Journal of Biomolecular NMR, **14:** 253–271, 1999.
KLUWER/ESCOM

An NMR solution study of the mega-oligosaccharide, rhamnogalacturonan II

Catherine Hervé du Penhoat[a,*], Claude Gey[a], Patrice Pellerin[b] & Serge Perez[a]
[a]*Centre de Recherches sur les Macromolécules Végétales, CNRS (associated with University Joseph Fourier), BP 53, F-38041 Grenoble Cedex 9, France;* [b]*Unité de Recherches des Biopolymères et des Arômes, Institut des Produits de la Vigne, Institut National de la Recherche Agronomique, 2 Place Viala, F-34060 Montpellier, France*

Received 23 March 1999; Accepted 3 June 1999

Key words: long-range nOes, mega-oligosaccharide, rhamnogalacturonan II, sequential nOes

Abstract

Rhamnogalacturonan II (RG-II) is a structurally complex pectic mega-oligosaccharide that is released enzymatically from the primary cell wall of higher plants. It contains roughly 30 monosaccharide units (MW ∼5 kDa) including very unusual residues such as Kdo, Dha, aceric acid and apiose. Previous studies have demonstrated that these monomers are arranged into four structurally well-defined oligosaccharide side chains (A–D), linked to a homogalacturonan mainchain, but the specific attachment sites of these branches on the pectic backbone have not yet been elucidated. In the present work, fairly complete assignments of the 750 MHz ^{1}H NMR spectra and partial assignments of the ^{13}C NMR spectra of the sodium-borohydride-reduced RG-II monomer were obtained for a 5 mM sample isolated from red wine. On the whole, these data corroborate the primary structures of the sidechains previously established by methylation analysis, partial hydrolysis and FAB-MS spectrometry but some heterogeneity has been demonstrated (partial substitution at B5, B6, and A5). The preferred orientations of the majority of the sidechain glycosidic linkages in the RG-II monomer have been determined from the sequential nOe data and the solution structure is generally in good agreement with the stable conformers previously obtained by molecular modeling (MM3) of the disaccharide and sidechain oligosaccharide building blocks. All of a two-residue, a three-residue, and a four-residue segment of the backbone have been tentatively identified from long range interactions between sidechain protons as well as in the mainchain. Taking into account the length of the 9-mer galacturonan mainchain described in prior work, these building blocks constitute almost the complete structure of RG-II (Scheme 2).

Abbreviations: RG-II, rhamnogalacturonan II; mRG-II-ol, saponified and borohydride-reduced rhamnogalacturonan II.

Introduction

Rhamnogalacturonan II (RG-II) is a structurally complex pectic mega-oligosaccharide that is released from the primary cell wall of growing plants by treatment with endo-α-1,4-polygalacturonase. Previous studies (Darvill et al., 1978; O'Neill et al., 1996) based on methylation analysis, partial hydrolysis and FAB-MS spectrometry have shown that RG-II contains 29 monosaccharide units (M_W ∼5 kDa) representing 12 different glycosyl configurations, including very unusual ones such as those of Kdo, Dha, aceric acid and apiose (Scheme 1). Several of the sugar residues are present in both anomeric (α and β) and ring (furanose and pyranose) forms while certain units carry 2-*O*-methyl, acetyl or methylester substituents. The monomers are arranged into four structurally well-defined oligosaccharide side chains (A–D) that are either 2- (A and B) or 3-linked (C and D) to a pectic backbone made up of 1,4-linked α-D-galacturonosyl

*To whom correspondence should be addressed. E-mail: penhoat@cermav.cnrs.fr.

residues. At present, the specific attachment sites of these branches on the pectic backbone have not been elucidated.

The biological roles of RG-II have yet to be determined. Control of plant supply of boron, which is an essential microelement for growth, has been evoked as RG-II is the only boron-containing polysaccharide that has been isolated from a biological source (Kobayashi et al., 1996; Matoh et al., 1993; Ishii and Matsunga, 1996). Borate ester cross-linked RG-II dimer formation has been suggested to involve C-2 and C-3 of one of the 3′-linked apiosyl residues of each monomer and the covalent cross-linking of primary cell wall RG-II, RG-I and homogalacturonan may lead to a pectic matrix which participates in the regulation of the rate of cell growth and eventually the flux of cell-wall components through the wall (O'Neill et al., 1996). More recently, it has been shown that boron, calcium and RG-II influence the porosity of the cell wall (Titel et al., 1998). The chelating properties of RG-II have been considered for reducing the levels of toxic cations in soil and water (Salt et al., 1995; Drake and Rayson, 1996; Jackson et al., 1990) and should current efforts reveal selective binding to particular cations, such as heavy metals and lanthanides, certain industrial applications may become feasible.

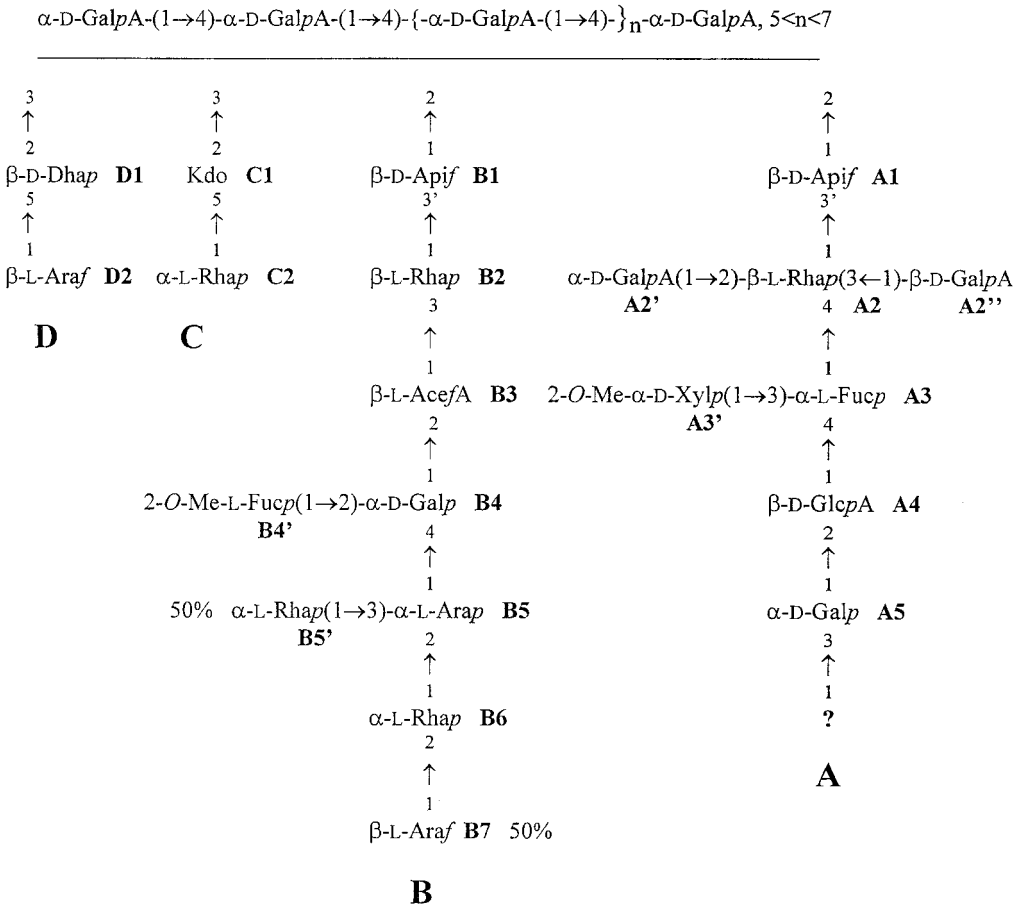
A molecular modeling study of the four side-chain oligomers of RG-II was conducted recently (Mazeau and Perez, 1998). The conformational behavior of the 18 different disaccharide segments contained in the RG-II monomer was evaluated using the flexible residue procedure of the MM3 (Allinger et al., 1990) molecular mechanics procedure. For each disaccharide, the adiabatic energy surface, along with the locations of the local energy minimum were established. These results were included in the database of the molecular builder POLYS (Engelsen et al., 1996), which was implemented to generate all stable conformers of the four RG-II sidechains. This investigation showed that the A-chain was much more rigid than the B-branch but the apiosyl *cis*-diol moiety was well-exposed in both cases. At present, the primary structure of RG-II is required in order to pursue these endeavors with the ultimate goal of constructing favorable 3D models.

In this study, the structural analysis of RG-II from red wine was undertaken by ^{1}H and ^{13}C NMR spectroscopy. Determination of the 3D structure of RG-II represents a real challenge in NMR analysis, comparable to that of elucidating the structure of a 15 kDa protein or a 30-mer nucleic acid. The spectral dispersion of glycans is extremely limited and, in the case of plant biopolymers, this handicap is amplified by natural structural heterogeneity or polydispersity (Aspinall, 1982). Indeed, the biosynthesis of plant polysaccharides is often controlled by many different genes (certainly a large number in the case of RG-II) with the possibility that many of the steps will be interrupted (incomplete) depending on various external factors (climate, soil conditions, etc). However, as investigations of RG-II from a wide range of species have demonstrated a well-conserved primary structure (O'Neill et al., 1990), sample heterogeneity was expected to be fairly limited. The goals of the present work were as follows: (i) to assign as much of the 750 MHz proton spectrum as possible, (ii) to evaluate the orientations of the glycosidic linkages of both the side- and mainchain residues, (iii) to confirm the primary structure determined with other physicochemical methods, (iv) to establish the branchpoints of the sidechains where possible, and (v) to obtain long range interchain nOe data for validating 3D models in future molecular modeling studies.

Materials and methods

Nomenclature

The usual nomenclature for the RG-II sidechain residues has been adopted (i.e, A1, A2....A5 with numbering increasing from the branchpoint apiosyl residue to the terminal α-D-Gal*p*A) (Albersheim et al., 1994). In cases where two sets of resonances are observed for a given residue (e.g., B5) due to heterogeneity in the substitution pattern, the set of resonances with the anomeric proton at lowest field is referred to as A and the one with the anomeric proton at highest field is designated as B (i.e., B5A and B5B for the B5 residue which is either 2-substituted or 2,3-disubstituted). The galacturonan mainchain residues are referred to as M_1 (galactonic acid) to M_T (terminal Gal*p*A) while M_A–M_D designates the corresponding branchpoint sugars and M_X, M_Y, M_{Z1}, and M_{Z2} refer to unassigned Gal*p*A units. It must be underlined that the absolute configurations indicated throughout this manuscript are those reported in previous work (Darvill et al., 1978; O'Neill et al., 1996, and references therein). Finally, the methylene protons with the low- and high-field chemical shifts are referred to as a and b, respectively.



Scheme 1. Schematic drawing of RG-II.

Sample preparation

Some heterogeneity, in the positions of acetyl and methylester substituents, has been reported for RG-II from red wine (Pellerin et al., 1996). In order to optimize the homogeneity of the sample used for the NMR studies, native RG-II, isolated as described previously (Pellerin et al., 1996), was saponified (by treatment with 50 mM NaOH at 4 °C during 2 h followed by neutralization and dialysis) and reduced with sodium borohydride (in 1 M NH_4OH) and then purified by size exclusion chromatography on a ACA 202 (IBF; 1–15 kDa, 400 mm × 16 mm, 13 ml h^{-1}) (Boulard et al., 1997). These steps, which remove the acetyl and methylester substituents, also reduce the anomeric center of the first galacturonosyl residue of the main chain (M_1) to afford 3-linked galactonic acid (see Table 2 for the numbering of galactonic acid). 20 mg of pure RG-II monomer (according to analytical HPLC) was lyophilized three times from D_2O (99.8%) followed by final dissolution in 0.7 ml of D_2O (99.96%) and the resulting 5 mM solution was sealed in a 5-mm NMR tube under argon. This material will be subsequently referred to as mRG-II-ol in order to designate the monomeric form of RG-II possessing a galactonic acid residue at M_1.

NMR spectroscopy

750 MHz ^1H NMR spectra were acquired on a Varian INOVA-750 spectrometer (triple-resonance H/C/N gradient probe). Both ^1H and $^1H/^{13}C$ correlation experiments were run at 500 MHz (F1 125 MHz) on a Varian UNITYPLUS-500 machine (double-resonance H/C gradient probe). Analyses were conducted at 40 °C, as at this temperature the residual H_2O signal overlapped with the fewest resonances and the linewidths had narrowed somewhat. The only exception concerned a DQCOSY experiment which was carried out at 20 °C to verify the number and intensity

of cross peaks located under the water signal in the former (40 °C) homonuclear spectra. Solvent suppression in the homonuclear spectrum was achieved by presaturation during the recycle delay (0.9 s). Quadrature detection in the non-acquisition dimensions was achieved by the hypercomplex method (States et al., 1982) for both homonuclear and HMQC experiments. The gHMBC spectrum was acquired in the magnitude mode (p-type selection).

750 MHz homonuclear 2D spectra were recorded with the following acquisition times (ms, F2) or sweep widths (Hz, F1) and numbers of complex points: DQF-COSY (Piantini et al., 1983; Rance et al., 1983), F1 4291.8 Hz and 400, and F2 469 ms and 4032 (128 transients); clean-TOCSY (Griesinger et al., 1988), F_1 4291.8 Hz and 400 pts, and F_2 469 ms and 4032 pts (64 transients); NOESY (Jeener et al., 1979; Macura and Ernst, 1980), F_1 4291.8 Hz and 400 pts, and F_2 469 ms and 4032 pts (64 transients). The mixing times for the TOCSY (MLEV17) and NOESY spectra were 70 and 200 ms, respectively. NOESY experiments were also run under similar conditions at 500 MHz with mixing times of 100, 200 and 350 ms. The homonuclear 2D data were processed with the program NMRPipe (Delaglio et al., 1995). Phase-shifted squared sinebell apodization functions were applied in both dimensions and the first points were scaled. Data were zero-filled to 4096 and 2048 points in F2 and F1 respectively. A polynomial baseline correction was applied in F2 after Fourier transformation.

Heteronuclear HMQC (Bax, 1983; Summers et al., 1986) and gHMBC (Darmstadt version) experiments were performed at 500 MHz (125 MHz in F1) with standard pulse sequences of the Varian library under the following conditions: HMQC, F_1 16250 Hz and 160 pts, and F_2 400 ms and 1984 pts (128 transients); gHMBC, F_1 25000 Hz and 230 pts, and F_2 400 ms and 2560 pts (560 transients). The heteronuclear 2D data were processed with VNMR, version 5.3 (Varian Associates). Zero-filling to 4096 points, a phase-shifted squared sinebell apodization function and Fourier transformation were applied in F2 while the data size in F1 was extended by linear prediction (1024 points).

Results and discussion

Strategy

Proton and carbon chemical shift assignments of mRG-II-OH relied on the following strategy: (i) initially, the usual criteria for identification of the spin systems of the various types of sugars were adopted (vicinal coupling constant patterns (Koerner et al., 1987), carbon and proton chemical shifts, and glycosylation shifts), (ii) then sequential assignment of the monomers through interglycosidic nOes was achieved in keeping with the primary structure of the A-D sidechains in Scheme 1, (iii) possible branchpoints for the unassigned sidechain sugars were evaluated, and (iv) finally, several blocks of the mainchain galacturonosyl residues (M) were assigned from interglycosidic interactions (nOes and long-range coupling) and long-range nOes with the sidechains. The ^{1}H and ^{13}C chemical shift and coupling data have been collected in Tables 1 (A-D) and 2 (M) while the nOe data are given in Table 3.

General features of the NMR spectra

The overall features of the quantitative 1D proton spectrum (Figure 1) corroborated the primary structure of RG-II established through methylation analysis, partial hydrolysis and FAB-MS spectra of the resulting oligomers (Darvill et al., 1978; O'Neill et al., 1996): a group of methyl resonances corresponding to the 5- and 6-deoxysugars (Ace*f*A, Rha*p* and Fuc*p*) between 1.05 and 1.43 ppm, broad methylene peaks between 1.75 and 2.16 ppm for $H3_{ax}$ and $H3_{eq}$ of Kdo and Dha*p*, a well-resolved doublet of doublets at 3.122 ppm, an extended overlapping region between 3.25 and 5.35 ppm containing most of the sugar resonances including an intense methoxyl methyl singlet between 3.38 and 3.39 ppm, a sharp doublet at 5.54 ppm, and finally broad multiplets near 5.60 ppm.

Initially, the various spin systems were extracted from the 750 MHz TOCSY spectrum which contained fairly complete sets of signals for the protons of the individual sugars and generally similar cross-peak structure was observed for all residues with the same sugar configuration (i.e., B7 and D2 β-L-Ara*f* residues). Then, the ring positions of the individual spins of a given system were determined from the DQ-COSY spectrum as the 3.5–4.2 ppm region (Figure 2) was amenable to analysis due to the enhanced resolution at 750 MHz. Finally, the vicinal coupling constant patterns of the various residues were established from

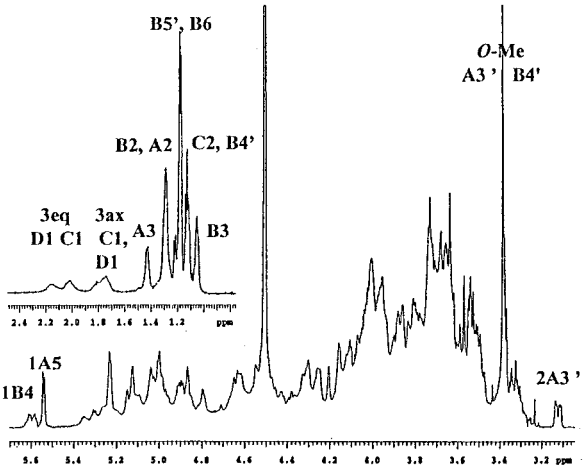


Figure 1. Quantitative 500 MHz 1D spectrum of a 5 mM solution of mRG-II-ol in D_2O at 40 °C. Signals and regions of interest have been labeled.

the fine structure of the crosspeaks in this latter 2D spectrum (digital resolution of ∼1 Hz in F2) according to the following classification: very large (*vl*, >10 Hz), large (*l*, 6–10 Hz), medium (*m*, 3–6 Hz), small (*s*, 1–3 Hz), and very small (*vs*, <1 Hz).

As regards the ^{13}C data, the HMQC spectrum was much more poorly resolved and individual crosspeaks could only be distinguished in the anomeric, the low-field proton (F2-δ_H > 4.15 ppm), the low-field carbon (F1-δ_C > 76.5 ppm) and the high-field carbon (F1-δ_C < 67 ppm) regions. Continuous bands of crosspeaks were detected at 68 (F2 – 3.46–3.80 ppm), 69 (F2 – 3.63–3.95 ppm), 70 (F2 – 3.32–4.05 ppm), 72 (F2 – 3.31–3.76 ppm) 74.4 (F2 – 3.50–3.70 ppm), and 75.6 (F2 – 3.50–3.80 ppm) ppm in F1 and therefore the corresponding carbon chemical shift assignments for resonances in these overlapping regions are not informative. Finally, heteronuclear correlations were not detected for the spins which gave rise to the weaker crosspeaks in the homonuclear 2D spectra due to the lower signal-to-noise of the HMQC experiment.

Classification of the monomer spin systems

Previous studies have shown that RG-II contains sugars with the *gluco* (Glc*p*A and Xyl*p*), *arabino* (Ara*f* and Ara*p*), *galacto* (Gal*p*, Fuc*p*, and Gal*p*A), and *manno* (Rha*p*) configurations along with unusual 3-deoxy hexopyranosides (Dha*p* and Kdo*p*) and 3-disubstituted furanosides (Ace*f*A and Api*f*) residues. Identification of the various spin systems belonging to each of these six groups was fairly straightforward whereas it was much more difficult to distinguish the various sugars within a given group.

The primary structure in Scheme 1 contains two residues with the *gluco* configuration, 2-*O*-Me-α-Xyl*p* (A3′) and β-Glc*p*A (A4). The well-resolved signal at 3.122 ppm was readily assigned to H-2(A3′) as a complete set of proton chemical shifts was extracted from the DQCOSY spectrum and the terminal A3′ sugar was the only RG-II pentose with the *gluco* configuration. A strong intraresidue nOe between H-1(A3′) and the methoxyl methyl resonating at 3.39 ppm as well as a heteronuclear long range coupling between these methoxyl methyl protons and C-2(A3′) (80.5 ppm) corroborated this assignment. The DQCOSY spectrum also contained two sets of H-1/H-2 crosspeaks with the appropriate fine structure and chemical shifts (4.64/3.57 and 4.46/3.555 ppm) for β-Glc*p*A (Skelton et al., 1991). Analysis of the corresponding spin systems revealed almost identical chemical shifts for the H-2–H-3–H-4–H-5 fragments in the DQCOSY spectrum (Figure 2). It was not possible to distinguish these residues in the HMQC spectrum as two sets of H-1/C-1 crosspeaks were observed for each of the anomeric protons ($\delta_{H\text{-}1}$ 4.63 − $\delta_{C\text{-}1}$ 101.3 or 98.1 ppm; $\delta_{H\text{-}1}$ 4.46 − $\delta_{C\text{-}1}$ 104.7 or 101.2 ppm).

The majority of sidechain (A2′, A2″, A3, A5, B4, B4′, B5, B7, and D2) and all of the mainchain residues in RG-II have been shown to have the *galacto* or the related *arabino* configuration. Complete sets of proton chemical shifts were obtained for three pentofuranosides which were assigned to terminal β-L-Ara*f* as these data were analogous to those reported for the corresponding methyl glycoside (Serianni and Barker, 1979).

Furthermore, the characteristic β-L-Ara*f* C-1 (∼103 ppm), C-3 (∼79 ppm), and C-4 (∼81 ppm) resonated in isolated regions of the HMQC spectrum in agreement with literature data for Me β-D-Ara*f* (Bock and Pedersen, 1983). The remaining two sets of pentoside signals stemming from pyranosyl *arabino* sugars were attributed to α-L-Ara*p* units. The chemical shift patterns of these latter sugars were similar to those of a 3-linked α-Ara*p* dimer (Joao et al., 1988) but the $^3J_{1,2}$ vicinal coupling constants (m) were smaller than the 8-Hz one reported for Me-α-D-Ara*p* (Bock and Thogersen, 1982). However, analogous $^3J_{1,2}$ values (4–6 Hz) have been reported for the α-L-Rha*p* (1→2)-α-L-Ara*p*(1→ disaccharide fragments of saponins and related model compounds (Castro et al., 1997).

The signals of the H-1–H-2–H-3–H-4 fragments for the residues in the *galacto* group were identified

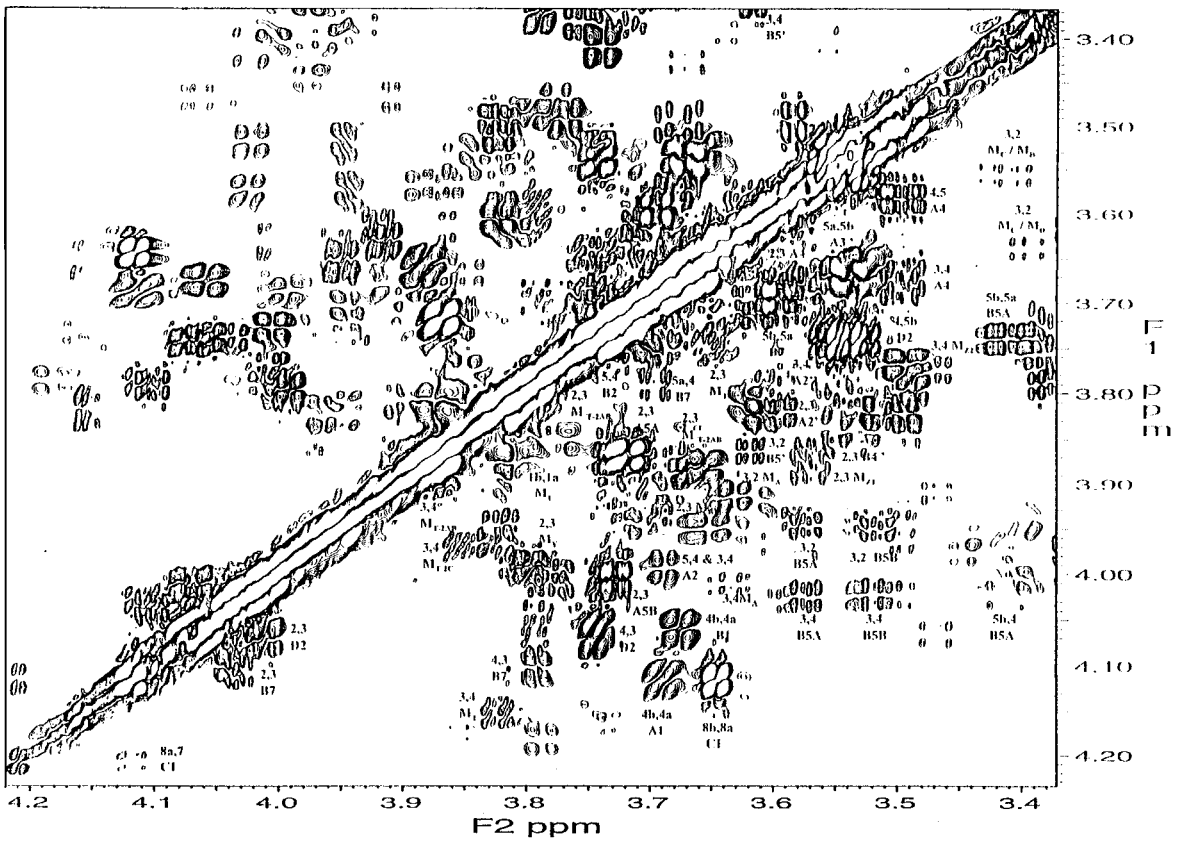


Figure 2. Fingerprint region of the 750 MHz DQCOSY spectrum of mRG-II-ol in D_2O at 40 °C containing the crosspeaks between the high field methine (and methylene) protons. Many of the crosspeaks for the assigned protons have been labeled.

through the distinctive fine structure in the homonuclear correlation spectra (m or g for $^3J_{1,2}$ depending on the anomeric configuration, g for $^3J_{2,3}$ and s for $^3J_{3,4}$). The H-1–H-2–H-3–H-4 connectivity was obtained from the TOCSY spectrum but the very small scalar coupling between H-4 and H-5 prevented further polarization transfer to H-5 and the exocyclic groups. Assembly of the H-1–H-2–H-3–H-4 segments with either H-5–H-6,6′ (α-D-Gal*p*), H-5–CH_3 (α-L-Fuc*p*) or H-5–COOH (D-Gal*p*A) fragments represented the major difficulty in the chemical shift assignments of this group. The only notable difference in proton chemical shifts for the fucosyl (or galactosyl) and galacturonosyl H-1–H-2–H-3–H4 segments concerns H-4 which resonates between 3.8–4.0 ppm (Bock and Thogersen, 1982; Backman et al., 1990; York et al., 1990; Baumann et al., 1991) for the former sugars and between 4.1–4.5 ppm (Nakahara and Ogawa, 1987; Hricovini et al., 1991; Cros et al., 1992) for the latter residues. In carbon spectra, the C-5 chemical shift ranges of unsubstituted fucosyl (66–68 ppm) (Backman et al., 1990; Baumann et al., 1991) and galacturonosyl (or galactosyl – 70–72 ppm) (Hricovini et al., 1991; Cros et al., 1992) residues are also fairly different.

Amongst the 6-deoxymethyl signals, three resonances could be assigned to either CH_3-H-5 fucosyl or CH_3–H-4 Ace*f*A fragments due to the limited polarization transfer from the methyl group in the TOCSY spectrum (cf., α or β-Rha*p*) (Figure 3A). The CH_3–H-4 Ace*f*A system was readily identified on chemical shift criteria (Spellman et al., 1983) and the CH_3-H-5 fragment of the 3,4-disubstituted fucosyl unit was characterized by numerous crosspeaks in the NOESY spectrum (Figure 3B). As regards B4′, long range heteronuclear coupling between the 6-deoxymethyl protons at 1.12 ppm and C-5(B4′) (66.4 ppm) corroborated the assignment of the CH_3–H-5 fragment. Many of the H-1–H-2–H-3–H-4 fragments of the *galacto* sugars displayed analogous NMR data to those of the related fucosyl sugar in the β-D-Glc*p*(1→4)-α-L-Fuc*p*-*O*-Me disaccharide (Backman et al., 1990)

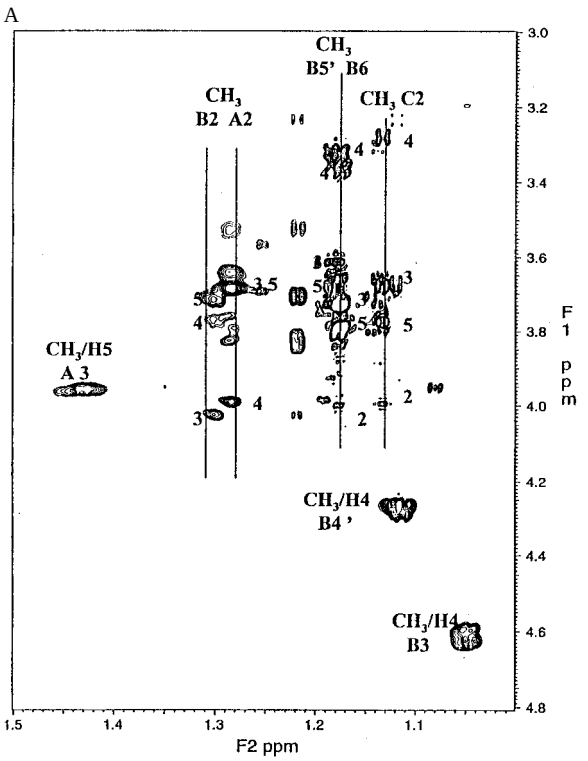


Figure 3. (A) Expansion of the 750 MHz TOCSY spectrum of mRG-II-ol in D_2O at 40 °C containing the region with the deoxymethyl crosspeaks. The more intense signals have been labeled. (B) Expansion of the 750 MHz NOESY containing the region with the deoxymethyl crosspeaks. The more intense nOes between assigned protons have been labeled.

(H-1–H-2–H-3–H-4 chemical shifts in D_2O at 70 °C: 4.80, 3.78, 3.81, and 4.01 ppm) precluding assignment of A3 on the basis of proton chemical shifts alone. The starting point for the identification of the B4′ H-1–H-2–H-3–H-4 segment was long range coupling between the 2-*O*-Me protons (3.38 ppm) and C-2(B4′) ($\delta_{C\text{-}2}$ 77.4 ppm). From the latter $\delta_{C\text{-}2}$ value, a fairly limited H-2 chemical shift range was deduced from the related HMQC crosspeaks (77.4 ppm in F1). Moreover, inspection of the chemical shift data reported for diversely methylated fucosyl residues (Toman et al., 1986) showed that the protons α to the methoxyl moiety resonate at high field (between 3.40 and 3.60 ppm). This in turn narrowed the range of suitable *galacto* H-1/H-2 crosspeaks down to three sets, all of which exhibited H-1–H-2–H-3–H-4 chemical shifts compatible with a terminal 2-*O*-Me-α-L-Fuc*p* structure (4.65/3.57/3.885/3.852, 4.645/3.57/3.87/3.86, and 4.62/3.57/3.86/3.923 ppm).

Three sets of proton signals were assigned to α-Gal*p* sugars and in agreement with data reported for A (Melton et al., 1986) and B (Whitcombe et al., 1995) sidechain oligomers, the resonances of the galactosyl anomeric protons were located at lowest field. The sharp doublet at 5.54 ppm, which was attributed to H-1(A5) displayed similar intensity to that of the H-2(A3′) signal in Figure 1. It should be kept in mind that two sets of resonances were observed for the A4 residue, and indeed inspection of the TOCSY crosspeaks for A5 indicated one set of signals for H-1, H-2, and H-4 (5.54/3.726/3.77 ppm) and two different ones for H-3 (3.87 and 4.005 ppm). Comparison of the C-3 and C-5 chemical shifts extracted from the HMBC spectrum through intracyclic long range heteronuclear coupling with the anomeric protons (terminal A5–C-3, 68.3 ppm; C-5, 69.6 ppm; 3-substituted A5–C-3, 74.2 ppm) suggested partial 3-substitution for A5 (downfield α-glycosylation shifts for C-3). The H-1(B4) resonances included two broad multiplets indicating that there was also some heterogeneity in the substitution pattern of the B-chain (5.61 and 5.59 ppm, integrating for 0.5 H each). Close inspection of the H-2(B4)/H-3(B4) crosspeaks in both the DQCOSY and TOCSY revealed multiple H-3(B4) frequencies but further coupling to H-4(B4) was not detected. Intracyclic long range heteronuclear cou-

pling with the anomeric protons (2,4-disubstituted B4: C-3, 67.8 ppm; C-5, 66.4 ppm) and the C-1 chemical shift (∼95 ppm) confirmed the substitution pattern in Scheme 1 for B4 (upfield β-glycosylation shifts for C-1, C-3 and C-5) and it was possible to assign H-5(B4) (δ 4.11 ppm) from the isolated H-5(B4)/C-5(B4) crosspeaks in the HMQC spectrum.

The remaining sets of signals with the α-*galacto* fine structure (∼12 residues) were assigned to galacturonosyl residues, Table 2. As observed for the sidechain residues, a few of these spin systems presented dual sets of signals with only a small shift in one or more proton frequencies. From the C-4 chemical shifts in the HMQC spectrum (∼71 ppm), it could be inferred that two of these sugars were not substituted at the 4-position and they were assigned to the terminal α-Gal*p*A residues in the mainchain (M_T, $\delta_{H\text{-}1}$ 5.012 with a weaker satellite signal at 4.94 ppm) and in the A sidechain (A2′, $\delta_{H\text{-}1}$ 4.865 ppm). Most of the protons of the α-Gal*p*A residues in the mainchain exhibited a fairly wide range ($\Delta\delta_{H\text{-}X}$ ∼0.8 ppm) of chemical shifts. However, all the H-3 resonated in a fairly narrow spectral band (3.77–3.99 ppm) with the exception of two residues ($\delta_{H\text{-}3}$ 3.39 and 3.40 ppm) which have been tentatively assigned to 3,4-disubstituted sugars, →3,4)-α-D-Gal*p*A-(1→ (M_C and M_D) (Figure 2). A residue with NMR data analogous to those of terminal β-D-Gal*p*A (Rinaudo and Ravanat, 1980) was also identified (A2″, $\delta_{H\text{-}1}$ 4.425 ppm) and this spin system was characterized by strong second order effects ($\Delta\delta_{H\text{-}2,H\text{-}3}$ ∼15 Hz).

In the case of the sugars with the *manno* configuration, two groups of TOCSY crosspeaks could be distinguished on the basis of the H-4 chemical shift (Figure 3A). At least 10 complete sets of chemical shifts were detected for residues with the H-4 proton resonating at high field ($\delta_{H\text{-}4} < 3.4$ ppm) whereas three partial sets of chemical shifts were observed for residues with the H-4 proton resonating at lower field ($\delta_{H\text{-}4} > 3.5$ ppm). Only the more intense sets of signals have been indicated in the rows of Table 1 (data for the weaker sets of signals have been given as a footnote). The proton chemical shifts of rhamnosyl sugars are very sensitive to the substitution pattern and, in a study of the solution conformations of oligosaccharides related to a *Shigella flexneri O*-antigen (Bock et al., 1982), ^{1}H and ^{13}C chemical shifts of terminal, 2- (2- and 3-linked) and 3-substituted (3-linked) α-rhamnosides were measured under similar conditions (D_2O, 37 °C) to those of the present study. The H-4 protons of all these glycans resonate at high field ($\delta_{H\text{-}4} < 3.4$ ppm) and the 2- and 3-linked terminal residues show significant upfield shifts for both H-1 (−0.15 to −0.30 ppm) and H-2 (−0.05 to −0.35 ppm) signals when compared to those of the 2-linked 2-substituted rhamnose units. Consequently, the rhamnosyl sugars with the H-1 resonating at 5.12 (H-2, 4.00 ppm) and 4.99 (H-2, 3.88 ppm; as well as numerous weaker H-1/H-2 signals such as 5.035/3.93, 5.02/3.91, 5.02/3.90, 4.975/3.865, and 4.955/3.86) have been assigned to 2-substituted and terminal α-Rha*p*, respectively (B6). The remaining rhamnosyl sugar with intense signals ($\delta_{H\text{-}1}$ and $\delta_{H\text{-}2}$ 5.15 and 3.995 ppm respectively) was assigned to C2 as will be discussed later.

Two strong sets of CH3-H-5-H-4··H3 signals of residues with the *manno* configuration with the H-4 proton resonating at lower field, >3.5 ppm were assigned to β-Rha*p*. Inspection of the proton chemical shifts of 4-substituted (De Bruyn et al., 1975) and 3,4-disubstituted (Peters and Weimar, 1994) β-L-rhamnose shows that the presence of substituents at C-3 and C-4 leads to significant downfield shifts of the surrounding protons as compared to those of monomeric β-L-rhamnose (De Bruyn et al., 1975) (4-substituted β-Rha*p*: for H-3–H-4–H5, +0.24, +0.28, and +0.10 ppm; 3,4-disubstituted β-Rha*p*: for H-2–H-3–H-4–H-5, +0.35, +0.33, +0.57 and +0.18 ppm). It should also be noted that the 6-deoxymethyl doublets of the aforementioned α-rhamnosides related to a *Shigella flexneri O*-antigen (Bock et al., 1982) are all found between 1.13 and 1.20 ppm while those of the 4-substituted β-L-rhamnose (De Bruyn et al., 1975), 3,4-disubstituted β-L-rhamnose (Peters and Weimar, 1994), and a persubstituted β-L-rhamnoside (De Marco et al., 1992) all resonate between 1.35–1.42 ppm. This tendency is also observed for the α- and β-Rha*p* residues of mRG-II-ol whose 6-deoxymethyl doublets resonate at higher (1.12–1.22 ppm) and lower field (1.28–1.30 ppm), respectively (Figure 3A). Extensive through-space interactions were observed for the 6-deoxymethyl groups of these spin systems while very few nOes were detected for the 6-deoxymethyl moieties of the α-Rha*p* sugars (Figure 3B). The H-1 chemical shifts of the β-Rha*p* residues were obtained from weak intraresidue nOes with 6-deoxymethyl moieties ($\delta_{H\text{-}1}$ 4.85 and 4.82 ppm). H-1/H-2 crosspeaks are not detected for these anomeric protons in the DQCOSY experiment (such correlations were detected for $\delta_{H\text{-}1}$ 4.82 ppm only in the TOCSY spectrum), indicating very small scalar coupling, but strong H-1/H-2 interactions are

Table 1. 750 MHz $^{1}H^{a}$ and 125 MHz ^{13}C NMR data of the sidechain residues of mRG-II-ol in D_2O (5 mM) at 313 K

Residue	C1 H1	C2 H2	C3 H3	C4 H4	C5 H5	C6 H6
A1 → 3′)βDApi*f*(1 → 2	110.12 5.00 nc	75.4 3.77	78.4 nap	74.1 4.11, 3.69	69.75 4.38, 3.63	nap
A2 → 2,34)βLRha*p*(1 → 3′	nd 4.85	81.2 4.025	78.2 3.69	81.6 3.99	68.0 3.685	17.6 1.285
A2′ αDGal*p*A(1→2	99 4.865	68.1 3.59	69.2 3.82	70.8 4.146	71.2 4.485	na
A2″ βDGal*p*A(1→3	104.84 4.425	68.1 3.53	70.0 3.55	67.7 3.81	72.6 4.485	na
A3 → 3,4)αLFuc*p*(1 → 4	99.72 4.865	68.5 4.03	74.2 na	78.6 4.01	67.6 3.96	15.9 1.43
A3′ 2-*O*- MeαDXyl*p*(1→3	97.55 5.23	80.76 3.122	74.01 3.638	71.8 3.32	61.2 3.525, 3.665	nap
A4A → 2′)βDGlc*p*A(1 → 4	101.3 4.64 na	77.10 3.576	73.4 3.67	71.8 3.495	75.6 3.58	na
A4B	4.46	3.555	3.67	3.49	3.58	
A5A αDGal*p*(1→2 **A5B**	97.5 5.54	69.9 3.726	69.2 3.87 4.005	69.6 3.765	70.5	62.5 3.41,3.74
B1 → 3′)βDApi*f*(1 → 2	109.8 5.10 nc	75.0 3.77	nap	74.2 4.06,3.68	69.75 4.31, 3.634	nap
B2 → 3)βLRha*p*(1 → 3′	99.7 4.82	71.0 4.11	77.20 4.015	3.765	3.72	17.1 1.30
B3 →2βLAc*f*A(1→ 3	108.8 4.87 nc	4.495	82.87	78.94 4.605 4.62	13.8 1.05	nap
B4A →2,4)αDGal*p*(1→ 2 **B4B** **B4C**	95.19 5.612 95.02 5.584	77.25 3.36 3.355	67.6 3.98 68.5 3.98,4.055, 4.085	3.73,3.74 3.705- 3.73	66.6 4.11	na
B4A′ 2-*O*-MeαLFuc*p*(1 →2	99.5 4.665	81.0 ∼3.57	68.2 3.885	71.0 3.852	66.53 ∼4.275	15.8 1.12 15.8
B4B′	4.645	∼3.57	3.87	3.862	∼4.275	1.11

Table 1 (continued)

Residue	C1 H1	C2 H2	C3 H3	C4 H4	C5 H5	C6 H6
B5A	103.2	81.0	74.3	71.5	70.2	nap
→2)αLAra*p*(1→4	4.888	3.943	3.575	4.025	3.41,3.74	
B5B	103.2	81.0	77.1	70.2	70.2	
→2,3)αLAra*p*(1→4	4.865	3.948	3.52	4.022	3.41,3.74	
B5′[b]	101.2	70.2	69.7	72.0		16.65
αLRha*p*(1→3	4.975	3.865	3.62	3.353	3.68	1.18
	4.95	3.86				
B6A[b]	101.3	81.3	69.7			16.65
→2)αLRha*p*(1→	5.12	4.00	3.735	3.328	3.805	1.177
2,3 + →2)αLRha*p*(1→2						
B6B[b]	99.4	70.1	69.7	72.0		16.65
αLRha*p*(1→2,3	4.99	3.885	3.735	3.328	3.805	1.174
+						
αLRha*p*(1→2						
B7	101.2	78.5	77.9	81.4	60.7	nap
βLAra*f*(1→2	4.985	4.03,	4.098, 4.11	3.79	3.69,	
		4.035			3.59;	
					3.663,3.598	
C1A			35.2		na	na
→5)αKdo*p*(2→3	nap	nap	1.744,	4.02		
			2.025;			
C1B			1.82, 2.02	4.02		
C2[b]	101.24	71.50				16.65
αLRha*p*(1→5	5.15	3.998	3.775	3.29	3.69	1.133
D1			35.5	71.2	78	
→5)βDDha*p*(2→3	nap	nap	1.744, and	4.085	4.035,	na
			2.155		4.105	
D2A	103.24	75.8	78.6	81.0	61.2	nap
βLAra*f*(1→5	4.918	4.032	4.062	3.735	3.53,3.67	
D2B	103.16					
	4.865	4.012	4.07	3.743	3.54, 3.66	

[a]Key: s – singlet, d – doublet; t – triplet; w – weak; br – broad; na not assigned; nc – not correlated; nd – not detected; nap not applicable; or – overlapping region. Fine structure (i.e., ^{3}JH1,H2, ^{3}JH2,H3, ^{3}JH3,H4, ^{3}JH4,H5, ^{3}JH5,H6a, ^{3}JH5,H6b, ^{2}JH6a,H6b where a denotes the low-field proton of the methylene group) in the DQCOSY (TOCSY for very small coupling) spectrum (very large – vl > 10 Hz, large – l 6–10 Hz, medium – m 3–6 Hz, small – s 1–3 Hz, and very small – vs <1 Hz): βAce*f*A nc,nap,nap,l; αApi*f*A nc,nap,nap ^{3}JH4a,H4b vl, ^{4}JH2,H3′a vs, ^{4}JH2,H3′b vs, ^{3}JH3′a,H3′b vl; βAra*f* m,l,l,s,or; αAra*p* m,l,l,nd,s,vl; βDha*p* ^{3}JH3a,H3b vl, ^{3}JH3b,H4 l; αGal*p* m,l,m,nd; αGal*p*A m,l,m,nd; βGal*p*A l,l,m,nd; βGlc*p*A m,l,m,nd; αKdo*p* ^{3}JH3a,H3b vl, ^{3}JH3b,H4 l; αRha*p* vs,s,l,l,l; βRha*p* nc and vs (A2 and B2, respectively),s,l,l,l; αXyl*p* m,l,l,l,s. Supplementary unassigned well-defined spin systems include: 1/ 4.815 s, 3.425 l, 3.67 l, 3.34 l, 4.40; 4.84 vs, 3.91 (3.925 w) s, 3.465 l, 4.065 s, ?.

[b]In the case of the B5′, B6, and C2, it should be noted that weaker signals for numerous other spin systems with the coupling graphs and chemical shifts of α-rhamnosyl residues were also detected (5.115, 3.935, 3.69&3.665, 3.235, 3.83, 1.12; 4.92, 4.025&4.055&4.068, 3.795, 3.37, 3.72, 1.174; 4.80, 4.11&4.135, 3.52&3.23, 3.71&3.83, 1.219 – 0.5 CH_3; ?, 3.895, 3.695, 3.318, 3.565, 1.256 w).

observed in the NOESY spectrum (4.025 for $\delta_{H\text{-}1}$ 4.85 and 4.105 ppm for $\delta_{H\text{-}1}$ 4.82 ppm, respectively) (Figure 4).

RG-II contains two unusual furanosyl residues, aceric acid and apiose, which are characterized by vanishingly small $^3J_{1,2}$ coupling constants and low field C-1 chemical shifts (~110 ppm). As Ace*f*A is substituted at the 2-position, its anomeric carbon would be expected to resonate at slightly higher field due to the β-glycosylation shift (−1.5 ppm). Accordingly, the three sets of signals associated with anomeric carbon signals at 108.8, 109.8 and 110.12 ppm were assigned to B3 ($\delta_{H\text{-}1}$ 4.865 ppm) and A1 or B1 (the latter two signals, $\delta_{H\text{-}1}$ 5.00 and 5.10 ppm, respectively). In an HMQC experiment conducted on a more concentrated sample of reduced RG-II (which displayed much less favorable spectral dispersion in the proton dimension), crosspeaks were observed between a carbon resonating at 88 ppm and a proton with a chemical shift near 4.50 ppm. Aceric acid is the only residue which would be expected to exhibit such low field methine carbon signals (considering the $\delta_{H\text{-}2}$ and $\delta_{C\text{-}2}$ of the monomer, 4.13 and 83.5 ppm, respectively (Spellman et al., 1983), carbon glycosylation shifts of 2–5 ppm would lead to $\delta_{C\text{-}2}$ values in the 85–88 ppm range). A strong nOe was detected for H-1(B3) and a proton resonating at 4.49 ppm, Figure 4, and this latter interaction was ascribed to H-2(B3). As previously stated, the CH_3-H-4 fragment of Ace*f*A was readily identified due to its unusual H-4 (4.595 ppm) and C-4 (78.94 ppm) chemical shifts when compared to those of the other 6-deoxysugars.

The apiose residues contain two isolated methylene groups, an endocyclic CH_2 moiety (H-4a and H-4b) and an exocyclic one (H-3′a and H-3′b). Four such methylene pairs were detected in the DQ-COSY (Figure 2) and NOESY (Figure 4) spectra (4.38/3.63, 4.31/3.634, 4.115/3.69, 4.06/3.645 ppm). The methylene carbons corresponding to the pairs with the least shielded protons exhibited chemical shifts around 70 ppm and therefore they were assigned to the exocyclic apiose methylene groups as the C-3′ and C-4 chemical shifts of terminal apiose are 64.1 and 74.3 ppm, respectively (Mbairaroua et al., 1994), and a 5–7 ppm glycosylation shift would be expected for C-3′. Long range heteronuclear coupling between the apiosyl anomeric proton that resonated at 5.00 ppm and carbons with chemical shifts of 74.1 ($^3J_{H\text{-}1,C\text{-}4}$) and 78.4 ($^3J_{H\text{-}1,C\text{-}3}$) ppm was interpreted in terms of intramolecular coupling. Both of the low-field exocyclic protons displayed long-range coupling in the TOCSY spectrum to protons resonating close to 3.77 ppm. Strong nOes were observed between all four exocyclic protons and these same spins which were identified as H-2 of the apiosyl sugars.

Partial spin systems for the 3-deoxy Kdo*p* (H-3_{ax}–H-3_{eq}–H4) and β-Dhap (H-3_{ax}–H-3_{eq}–H-4–H-5) could be obtained from the TOCSY experiments as the H-3_{ax}/H-3_{eq} methylene protons resonate in an isolated region of the proton spectrum (Figure 1). The H-3_{ax}/H-3_{eq} signals presented chemical shift and coupling constant data analogous to those reported for the α-Kdo*p* (1.75 and 2.03 ppm, C1A; 1.82 and 2.02 ppm, C1B) and β-Dha*p* (1.744 and 2.155 ppm, D1) monosaccharides in the $^5C_2$ pyranose form (Hofinger et al., 1993). Due to very small coupling between H-5 and H-6 of the 3-deoxy sugars, it was not possible to unambiguously identify the signals of the Kdo*p* exocylic group (H-5–H-6–H-7–CH_2OH fragments). However a partial spin system with similar chemical shifts (4.205, 4.115 and 3.645 ppm) to those reported for the H-7–H-8a,H-8b protons of the pendant group of α-Kdo*p*-(2→ or →4,5)-α- Kdo*p*-(2→ (Fukuoka et al., 1997) was detected (Figure 2).

Finally, the sodium borohydride reduction of RG-II was expected to transform the reducing galacturonosyl residue into galactonic acid. In aqueous solution, unsubstituted galactonic acid exists as a mixture of linear (free acid, minor product) and cyclic (aldono-1,4-lactone, major product) forms (Angelotti et al., 1987) and it was not clear which form of galactonic acid would predominate in the case of mRG-II-ol (i.e. M_1). A partial spin system with the characteristics (4.72-*l*, 4.465-*l*, 4.40-*s*) of the aldono-1,4-lactone H-2–H-3–H-4 fragment (this segment corresponds to H-5–H-4–H-3 of the α-Gal*p*A sugar prior to reduction) was identified but these signals were very weak and the resonances for the corresponding carbons were not detected in the HMQC experiment. Through space interactions were observed between H-3 and a second group of signals with the characteristics of the H-5-H6a,H-6b fragment (Figure 4), allowing complete assignment of the aldono-1,4-lactone spin system. It was not possible to identify the signals of the free acid.

This concludes the classification of the signals of the monomer residues according to chemical shift data and scalar coupling patterns. Two sets of sharp resonances with unusual coupling patterns (footnote to Table 1) and a group of extremely broad resonances have been detected which do not appear to correspond to the sugars of the RG-II 29-mer in Scheme 1. Otherwise, the NMR data of the vast majority of the

Table 2. 750 MHz ^{1}H[a] and 125 MHz ^{13}C NMR data of the mainchain residues of mRG-II-ol in D_2O (5 mM) at 313 K

Residue	C1 H1	C2 H2	C3 H3	C4 H4	C5 H5
M_1[b]					
→ 4)Galactonic acid	na	nd	nd	nd	na
Aldono-1,4-lactone		4.72	4.465	4.40	3.905
Free acid		na	na	na	na
M_A	99.72	77.0		78.6	na
→2,4)αDGal*p*A(1→4	5.235	3.91	3.63	4.01	4.26
M_B	99.5	75.0	68.2	75.8	71.3
→2,4)αDGal*p*A(1→4	5.351	3.948	3.65	4.162	4.61
	99.33	75.0	68.2	75.8	
	5.335	3.952	3.65	4.164	4.625
M_C or **M_D**	99.3	68.0	71.6	77.0	nd
→3,4)αDGal*p*A(1→4	4.79	3.555	3.40	4.51	5.31
M_C or **M_D**	99.2	68.0	71.8	78.8	nd
→3,4)αDGal*p*A(1→4	4.62	3.64	3.39	4.485	5.31
M_T	98.4	68.1	68.0	71.2	71.2
αDGal*p*A(1→4	5.012	3.608	3.822	4.155	4.55
	4.94w	3.608			
$M_{T\text{-}2}$	99.1	80.76	74.01	72.0	nd
→4)αDGal*p*A(1→4	4.58	3.685	3.875	3.902	4.70
	4.57w	3.665	3.87	3.91	4.74
	97.5	69.9	69.2	69.6	70.5
	4.54	3.726	3.852	3.965	4.42
$M_{X\text{-}1}$	98.3	68.2	68.9	77.62	71.0
→4)αDGal*p*A(1→4	5.04	3.72	3.865	4.325	4.24
M_X	98.8	68.1	68.0	nd	nd
→4)αDGal*p*A(1→4	4.968 br	3.66	3.88		
M_Y	99.0	68.0	70.2	77.4	
→4)αDGal*p*A (1→4	4.91w	3.81	3.99	4.30	4.50
	97.5	68.0			71.2
	4.89	3.80	3.99	4.298	4.53
M_{Z1}	99.72	68.1	74.4	78.6	nd
→4)αDGal*p*A(1→4	4.62	3.57	3.86	3.923	
M_{Z2}	4.61w	3.495	3.765	4.74	nd
→4)αDGal*p*A(1→4	4.615w	3.47	3.775		

[a]Key as in Table 1.
[b]The numbering of galactonic acid begins with the carboxyl carbon so that the H-2–H-3–H-4–H-5–H6a,H6b fragment corresponds to H-5–H-4–H-3–H-2–H1–$C_{anomeric}$ of the α-Gal*p*A sugar prior to reduction. The chemical shifts for the reduced center are as follows: C6 59.1 ppm, H6a 3.875 ppm, and H6b 3.81 ppm.

Table 3. Comparison of the glycosidic linkage orientations of mRG-II-ol defined from the 750 MHz NOESY spectrum (mixing time of 200 ms) compared with those of the minima of the MM3 adiabatic energy maps

Residue	H1 nOes	Other reporter groups	Reporter group nOes	Glycosidic linkage orientation rank and (Φ,Ψ) values in the MM3 force field
A1 →3βDApi*f*(1→2	H2(M_{T-2}) br S, H5(M_T) w, H1(A2′) w,3.955 s S	H3′a	H3′b(A1) S, H2(A1) S, H2(M_{T-1}) S H1(M_{T-1}) w;	5(60,120)
A2 →2,3,4)βLRha*p*(1→3′	H2(A2) S, H5(A2) M, CH_3(A2) w	CH_3	H5(A2) S, H5(A3) S, H1(A3) S, H1(A2) w, H1(M_A) w	2 (300,170)
A2′ αDGal*pA*(1→2	H2(A2) or, H2(A2′) M, H1(A1) w, H2(A2) S			2(100,300)
A2″ β3DGal*A*(1→3	3.965 S			3(60,120)
A3 →3,4)αLFuc*p*(1→4	or - H2(A3) S, CH_3(A2)M, CH_3(A3) w	CH_3	H1(A4) S, H3(A4) S, H5(A4) S, H5(A3) S, H4(A3) S, H3′a(A1)	2(280,60)
A3′ 2-O-MeαDXyl*p*(1→3	H2(A3′) S, H2(A3′) M,OMe(A3′) S	OMe	H1 (A3′) S, H2(A3′) S	3(280,140)
A4A →2)βDGlc*A*(1→4	H4(A3) S, H3(A4) S, H5(A4) S, CH_3(A3) S, H1(A5) w			1(280,120)
A4B	H3(A4) S, H5(A4) S; H1(A5) w			2(60,120) 3(280,320)
A5 αDGal*p*(1→2	H2(A5) S, H1(A4A) S, H1 (A4B) S			1(80,120) 2(100,160)
B1 →3′βDApi*f*(1→2	3.58 M, 4.075 M	H3′a(B1)	H3′b(B1) S, H2(B1) S, H2(M_B) S	5(160, 120)
B2 →3(βLRha*p*(1→3′	H2(B2) S, CH_3(B2) M	CH_3	H5(B2) M, H4(B2) M, H1(B2) S, H1(B3) M, H1(M_{B-1}) S, H1(M_B) w	2(300, 170)
B3 →2)βLAce*Af*(1→3	H2(B3) S, H3(B2) or	CH_3	H4(B3)	2(80, 160) 4(300,100)

Table 3 (continued)

Residue	H1 nOes	Other reporter groups	Reporter group nOes	Glycosidic linkage orientation rank and (Φ,Ψ) values in the MM3 force field
B4A+B →2,4)αDGal*p*(1→2	H2(B4) S, OMe(B4′) S; H2(B4′) w,			1(80,220) (*40*, *160* or *180*, *280*)
B4′ 2-O-MeαLFuc*p*(1 →2	H5(B4′), 3.41 br, 3.955	OMe	H1(B4) S, H3(B4′) w, CH_3(B4′) S, H1(B4′ br	2(260,280)
B5A →2,3)αLAra*p*(1→4	H2(B5A) M,			1(280,260)
B5B →2)αLAra*p*(1→4	H4(B4B) M H3(B4A) M, H4(B4A) H2(B5) S, H3(B5)			
B5′ αLRha*p*(1→3	H2(B5′)S			not simulated
B6A →2)αLRha*p*(1→ 2,3 + →2)αLRha*p*(1→2	H2(B6) S; H2(B7) w	CH_3	H1(B4A+B)	4(60, 120)
B6B αLRha*p*(1→2,3 + αLRha*p*(1→2	H2(B6) S			
B7 βLAra*f*(1→2	H2(B7) S, H2(B6) S			1(80,240)
C1A →5)αKdo(2→3		H3a	H1(C2A)	nnd
C1B		H3a	H2(C2B)	
C2A αLRha*p*(1→5	H2(C2A) M, H1($M_b \pm 1$) M, H4($M_b \pm 1$) S			4(280,60)
D1 →5)βDDha(2→3		H3a H4	H4(D1) S; 4.71, 4.485 H3a(D1) M, H3b(D1) M, CH_3(B2) M, H1(D2B) S	nnd
D2A βLAra*f*(1→5	H2(D2A) S, H5(D1) S			1(80,80)

Key: nnd (no nOe data), S (strong), M (medium), w (weak) nOes when compared to the corresponding H1/H2 (strong) intraresidue interactions and otherwise as in Tables 1 and 2. Linkage orientations (Φ,Ψ) correspond to the best fit to the minimum energy structures of the MM3 potential energy surfaces (Mazeau and Perez, 1998).

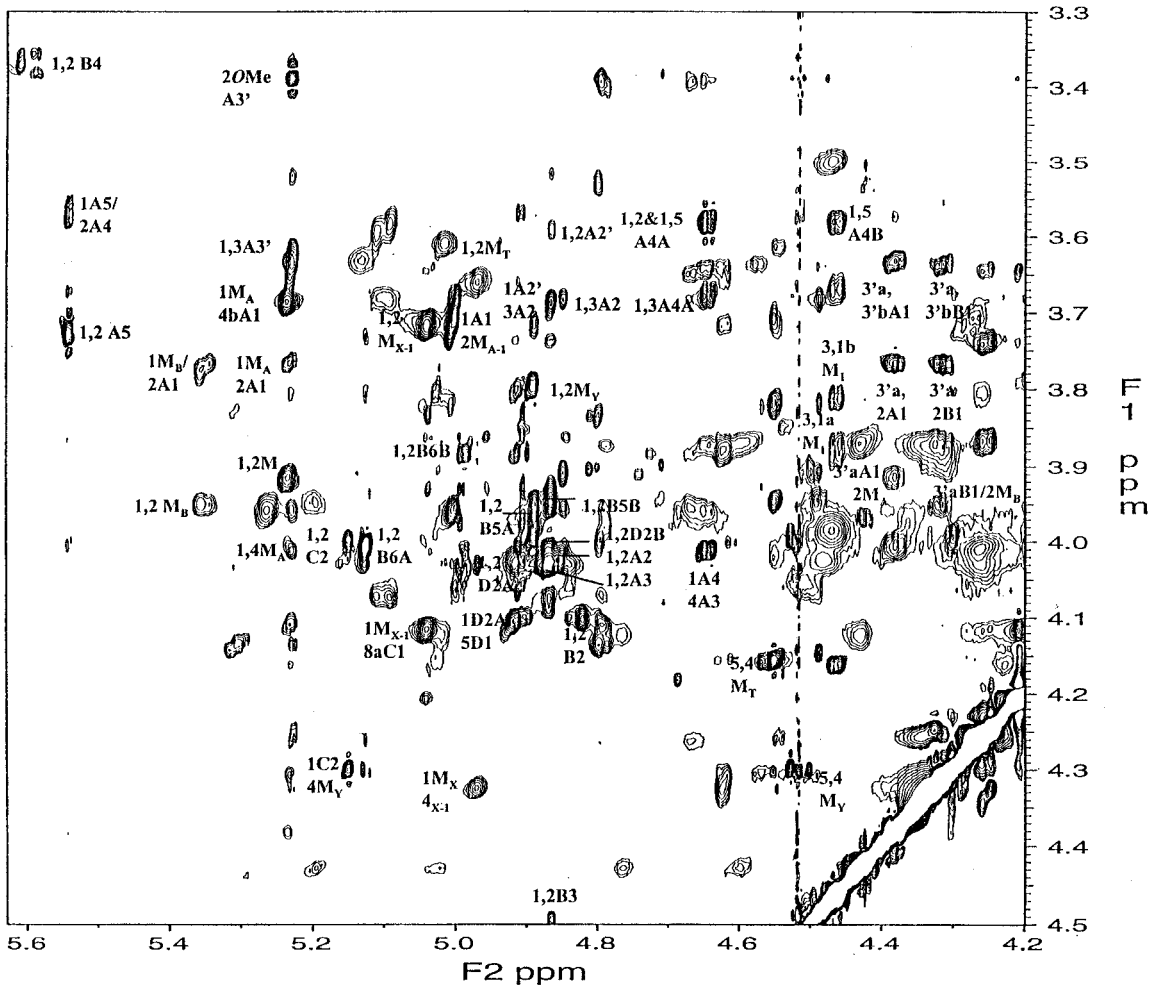


Figure 4 Anomeric region of the 750 MHz NOESY spectrum of mRG-II-ol in D_2O at 40 °C containing the crosspeaks with methine and methylene protons. Many of the more intense nOes between assigned protons have been labeled.

spin systems observed in the 2D correlation spectra (>90%) have been accounted for in Tables 1 and 2. It is to be noted that many more spin systems have been detected than would be expected on the basis of the structure in Scheme 1. As regards the A sidechain, partial substitution at C-3(A5) would appear highly likely considering the dual shifts detected for H-3(A5), C-3(A5), and H-1(A4). Similarly, some heterogeneity must exist in the substitution pattern of the B sidechain as at least two or more sets of proton chemical shifts were observed for B4, B5, and B6. Indeed, it has been shown by FAB-MS analyses that RG-II from red wine is partially branched at the 2-position of B6 by β-L-Ara*f* (B7) (Pellerin et al., 1996) while B5 is partially substituted at the 3-position by α-L-Rha*p* (B5′) (Shin et al., 1997). Finally, two sets of H-3_{ax}/H-3_{eq} signals were detected for Kdo*p* and several of the sets of resonances of the galacturonan backbone appeared in pairs with only a slight shift in the anomeric proton frequency, undoubtedly translating either the heterogeneity in the sidechains or variability in the backbone (length or branching). Unfortunately, four of the key residues (aceric acid, apiose, and one of the β-rhamnose units) did not exhibit H-1/H-2 crosspeaks in the 2D correlation spectra and these H-2 assignments therefore rely on correct interpretation of the nOe data. The H-5–H-6a,H-6b fragments of the galactosyl sugars have also eluded assignment but it could be seen from the HMQC experiment that the chemical shift range of the methylene protons was fairly restricted (3.40–3.75 ppm). Finally, several correlations for sev-

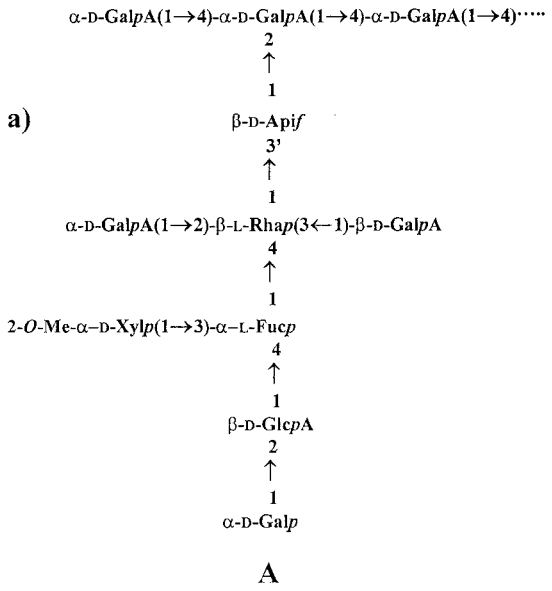
eral α-Gal*p*A H-5 signals were missing in the HMQC spectrum.

Sequential resonance assignments

The nOes of the anomeric protons have been collected in Table 3 along with those of various reporter groups (i.e, the deoxymethyl groups and the branchpoint protons with signals in non-overlapping regions). These interactions have been classified as strong, medium or weak according to their relative intensity with respect to the strongest intraresidue nOe (Figure 4). This internal ruler was generally the H-1/H-2 nOe (α-anomers) which corresponds to an interproton distance of ∼2.5 Å (the H-1/H-3 nOe corresponds to a similar distance in the case of A2″ and A4). As regards the A-chain, strong sequential nOes were detected for the A5-A4A-A3-A2 segment (H-1(A5)/H-2(A4), H-1(A4A)/H-4(A3), H-5(A3)/CH_3(A2), H-1(A3)/CH_3(A2)) which corroborated the chemical shift assignments and confirmed the primary structure indicated in Scheme 1. Contacts were not observed between the A3 and A4B protons so that this β-Glc*p*A residue may be located elsewhere. Assignment of the A1 apiosyl signals was based on long range interactions between CH_3(A3) and H-3′a(A1) and between H-1 (A1) and two protons belonging to two different terminal α-Gal*p*A residues (H-1(A1)/H-5(M_T) and H-1(A1)/H-1(A2′)) whereas that of M_A was based on two interglycosidic nOes (H-2(A1)/H-2(M_A) and CH3(A2)/H-1(M_A)). Inspection of the structure in Scheme 1 shows that the only unsubstituted α-Gal*p*A units were A2′ and the terminal sugar of the mainchain, M_T. Therefore, these latter nOes indicated that the A-chain was linked to the second last residue of the galacturonan mainchain ($M_A = M_{T-1}$). Furthermore, the strong nOes between H-1(A1) and the H-2 of another monosubstituted α-Gal*p*A sugar indicated that the M_{T-2} residue was not a branchpoint. The M_T – M_A – M_{T-2} ... fragment has been described recently in a study of the location of borate esters in the RG-II-borate complex (Ishii et al., 1998). The rank of the MM3 minima (the first being the global minimum) along with the corresponding Φ,Ψ values (Mazeau and Perez, 1998) which best reproduced the nOes for each disaccharide segment have also been given in Table 3. The interglycosidic interactions of the A-chain are all compatible with one or more of the low-energy minima of the MM3 isoenergy contour maps, suggesting that the so-called *independent residue approximation* is a very reasonable approach to carbohydrate structure even in the case of molecules as complex as the RG-II mega-oligosaccharide.

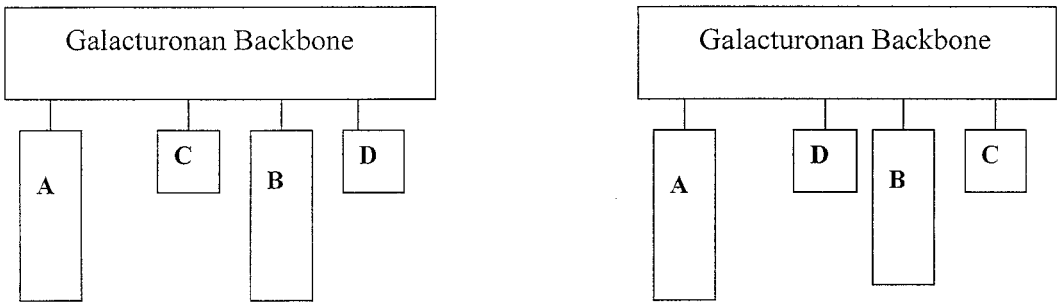
As regards the B-chain, assignment of B7 was based on interglycosidic long range heteronuclear scalar coupling between H-1(B7) and C-2(B6) and an interresidue nOe between H-1(B6) and H-2(B7) while interresidue contacts (H-1(B5B)/H-3(B4A)&H4(B4A)) corroborated the B5-B4 connectivity. Both the fucosyl (B4′) and the galactosyl (B4) anomeric protons exhibited nOes with the B4′-2-*O*-Me pendant group in agreement with 2-substitution of B4 by B4′. A long range nOe between CH_3(B2) and H-1(B3) corroborated the assignments of the B2 and B3 sugars. Finally, a weak nOe between H-1(M_B) and CH_3(B2) was in agreement with the assignment of the α-Gal*p*A branchpoint residue. However, sequential interactions were not detected for either the B6-B5 or the B4-B3 linkages and due to strong overlapping in the 4.0–4.1 ppm region it was not possible to unambiguously identify a nOe between H-1 (B3) and H-3(B2). Clearly, the sequence of the B-chain was not as well defined by interglycosidic nOes as the A-chain. Examination of certain low-energy conformers of the MM3 energy maps for the disaccharides which constitute the B-chain (in particular, the B6-B5, B4-B3, and B3-B2 maps) reveals that none of the interglycosidic distances are shorter than 3.5–4.0 Å. Thus, some of the preferred conformers would be expected to exhibit very weak (or undetectable) interglycosidic nOes. The added branching (particularly when it concerns 2-substitution or 2,3-disubstitution) must generate considerable steric hindrance as it is accompanied by a significant change in the conformational preferences of the nearby glycosidic linkages (chemical shifts and coupling constants). As a result the RG-II B-chain contains at least four major components (B7-B6-(B5′)-B5-B4-B3-B2-B1, B7-B6-B5-B4-B3-B2-B1, B6-(B5′)-B5-B4-B3-B2-B1 and B6-B5-B4-B3-B2-B1) and the concomitant decrease in the concentration of each species has undoubtedly hampered detection of weaker nOes.

Interresidue nOes were observed between H-3_{eq}(Cl) and H-1(C2A) (or H-2(C2B)) and H-5(D1) and H-1(D2A). However, interactions between the branchpoint residues C1 or D1 and the galacturonan backbone were not detected, undoubtedly due to the presence of quaternary carbons at these linkages. As for the longer chains, MM3 minimas were compatible with the experimental nOe data for the C1-C2 and the D1-D2 connectivity. Finally, only one of the linkages of the galacturonan mainchain exhibited a strong



c) $\cdots M_X - M_{X-1} \cdots$

b) $\cdots M_C - M_Y - M_B - M_D \cdots$ or $\cdots M_D - M_B - M_Y - M_C \cdots$



Scheme 2 Top: Schematic drawing of (a) the 3-residue block, (b) the 4-residue block, and (c) the 2-residue block of RG-II containing the branchpoints and/or the galacturonan backbone characterized by NMR. *Bottom*: Cartoons of the two possible arrangements of the RG-II building blocks characterized by NMR.

H-1/H-4′ interresidue nOe (M_X, δ_{H-1} 4.97 ppm with M_{X-1}, $\delta_{H-4'}$ 4.32 ppm) suggesting that the two-fold helical arrangement which engenders short (~2.2 Å) H-1/H-4′ distances (Cros et al., 1992) does not represent the majority of the conformations adopted by the RG-II backbone residues.

Several other low energy conformers which do not display any short (<3 Å) interresidue contacts can be adopted by homogalacturonan chains (Cros et al., 1992).

Long range nOes

Several strong long range nOes were detected between protons belonging to sidechain residues and those of the galacturonan backbone. Contacts between H-1(C2) and both H-1 and H-4 of an αGal*p*A residue, M_Y (δ_{H-1} 4.89 and δ_{H-4} 4.30 ppm, respectively) revealed that the C-chain was folded back onto the mainchain. The anomeric proton of M_Y was also in the vicinity of the CH_3(B2) moiety and this latter 6-deoxymethyl group also exhibited a moderately strong nOe with H-4(D1). This succession of long range contacts pointed to either of the following 4-residue

blocks, ...$M_C - M_Y - M_B - M_D$ ... or ... $M_D - M_B - M_Y - M_C$....

Furthermore, the M_{X-1} residue of the galacturonan backbone displayed strong interactions with the protons of the fragment that has been tentatively identified as the Kdo pendant group which would allow extension of the 4-residue segment to a 6-unit one. However, convincing evidence for the assignment of the C1 H-7-H-8a,H-8b fragment is lacking.

In summary, the 750 MHz NOESY data have revealed the primary structure of three different segments of the RG-II backbone: the terminal end ($M_T - M_A$ - M_{T-2} ...), a 4-residue block (... $M_C - M_Y - M_B - M_D$ ... or ... $M_D - M_B - M_Y - M_C$...) and a 2-unit fragment (... $M_X - M_{X-1}$ ...) (Scheme 2, top). It has not been possible to unambiguously establish the primary sequence for the reduced end (galactonic acid terminal) conceivably because of chemical exchange between the free acid and the aldono-1,4-lactone forms. In keeping with the 9-residue backbone which has been characterized by mass spectrometry (O'Neill et al., 1996), the primary structure of RG-II should therefore be reasonably close to one of the cartoons depicted in Scheme 2 (bottom). It is to be noted that the majority of the resonance frequencies of the sidechain and mainchain protons participating in the key long range interactions were unique (H-1(A1), CH_3(B2), and H-1(C2)). In the two cases of overlapping signals (H1(M_Y) overlaps with H-1(B5A) and H-4(D1) resonates near the same frequency as H-3(D2B)) the nOe data was easily interpreted in terms of the primary structure established previously (O'Neill et al., 1995).

Conclusions

The vast majority of the resonances in the 750 MHz proton spectrum corresponding to the 29-mer has been identified corroborating the partial primary structure established previously by methylation analysis, partial hydrolysis and FAB-MS spectrometry of the resulting oligomers. Sequential nOes have largely confirmed the structure of the A, G, and D sidechains and to a lesser extent that of the B branch. A single interglycosidic nOe has been detected for the galacturonan backbone and the most probable sequence has been determined through long range interactions between sidechain and mainchain residues (Scheme 2). Incomplete substitution at several sugars has led to multiple signals in all the 2D correlation spectra with concomitant loss in both the resolution and intensity of the crosspeaks. The use of ^{13}C labeling to overcome the sensitivity problem in the heteronuclear experiments is not feasible due to the large number of genes involved in the biosynthesis of RG-II. It is possible that future NMR studies of RG-II oligomers will reveal inversions in some of the assignments (particularly in the case of the *galacto* sugars) as analysis of such a large system is handicapped by overlapping sets of signals even at 750 MHz. Moreover, secure assignments of smaller systems may eventually allow a more complete interpretation of the present data.

Even if it was not possible to unambiguously establish a unique primary structure for RG-II, this investigation has dramatically reduced the number of structures that will have to be considered in future modeling studies. A comprehensive molecular modeling study will be required to generate low energy RG-II conformers for both types of the primary structures in Scheme 2 and to determine which of these sequences is the more energetically favored (i.e, obtain all the desired contacts with interproton distances of $\sim$3 Å without creating any serious steric conflicts). Then, it should be possible to distinguish these two solutions through back-calculation of theoretical NOESY volumes. Optimization of 3D models under nOe constraints is underway in our laboratory and this in turn should help complete the NMR assignments as it should be possible to more fully interpret the nOe data. Construction of 3D models that have been validated by NMR data should reveal the nature of the apiosyl/borate ester complex and ultimately lead to models of the cross-linked dimer. Furthermore, RG-II contains many potential chelating sites and theoretical studies of its interactions with various cations may ultimately inspire the synthesis of possible mimics with selective binding sites.

Acknowledgements

This study made use of a Varian UNITYPLUS-500 NMR spectrometer at the CGRM in Grenoble, and the Varian INOVA-750 NMR spectrometer at the European SON NMR Large-Scale Facility in Utrecht, The Netherlands. The latter work benefited from financial support from EC program Training and Mobility in Research (Contract ERBFMGECT-950032). The authors would like to thank Dr. Michael Czisch for running the 750 MHz NMR experiments, Stephane Vidal for his participation in the preparation of the

Rg-II sample, Vladan Krstic for his participation in preliminary work, and Dr. Alain Heyraud for his advice concerning the chromatography of RG-II. Fruitful discussions with Dr. Karim Mazeau are also acknowledged.

References

Albersheim, P., An, J., Freshour, G., Fuller, M.S., Guillen, R., Ham, K.-S., Hahn, M.G., Huang, J., O'Neill, M., Whitcombe, A., Williams, M.V., York, W.S. and Darvill, A.G. (1994) *Biochem. Soc. Trans.*, **22**, 374–378.

Allinger, N.L., Rhaman, M. and Lii, J.-H. (1990) *J. Am. Chem. Soc.*, **112**, 120–140.

Angelotti, T., Krisko, M., O'Connor, T. and Serianni, A.S. (1987) *J. Am. Chem. Soc.*, **109**, 4464–4472.

Aspinall, G.O. (1982) In *Molecular Biology, An International Series of Monographs and Textbooks, The Polysaccharides*, Vol. 1, Academic Press Inc., London, pp. 1–34.

Backman, I., Jansson, P.E. and Kenne, L. (1990) *J. Chem. Soc., Perkin Trans I*, 1383–1388.

Baumann, H., Jansson, P.-E. and Kenne, L. (1991) *J. Chem. Soc., Perkin Trans I*, 2229–2232.

Bax, A., Griffey, R.H. and Hawkins, B.L. (1983) *J. Magn. Reson.*, **55**, 301–315.

Bock, K., Josephson, S. and Bundle, D.R. (1982) *J. Chem. Soc., Perkin Trans II*, 59–70.

Bock, K. and Pedersen, C. (1983) *Adv. Carbohydr. Chem. Biochem.*, **41**, 27–66.

Bock, K. and Thogersen, H. (1982) *Annu. Rep. NMR Spectrosc.*, **13**, 2–57.

Bourlard, T., Pellerin, P. and Morvan, C. (1997) *Plant Physiol. Biochem.*, **35**, 623–629.

Castro, V.H., Ramirez, E., Mora, G.A., Iwase, Y., Tsuneatsu, N., Okabe, H., Matsunaga, H., Katano, M. and Mori, M. (1997) *Chem. Pharm. Bull.*, **45**, 349–358.

Cros, S., Herve du Penhoat, C., Bouchemal, N., Imberty, A. and Perez, S. (1992) *Int. J. Biol. Macromol.*, **14**, 313–320.

Darvill, A.G., McNeil, M. and Albersheim, P. (1978) *Plant Physiol.*, **62**, 418–422.

De Bruyn, A., Anteunis, M., De Gussem, R. and Dutton, G.G.S. (1975) *Carbohydr. Res.*, **19**, C9–C11.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.

De Marco, A., Gariboldi, P., Molinari, H. and Verotta, L. (1992) *Carbohydr. Res.*, **226**, 15–27.

Drake, L.R. and Rayson, G.D. (1996) *Anal. Chem.*, **69**, 22A–27A.

Engelsen, S.B., Cros, S., Mackie, W. and Perez, S. (1996) *Biopolymers*, **39**, 417–433.

Fukuoka, S., Knirel, Y.A., Lindner, B., Moll, H., Seydel, U. and Zahringer, U. (1997) *Eur. J. Biochem.*, **250**, 55–62.

Griesinger, C., Otting, G., Wüthrich, K. and Ernst, R.R. (1988) *J. Am. Chem. Soc.*, **110**, 7870–7872.

Hofinger, A., Kosma, P., Christian, R., Bock, K. and Brade, H. (1993) *Carbohydr. Res.*, **243**, 273–291.

Hricovini, M., Bystricky, S. and Malovikova, A. (1991) *Carbohydr. Res.*, **243**, 23–31.

Ishii, T. and Matsunaga, T. (1996) *Carbohydr. Res.*, **284**, 1–9.

Ishii, T., Matsunaga, T., Shimokawa, T., O'Neill, M., Darvill, A. and Albersheim, P., 8th International Cell Wall Meeting, Norwich, 1998.

Jackson, P.L., Torres, A.T., Delhaize, E., Pack, E. and Bolender, S.L. (1990) *J. Environ. Qual.*, **19**, 644–648.

Jeener, J., Meier, B.H., Bachmann, P. and Ernst, R.R. (1979) *J. Chem. Phys.*, **71**, 4546–4553.

Joao, H.I., Jackson, G.E., Ravenscroft, N. and Stephen, A.M. (1988) *Carbohydr. Res.*, **176**, 300–305.

Kobayashi, M., Matoh, T. and Azuma, J.-I. (1996) *Plant Physiol.*, **110**, 1017–1020.

Koerner, T.A.W., Prestegard, J.H. and Yu, R.K. (1987) *Methods Enzymol.*, **138**, 38–59.

Macura, S. and Ernst, R.R. (1980) *Mol. Phys.*, **41**, 95–117.

Matoh, T., Ishigaki, K.-I., Kaori, O. and Azuma, J.-I. (1993) *Plant Cell Physiol.*, **34**, 639–642.

Mazeau, K. and Perez, S. (1998) *Carbohydr. Res.*, **311**, 203–217.

Mbairaroua, O., Ton-That, T. and Tapiero, C. (1994) *Carbohydr. Res.*, **253**, 79–99.

Melton, L.D., McNeil, M., Darvill, A.G., Albersheim, P. and Dell, A. (1986) *Carbohydr. Res.*, **146**, 279–305.

Nakahara, Y. and Ogawa, T. (1987) *Carbohydr. Res.*, **167**, C1–C7.

O'Neill, M.A., Warrenfeltz, D., Kates, K., Pellerin, P., Doco, T., Darvill, A.G. and Albersheim, P. (1996) *J. Biol. Chem.*, **271**, 22923–22930; ibid **272**, 3869.

O'Neill, M., Albersheim, P. and Darvill, A.G. (1990) *Methods Plant Biochem.*, **2**, 415–441.

Pellerin, P., Doco, T., Vidal, S., Williams, P., Brillouet, J.-M. and O'Neill, M.A. (1996) *Carbohydr. Res.*, **290**, 183–197.

Peters, T. and Weimar, T. (1994) *J. Biomol. NMR*, **4**, 97–116.

Piantini, U., Sørensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R. and Wüthrich, K. (1983) *J. Am. Chem. Soc.*, **104**, 6800–6801.

Rance, M., Sørensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R. and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.*, **117**, 479–485.

Rinaudo, M. and Ravanat, G. (1980) *Makromol. Chem.*, **181**, 1059–1070.

Salt, D.E., Blaylock, M., Kumar, N.P.B.A., Dushenkov, V., Ensley, B.D., Chet, I. and Rashin, I. (1995) *Biotechnology*, **13**, 468–474.

Serianni, A.S. and Baker, R. (1979) *Can. J. Chem.*, **57**, 3160–3167.

Shin, K.-S., Kiyohara, H., Matsumoto, T. and Yamada, H. (1997) *Carbohydr. Res.*, **300**, 239–249.

Skelton, M.A., Cherniak, R., Poppe, L. and van Halbeek, H. (1991) *Magn. Reson. Chem.*, **29**, 786–793.

Spellman, M.W., McNeil, M., Darvill, A.G., Albersheim, P. and Henrick, K. (1983) *Carbohydr. Res.*, **122**, 115–129.

States, D.J., Haberkorn, R.A. and Ruben, D.J. (1982) *J. Magn. Reson.*, **48**, 286–292.

Summers, M.F., Marzilli, L.G. and Bax, A. (1986) *J. Am. Chem. Soc.*, **108**, 4285–4294.

Titel, C., Erfurt, J. and Ehwald, R., 8th International Cell Wall Meeting, Norwich, 1998.

Toman, R., Rosik, J. and Alfoldi, J. (1986) *Carbohydr. Res.*, **158**, 236–244.

Whitcombe, A.J., O'Neill, M.A., Steffan, W., Albersheim, P. and Darvill, A.G. (1995) *Carbohydr. Res.*, **271**, 15–29.

York, W.S., van Halbeek, H., Darvill, A.G. and Albersheim, P. (1990) *Carbohydr. Res.*, **200**, 9–31.