picture of RNE molecules as flexible rods that are capable of end-to-end conjugation. The control of these RNE interconversions could have important implications for the biomechanics of infection thread growth.

As discussed recently (Brewin 2004), our current model proposes the following elements in order to sustain a steady state of infection thread growth, without leading to abortion. The infection thread is bounded by a cylindrical primary plant cell wall that grows at its apex as an intrusive tube within the plant cytoplasm. The plant cytoskeleton controls the orientation of apical growth by focusing the synthesis of cellulose microfibrils at the apex and by targeting cytoplasmic vesicles (containing pectins and extracellular matrix glycoproteins) to this growth point (Rae et al. 1992). Rhizobial cells within the lumen of an infection thread are embedded in a plant extracellular matrix comprising RNE (Rathbun et al. 2002) and other plant glycoproteins that are secreted into the lumen (Brewin 2004). Initially, the matrix is a fluid phase and the bacterial cells are able to grow and divide predominantly at the growing tip of the infection thread (Gage 2002, 2004). In *M. truncatula*, rhizobial cells progressively reduce their rate of cell division until all division has ceased approximately 30 to 60 μm behind the growing point. At the same time, the luminal matrix progressively becomes solidified as a result of protein crosslinking, as was shown by Higashi and associates (1987) for *Astragalus* spp.

Hydrogen peroxide is perceived to be the agent of protein crosslinking in the lumen of infection threads (Herouart et al. 2002) and the high Tyr content of RNE suggests that it may be the main target for oxidative crosslinking. Furthermore, cleavage of the AG side chains would convert RNE from a soluble glycoprotein into a more conventional extensin that probably would be more susceptible to protein crosslinking by peroxide, thus promoting the fluid-to-solid transition in the infection thread matrix. Feedback control systems that regulate the rate of protein crosslinking by peroxide could modulate infection thread development (Campbell et al. 2003; Shaw and Long 2003). Similarly, feedback control systems involving glycosidases that regulate the removal of AG blocks from RNE could regulate the fluid-to-solid phase transition in the infection thread lumen (Mitra and Long 2004).

In summary, RNE glycoproteins appear to be a family of rod-shaped macromolecules of variable length that are secreted into the lumen of the infection as a glycan-rich gum. When held as an emulsion within this fluid matrix, the colonizing rhizobial cells are able to grow and divide at the apex of the infection thread. A controlled level of tyrosine-crosslinking within the RNE molecules would serve to strengthen the molecular rods, whereas end-to-end linkage would serve to enhance the lubricant properties of these rigid rod-shaped quasi-linear molecules at the plant-bacterial cell interface. There also is the potential for a peroxide-driven gum-to-resin transition which would serve to solidify the infection thread matrix and prevent further growth of the bacterial cells. However, it is possible that this side-to-side crosslinking of RNE could occur only after removal of the AG blocks which would otherwise prevent close interaction of the RNE polypeptide backbones.

Further studies of the biochemistry of RNE macromolecules and epitope-tagged variants could lead to a better understanding of the biophysics of infection thread growth. In addition, studies of the mechanism whereby RNE is targeted into the lumen of the infection thread could help to reveal how cell wall architecture and infection thread orientation are coupled to cytoskeletal orientation and the control of the cell cycle (Brewin 2004).

MATERIALS AND METHODS

RNE sequences.

PsRNE1, PsRNE2, and PsRNE3 are full-length coding sequences corresponding to clones PsExt5.5, PsExt.RT1, and PsExt.RT3, respectively, obtained by reverse transcription-PCR (Rathbun et al. 2002). Following deletion of the leader

peptide (27 residues), the predicted molecular mass of the (un-glycosylated) PsRNE1 polypeptide is 20 kDa with the leader peptide and 17 kDa without.

Transformation of insect cells.

Sequences encoding complete ORFs for PsRNE1, PsRNE2, and PsRNE3 (Fig. 1) were cloned into pFastBac 1 vector for expression in sf21 (*Spodoptera frugiperda* ovarian cells) using the Bac-to-Bac Baculovirus system following the instructions of the manufacturer (Invitrogen Life Technologies, Basingstoke, U.K.). The inserts (including the signal peptide) first were amplified from cDNA clones by PCR using gene-specific primers. PER20 (GCGAATTCAATGAGGTCCCTAATGG), containing an *Eco*R1 site, was used as the forward primer, and the reverse primer PER22 (CGTCTAGATGCCAATGATGTGC) was used with an *Xba*I site. PCR products were cloned into pFastBac using these restriction sites. The constructs were confirmed by sequencing and transformed (plus vector alone) into *Escherichia coli* DH5 α containing bacmid DNA. White colonies containing the recombinant bacmids (plus some controls with vector alone) were isolated and checked for the presence of the cloned DNA by PCR and restriction digests. Bacmid DNA was isolated from each culture and used to transfect sf21 cells in monolayer culture. The culture media, which now contained recombinant virus particles, was used to re-infect more insect cells. Infected cells were titrated and recombinant virus produced from single plaques to give the optimum multiplicity of infection. Following sf21 expression of RNE, the cells were washed to remove culture medium and protein released from the cellular pellet with either Tris-buffered saline containing NP40 (1% vol/vol), with or without dithiothreitol and ascorbate (both at 10 mM), or the commercial cell lysis product, Cytobuster (Novagen, Merck KGaA, Darmstadt, Germany). Proteins were examined with SDS-PAGE by dot assay and immunoblots with both R82 and MAC265.

Transformation of tobacco cells.

In order to construct the vector for plant transformation, sequences encoding PsRNE1 and PsRNE3 were cloned into the enhanced 35S/CaMV cassette of pGreen binary vector using the *Eco*RI restriction site (Hellens et al. 2000). The T-DNA of this vector also contained the *NptII* gene behind the Nos promoter (a derivative constructed by B. Field, *personal communication*). For the agroinfiltration procedure, cultures of *Agrobacterium tumefaciens* were grown to stationary phase at 28°C in L-broth supplemented with kanamycin (50 μ g/ml) and tetracycline (5 μ g/ml). Bacterial cells were collected by centrifugation at 5,000 $\times g$ for 15 min at room temperature and resuspended in water containing 10 mM MgCl₂ and acetosyringone at 150 μ g/ml. Cells were left in this medium for 3 h and then infiltrated into the abaxial air spaces of leaves from 2- to 4-week-old plants of *Nicotiana benthamiana* (Kapila et al. 1997). This protocol utilized a suppressor of silencing to permit the accumulation of the transiently expressed proteins until the senescence of the leaf (Voinnet et al. 2003). *Agrobacterium* strains (GV3101 containing pSoup) harboring RNE-derivative sequences cloned into a pGreen binary vector were co-inoculated with an *Agrobacterium* strain (C₅₈C₁) containing the sequence of the suppressor of silencing, p19, cloned behind the 35S promoter of pBin61 binary vector. (O. Voinnet provided the *Agrobacterium* strain.) The proteins were extracted in phosphate buffer (20 mM, pH 8) 8 days following infiltration of leaves.

Plant growth and inoculation.

Germinating seed of pea (*Pisum sativum* var. Frisson) were inoculated with cultures of *Rhizobium leguminosarum* bv. *viciae* 3841 as previously described (Rathbun et al. 2002). Seedlings

were germinated and maintained in a growth cabinet at day and night temperatures of 20 and 16°C, respectively, with a photoperiod of 16 and 8 h, respectively.

Antibodies and antisera.

R82 antiserum from rabbits was raised against the synthesized peptide HypTPHKPYKYPShyp-KLH (Sigma-Genosys Ltd., Cambridge, U.K.). This apparently is a unique motif of PsRNE1 and related polypeptides. The following rat monoclonal antibodies were used in this study: MAC265, MAC236, MAC204, and LM1, as previously described (Rathbun et al., 2002). The GFP monoclonal antibody (#8371) was obtained from Clontech (Reading, U.K.).

Protein separation and Western blotting.

SDS-PAGE gels were adapted from the protocol of Sambrook and associates (1989) and adjusted to different compositions according to the size of protein being analyzed. Western blotting was carried out as previously described (Rathbun et al. 2002). Procedures for Western blotting and for purification of RNE from pea seedling roots were carried out as previously described (Rathbun et al. 2002).

Computational methods.

Systematic conformational searches on the arabinose chains attached to a polyhydroxyproline backbone, assembly of the complete RNE model, and molecular mechanics geometry optimizations were done using Insight II software (Accelrys Ltd., Cambridge). Conformational searches on the YAY moiety were done with Chem-X software (Oxford Molecular Ltd., Oxford), following the procedure of Davies and associates (1999). The resulting isodityrosine-linked macrocycles were imported into Insight II for final geometry optimization, and the lowest-energy conformer was subsequently used in model building. The initial complete RNE model was horseshoe-shaped; it was converted into a more linear form by manipulation of the ϕ/ψ angles of selected residues within the allowed regions of the Ramachandran plot, before geometry optimization using the consistent valence force field and conjugate gradient algorithm, to an overall derivative of 1.0 kcal mol⁻¹ Å⁻¹.

Studies of sequence homology.

The C-terminal part of the PsRNE2 (Q94ES6) sequence (18 residues) was used as query in WU-Blast2 (W. Gish, *personal communication*) search of the UniProt and all EST databases (as available at the EMBL-EBI website, June 2002), using the WU-blastp and WU-tblastn algorithms, respectively, with default parameters. Sequence alignment was made with ClustalW 1.83 (Thompson et al. 1994). Additionally, all *L. japonicus*, *M. truncatula*, and *G. max* ESTs were collected from the EMBL database and stored locally (May 2004). These EST databases were searched using the tblastn algorithm (Altschul et al. 1997) at the threshold e-value of 10⁻⁴ or better, using the same query sequence as above.

ACKNOWLEDGMENTS

This work was supported by the John Innes Foundation (S. Gucciardo, S. Jenkyns, and L. Mak) and by a grant-in-aid from UK-BBSRC. We thank O. Voinnet and B. Field for help with tobacco leaf transformation, and B. Arnold for EST sequence retrieval.

LITERATURE CITED

- Akiyama, Y., Mori, M., and Kato, K. 1980. 13-C-NMR analysis of hydroxyproline arabinosides from *Nicotiana tabacum*. Agric. Biol. Chem. Tokyo 44:2487-2489.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W., and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST:

- A new generation of protein database search programs. *Nucleic Acids Res.* 25:3389-3402.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. 2000. The protein data bank. *Nucleic Acids Res.* 28:235-242.
- Brady, J. D., and Fry, S. C. 1997. Formation of Di-isodityrosine and loss of isodityrosine in the cell walls of tomato cell-suspension cultures treated with fungal elicitors or H₂O₂. *Plant Physiol.* 115:87-92.
- Brewin, N. J. 1991. Development of the legume root nodule. *Annu. Rev. Cell Biol.* 7:191-226.
- Brewin, N. J. 1998. Tissue and cell invasion by *Rhizobium*: The structure and development of infection threads and symbiosomes. Pages 417-429 in: *The Rhizobaceae*. H. P. Spaink, A. Kondorosi, and P. J. J. Hooykaas, eds. Kluwer, Dordrecht, The Netherlands.
- Brewin, N. J. 2004. Plant cell wall remodeling in the *Rhizobium*-legume symbiosis. *Crit. Rev. Plant Sci.* 25:1-24.
- Campbell, G. R. O., Sharypova, L. A., Scheidle, H., Jones, K. M., Niehaus, K., Becker, A., and Walker, G. C. 2003. Striking complexity of lipopolysaccharide defects in a collection of *Sinorhizobium meliloti* mutants. *J. Bacteriol.* 185:3853-3862.
- Davies, M. K., Durrant, M. C., Levason, W., Reid, G., and Richards, R. L. 1999. Synthesis, spectroscopic and structural studies on transition metal carbonyl complexes of cyclic and tetra-selenoether ligands. *J. Chem. Soc. Dalton Trans.* 84:1077-1083.
- de Lorenzo, C. A., Fernandez-Pascual, M. M., and de Felipe, M. R. 1998. Subcellular localization of glycoprotein epitopes during the development of lupin root nodules. *Protoplasma* 201:71-84.
- Gage, D. J. 2002. Analysis of infection thread development using Gfp- and DsRed-expressing *Sinorhizobium meliloti*. *J. Bacteriol.* 184:7042-7046.
- Gage, D. J. 2004. Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. *Microbiol. Mol. Biol. Rev.* 68:280-299.
- Goodrum, L. J., Patel, A., Leykam, J. F., and Kieliszewski, M. J. 2000. Gum arabic glycoprotein contains glycomodules of both extensin and arabinogalactan-glycoproteins. *Phytochemistry* 54:99-106.
- Hellens, R. P., Edwards, E. A., Leyland, N. R., Bean, S., and Mullineaux, P. M. 2000. pGreen: A versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. *Plant Mol. Biol.* 42:819-832.
- Herouart, D., Baudouin, E., Frendo, P., Harrison, J., Santos, R., Jamet, A., Van de Sype, G., Touati, D., and Puppo, A. 2002. Reactive oxygen species, nitric oxide and glutathione: A key role in the establishment of the legume-Rhizobium symbiosis? *Plant Physiol. Biochem.* 40:619-624.
- Higashi, S., Kushiya, K., and Allan, D. 1987. Electron-microscopic observations of infection threads in driselase-treated nodules of *Astragalus sinicus*. *Can. J. Microbiol.* 32:947-952.
- Kapila, J., DeRycke, R., VanMontagu, M., and Angenon, G. 1997. An Agrobacterium-mediated transient gene expression system for intact leaves. *Plant Sci.* 124:227-237.
- Kieliszewski, M. J. 2001. The latest hype on Hyp-O-glycosylation codes. *Phytochemistry* 57:319-323.
- Limpens, E., and Bisseling, T. 2003. Signaling in symbiosis. *Curr. Opin. Plant Biol.* 6:343-350.
- Mitra, R. M., and Long, S. R. 2004. Plant and bacterial symbiotic mutants define three transcriptionally distinct stages in the development of the *Medicago truncatula*/*Sinorhizobium meliloti* symbiosis. *Plant Physiol.* 134:595-604.
- Rae, A. L., Bonfantefasolo, P., and Brewin, N. J. 1992. Structure and growth of infection threads in the legume symbiosis with *Rhizobium leguminosarum*. *Plant J.* 2:385-395.
- Rathbun, E. A., Naldrett, M. J., and Brewin, N. J. 2002. Identification of a family of extensin-like glycoproteins in the lumen of *Rhizobium*-induced infection threads in pea root nodules. *Mol. Plant-Microbe Interact.* 15:350-359.
- Sambrook, J., Fritsch, E. F., and Maniatis, T. 1989. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, U.S.A.
- Shaw, S. L., and Long, S. R. 2003. Nod factor inhibition of reactive oxygen efflux in a host legume. *Plant Physiol.* 132:2196-2204.
- Stafstrom, J. P., and Staehelin, L. A. 1986a. Cross-linking patterns in salt-extractable extensin from carrot cell walls. *Plant Physiol.* 81:234-241.
- Stafstrom, J. P., and Staehelin, L. A. 1986b. The role of carbohydrate in maintaining extensin in an extended conformation. *Plant Physiol.* 81:242-246.
- Tan, L., Qiu, F., Lampion, D. T. A., and Kieliszewski, M. J. 2004. Structure of a hydroxyproline (Hyp)-arabinogalactan polysaccharide from repetitive Ala-Hyp expressed in transgenic *Nicotiana tabacum*. *J. Biol. Chem.* 279:13156-13165.
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. 1996. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positive-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673-4680.
- Trinh, T. H., Ratet, P., Kondorosi, E., Durand, P., Kamate, K., Bauer, P., and Kondorosi, A. 1998. Rapid and efficient transformation of diploid *Medicago truncatula* and *Medicago sativa* ssp *falcata* lines improved in somatic embryogenesis. *Plant Cell Rep.* 17:345-355.
- VandenBosch, K. A., Bradley, D. J., Knox, J. P., Perotto, S., Butcher, G. W., and Brewin, N. J. 1989. Common components of the infection thread matrix and the intercellular space identified by immunocytochemical analysis of pea nodules and uninfected roots. *EMBO (Eur. Mol. Biol. Organ.) J.* 8:335-341.
- Voinnet, O., Rivas, S., Mestre, P., and Baulcombe, D. 2003. An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J.* 33:949-956.
- Wisniewski, J.-P., Rathbun, E. A., Knox, J. P., Brewin, N. J. 2000. Involvement of diamine oxidase and peroxidase in insolubilization of the extracellular matrix: Implications for pea nodule initiation by *Rhizobium leguminosarum*. *Mol. Plant-Microbe Interact.* 13:413-420.