Available online at www.sciencedirect.com
SCIENCE DIRECT®
Carbohydrate Research 338 (2003) 237–250
CARBOHYDRATE RESEARCH
www.elsevier.com/locate/carres

Structural determination of the lipo-chitin oligosaccharide nodulation signals produced by *Rhizobium giardinii* bv. giardinii H152

M. Eugenia Soria-Díaz,[a] Pilar Tejero-Mateo,[a] José L. Espartero,[b] Miguel A. Rodríguez-Carvajal,[a] Belén Morón,[c] Carolina Sousa,[c] Manuel Megías,[c] Noelle Amarger,[d] Jane Thomas-Oates,[e] Antonio M. Gil-Serrano[a,*]

[a]*Departamento de Química Orgánica, Facultad de Química, Universidad de Sevilla, 41071 Sevilla, Spain*
[b]*Departamento de Química Orgánica y Farmacéutica, Facultad de Farmacia, Universidad de Sevilla, 41071 Sevilla, Spain*
[c]*Departamento de Microbiología y Parasitología, Facultad de Farmacia, Universidad de Sevilla, 41071 Sevilla, Spain*
[d]*Laboratory of Microbiology of the Soils, INRA, 21034 Dijon, France*
[e]*Michael Barber Centre for Mass Spectrometry, Department of Chemistry, UMIST, Manchester M60 1QD, UK*

Received 30 July 2002; accepted 18 October 2002

Abstract

Rhizobium giardinii bv. giardinii is a microsymbiont of plants of the genus *Phaseolus* and produces extracellular signal molecules that are able to induce deformation of root hairs and nodule organogenesis. We report here the structures of seven lipochitooligosaccharide (LCO) signal molecules secreted by *R. giardinii* bv. giardinii H152. Six of them are pentamers of GlcNAc carrying $C_{16:0}$, $C_{18:0}$, $C_{20:0}$ and $C_{18:1}$ fatty acyl chains on the non-reducing terminal residue. Four are sulfated at C-6 of the reducing terminal residue and one is acetylated in the same position. Six of them are *N*-methylated on the non-reducing GlcN residue and all the nodulation factors are carbamoylated on C-6 of the non-reducing terminal residue. The structures were determined using monosaccharide composition and methylation analyses, 1D- and 2D-NMR experiments and a range of mass spectrometric techniques. The position of the carbamoyl substituent on the non-reducing glucosamine residue was determined using a CID-MSMS experiment and an HMBC experiment.

Keywords: *Rhizobium giardinii* bv. giardinii; Nod factors; Structure

1. Introduction

Soil bacteria of the genus *Rhizobium* are nitrogen-fixing bacteria that are able to invade the roots of leguminous plants and trigger the formation of the nodule that contains the nitrogen-fixing microsymbiont. Infection and nodule development are highly species specific and are largely controlled by reciprocal signal exchange between the symbiotic partners. The host plant roots secrete specific flavonoids or isoflavonoids that induce the rhizobial nodulation genes (*nod*, *nol* and *noe* genes). These genes have been shown to be involved in the synthesis and secretion of bacterial nodulation signals called nodulation factors or lipochitooligosaccharides (LCOs) because of their chemical structure. This chemical structure consists of a backbone of two to six β-(1→4)-linked GlcNAc residues bearing an amide bound fatty acyl residue (saturated or unsaturated) on the non reducing terminal GlcN residue. This basic structure has variations that are dependent on each strain or species and determine the host-specificity. The C-6 of the reducing GlcNAc residue may either be

Abbreviations: LCO, lipo-chitin oligosaccharide or lipochitooligosaccharide; CID-MSMS, collision induced dissociation mass spectrometry mass spectrometry; LSIMS, liquid secondary ions mass spectrometry; ES-Q-o-TOF MS, electrospray quadrupole orthogonal time of flight mass spectrometry; DBF-COSY, double band filtered correlation spectroscopy; HSQC, ^{1}H-detection mode via single quantum coherence; HMBC, heteronuclear multiple bond correlation.

* Corresponding author. Fax: +34-954-624960
E-mail address: agil@us.es (A.M. Gil-Serrano).

0008-6215/03/$ - see front matter
PII: S0008-6215(02)00434-2

unsubstituted, sulfated, acetylated, fucosylated, acetylfucosylated, sulfofucosylated, or 2-*O*-methyl fucosylated. The C-3 of the same residue may be substituted by an arabinose residue. In addition to the amide bound fatty acyl residue, the non-reducing GlcN may be also *N*-methylated and substituted at different positions by acetyl or carbamoyl groups (for reviews see references).[1–4]

The common bean (*Phaseolus vulgaris*) is the most important legume in human nutrition, and is known to be promiscuous.[5] This plant has the ability under controled conditions to form nodules with a broad range of rhizobia including strains belonging to different genera and species, and strains which can not been assigned to any of the recognized species. In the field, the rhizobia which nodulate cultivated bean have been found, so far, to belong to five species of *Rhizobium*: *R. leguminosarum* bv. phaseoli, *R. etli*, *R. tropici*, *R. gallicum* and *R. giardinii*.[6–9] Amarger *et al.*[9] have proposed that *R. giardinii* should be subdivided into two biovars, as follows: *R. giardinii* biovar giardinii and *R. giardinii* biovar phaseoli.

In this paper we describe the structure of the Nod factors produced by *R. giardinii* bv. giardinii H152. This strain was isolated from root nodules of *P. vulgaris* originating from France.[9] We have in particular studied the different functional groups attached to the Nod factors in order to determine those groups that are absolutely necessary for nodulation of this strain's normal host plant, the bean. On the basis of mass spectrometry and 1D- and 2D-NMR results we conclude that all these Nod factors are carbamoylated on the non-reducing glucosamine residue. The position of the substituent was determined by a CID-MSMS experiment performed on the B_1 fragment[10] and by an HMBC experiment.

2. Experimental

2.1. Bacterial strain, cell cultures and TLC analysis of Nod factors

Rhizobium giardinii bv. giardinii strain H152 was grown in a minimal B^- medium[11] until the cultures reached an A_{600} of 0.6–0.8. The cells were induced to produce Nod factors by the addition of 1 μM apigenin solution. Nodulation factors were labelled *in vivo* and analysed by TLC using the procedure described by Spaink *et al.*[11] Briefly, *R. giardini* was grown in 1 mL of minimal B^- medium supplemented with apigenin and was radioactively labelled by the addition of 0.2 μCi of [^{14}C]-glucosamine hydrochloride (50–62 mCi/mmol) from Amersham Pharmacia Biotech. The cultures were grown to stationary phase and the supernatant was extracted with 0.5 mL of water saturated with *n*-butanol. The solvent was evaporated to dryness and the residue was re-suspended in 40 μL of *n*-butanol. Five microlitres of the solution were applied to the TLC plates (RP-18 F254S, from Merck) where the Nod factors were separated using acetonitrile–water (1:1, v/v) as the mobile phase. TLC plates were dried and exposed to Kodak BioMax MR-2 film for 15 days at room temperature.

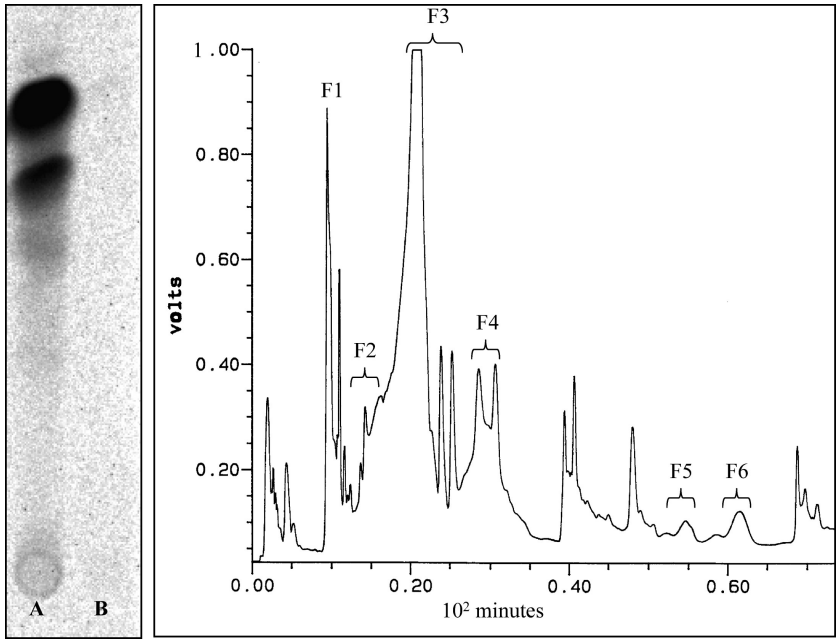


Fig. 1. TLC profile of radiolabelled Nod factors and HPLC profile of *n*-butanol extract of culture from *R. giardinii* bv. giardinii H152. Absorbance at 206 nm. Fractions pooled as indicated.

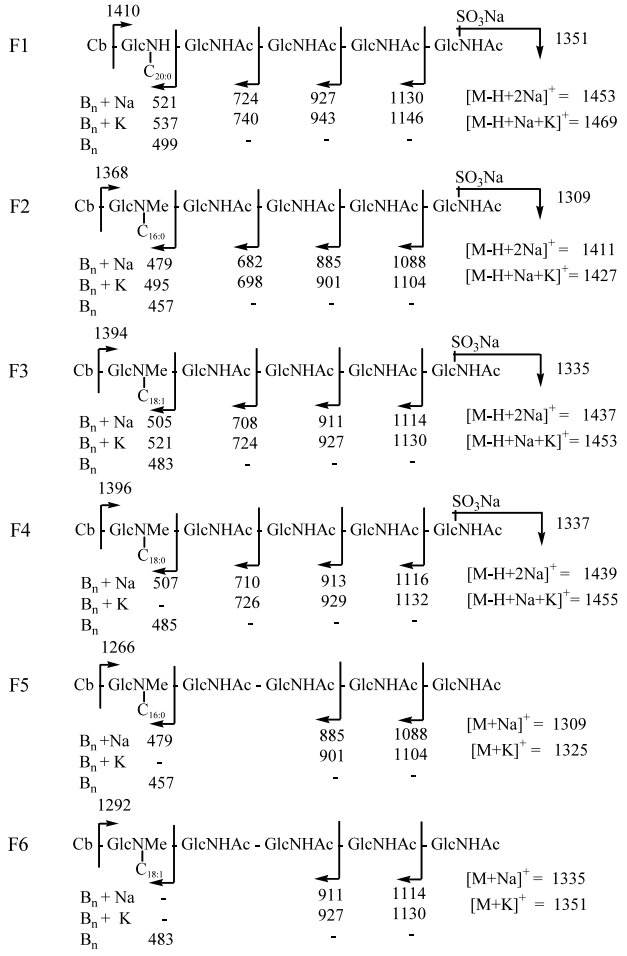


Fig. 2. Results of the LSIMS analysis of fractions F1–F6. Fragmentation scheme of the six Nod factors identified in the HPLC fractions.

2.2. Nod factor purification

Apigenin-induced culture medium (9 L) was extracted with *n*-butanol (3 L). The butanol extract was evaporated to dryness under vacuum and the residue was suspended in 45 mL of acetonitrile–water (3:2, v/v) which was then brought to acetonitrile–water (1:4, v/v). A pre-purification step was performed by passing the crude extract through a C_{18} cartridge from Resprep, from which the Nod factors were eluted using different acetonitrile–water solutions (20, 45, and 60%). The Nod factors in these fractions were further fractionated by HPLC on a semi-preparative C_{18} reversed phase column (250 × 7.5 mm, Spherisorb ODS2, 5 μm, Tracer) using isocratic elutions of acetonitrile–water (20%, 5 min; 30%, 30 min; 40%, 30 min; 60%, 15 min), and then using a linear gradient for 10 min from 60 to 100% acetonitrile. The eluent from the HPLC was monitored at 206 nm and 2.5 mL fractions collected. These fractions were then pooled in order to obtain fractions corresponding to the major HPLC peaks.

2.3. Monosaccharide composition and methylation analysis

Glycosyl composition analysis was carried out after methanolysis using anhydrous methanolic 0.625 M HCl (16 h, 80 °C). The samples were re-*N*-acetylated using Ac_2O–pyridine (1:1, v/v), then trimethylsilylated with pyridine–BSTFA (1:1, v/v) and analysed by GLC-MS.[12] Glucosamine was assigned as D-GlcN following GLC-MS analysis of its trimethylsilylated 2-butyl glycosides prepared by using (+)-2-butanol and (±)-2-butanol as above.[13] The Nod factors were permethylated by the method of Ciucanu and Kerek[14] modified as described by Cardenas *et al.*[15] in order to remove the base-sensitive substituents, and the samples were hydrolysed with 2 M CF_3COOH (120 °C, 1 h), reduced and acetylated by the method of Blakeney.[16] The permethylated alditol acetates were analysed by GLC-MS.

GLC-MS was performed with a Micromass Auto Spec-Q instrument fitted with a Fisons GC 8065MS gas chromatograph with an OV-1 column (25 m × 0.25 mm). The temperature programme for separating the trimethylsilylated methyl glycosides was isothermal at 150 °C for 2 min, followed by a 10 °C/min gradient up to 250 °C, whereas the programme for the partially methylated alditol acetates was isothermal at 120 °C for 1 min followed by an 8 °C/min gradient up to 250 °C. The protocol for the trimethylsilylated 2-butyl glycosides was isothermal at 130 °C followed by a 2 °C/min gradient up to 250 °C. The ionisation potential was 70 eV.

2.4. Fatty acid analysis

The fatty acids were identified as their methyl esters. These were prepared by methanolysis in methanolic 1 M HCl at 85° for 4 h and were identified by GLC-MS. The location of the double bond in the unsaturated fatty acyl residue was determined by the preparation of the dimethyl disulfide ethers[17] of its methyl ester followed by GLC-MS. The chromatographic conditions are those described for the trimethylsilylated methyl glycosides.

2.5. Mass spectrometry

Positive ion mode LSIMS was performed on a Micromass AutoSpecQ instrument with a cesium ion gun at an acceleration voltage of 8 kV. Either glycerol–*m*-nitrobenzyl alcohol (1:1), or thioglycerol containing NaI as the cationizing agent, were used as matrices. The samples were dissolved in 0.2 ml Me_2SO. Positive ion mode ES-Q-o-TOF MS was performed using a Micromass Q-TOF hybrid tandem mass spectrometer equipped with a Z-spray sample introduction system

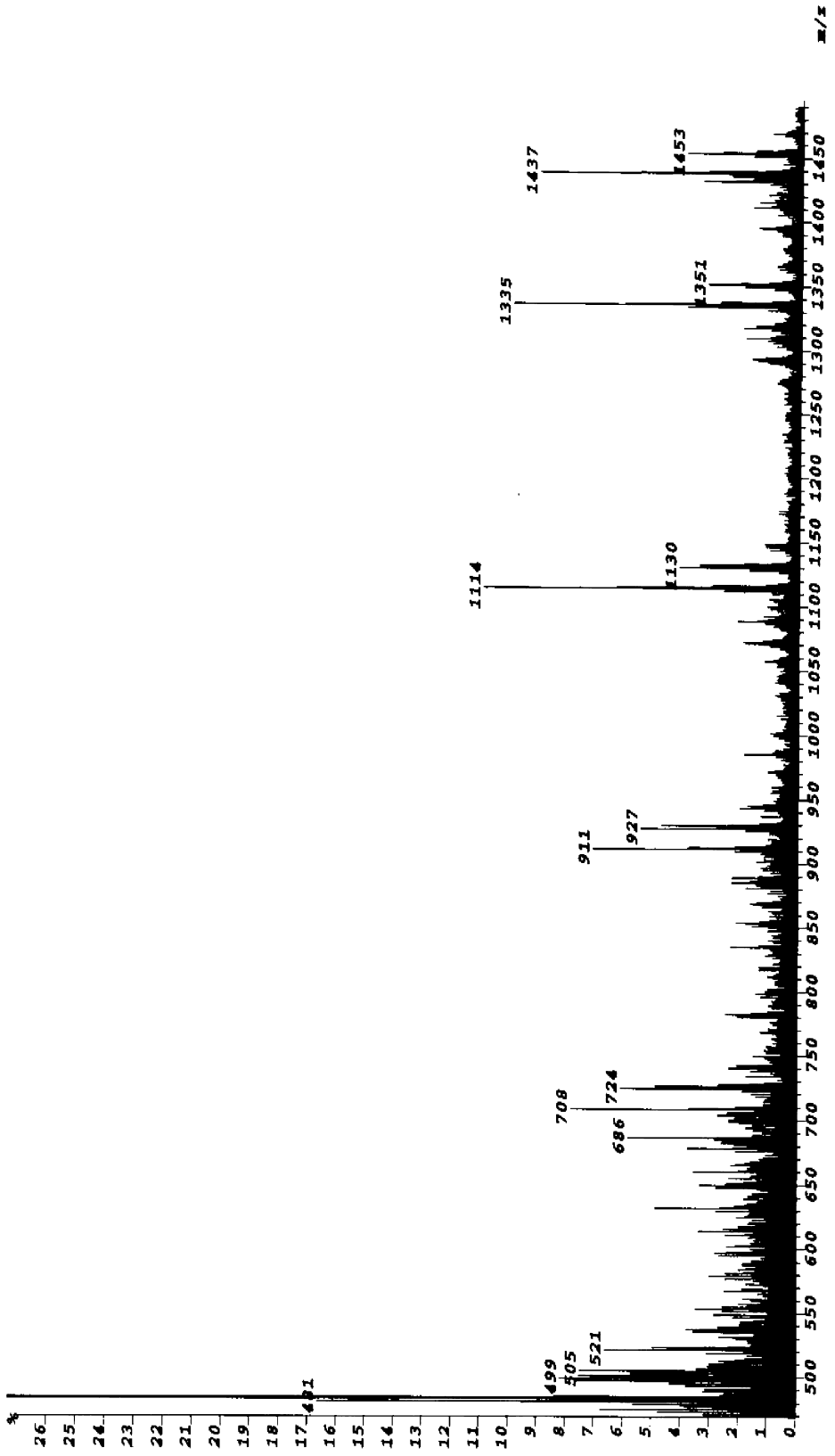


Fig. 3. Positive liquid secondary-ion mass spectrum of the Nod factor in the HPLC fraction F3. Cationised pseudomolecular ions and cationised B fragment series are present.

and a continuous infusion sample delivery system. The syringe was loaded with the HPLC fraction and the sample solution infused into the source through a fused silica capillary at a rate of 1 μL min^{-1}. A potential of 3500 V was applied to the capillary tip. The source temperature was 80 °C. The quadrupole was used in the Rf-only mode. TOF mass spectra were integrated every 2.4 s over the m/z 100–2000 range and recorded using a cone voltage varying between 40 and 130 V. The data were recorded and processed using the MassLynx software, version 3.1. CID-MS/MS was performed using the quadrupole to select the precursor ion for fragmentation in the hexapole collision cell. Argon was used as collision gas and the collision energy offset setting varied between 10–25 V. Product ions were analysed using the orthogonal TOF analyser. The samples were dissolved in acetonitrile–water–formic acid (50:50:0.1, v/v/v).

2.6. NMR analysis

The samples were deuterium exchanged several times by freeze drying from D_2O and then examined as solutions (3 mg/mL) in d_6-Me_2SO containing the highest amount of D_2O possible before causing precipitation of the samples (approximately 8% in D_2O). Spectra were recorded at 303 or 343 K on a Bruker AMX500 spectrometer operating at 500.13 MHz (^1H) and 125.75 MHz (^{13}C). Chemical shifts are given in ppm, using the Me_2SO signals (2.49 ppm, ^1H; and 39.5 ppm, ^{13}C) as references. The two-dimensional homonuclear proton double-band filtered selective correlation experiment (DBF-COSY)[18] was registered by application of two BURP pulses based on the DANTE-Z scheme[19] ($n = 10$, $\tau = 50$ μs, $\theta = 10°$). A data matrix of 512×1024 points was used to resolve a spectral width of 1500 Hz; 64 scans were used per increment. The two-dimensional heteronuclear one-bond proton carbon correlation experiment[20] was registered in the ^1H-detection mode via single-quantum coherence (HSQC). A data matrix of 256×1024 K points was used to digitise a spectral width of 3000 and 15 000 Hz in F_2 and F_1. 64 scans were used per increment with a delay between scans of 1 s and a delay of 3.4 ms corresponding to a J value of 150 Hz. In this experiment, ^{13}C decoupling was achieved by the globally optimised alternating-phase rectangular pulses (GARP) scheme. Squared cosine-bell functions were applied in both dimensions and zero-filling was used to expand the data to 512×1024 points. This experiment was slightly modified by the implementation of an editing block in the sequence.[21] The heteronuclear multiple-bond correlation (HMBC) experiment was performed using the Bruker standard sequence with 256 increments of 2048 real points to

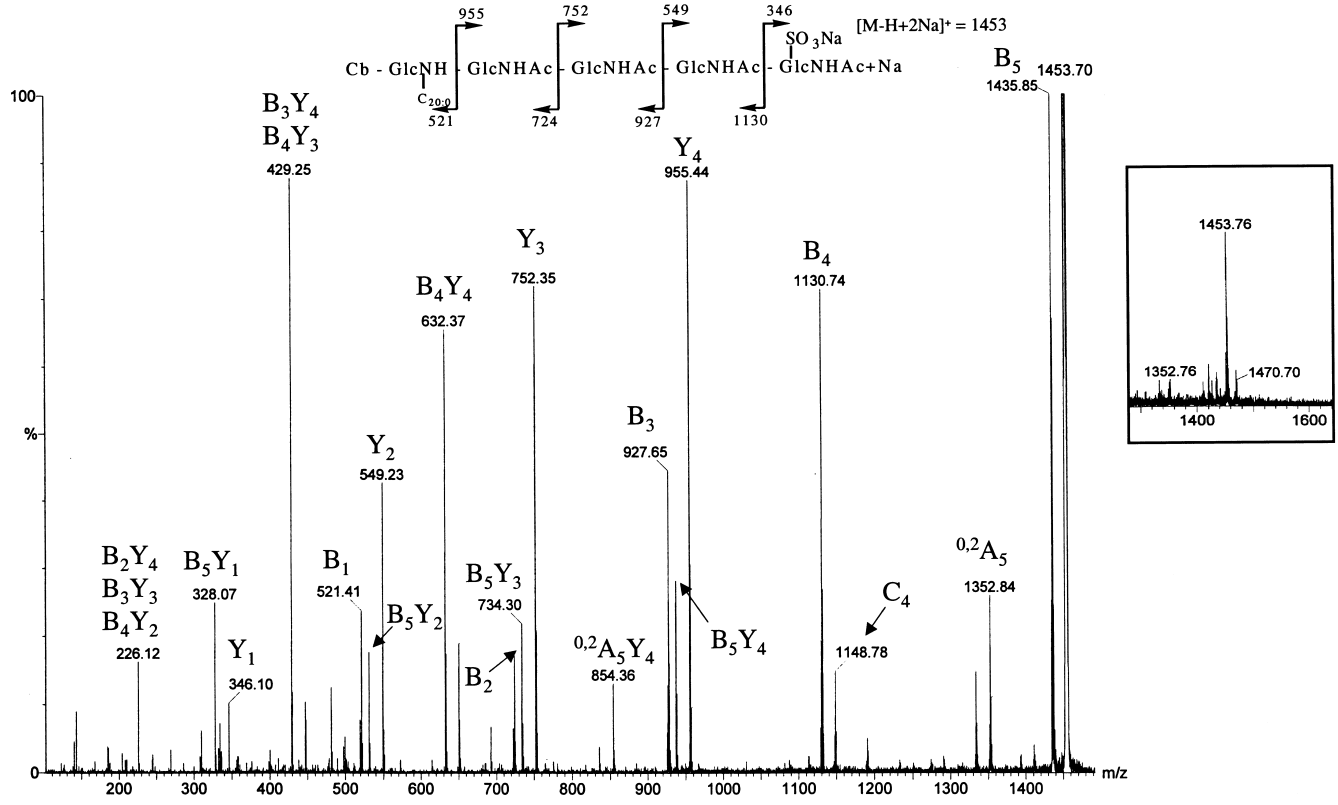


Fig. 4. CID ES-Q-o-TOF mass spectrum and fragmentation scheme of the Nod factor in HPLC fraction F1. The spectrum was obtained on collision of m/z 1453. Inset: Pseudomolecular ion region in the positive ion mode ES-Q-o-TOF mass spectrum.

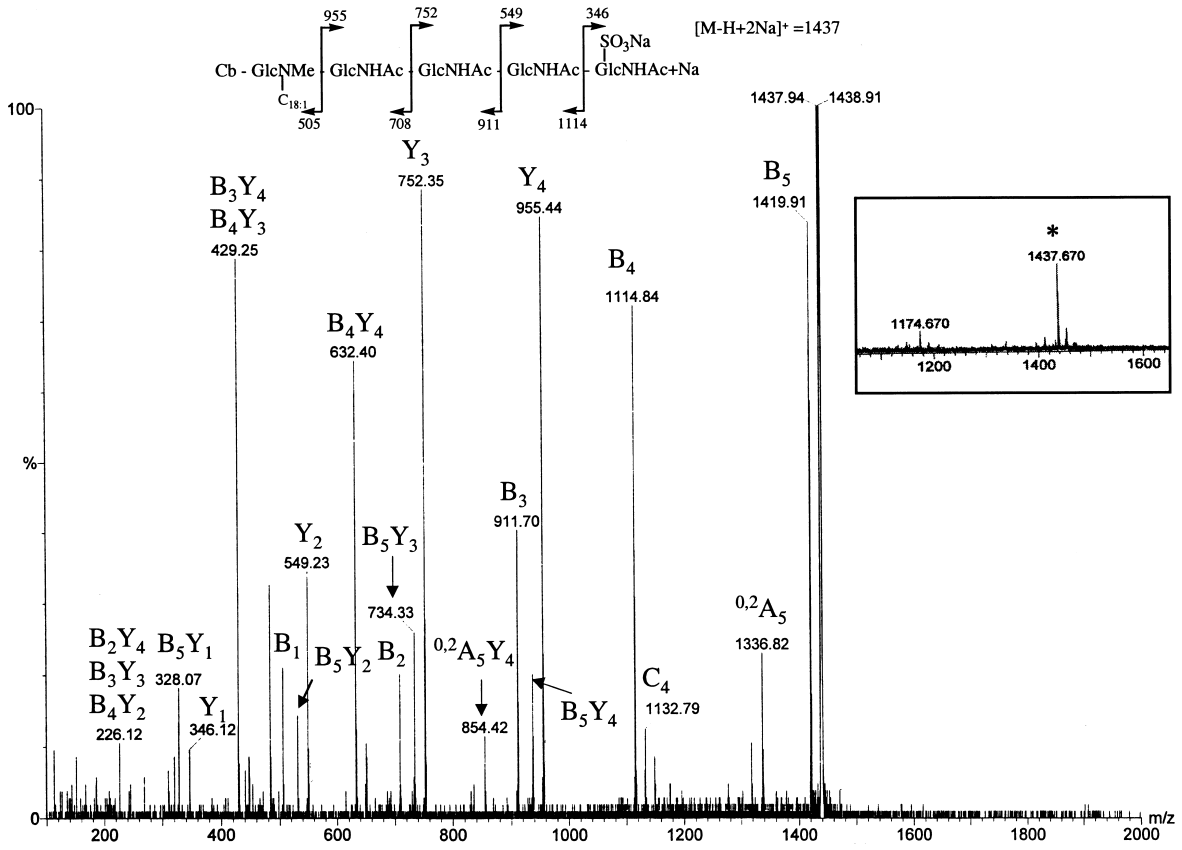


Fig. 5. CID ES-Q-o-TOF mass spectrum and fragmentation scheme of the major Nod factor in HPLC fraction F3. The spectrum was obtained on collision of m/z 1437. Inset: Pseudomolecular ion region in the positive ion mode ES-Q-o-TOF mass spectrum.

digitise a spectral width of $3000 \times 28\,000$ Hz. 192 scans were acquired per increment with a delay of 50 ms for the evolution of long-range coupling. Selective excitation one-dimensional experiments were performed by application of the DANTE-Z pulse train ($n = 300$, $\tau = 100$ μs, $\theta = 0.3°$).[22] This train was concatenated to a TOCSY sequence (at different isotropic mixing times, $\pi/2$ pulse width 49 μs)[23] to yield the 1D-TOCSY subspectra. The number of accumulated scans was 512 in each case. 1D-NOESY subspectra were obtained by allowing the free evolution of magnetisation over a mixing time of 200 ms after the selection.

3. Results

3.1. Production and purification of the Nod factors

Rhizobium giardinii bv. giardinii H152 was grown either in the presence or absence of apigenin as inducer, with the addition of [^{14}C]-glucosamine hydrochloride. TLC analysis (Fig. 1) showed two major Nod factors and four additional spots. For preparative purposes, 9 litres of medium from the apigenin-induced culture were extracted with *n*-butanol. The *n*-butanol extract was pre-purified using a reversed-phase C_{18} cartridge and eluting with different acetonitrile–water compositions (20, 45 and 60%, v/v). The 45% fraction was subjected to HPLC fractionation using a semipreparative C_{18} column. The HPLC profile (Fig. 1) is shown. Six fractions (F1-F6) were obtained. Fractions F1–F4 were eluted in 30% acetonitrile, while fractions F5 and F6 were eluted in 40% acetonitrile.

3.2. Mass spectrometric analysis: LSIMS analysis of fractions F1–F6

The results of the LSIMS analysis in the positive mode of fractions F1–F6 are summarised in Fig. 2. The mass spectrum obtained from fraction F1 (data not shown) exhibited a complex pattern in the molecular ion region. Signals at m/z 1453 and 1469 correspond to the sodium and potassium adducts of the molecular species. The signal at m/z 1410 was assigned to the loss of a carbamoyl group and the signal at m/z 1351 was assigned to the loss of a $NaSO_3$ group. The signal at m/z 1333 corresponds to the loss of $NaSO_3$ from the B_5 oxonium ion. In addition, the mass spectrum contained signals at m/z 1130 and 1146 (assigned as the sodium and potassium cationised B_4 ion); m/z 927 and 943 (B_3 plus sodium and potassium); m/z 724 and 740 (B_2 plus sodium and potassium); and m/z 499, 521 and 537

which correspond to the B_1 ion and its cationised species. The odd-number value of the m/z of B_1 indicates the presence of a nitrogen-containing substituent on the non-reducing glucosamine residue, which is consistent with a carbamoyl substitution. Furthermore, the B_1 mass value indicates the presence of a $C_{20:0}$ *N*-acyl substituent. We propose that the ion at m/z 1453 corresponds to the $[M-H+2Na]^+$ pseudomolecular ion of a Nod factor having the following structure: V ($C_{20:0}$, Cb, S). This is the representation of a nodulation factor with a backbone of five GlcNAc residues, a 20-carbon fatty acyl chain with no double bonds, a carbamoyl group and a sulfate group as backbone substituents.

The mass spectrum from fraction F2 (data not shown) contained signals corresponding to $[M-H+2Na]^+$ and $[M-H+Na+K]^+$ pseudomolecular ions at m/z 1411 and 1427, respectively, the loss of carbamoyl group at m/z 1368 and the loss of $NaSO_3$ at m/z 1309. In addition, the mass spectrum contained signals corresponding to the salts adducts (Na and K) of the B_4 to B_1 ions (1088, 1104; 885, 901; 682, 698; and 479, 495) and to the B_1 fragment at m/z 457. This B_1 ion is consistent with Nod factors bearing a carbamoyl, an *N*-methyl, and a $C_{16:0}$ fatty acyl substituents on its non-reducing residue. The Nod factor in fraction F2 was thus identified as V ($C_{16:0}$, NMe, Cb, S).

The mass spectrum from fraction F3 (Fig. 3) contains signals corresponding to $[M-H+2Na]^+$ and $[M-H+Na+K]^+$ pseudomolecular ions at m/z 1437 and 1453, respectively, and the loss of $NaSO_3$ at m/z 1335. The fragment series ended at m/z 483. The mass value of B_1 is consistent with the presence of a carbamoyl, an *N*-methyl and a $C_{18:1}$ fatty acyl substituents. The Nod factor in fraction F3 was identified as V ($C_{18:1}$, NMe, Cb, S).

The mass spectrum from F4 (data not shown) contained a pseudomolecular ion $[M-H+2Na]^+$ at m/z 1439 and the loss of carbamoyl and $NaSO_3$ groups. The fragments series corresponding to the oxonium type ions ended at m/z 485. The Nod factor was identified as V ($C_{18:0}$, NMe, Cb, S). In the nod factors present in fractions F1–F4, M correspond to the molecular ion considering the sulfate group as SO_3H.

The mass spectrum from F5 (data not shown) revealed the presence of pseudomolecular ions $[M+Na]^+$ and $[M+K]^+$ at m/z 1309 and 1325, respectively. The presence of a carbamoyl group was detected by the loss of 43 mass units. The adducts (Na and K) of the oxonium type ions were also present and ended at m/z 457 (non-cationised B_1 ion). The mass value of B_1 is consistent with the presence of a carbamoyl, an *N*-methyl and a $C_{16:0}$ fatty acyl substituents. The Nod factor was identified as V ($C_{16:0}$, NMe, Cb).

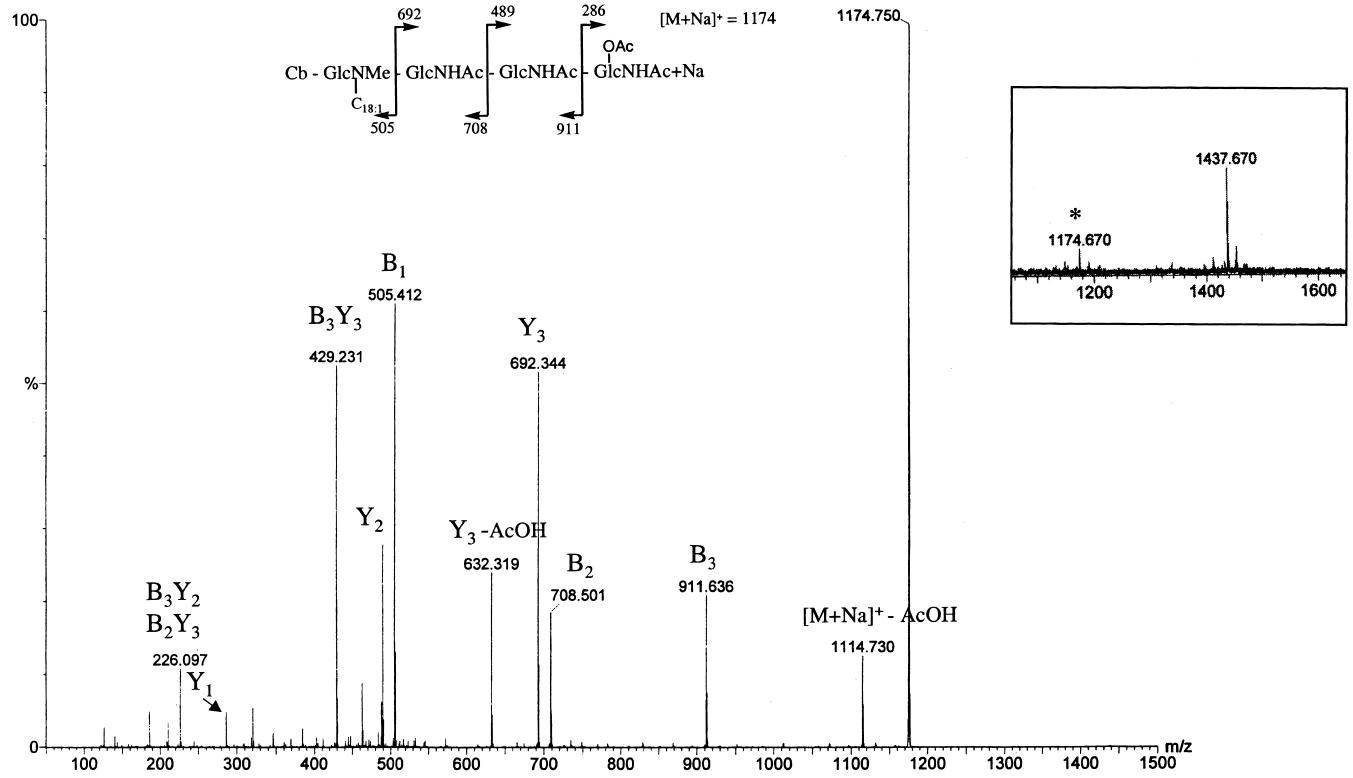


Fig. 6. CID ES-Q-o-TOF mass spectrum and fragmentation scheme of the minor Nod factor in HPLC fraction F3. The spectrum was obtained on collision of m/z 1174. Inset: Pseudomolecular ion region in the positive ion mode ES-Q-o-TOF mass spectrum.

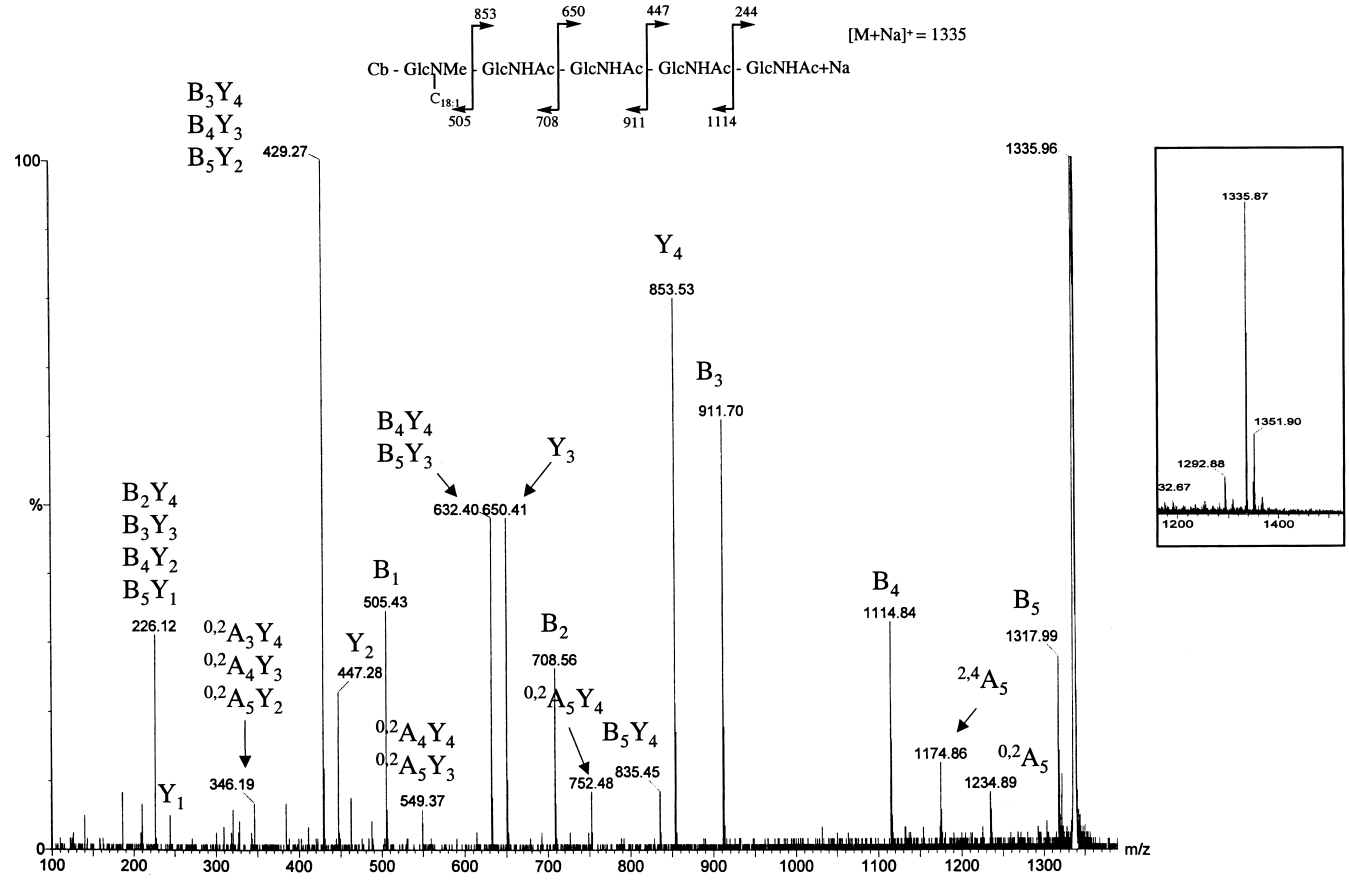


Fig. 7. CID ES-Q-o-TOF mass spectrum and fragmentation scheme of the Nod factor in HPLC fraction F6. The spectrum was obtained on collision of m/z 1335. Inset: Pseudomolecular ion region in the positive ion mode ES-Q-o-TOF mass spectrum.

Finally, the mass spectrum of fraction F6 (data not shown) contained pseudomolecular ions $[M + Na]^+$ and $[M + K]^+$ at m/z 1335 and 1351, respectively. The loss of 43 mass units indicated the presence of a carbamoyl group. The B-type fragments series is also present and ended at the non-cationised B_1 fragment at m/z 483. The Nod factor was identified as V ($C_{18:1}$, NMe, Cb).

3.3. ES-Q-o-TOF mass spectrometry

To exploit its greater sensitivity in order to identify any minor components, and in order to link fragment ions to precursors and thus confirm assignments made on the basis of LSIMS data, we performed ES-Q-o-TOF and CID-MSMS analysis on fractions and signals representing the different types of Nod factors produced by *R. giardinii* bv. giardinii H152. Fig. 4 shows the pseudomolecular ion region of the positive ES-Q-o-TOF mass spectrum of fraction F1 (inset) and the CID tandem mass spectrum obtained from the Nod factor component V ($C_{20:0}$, Cb, S) on collision of pseudomolecular ion $[M - H + 2Na]^+$ at m/z 1453. A range of different fragment ions corresponding to glycosidic bond cleavage that results in the formation of B-, Y- and C-type ions, double cleavage ions (BY-type ions) and cross-ring cleavage and double-cleavage ions (A- and AY-type ions) is present, allowing the structure to be identified (see fragmentation scheme). The series of B_n and Y_n ions observed at m/z 521, 724, 927, 1130, and 1435; and m/z 346, 549, 752, and 955, respectively, shows that a sulfate group is located on the reducing glucosamine residue and a carbamoyl and a $C_{20:0}$ fatty acyl residue are located on the non-reducing glucosamine residue. These results confirm the structure propose for the Nod factors in fraction F1.

The Nod factor from HPLC fraction F3 represents a second Nod factor structural type produced by *R. giardinii*, bearing an *N*-methyl group in the non-reducing terminal residue. Fig. 5 shows the pseudomolecular ions region of the ES-Q-o-TOF mass spectrum in the positive mode (inset). The spectrum contains signals corresponding to a major Nod factor having a pseudomolecular ion $[M - H + 2Na]^+$ at m/z 1437, and a minor Nod factor with a psudomolecular ion at m/z 1174 that was not detected in the LSIMS analysis. The CID tandem mass spectrum on collision of the ion at m/z 1437 contains product ions corresponding to the

different types of fragments ions. The series of B_n and Y_n ions observed at m/z 505, 708, 911, 1114 and 1419 and m/z 346, 549, 752, and 955, respectively, shows that a sulfate group is located on the reducing glucosamine residue, and that a carbamoyl, an *N*-methyl groups and a $C_{18:1}$ fatty acyl residue are located on the non-reducing terminal glucosamine residue. This substitution pattern was confirmed by ^1H and ^{13}C NMR analysis (see the discussion of the NMR results) and by the results of the methylation analysis.

The CID tandem mass spectrum obtained from the minor Nod factor component (Fig. 6) on collision of m/z 1174 contains the B_n and Y_n ion series at m/z 505, 708, and 911 and m/z 286, 489, and 692, in addition to a double cleavage BY-series. Two fragments arising from the loss of AcOH from the pseudomolecular ion and from the Y_3 ion at m/z 1114 and 632, respectively, were also detected. These results are consistent with the presence of a chitotetrameric Nod factor containing an acetyl group on the reducing glucosamine residue and a carbamoyl, an *N*-methyl, and a $C_{18:1}$ fatty acyl substituents on the non-reducing glucosamine residue.

The fourth type of Nod factor, unsubstituted on the reducing terminal residue, is represented by the species in fraction F6 (Fig. 7). The CID tandem mass spectrum obtained on collision of m/z at 1335 contains B- and Y-type ions, cross-ring cleavage ions (A-type), double cleavage ions (BY), and cross-ring double cleavage ions (AY), allowing the structure to be identified as V ($C_{18:1}$, NMe, Cb).

3.4. Location of carbamoyl substituents

To locate the carbamoyl group substituted on the non-reducing glucosamine residue in the major Nod factor V ($C_{18:1}$, NMe, Cb, S) from HPLC fraction F3, we recorded a CID mass spectrum following selection and collision of the B_1 ion at m/z 483[10] (Fig. 8). The spectrum did not show signals at m/z 126 (ion a), 144 (ion b), and 138 (ion c) that are the diagnostic ions proposed for differentiation of positional isomers of acetylated or carbamoylated non-reducing glucosamine residues in Nod factors. However, we detected ions at m/z 158 and m/z 140 which are the equivalent of ion b and ion a but shifted upwards by 14 Th because of the presence of a methyl substituent on the amidic nitrogen. The low ratio we observe for abundance of 158/abundance of 140 (0.08) is characteristic of 4-*O*- and 6-*O*-substitution. In addition, the low ratio of abundance of 152 (methylated ion c)/abundance of 140 (0.0) that we observe is characteristic of 6-*O*-substitution for the carbamoyl substituents on the non-reducing terminal residue.[10] We thus assign the non-reducing carbamoyl substituent as being 6-O-substituted in Nod factor V ($C_{18:1}$, NMe, Cb, S) in fraction F3.

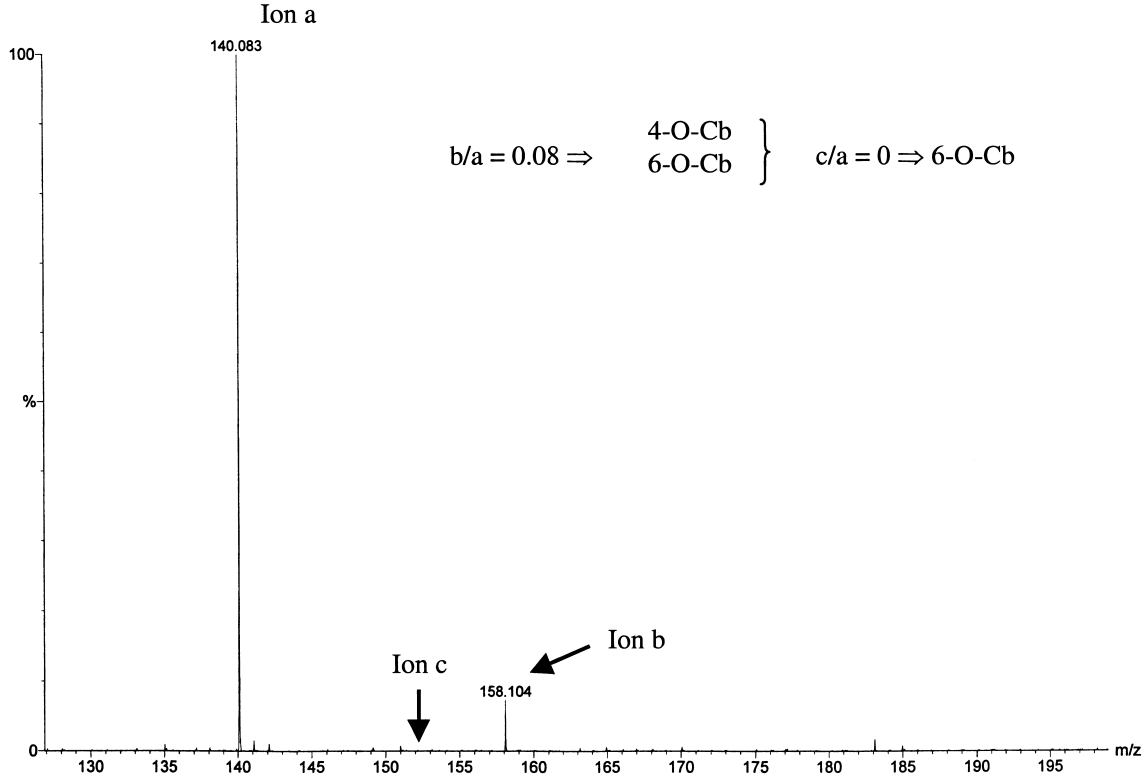


Fig. 8. CID ES-Q-o-TOF mass spectrum of the major Nod factor in HPLC fraction F3. The spectrum was obtained on collision of B_1 fragment at m/z 483. The ratio of abundance between the different diagnostic fragments indicated a 6-*O*-substitution for the carbamoyl substituent.

Table 1
^{1}H and ^{13}C NMR chemical shifts (δ, ppm) and coupling constants (*J*, Hz) for the major Nod factor in fraction F3

Residue	Position							
	1	2	3	4	5	6a	6b	
6-*O*-Cb-β-D-GlcNMeAcyl-(1→								
δ_H	4.74	3.36	3.52	3.23	3.50	4.17	3.95	
J	(1,2) 8.2	(2,3) 9.6	(3,4) 8.2	(4,5) 8.2	(5,6) n.d.	(6a,6b) −11.1		
δ_C	98.7	62.3	70.1	70.1	73.7	62.9		
→4)-β-D-GlcNAc-(1→								
δ_H	4.31–4.45	3.44–3.60	3.45–3.58	3.20–3.40	3.20–3.27	3.61	3.36	
δ_C	101.3–101.5	53.9–54.7	68.4–72.1	80.3–81.0	74.3–74.8	59.9		
→4)-6-*O*-$NaSO_3$-α-D-GlcNAc								
δ_H	4.86	3.59	3.29	3.29	3.77	3.97	3.79	
J	(1,2) 2.1	n.d.	n.d.	n.d.	n.d.	n.d.		
δ_C	90.2	53.5	73.0	81.5	67.5	64.4		
→4)-6-*O*-$NaSO_3$-β-D-GlcNAc								
δ_H	4.43							
δ_C	95.0							
Side chains	1	2	3	4–9	10, 13	11, 12	14–17	18
Fatty acyl ($C_{18:1}$)								
δ_H		2.24	1.43	1.21	1.95	5.30	1.21	0.82
δ_C	173.5	32.5	24.6	22.0–31.0	26.5	130.0	22.0–31.0	13.9
Methyl								
δ_H	2.71							
δ_C	27.3							
Acetyl								
δ_H		1.80–1.85						
δ_C	169.4–169.6	22.0–22.8						
Carbamoyl								
δ_C	156.6							

3.5. Composition and methylation analysis

GLC-MS of the trimethylsilyl ethers of methyl glycosides showed that only glucosamine was present. Glucosamine was assigned as D-GlcN following GLC-MS analysis of its trimethyl silylated (+)-2-butyl and (±)-2-butyl glycosides. Methylation analysis of the Nod factor in fraction F3 showed the presence of 1,5-di-*O*-acetyl-3,4,6-tri-*O*-methyl-*N*-acetyl-*N*-methyl glucosaminitol from the non-reducing terminal glucosamine residue unsubstituted or substituted by a base sensitive groups, 1,4,5-tri-*O*-acetyl-3,6-di-*O*-methyl-*N*-acetyl-*N*-methyl glucosaminitol from the internal 4-substituted glucosamine residues, and 1,4,5,6-tetra-*O*-acetyl-3-*O*-methyl-*N*-acetyl-*N*-methyl glucosaminitol from the reducing terminal glucosamine residue substituted on O-6. The carbamoyl substituent on the non-reducing residue is removed by the strongly basic medium used in the methylation step, and the OH esterified initially by the carbamoyl group is therefore methylated. On the contrary, the sulfate substituent on the reducing glucosamine residue resists the methylation step, but it is removed during the hydrolysis step. The initially sulfated OH of this glucosamine is thus acetylated at the end of the methylation analysis. The fatty acid present in fraction F3 was identified as octadecenoic acid by GLC-MS of its methyl ester. The mass spectrum of the dimethyl disulfide derivative of the unsaturated fatty acid methyl ester, contains signals at *m*/*z* 390 (molecular ion), *m*/*z* 145 and *m*/*z* 245, arising from the fragmentation between the carbons that carry the dimethyl disulfide groups, which demonstrates that the double bond is located at carbon number 11.

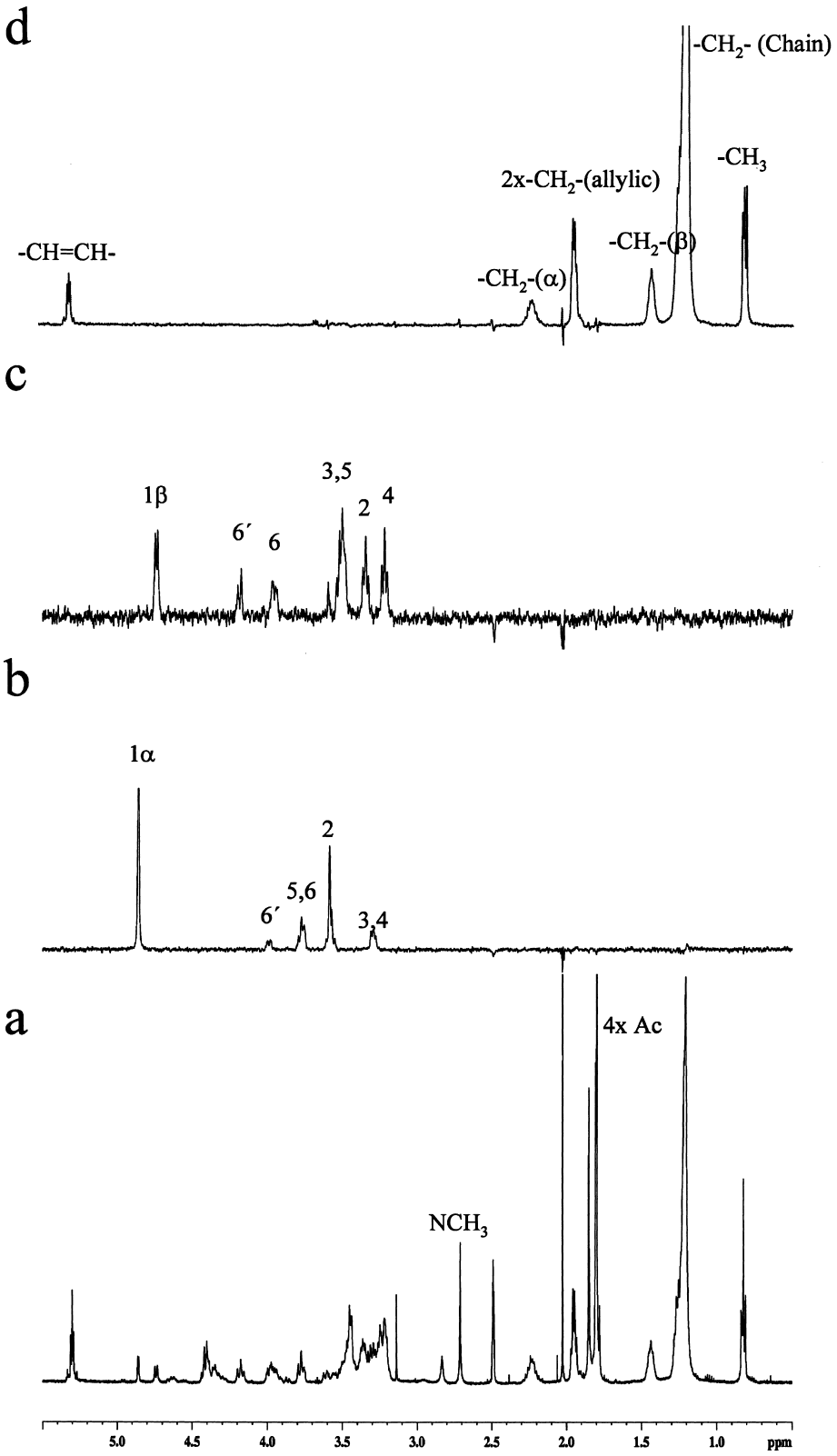


Fig. 9. 500 MHz ^{1}H NMR spectrum (a) and 1D-TOCSY subspectra obtained by selective excitation of the signals at δ_H 4.86 ppm (b); δ_H 4.74 ppm (c); and δ_H 1.21 ppm (d); obtained for the major Nod factor in fraction F3. Subspectrum (b) corresponds to the α-anomer of the reducing *N*-acetylglucosamine residue. Subspectrum (c) corresponds to the non-reducing *N*-acylglucosamine residue. Subspectrum (d) corresponds to the $C_{18:1}$ fatty acyl residue.

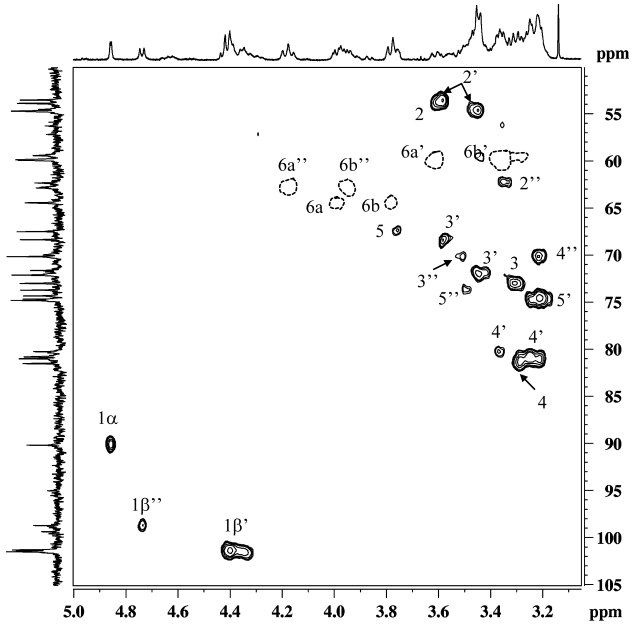


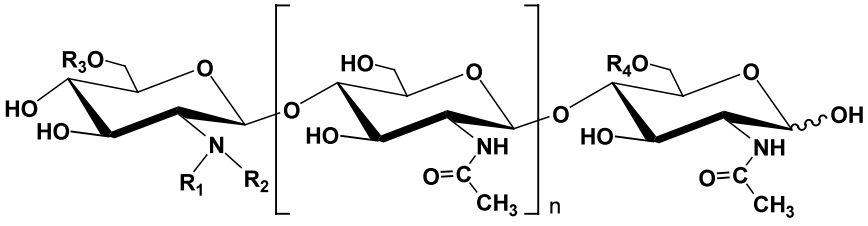
Fig. 10. Multiplicity-edited 2D-heteronuclear one bond proton–carbon correlation experiment registered in the HSQC mode for the major Nod factor in fraction F3 (only the sugar protons region is shown). Positive peaks (solid lines) correspond to methine and methyl groups, whereas negative peaks (broken lines) correspond to methylenic groups. Numbers with (') correspond to the internal glucosamine residues and numbers with ('') correspond to the non-reducing glucosamine residue. Numbers without labels correspond to the α-anomer of the reducing glucosamine residue.

3.6. NMR analysis

The chemical shifts for the ^{1}H and ^{13}C NMR signals of the major HPLC fraction F3 [mainly V ($C_{18:1}$, NMe, Cb, S)] were assigned from DBF-COSY, 1D-TOCSY, 1D-NOESY and edited HSQC experiments (Table 1). The ^{1}H NMR shows (Fig. 9a) a signal at δ_H 5.32 ppm ($^{3}J = 4.6$ Hz) correlated in the HSQC spectrum (data not shown) with the signal at δ_C 130.0 ppm, that was assigned to the olefinic protons of the unsaturated acyl residue. The signal coupling constant value allowed the identification of the *cis*-isomer. The doublet at δ_H 4.86 ppm ($^{3}J = 2.1$ Hz) correlated in the HSQC spectrum (Fig. 10) with the signal at δ_C 90.2 ppm was assigned to the anomeric proton of the α-anomer of the reducing glucosamine residue of the oligosaccharide. The signals at δ_H 4.74 ppm ($^{3}J = 8.2$ Hz) and δ_C 98.7 ppm were assigned to the anomeric proton and carbon, respectively, of the β-linked, non-reducing glucosamine residue, shifted downfield as a consequence of the *N*-methyl substitution. This *N*-methyl substitution was demonstrated by the intra-residue NOE observed between the *N*-methyl group (δ_H 2.71 and δ_C 27.3 ppm), and the anomeric proton at δ_H 4.74 ppm in a 1D-NOESY experiment (data not shown). The resonances at δ_H 4.31–4.45 and δ_C 101.3–101.5 ppm are due to the anomeric protons and carbons of the β-linked internal glucosamine residues. In the HSQC spectrum (data not shown) we observed a cross peak at δ_H 4.45 and δ_C 95.0 ppm which was assigned to the β-anomer of the reducing glucosamine residue. Resonances between δ_H 4.17–3.22 ppm were assigned to the sugar protons. The signals between δ_H 2.24–0.82 ppm were assigned to the acetyl and fatty acyl substituents. In addition, the ^{13}C NMR spectrum (data not shown) shows signals at δ_C 173.5, 169.4–169.6, and 156.6 ppm that were assigned to the carbonyl carbons of the fatty acyl, acetyl, and carbamoyl substituents, respectively.

In order to completely assign the signals corresponding to the reducing and non-reducing glucosamine

Table 2
Nodulation factors induced by *R. giardinii* bv. giardinii H152



Fractions	$[M+Na]^+$ (m/z)	$[M-H+2Na]^+$ (m/z)	n	R_1	R_2	R_3	R_4	Designation
F_1		1453	3	$C_{20:0}$	H	$CONH_2$	SO_3^-	Rg V($C_{20:0}$,Cb,S)
F_2		1411	3	$C_{16:0}$	Me	$CONH_2$	SO_3^-	Rg V($C_{16:0}$,NMe,Cb,S)
F_3		1437	3	$C_{18:1}$	Me	$CONH_2$	SO_3^-	Rg V($C_{18:1}$,NMe,Cb,S)
	1174		2	$C_{18:1}$	Me	$CONH_2$	Ac	RgIV($C_{18:1}$,NMe,Cb,Ac)
F_4		1439	3	$C_{18:0}$	Me	$CONH_2$	SO_3^-	Rg V($C_{18:0}$,NMe.Cb,S)
F_5	1309		3	$C_{16:0}$	Me	$CONH_2$	H	Rg V($C_{16:0}$,NMe,Cb)
F_6	1335		3	$C_{18:1}$	Me	$CONH_2$	H	Rg V($C_{18:1}$,NMe,Cb)

residues and the fatty acyl moiety, selective excitation 1D experiments were carried out. The 1D-TOCSY subspectra from selective inversion of the anomeric proton of the α-anomer of the reducing glucosamine residue at δ_H 4.86 ppm, via a DANTE-Z pulse train, were obtained at different isotropic mixing times of 23, 46, 92 (data not shown), and 276 ms (Fig. 9b), and allowed the observation of all protons belonging to the reducing moiety, which were easily assigned. The chemical shifts for the ^{13}C resonances were assigned from the edited HSQC experiment. The signal corresponding to C-4 is shifted downfield (δ_C 81.5 ppm), indicating that position 4 is substituted. Carbon C-6 was easily identified because of its methylenic nature, showed by the negative sign (broken lines) of their one-bond proton–carbon correlation cross peak in the edited HSQC spectrum. H-6 and C-6 are shifted downfield from their expected values, indicating that position 6 in this unit is substituted. These results suggest that the sulfate group identified in the mass spectrometric analysis, is linked to the reducing glucosamine residue at O-6. This conclusion is in agreement with the results of the methylation analysis.

The same procedure was followed for the non-reducing glucosamine residue, by selecting the anomeric proton at δ_H 4.74 ppm (Fig. 9c). The δ_C of C-4 at 70.1 ppm indicated that this position is not substituted. The values for H-6 and C-6 at δ_H 3.95, 4.17 and δ_C 62.9 ppm, are shifted downfield, indicating that position 6 is substituted in this case by a carbamoyl group. This supposition was confirmed by an HMBC experiment (data not shown). The cross peak between the carbonyl carbon from the carbamoyl group and H-6, showed clearly that this substituent is located at C-6 of the non-reducing glucosamine residue. This conclusion is consistent with the results of the CID-MSMS experiment performed on the B_1 fragment.

Finally, by selecting the signal at δ_H 1.21 ppm, the spin system corresponding to the fatty acyl moiety was obtained and totally assigned from this subspectrum (Fig. 9d) and from the HSQC experiment.

4. Discussion

The structures of seven nodulation factors isolated from apigenin-induced cultures of *Rhizobium giardinii* bv. giardinii H152 have been determined. These Nod factors were present in six C_{18}-HPLC fractions and were studied by monosaccharide composition and methylation analysis, LSIMS, ES-Q-o-TOF, CID-MSMS and 1D- and 2D-NMR experiments. From the results of this study, we concluded that *R. giardinii* bv. giardinii H152 produces four different types of Nod factors, all of them carbamoylated.

One type is a pentamer carrying a sulfate group on the reducing residue and an *N*-fatty acyl substituent (icosanoic acid) on the non-reducing glucosamine residue. Another type is a pentamer substituted by a sulfate group on the reducing residue, and *N*-methylated on the non-reducing terminal residue. The non-reducing residue is *N*-acylated with different fatty acid residues: octadecanoic acid, *cis*-octadec-11-enoic acid and hexanoic acid. The third type of Nod factors has a tetrameric backbone of chitin carrying an acetyl group on the reducing glucosamine residue and an *N*-methyl and octadec-11-enoyl residues on the non-reducing glucosamine residue. Finally, the fourth type comprises two non-sulfated pentamers carrying an *N*-methyl group and two different fatty acids: vaccenic acid and palmitic acid. These results are summarised in Table 2.

In Section 1, we mentioned that *Phaseolus vulgaris* is known to be promiscuous and it can be nodulated by a large number of *Rhizobium*, some of which are strain specific, and induce the formation of nitrogen-fixing nodules. Among these, we find *Rhizobium leguminosarum* bv. phaseoli, *Rhizobium etli*, *Rhizobium tropici*, *Rhizobium gallicum*, and *Rhizobium giardinii*, the latter producing non-fixing nodules. On the other hand *Phaseolus vulgaris* can be inefficiently nodulated by other Rhizobia such as *Mesorhizobium loti*, *Sinorhizobium saheli*, *Sinorhizobium teranga* and *Sinorhizobium fredii* NGR234.[3]

The Nod factors produced by all these strains show common features, such as the presence of a methyl group on the nitrogen of the non-reducing GlcN residue, the presence in all cases of cis-vaccenic acid, sometimes accompanied by other fatty acids, and finally, in most cases, a pentameric GlcNAc backbone is found. Waelkens *et al*. indicate that the methyl group is necessary for the nodulation of *Phaseolus vulgaris* by *Rhizobium tropici* CIAT899.[24] The other substituents vary according to the different strains, with some having been identified as AcFuc, Fuc, MeFuc and sulfate on the *O*-6 of the reducing GlcN residue and Cb on *O*-3, *O*-4 or *O*-6 of the non-reducing GlcN residue. The Nod factors produced by *R. giardinii* bv. giardinii H152 show the common features of the other Nod factors produced by other strains which are able to nodulate *Phaseolus vulgaris*.

In addition, the structures of the Nod factors isolated form *R. giardinii* bv. giardinii H152 were found to be similar to the Nod factors produced by *Sinorhizobium teranga* bv. acaciae ORS1073 and *Rhizobium* sp. strain ORS1002, both symbionts of particular *Acacia* species.

Acknowledgements

We thank the Comisión Interministerial de Ciencia y Tecnología (grant no. BIO99-0614-C03-02) and INCO-

DC ERBIC18-CT98-0321 for financial support. J.T.O. gratefully acknowledges financial support (to S. Gaskell and R. Beynon) from the Higher Educational Funding Council for England, and additional support from SmithKline Beecham, UMIST, University of Wales College of Medicine, and Chugai Pharmaceuticals, for the purchase of the Q-Tof.

References

1. Bladergroen, M. R.; Spaink, H. P. *Curr. Opin. Plant Biol.* **1998**, *1*, 353–359.
2. Dénarié, J.; Debellé, F.; Promé, J.-C. *Annu. Rev. Biochem.* **1996**, *65*, 503–535.
3. Downie, J. A. In *The Rhizobiaceae*; Spaink, H. P.; Kondorosi, A.; Hookyas, P. J. J., Eds.; Kluwer Academic Publishers: Dordrecht, 1998; pp 387–402.
4. Pacios Bras, C.; Spaink, H. P.; Stuurman, N. In *Prokaryotic Nitrogen Fixation: A Model System for Analysis of Biological Process*; Triplett, E. W., Ed.; Horizon Scientific Press: Wymondham, 2000; pp 365–383.
5. Michiels, J.; Dombrecht, B.; Vermeiren, N.; Xi, C.-W.; Luyten, E.; Vaderleyden, J. *FEMS Microbiol. Ecol.* **1998**, *26*, 193–205.
6. Jordan, D. C. In *Bergey's Manual of Systematic Bacteriology*; Krieg, N. G.; Holt, J. G., Eds.; Williams & Wilkins: Baltimore, USA, 1984; pp 235–244.
7. Martínez-Romero, E.; Segovia, L.; Mercante, F. M.; Franco, A. A.; Graham, P.; Pardo, M. A. *International Journal of Systematic Bacteriology* **1991**, *41*, 417–426.
8. Segovia, L.; Young, J. P. W.; Martínez-Romero, E. *International Journal of Systematic Bacteriology* **1993**, *43*, 374–377.
9. Amarger, N.; Macheret, V.; Laguerre, G. *International Journal of Systematic Bacteriology* **1997**, *47*, 996–1006.
10. Treilhou, M.; Ferro, M.; Monteiro, C.; Poinsot, V.; Jabbouri, S.; Kanony, C.; Promé, D.; Promé, J.-C. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 301–311.
11. Spaink, H. P.; Aarts, A.; Stacey, G.; Bloemberg, G. V.; Lugtenberg, B. J. J.; Kennedy, E. P. *Mol. Plant-Microbe Interact.* **1992**, *5*, 72–80.
12. Chaplin, M. F. *Anal. Biochem.* **1982**, *123*, 336–341.
13. Gerwig, G. J.; Kamerling, J. P.; Vliegenthart, J. F. G. *Carbohydr. Res.* **1978**, *62*, 349–357.
14. Ciucanu, I.; Kerek, F. *Carbohydr. Res.* **1984**, *131*, 209–217.
15. Cárdenas, L.; Domínguez, J.; Quinto, C.; López-Lara, I. M.; Lugtenberg, B. J. J.; Spaink, H. P.; Rademaker, G. J.; Haverkamp, J.; Thomas-Oates, J. E. *Plant Mol. Biol.* **1995**, *29*, 453–464.
16. Blakeney, A. B.; Harris, P. J.; Henry, R. S.; Stone, B. A. *Carbohydr. Res.* **1983**, *113*, 291–299.
17. Buser, H. R.; Arn, H.; Guerin, P.; Rauscher, S. *Anal. Chem.* **1983**, *55*, 818–822.
18. Kupce, E.; Freeman, R. *J. Magn. Reson.* **1995**, *112*, 134–137.
19. Roumestand, C.; Canet, D.; Mahieu, N.; Toma, F. *J. Magn. Reson.* **1994**, *106*, 168–181.
20. Bodenhausen, G.; Ruben, D. J. *Chem. Phys. Lett.* **1980**, *69*, 185–189.
21. Parella, T.; Sánchez-Ferrando, F.; Virgili, A. *J. Magn. Reson.* **1997**, *126*, 274–277.
22. Boudot, D.; Canet, D.; Brondeau, J.; Bouble, J. C. *J. Magn. Reson.* **1989**, *83*, 428–433.
23. Boudot, D.; Roumestand, C.; To, F.; Canet, D. *J. Magn. Reson.* **1990**, *90*, 221–227.
24. Waelkens, F.; Voets, T.; Vlassak, K.; Vanderleyden, J.; van Rhijn, P. *Mol. Plant-Microbe Interact.* **1995**, *8*, 147–154.