A central domain of *Rhizobium* NodE protein mediates host specificity by determining the hydrophobicity of fatty acyl moieties of nodulation factors

Guido V. Bloemberg,¹ Eric Kamst,¹
Marga Harteveld,¹ Koen M. G. M. van der Drift,²
Johan Havenkamp,² Jane E. Thomas-Oates,²
Ben J. J. Lugtenberg¹ and Herman P. Spaink¹

¹Leiden University, Institute of Molecular Plant Sciences,
Wassenaarseweg 64, 2333 AL Leiden, The Netherlands.
²Bijvoet Center for Biomolecular Research, Department of
Mass Spectrometry, Utrecht University, Sorbonnelaan 16,
3584 CA Utrecht, The Netherlands.

Summary

Previously, we have shown that the *nodE* gene is a
major determinant of the difference in host range
between *Rhizobium leguminosarum* biovars *viciae*
and *trifolii*. A new genetic test system for stringent
functional analysis of *nodE* genes was constructed.
By testing chimeric *nodE* genes constructed by
the exchange of polymerase chain reaction (PCR)-
generated restriction cassettes, we show that a cen-
tral domain, containing only 44 non-conserved amino
acid residues, determines the host specificity of the
NodE protein (401 amino acid residues). Mass spec-
trometric analysis of the lipo-chitin oligosaccharides
(LCOs) produced by the new test strain containing
the biovar *viciae nodE* gene shows that molecules
containing a polyunsaturated C18:4 (trans-2, trans-4,
trans-6, cis-11-octadecatetraenoic) fatty acyl moiety
are produced, as is the case for wild-type *R. leguminos-
arárum* bv. *viciae*. The LCOs determined by the biovar
*trifolii nodE* gene, which was overproduced in our
test strain, carry C18:2 and C18:3 fatty acyl chains con-
taining two or three conjugated *trans* double bonds,
respectively. Therefore, the main difference between
the *nodE*-determined LCOs of biovar *viciae* and *trifolii*
in this system is the presence or absence of one *cis*
double bond, resulting in the very different hydro-
phobicity of the LCOs. Using a newly developed spot
application assay, we show that the C18:2- and
C18:3-containing LCOs are able to induce the forma-
tion of nodule primordia on roots of *Trifolium pra-
tense*. On the basis of these and other recent results,
we propose that the host range of nodulation of the
*R. leguminosarum* biovars *viciae* and *trifolii* is deter-
mimed by the degree of hydrophobicity of the poly-
unsaturated fatty acyl moieties of their LCOs, which
is mediated by the host-specific central domain of
the NodE protein.

Introduction

The symbiosis between rhizobial bacteria and leguminous
plants, resulting in the formation of nitrogen-fixing root
nodules, is a host-specific process (for reviews see
have a broad host range, for example *Rhizobium* sp.
NGR234 nodulates more than 35 plant genera (Lewin et
al., 1987). Examples of rhizobia with a narrow host range,
which nodulate only a single or a few plant genera, are
*Rhizobium leguminosarum* bv. *viciae*, which nodulates
plants of the genera *Vicia, Pisum* and *Lens*, and the
closely related *Rhizobium leguminosarum* bv. *trifolii*,
which nodulates plants of the genus *Trifolium*. Host spe-
cificity is mediated by signal molecules from the plant and
the bacterium. Flavonoids secreted by the host plant
induce the transcription of the *nod* and *nol* genes (for
a review see Schlamann et al., 1992). Many of the *nod*
and *nol* gene products are involved in the biosynthesis
of lipo-chitin oligosaccharides (LCOs), which are secreted
by the bacterium and may carry biovar-specific substitu-
teus (for a review see Dénarié and Cullimore, 1993).
In the case of *R. leguminosarum* bv. *viciae*, LCO molecules
contain a NodE-dependent C18:4 acyl chain, the presence
of which is essential for host-specific biological activity on
the plant, including the formation of pre-infection threads
and nodule primordia (Spaink et al., 1991; Van Brussel
et al., 1992).

For *R. leguminosarum* biovar *viciae* we have shown that
the operons *nodABCJU* and *nodFEL* are sufficient for
the production of wild-type LCOs (Spaink et al., 1991).
Furthermore, the *nodABC* genes alone are sufficient to produce a
basic LCO structure. However, in the absence of the
*nodFEL* genes, no specific modifications such as the poly-
unsaturated (C18:4) fatty acyl group or the O-acetyl group
are present. The presence of the C18:4 group appears to
be determined by the *nodFE* genes, whereas the pre-

¹For correspondence. E-mail bloemberg@rulsb.
leidenuniv.nl; Tel. 71 275072; Fax 71 275088.
gene which encodes a transacylase enzyme (Spaink et al., 1991; Bloemberg et al., 1994). On the basis of homologies with genes of known function, the genes nodF and nodE are hypothesized to encode an acyl carrier protein and a β-ketoacyl synthase, respectively (Shearman et al., 1986; Bibb et al., 1989; Hopwood and Shearman, 1990; Geiger et al., 1991).

The difference in host specificity between R. leguminosarum bv. viciae and R. leguminosarum bv. trifolii is mainly determined by the nodE gene (Spaink et al., 1991). Recently, we have elucidated the structures of NodE-dependent LCOs of R. leguminosarum bv. trifolii, which carry C18:3, C20:3 and C20:4 fatty acyl chains (Spaink et al., 1995). The NodE-dependent LCOs of R. leguminosarum bv. trifolii are more hydrophobic and are produced in relatively low quantities when compared with the C18:4-containing LCOs of R. leguminosarum bv. trifolii.

To analyse the function of the nodE genes we have constructed a new test system which contains only the nodD, nodABCJ and nodFEL genes. This limited set of nod genes is sufficient for the production of LCOs and nodulation. The use of our test system has made it possible to (i) determine the biological role of the nodE genes of R. leguminosarum bv. viciae and trifolii, in the absence of the nodO gene which can compensate for the lack of the nodFEL genes (Downie and Surin, 1990); (ii) obtain a high expression level of NodE protein, leading to overproduction of nodE-dependent LCOs; and (iii) test chimeric nodE genes, which were constructed by the exchange of PCR-generated restriction cassettes. The results show that a central domain of the NodE protein determines both the host specificity of nodulation as well as the hydrophobicity of the acyl chain of lipo-chitin oligosaccharide signal molecules, presumably through a causal relationship.

Results

A test system for analysis of nodE function

In R. leguminosarum biavars viciae and trifolii, the nodF, nodE and nodD genes are genetically organized in one operon (Djordjevic and Weinman, 1991; Downie, 1991). In order to study the biological function of these genes,

![Fig. 1. A. Schematic representation of genetic system for analysing nodF, nodE and nodD gene functions. B. Construction of plasmids. pMP258 and pMP263 contain the entire nodE gene of R. leguminosarum bv. viciae and R. leguminosarum bv. trifolii, respectively, preceded by a 3'-terminal moiety of the nodF gene (Spaink et al., 1989). pMP3440 is a general construct for replacing and analysing the function of the central moiety of nodE genes. PCR was performed with the primers shown in (C) and these are indicated by arrows in (B) showing their polarity. The positions of the newly introduced KpnI and ClaI sites in the nodE genes are indicated. Hatched boxes indicate nodF gene sequences; dotted boxes indicate nodE sequences of R. leguminosarum bv. viciae; open boxes with an arrow on top represent the nodD promoter (pA). Tc⁺, tetracycline-resistance gene; Cb⁺, carbenicillin-resistance gene. Restriction sites: S, Sphi; K, KpnI; C, ClaI; B, BamHI; H, HindIII; P, PstI; XbaI, XbaI; E, EcoRI; Bg, BgII; Xh, Xhol. C. Primers used for PCR. Asterisks indicate mutations from the wild-type sequences. The Shine and Dalgarno sequence of the nodF gene is indicated by a dotted line and the abbreviation SD.](1995 Blackwell Science Ltd, Molecular Microbiology, 16, 1123–1136)
with the same efficiency as the wild-type R. leguminosarum bv. viciae genes were brought together and the nodF, nodE, nodL genes from Rhizobium strain RBL5900, of different incompatibility groups (Fig. 1A). As a chromosomal background we used Rhizobium strain RBL5900, which contains the nodD and nodAABCDE genes of biovar viciae as the only nod genes. The plasmids containing the nodF, nodE and nodL genes were brought together into strain RBL5900, resulting in strains which differ only in the origin of the nodE gene (Table 1). Since the nodD gene of biovar viciae has been shown to limit the host range of nodulation for various Trifolium species (Spaink et al., 1987), the nodD of biovar trifolii was included in the nodL-containing plasmid (Fig. 1A).

The results of the nodulation experiments (Table 2) show that the strain harbouring pMP258 containing the biovar viciae nodE gene induces nodules on Vicia sativa with the same efficiency as the wild-type biovar viciae strain 248. In the absence of the nodE gene no nodulation is observed. These results confirm the conclusion of Downie and Surin (1990) that, in the absence of the nodD gene, the presence of the nodE gene is essential for nodulation of Vicia plants. Nodulation of Trifolium pratense plants is observed at a low frequency of 15±5% both in the presence of the biovar viciae nodE as well as in the absence of a nodE gene. In the presence of plasmid pMP263 containing the biovar trifolii nodE gene, a nodulation frequency of 40% is observed on T. pratense, whereas no nodulation is observed on V. sativa plants. In plasmid pMP263, the cloned nodE gene is preceded by a 3'-terminal moiety of the nodF gene (Fig. 1B), which could explain why this plasmid has a low nodE expression level (Spaink et al., 1989). The 3' moiety of nodF in pMP3434 was, therefore, deleted and the Shine and Dalgarno sequence of nodF was introduced in front of nodE (Fig. 1B). Immunoanalysis of a total cell lysate of RBL5900(pMP2368, pMP2109) harbouring pMP3434 shows that upon induction with naringenin a NodE protein is produced migrating at the expected position (Fig. 2, lane 10). The intensity of the NodE protein band is about ten times greater in comparison with the NodE band of the strain containing pMP263 (Fig. 2, lanes 8 and 10). The nodulation efficiency of RBL5900(pMP2368, pMP3434, pMP2109) on T. pratense is 90%, which is comparable to the wild-type R. leguminosarum bv. trifolii situation (Table 2). These results show that the low nodulation frequency on T. pratense in the presence of plasmid pMP263 results from the low nodE expression level of plasmid pMP263. In conclusion, comparison of the nodulation results of strain RBL5900(pMP2368, pMP2109) containing either pMP258 or pMP3434 shows that in our system the nodE gene is the only nod gene determining the difference in specificity for Vicia or Trifolium plants.

Functional analysis of chimeric nodE genes

Nodulation results from strains containing hybrid nodE genes, resulting from homologous recombination, indicated that a central region of the nodE gene determines the difference in host specificity between the R. leguminosarum biovars viciae and trifolii (Spaink et al., 1989). In order to test this hypothesis further and to construct a test system for further analysis of the structure-function relationship, new restriction cassettes were introduced in the nodE genes using polymerase chain reaction (PCR) technology.

Use was made of the fact that in the nodE genes of both

![Fig. 2. Western blot analysis of total cell proteins of Rhizobium strains using antibodies raised against NodE protein. Approximately 10^9 Rhizobium cells were used for this analysis. Positions of NodE proteins of R. leguminosarum bv. viciae and R. leguminosarum bv. trifolii are indicated by an arrow and capital V and T, respectively. Odd and even lanes contain proteins from cells grown in the absence or presence of naringenin, respectively. Lanes 1 and 2, R. leguminosarum bv. trifolii wild-type strain 843; lanes 3 and 4, R. leguminosarum bv. viciae wild-type strain 248; lanes 5 and 6, RBL5900(pMP2368, pMP2109); lanes 7 and 8, RBL5900(pMP2368, pMP263, pMP2109); lanes 9 and 10, RBL5900(pMP2368, pMP3434, pMP2109); lanes 11 and 12, RBL5900(pMP2368, pMP268, pMP2109); lanes 13 and 14, RBL5900(pMP2368, pMP3444, pMP2109).](image-url)
biovars *viciae* and *trifolii* *Clai* and *KpnI* restriction sites can be introduced at positions 290 and 884, respectively, without affecting the amino acid sequence of the gene products. Plasmid pMP3434 containing the biovar *trifolii nodE* gene with a *Clai*--*KpnI* restriction cassette was used for the expression and nodulation assays described above. Plasmid pMP3444 only differs from plasmid pMP3434 in that it contains the 594 bp central region of the biovar *viciae nodE* gene. Immunoanalysis of cell lysates of RBL5900(pMP2368, pMP3444, pMP2109) shows that upon induction with naringenin a NodE protein is produced that migrates on SDS–PAGE in between the wild-type NodE proteins of biovars *viciae* and *trifolii* (Fig. 2, lanes 2 and 4) and at the same position as a weakly cross-reacting protein which is present in all samples (Fig. 2, lane 14). The migration on SDS–PAGE of the chimeric nodE protein is consistent with the conclusion of Spaink et al. (1989) that several domains of the NodE protein are responsible for the difference in migration of the biovar *viciae* and *trifolii* NodE proteins. As expected, pMP3434 and pMP3444 produce equal amounts of NodE protein (Fig. 2, lanes 10 and 14). The strain harbouring the chimeric nodE gene of pMP3444 nodulates *V. sativa* with a frequency of 90%, whereas the nodulation frequency on *T. pratense* is at the same basic level as seen in the absence of *nodE*. These results show that a central domain of the NodE protein determines the difference in host specificity between the biovars *viciae* and *trifolii*. This domain contains only 44 non-conserved amino acids (Spaink et al., 1989).

**Chemical analysis of lipo-chitin oligosaccharide molecules**

The isogenic strains derived from RBL5900(pMP2109, pMP2368) that vary in their *nodE*-containing plasmids were analysed for the production of LCOs using high-pressure liquid chromatography (HPLC) coupled with diode array u.v.-spectroscopic detection. The LCOs from wild-type *R. leguminosarum* biovars *viciae* and *trifolii* were used as controls. In the HPLC chromatogram, u.v.-absorption maxima at 220, 260, 303 and 330 nm suggest the presence of LCOs containing fatty acyl moieties with 1, 2, 3 or 4 trans double bonds in conjugation with the carbonyl group, respectively (Lerouge et al., 1990; Spaink et al., 1991; 1995; Schultze et al., 1982; Dénarié and Cuillermore, 1993; Geiger et al., 1994). The strain harbouring plasmid pMP263, containing the biovar *trifolii nodE*, does not produce detectable amounts of polysaturated fatty acyl-containing LCOs (data not shown). In contrast, the strain harbouring pMP3434, which overproduces the biovar *trifolii* NodE protein, produces LCOs with characteristic u.v.-absorption maxima at 260 and 303 nm. The HPLC pattern (Fig. 3B) shows several differences from that obtained from the LCOs of wild-type biovar *trifolii* (Fig. 3A): (i) production of higher total amounts of polysaturated fatty acyl-containing LCOs (relative to the amounts of monounsaturated fatty acyl-containing LCOs) than in the wild type, (ii) the peaks absorbing at 303 and 330 nm with retention times longer than 30 min that were observed in the wild type are not present, (iii) the relative intensities of the peaks for the polysaturated fatty acyl-containing LCOs in the two strains (running time 20–30 min) are very different. To identify the structures of the NodE-dependent LCOs, the HPLC fractions corresponding to the u.v. absorption maxima were subjected to positive ion mode fast atom bombardment mass spectrometric (FAB-MS) analysis. The fractions having characteristic absorption maxima at 303 and 260 nm (fractions 24, 26, 27 and 29) yield [M + H]+ pseudomolecular ions at m/z 1294, 1296, 1091 and 1093 (Table 3). These ions indicate the presence of tetra- and pentasaccharide-containing LCOs with C18:3 and C18:2 fatty acyl chains. Collision-induced dissociation (CID MS) analysis, carried out on collision of these precursor ions with air to generate fragments, supports our assignment of these structures (Fig. 4; Table 3). The results, summarized in Table 3 and shown in Fig. 3, show that LCOs from the strain harbouring pMP3434 that overproduces the biovar *trifolii* NodE protein bear a less complex mixture of fatty acyl chains than those from wild-type biovar *trifolii* (Spaink et al., 1995). The main difference is the absence of the C20 fatty acyl-containing LCOs. Furthermore, the oligosaccharide backbones are predominantly pentasaccharides instead of tetrasaccharides. The polysaturated fatty acyl moieties of the biovar *trifolii* wild-type and the NodE-overproducing strain have in common that they are both more hydrophobic than the biovar *viciae* NodE-dependent C18:4 fatty acyl moiety. In the NodE-overproducing situation, this is apparently mainly the result of the lack of the cis double bond at position 11, which is invariably present in the NodE-dependent biovar *viciae* fatty acyl moieties (Spaink et al., 1991).

The HPLC profile of the LCOs produced by the strain harbouring pMP3444 (Fig. 3C), containing the chimeric *nodE* gene, is very similar to that of the LCOs from wild-type biovar *viciae* (Spaink et al., 1991). The NodE-dependent LCOs with a characteristic u.v.-absorption maximum at 303 nm have retention times identical to the wild-type NodRlv-IV(C18:4, Ac) and NodRlv-IV(C18:4, Ac). The predicted identity of the LCOs produced by this strain with those of the wild-type *viciae* was confirmed by mass spectrometry (Table 3). Mass spectrometric analyses of fractions C24, C26, C27 and C29 show that these do not contain the LCOs bearing polysaturated fatty acyl chains (data not shown), which are present in the biovar *trifolii* (Table 3).

In conclusion, the central domain of the NodE proteins...
Fig. 3. C18-reversed phase HPLC analysis of LCOs using a photodiode array detector. Isograms and corresponding absorbance profiles at 200 nm of LCO-containing n-butanol extracts from strains LPR5045(pRIF101) (A), RBL5900(pMP2368, pMP3444, pMP2109) (B) and RBL5900(pMP2368, pMP3444, pMP2109) (C). The elution gradient is shown schematically at the top of the figure. Fractions (1 ml) were collected at 1 min intervals. Positions of NodRI-V(C18:1, Ac) and NodRI-V(C18:1, Ac) are indicated in the 200 nm absorption profiles by an arrowhead and roman numerals V and IV, respectively.
The presence of an oxonium ion at \( m/z \) 1077, together with the NodE-dependent lipo-chitin oligosaccharide molecules from Table 3, saccharide-containing LCO, bearing a C18:1 fatty acyl chain on the NodR(-V'(G18:3, Ac) designations LCO, PMP2368, pMP3444, characterized by CID MS², RBL5900(pMP2109, pMP2368, pMP3444), containing the central nodE bv. trifoli region and RBL5900(pMP2109, pMP2368, pMP3444), containing the central nodE bv. viciae region, the common LCOs such as NodRI-V(C18:1-1, Ac) NodRI-V(C18:1, Ac), NodRI-V(C18:1, Ac), NodRI-III(C18:1, Ac) were detected (Spaink et al., 1995). Additionally, fraction 27 of both strains yields an \([M+H]^+\) pseudomolecular ion at \( m/z \) 1123. Surprisingly, CID MS analysis generates fragment ions which indicate an unusual tetrasaccharide-containing LCO, bearing a C18:1 fatty acyl chain on the non-reducing terminal, and bearing an additional group, having a mass of 45 amu, on C-1 of the reducing-terminal HexNAc residue. The presence of an oxonium ion at \( m/z \) 1077, together with the inability to remove the group under de-esterifying conditions, suggests that it probably corresponds to an ethyl glycoside. Biological activities of LCOs

The biological activities of the LCOs produced by the strains derived from RBL5900(pMP2109, pMP2368, pMP2109) and varying in the nodE-containing plasmid were tested on V. sativa and T. pratense plants. For V. sativa we made use of the INI assay (Van Brussel et al., 1990). In the INI assay, induction of de novo flavonoid synthesis by the plant is tested using an indicator bacterial strain. Previously it has been shown that the biovar viciae NodE-dependent C18:4 acyl chain is required to induce an INI response on Vicia plants (Spaink et al., 1991). The results (Table 4) show that the biovar trifolii NodE-dependent LCOs are not able to induce an INI response on Vicia plant roots. In contrast, LCOs isolated from the strain harbouring pMP3444 (chimeric nodE gene) induce an INI response comparable to the indolucul effect of wild-type biovar viciae LCOs, as was expected from its production of C18:4-containing LCOs.

For T. pratense, a new spot application procedure was developed for the induction of nodule primordia by LCOs. The results show that a mixture of LCOs isolated from the strain harbouring pMP3434 (biovar trifolii nodE gene) containing the C18:3- and C18:2-bearing LCOs is able to elicit node primordia on T. pratense (Fig. 5). The only difference between LCO- and Rhizobium-induced nodule primordia is the exaggerated root hair deformation on LCO-induced primordia (Fig. 5B), whereas marked curling of root hairs is observed after inoculation with Rhizobium bacteria (Fig. 5A). A mixture of common LCOs isolated from a strain lacking the nodE gene and applied at equivalent concentrations does not elicit nodule primordia on T. pratense.

Discussion

The nodE genes are involved in the synthesis of polyunsaturated fatty acyl moieties and play an essential role in determining the host-specificity of R. leguminosarum bv. viciae and R. leguminosarum bv. trifolii (Spaink et al., 1989; 1991; Geiger et al., 1994). The aim of this work was to analyse the role of nodE in host-specific nodulation and LCO production in more detail and to study the function of a central region of the nodE protein.

In order to study the function of nodE we have developed a new genetic test system containing a defined number of nod genes sufficient for nodulation and LCO production. The main advantages of this system are (i) the lack of the nodO gene, which can complement for the nodFe genes of biovar viciae in nodulation (Downie and Surin, 1990) via a fatty acid biosynthesis-independent mechanism which is still not understood, and (ii) the presence of the genes of the nodFEL operon on plasmids of different compatibility groups, which allows the functions of these genes to be studied without polar effects on transcription.

For an efficient nodulation phenotype of the strain containing the nodE gene of biovar trifolii, the production of the NodE protein had to be increased (Table 2). This also increased the production of NodE-dependent LCOs, showing that the NodE protein level in the cell can be a limiting factor for the production of polyunsaturated fatty acyl-containing LCOs. However, it should be noted that the HPLC pattern of LCOs from the biovar trifolii NodE-overproducing strain differs from that of the wild-type strain (Fig. 3A, B).

| Table 3. NodE-dependent lipo-chitin oligosaccharide molecules from RBL5900(pMP2109, pMP2368, pMP3444) and RBL5900(pMP2109, pMP2368, pMP3444) characterized by CID MS². |
|---|---|---|
| LCO designation | Detected in fractions | Oxonium-type fragment ions |
| NodRI-V(C18:4, Ac) | C13 | 1292, 1071, 866, 665, 462 |
| NodRI-V(C18:4, Ac) | C17 | 1089, 858, 665, 462 |
| NodRI-V(C18:3, Ac) | B24 | 1294, 1073, 870, 667, 464 |
| NodRI-V(C18:2, Ac) | B26 | 1296, 1075, 872, 669, 466 |
| NodRI-V(C18:2, Ac) | B27 | 1091, 870, 667, 464 |
| NodRI-V(C18:2, Ac) | B29 | 872, 669, 466 |

a. In both strains RBL5900(pMP2109, pMP2368, pMP3444), containing the central nodE bv. trifoli region and RBL5900(pMP2109, pMP2368, pMP3444), containing the central nodE bv. viciae region, the common LCOs such as NodRI-V(C18:1-1, Ac) NodRI-V(C18:1, Ac), NodRI-V(C18:1, Ac), NodRI-III(C18:1, Ac) were detected (Spaink et al., 1995). Additionally, fraction 27 of both strains yields an \([M+H]^+\) pseudomolecular ion at \( m/z \) 1123. Surprisingly, CID MS analysis generates fragment ions which indicate an unusual tetrasaccharide-containing LCO, bearing a C18:1 fatty acyl chain on the non-reducing terminal, and bearing an additional group, having a mass of 45 amu, on C-1 of the reducing-terminal HexNAc residue.

b. B and C fractions correspond to the 1 ml HPLC fractions collected from RBL5900(pMP2109, pMP2368, pMP3444) and RBL5900(pMP2109, pMP2368, pMP3444), respectively (Fig. 3, B and C).

determines the difference in hydrophobicity of the polyunsaturated fatty acyl moieties on LCOs.

| Table 4. INI effect of LCO isolates on Vicia plants. |
|---|---|---|
| LCO isolates | Beta-galactosidase activity |
| Absent | 427 |
| Purified NodRIV(C18:4, Ac) | 7162 |
| RBL5900(pMP2368, pMP2109) | 728 |
| RBL5900(pMP2368, pMP2109, pMP3444) | 7183 |
| RBL5900(pMP2368, pMP2109, pMP3444) | 7277 |

a. Four days after adding LCO isolates to test tubes containing six Vicia sativa subsp. nigra plants, INI was determined by measuring β-galactosidase activity of the indicator strain RBL5284 in diluted plant exudates (1:30) (Van Brussel et al., 1990).
A central domain of NodE determines acyl hydrophobicity.

Fig. 4. CID tandem mass spectra of LCOs from strain RBL5900(pMP2368, pMP3434, pMP2109).
A HPLC fraction B24.
B. HPLC fraction B26.
The ions marked 'TG' arise by the loss of thioglycerol from the precursor ions. The precursor ions, therefore, correspond coincidentally to a mixture of the LCO pseudomolecular species and a matrix cluster ion.
Mass spectrometric analysis of the LCOs produced by the biovar *trifolii* wild type and the NodE-overproducing strain shows (i) the predominant production of LCOs with a pentasaccharide backbone in the NodE-overproducing strain instead of a tetrasaccharide backbone as seen in the wild type, and (ii) the absence of C20 fatty acyl-containing LCOs in the NodE-overproducing strain. An explanation for these differences could be that in the biovar *trifolii* NodE-overproducing situation all other nod genes are biovar *viciae* genes. The biovar *viciae* nodA, nodB and nodC genes are the major candidates responsible for the predominant production of pentasaccharide-containing LCOs. NodC has been suggested to function as a UDP-N-acetylglucosamine transferase involved in the production of the glycan backbone of LCOs (Spaink et al., 1993; 1994; Geremia et al., 1994). The NodB protein removes the N-acetyl group from the non-reducing terminal GlcNAc of the chitin-oligosaccharide backbone to generate a free amino group to which the fatty acyl moiety is transferred (John et al., 1993; Spaink et al., 1994). The NodA protein has been suggested to be involved in the transfer of the acyl moiety to the de-N-acetylated sugar backbone of the LCOs (Röhrig et al., 1994; Spaink et al., 1994; Atkinson et al., 1994). It is of great interest to determine which of these genes determines the difference in oligosaccharide chain length. A candidate for the protein responsible for the lack of C20 fatty acyl moieties is the ACP homologue NodF (Sherman et al., 1986; Geiger et al., 1991), the presence of which is required for the production of the polyunsaturated fatty acyl moieties (Demont et al., 1993; Ritsma et al., 1994). However, experiments in which the nodF gene of biovar *trifolii* was present instead of the biovar *viciae* nodF showed an HPLC pattern similar to that of the NodE-overproducing strain used in Fig. 3B (data not shown). A better candidate is the nodA gene. Preliminary results suggest that the NodA proteins of biovar *viciae* and *trifolii* might indeed have different specificities for polyunsaturated fatty acyl chains (T. Ritsma, unpublished results). An alternative explanation for the absence of C20 fatty acyl groups is that because of the overproduction of the NodE protein another chromosomally encoded enzyme, such as a β-keto-acyl reductase or hydroxy-acyl dehydratase, becomes limiting for their production.

Notwithstanding the differences observed in LCO production between the biovar *trifolii* NodE-overproducing strain and the wild-type strain, the nodulation behaviour of the two strains was indistinguishable (Table 2). The fact that the NodE-overproducing strain produces relatively large amounts of polyunsaturated fatty acyl-containing LCOs compared with the wild-type strain made it possible to obtain sufficient amounts to test their biological activities. LCOs containing biovar *trifolii* nodE-dependent C18:2 and C18:3 fatty acyl moieties appeared...
to be able to induce nodule primordia on *T. pratense* (Fig. 5). In contrast, these molecules are not able to induce the INI effect on *V. sativa* (Table 4), showing their host-specific character.

By testing hybrid *nodE* genes, consisting of a 5′ moiety of biovar *viciae* and a 3′ moiety of biovar *trifolii*, indications were obtained that a central domain of the *nodE* gene determines the specificity of nodulation (Spaink *et al.*, 1989). By testing a chimeric *nodE* gene consisting of 5′ and 3′ moieties of biovar *trifolii* and a 594 bp central moiety of biovar *viciae*, we have now proved that this central domain determines host specificity (Table 2). Furthermore, we show that this domain determines the hydrophobicity of the polyunsaturated acyl moieties of the LCOs.

The NodE proteins show homology with β-ketoacylsynthases (KAS) (Bibb *et al.*, 1989; Hopwood and Sherman 1990; Downie 1991). For one of the KAS proteins, FabB of *Escherichia coli*, it has been indicated that cysteine 163 is the fatty acyl-binding site (Kauppinen *et al.*, 1988). This cysteine is conserved in all the proposed KAS proteins, including NodE, and even in the distantly related chalcone synthases. Since this conserved cysteine is located centrally in the host specificity-determining domain of NodE, it is likely that this domain is involved in substrate recognition. Considering the model for the synthesis of the NodE-dependent fatty acids (Spaink, 1992), we postulate that the central domain of the biovar *trifolii* NodE protein recognizes more hydrophobic fatty acid intermediates than the biovar *viciae* NodE. The restriction cassette we have constructed will be used in future research to carry out a more detailed structure–function analysis of the central domain of NodE protein. We will also construct chimeric *nodE* genes which contain the central domain of various *kas* genes, such as the highly homologous KAS1 proteins of *Streptomyces*. The various KAS proteins of *Streptomyces* are involved in the synthesis of β-ketide antibiotics having carbon chains of various length (Hopwood and Sherman, 1990). These experiments could show whether the central domain of KAS has a general role in determining the diversity of the keto-acyl/products.

### Experimental procedures

#### Bacterial strains

The *E. coli* and *Rhizobium* strains used in this study are listed in Table 1. *E. coli* strains were grown in Luria–Bertani (LB) medium at 37°C. For isolation of lipo-chitin oligosaccharides *Rhizobium* strains were grown at 28°C in B− medium (Spaink *et al.*, 1989), whereas for analysis of NodE protein production they were grown in tryptone–yeast (TY) medium supplemented with 20% B− medium. For induction of *nod* genes, naringenin was added to a final concentration of 3.75 μM. For strains harbouring plasmids, the growth medium was supplemented with appropriate antibiotics.

#### Construction of plasmids

Molecular genetic techniques were carried out according to Sambrook *et al.* (1989), PCR was performed with *Pfu*-polymerase (Stratagene) using standard methods (Innis *et al.*, 1990). Plasmids and primers used in this study are listed in Table 1 and Fig. 2C, respectively.

The following strategy was employed to construct a general plasmid for studying the function of the central domain of NodE proteins (Fig. 2B). A Sphi–BamHII fragment from pMP263 (Spaink *et al.*, 1989), containing the *nodE* gene of biovar *trifolii*, was cloned in pC20R resulting in pMP3404 and it was used as a template for PCR with primers oMP72, oMP50, oMP79 and oMP74 to obtain the 3′ and 5′ ends of the *nodE* gene of biovar *trifolii*. To obtain a resistant marker situated in between the NodE terminal moieties, the fragments were cloned together with pC20H (Marsh *et al.*, 1984) in the general expression vector pMP1070 (Bloemberg *et al.*, 1994) resulting in plasmid pMP3440. Restriction cassettes of the central regions of the *nodE* genes were made by PCR using pMP263 and pMP258 (Spaink *et al.*, 1989) as templates with primers oMP47, oMP48, oMP45 and oMP70. Both cassettes were introduced in pMP3440 by substitution of the KpnI–ClaI pC20H part, resulting in pMP3434 and pMP3444.

pMP2109 was constructed by cloning a 1.8 kb *HindIII* fragment from pMP2107 (Bloemberg *et al.*, 1994), containing the *nodD* gene from biovar *trifolii* and the biovar *viciae nodD* gene under the control of the *nodA* promoter, into the IncW vector pRL140 (Innes *et al.*, 1988). The *nodF* gene of biovar *viciae* under the control of the T7 promoter was cloned in pMP190 (Spaink *et al.*, 1987) resulting in pMP2368.

#### NodE protein analysis

Cells of the *Rhizobium* strains were grown in liquid culture medium for 2 days. Subsequently the cultures were diluted to an OD_{600} of 0.05 and grown for 16 h in the absence or presence of naringenin. Total cell proteins were separated using 11% SDS–PAGE (Lugtenberg *et al.*, 1975) and transferred to nitrocellulose using a semi-dry electrobolting apparatus (LKB). Immunoanalysis was carried out as described previously (De Maagd and Lugtenberg, 1986). To suppress cross-reactivity, antiserum against the NodE protein (Spaink *et al.*, 1989) was pre-incubated with a total cell lysate of *E. coli* strain JM101. The blot was incubated with the antiserum at a final dilution of 1:6000 in 0.05% Tween20 buffer. NodE proteins were visualized using alkaline phosphatase-conjugated goat anti-rabbit immunoglobulins (Sigma) as recommended by the manufacturers.

#### Bio-assays

Germinated seeds of *V. sativa* subsp. *nigra* (produced in our laboratory) and *T. pratense* (Kieft Boomzadens) were inoculated with rhizobial bacteria and grown on 1.8% agar Jensen medium as described previously (Van Brussel *et al.*, 1982).
Plants were monitored for root nodules 20 days after inoculation.

NIL tests using V. sativa subsp. nigra plants were performed according to Van Brussel et al. (1990). To each plant test-tube, an amount of LCOs was added equivalent to an amount produced by a 10 ml Rhizobium cell culture after induction with naringin. The LCOs were purified using disposable octadecyl silica columns (Baker) as described below. Purified NodRlv-V(C18:4,Ac) was used as a control.

For spot application of LCO to T. pratense plants, quartz sand (0.1–0.3 mm) was coated with LCO according to López-Lara et al. (1995). Before coating, 100 mg of sand was washed several times with acetonitrile/water (60/40, v/v) dried and sterilized. Coating was performed by adding 0.5 ml of an LCO solution in acetonitrile/water (60/40, v/v) to the sand followed by drying under vacuum. Approximately 5 mg of the sand was placed on the root tip of a T. pratense plant which had been growing in an agar tube (Van Brussel et al., 1982) for 1–2 days after germination. Eight days after the application, plants were scored for the presence of nodule meristems.

Isolation and analysis of LCOs

Cells were grown for 2 days in liquid B– medium and subsequently diluted to an OD260 of 0.04 in 11 of the same medium supplemented with naringin. After 16 h of growth, the total culture was extracted for isolation of LCOs with 350 ml n-butanol. After phase separation, the n-butanol phase was collected and evaporated under vacuum. The extracted hydrophobic compounds were dissolved in 5 ml 60% acetonitrile/water (60/40, v/v), as shaking for 16 h. As a pre-purification step, the dissolved compounds were passed through a disposable octadecyl silica (ODS) column (6 ml HC, J. T. Baker) equilibrated with acetonitrile/water (60/40, v/v). HPLC with a Pep-S column (5 μm, 5 × 250 mm; Pharmacia LKB Biotechn. Inc.) was used to separate the lipochitooligosaccharides. Before application to the column, 2 ml of each sample was diluted 1:1 with water. The column was eluted with a block-gradient of acetonitrile/water (40/60, v/v), acetonitrile/water (45/55, v/v) and acetonitrile/water (60/40, v/v) at a flow of 1 ml min–1. The eluent was monitored using a photodiode array detector (Pharmacia LKB Biotechn. Inc.). Fractions of 1 ml were collected and stored at –20°C under argon.

FAB-MS and CID MS analyses

Positive ion FAB-mass spectra were obtained using MS 1 of a JEOL JMS-SX/SX 102A tandem mass spectrometer operated at 10 kV accelerating voltage and using mono-thioglycerol as matrix. The FAB gun was operated at 6 kV accelerating voltage with an emission current of 10 mA and using xenon as the bombarding gas. Spectra were scanned at a speed of 30 s for the full mass range specified by the accelerating voltage used and were recorded and averaged using a Hewlett-Packard HP 9000 data system running JEL COMPLEMENT software. CID tandem mass spectra were obtained using all four sectors of the same instrument under similar conditions, and using air as the collision gas in the third-field free region collision cell at a pressure sufficient to reduce the parent ion to one third of its original intensity.

Acknowledgements

We thank Jeanette Beelen-Van Duin and Teun Tak for technical assistance with some of the experiments. The work of both the Leiden and the Utrecht groups was supported by the Netherlands Foundation for Chemical Research, with financial aid from the Netherlands Organization for Scientific Research (NWO). H.P.S. was supported by the Royal Netherlands Academy of Arts and Sciences and a Pionier Grant from the NWO.

References


This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.