Available online at www.sciencedirect.com
SCIENCE DIRECT®
Carbohydrate Research 338 (2003) 651–671
CARBOHYDRATE RESEARCH
www.elsevier.com/locate/carres

The three-dimensional structure of the mega-oligosaccharide rhamnogalacturonan II monomer: a combined molecular modeling and NMR investigation

Miguel A. Rodríguez-Carvajal,[a] Catherine Hervé du Penhoat,[a] Karim Mazeau,[a] Thierry Doco,[b] Serge Pérez[a,*]

[a] *Centre de Recherches sur les Macromolécules Végétales, CNRS (associated with University Joseph Fourier), BP 53, 38041 Grenoble Cedex 9, France*

[b] *Unité Mixte de Recherches Sciences Pour l' Œnologie, Laboratoire des Biopolymères, 2 Place Viala, 34060 Montpellier, France*

Received 2 October 2002; accepted 17 December 2002

Abstract

In this study, we describe the first optimized molecular models of the mega-oligosaccharide rhamnogalaturonan II, that is found in the primary cell walls of all higher plants. The 750 MHz ^{1}H NMR data previously reported and new heteronuclear correlation spectra (sensitivity-enhanced HSQC and HSQC–TOCSY) were first reassigned in light of the modifications in the primary structure. In turn, the experimental NMR data revealed the presence of an additional sugar, α-Ara*f* (E-chain), and also the disaccharidic repeating unit of RG-I, another component of the pectic matrix. Due to a fuller picture of the primary structure of RG-II, a much more complete assignment of the NOE data has been achieved. A systematic computational study based on these NOEs lead us to a realistic three-dimensional description of the RG-II, in excellent agreement with the molecular dimensions obtained from various experimental methods.

Keywords: Rhamnogalacturonan II; Conformation analysis; Molecular modeling; NMR spectroscopy

1. Introduction

Rhamnogalacturonan II (RG-II), a 30-mer mega-oligosaccharide that is present in the primary cell walls of all higher plants, has been the target of numerous structural studies over the past two decades.[1] It is by far the most complex polysaccharide in the plant kingdom with more than 18 different types of glycosidic linkages indicating that its biosynthesis necessarily requires a very impressive number of genes. One of its biological roles is thought to be control of cell supply of boron[2] as it forms a dimer cross-linked by a borate diester in the presence of this essential microelement.[3] Recently, physiological effects of boron-deficiency have been correlated to RG-II/borate diester formation in pumpkin tissue.[4]

Current knowledge of the primary structure of RG-II is largely based on two approaches: (1) glycosyl linkage composition (determined by methylation analysis); and (2) identification of various oligosaccharide fragments with mass spectrometry (obtained by mild acid and/or enzymatic hydrolysis of RG-II followed by chromatographic purification). RG-II is best described as a homogalacturonan backbone of roughly seven to nine residues bearing four oligosaccharide sidechains, A–D, Scheme 1.[5] Some of the backbone galacturonic acid residues are methyl-esterified and the B-chain aceric acid (**B3**) and 2-*O*-Me-Fuc*p* residues (**B4′**) are acetylated.[6,7] Recently, Vidal and co-workers[8] have published a convincing NMR investigation of an oligosaccharide fragment of RG-II from red wine that contains a hexasaccharide mainchain bearing the B and D sidechains. Most of the proton assignments were reported in this study and it has been demonstrated that the original configurations established for the anomeric centers of two residues in the B-chain, **B3** and

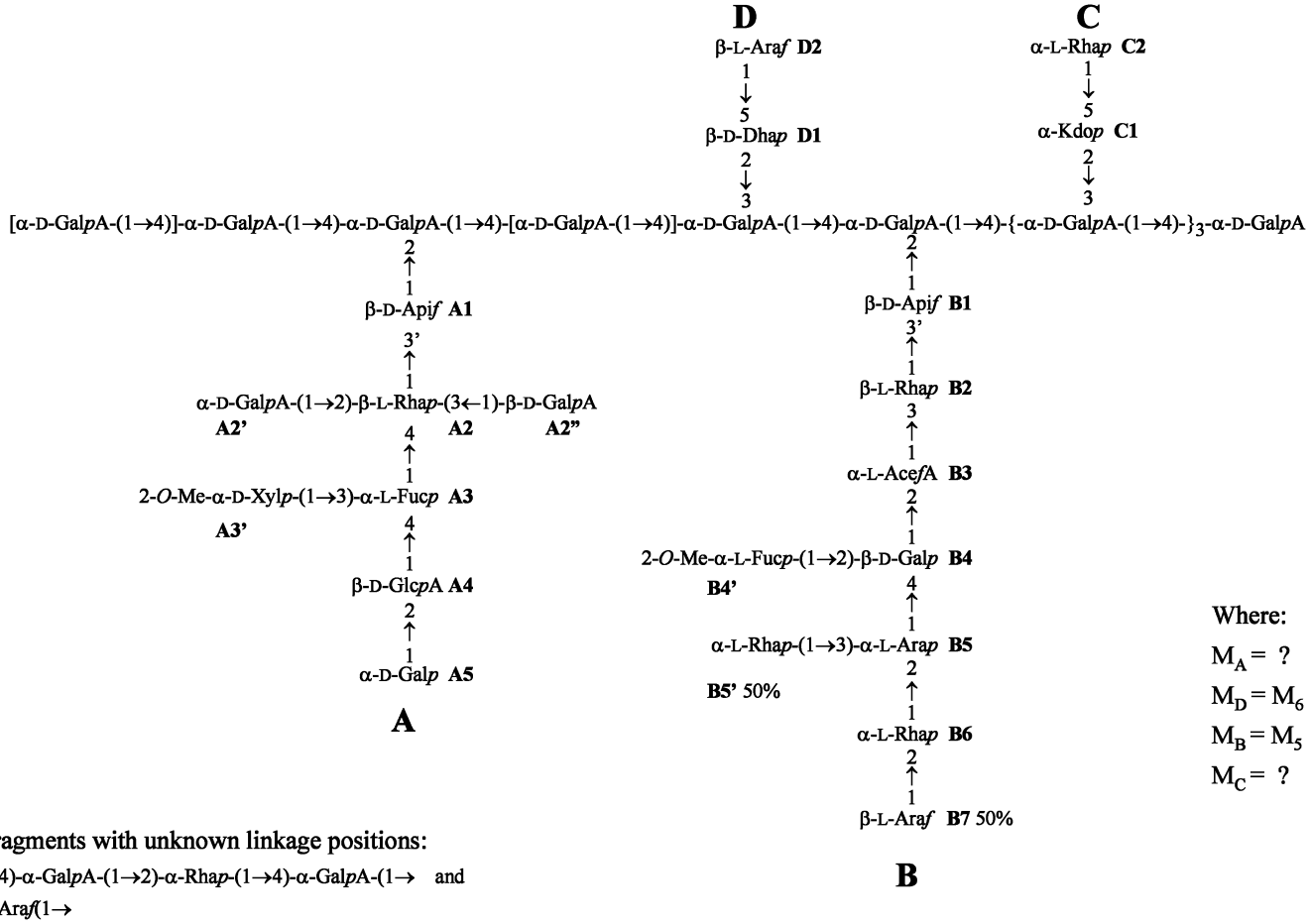
* Corresponding author. Fax: +33-4-76037629
E-mail address: serge.perez@cermav.cnrs.fr (S. Pérez).

0008-6215/03/$ - see front matter
doi:10.1016/S0008-6215(03)00003-X

B4, were incorrect. Moreover, they showed that the B- and D-sidechains were located at the fifth ($\mathbf{M_B} = \mathbf{M_5}$) and sixth ($\mathbf{M_D} = \mathbf{M_6}$) α-Gal*p*A residue from the reducing end, respectively. The Kdo-containing disaccharide C-sidechain was detected in non-stoichiometric amounts in this oligosaccharide fragment and, consequently, it was removed to increase homogeneity and facilitate analysis with the result that the exact branch point of the C-chain is still unknown. Finally, some heterogeneity has been reported for RG-II from various sources including: further substitution by rhamnosyl (3-linked to **B5**)[9,10] and arabinosyl sugars (3-linked to the galacturonan backbone[11]) and incomplete substitution by several of the terminal sugars (**A5**, **B5′**, and **B7**).

The first molecular mechanics study[12] of the four oligosaccharide sidechains of RG-II explored the stable orientations of all of the glycosidic linkages of the sidechain sugars as well as the puckering preferences of the apiosyl and aceric acid five-membered rings. The preferred conformers were partially validated by the sequential nuclear Overhauser effects (NOEs) described in the first NMR investigation[13] of the intact saponified and borohydride-reduced RG-II monomer, mRG-II-ol. This latter study placed the A-sidechain at the penultimate α-Gal*p*A residue from the non-reducing end and revealed some long-range contacts. However, spectral assignments relied heavily on the available primary structure which has been modified since.[8] These events prompted us to undertake a molecular modeling investigation of the complete RG-II mega-oligosaccharide. Concomitantly, better quality proton-carbon chemical shift correlation data would be acquired to allow more secure assignment of many of the spin systems.

Our goal was to produce the first conformational models of RG-II monomer consistent with the 750 MHz NOESY data previously reported.[13] The long-range interactions and the exact position of the A-, B-, and D-sidechains were only expected to be compatible with a small number of low-energy model structures. Using an iterative combined NMR and molecular modeling approach, it might be possible to identify some of the additional sugars demonstrated in the previous NMR study and to obtain a much more complete assignment of the NOE crosspeaks. The stable conformers of the RG-II monomer would subsequently constitute the starting point for an investigation of the biologically-relevant borate diester RG-II dimer.



Scheme 1. Current primary structure of RG-II. The fragments enclosed in brackets may not exist.

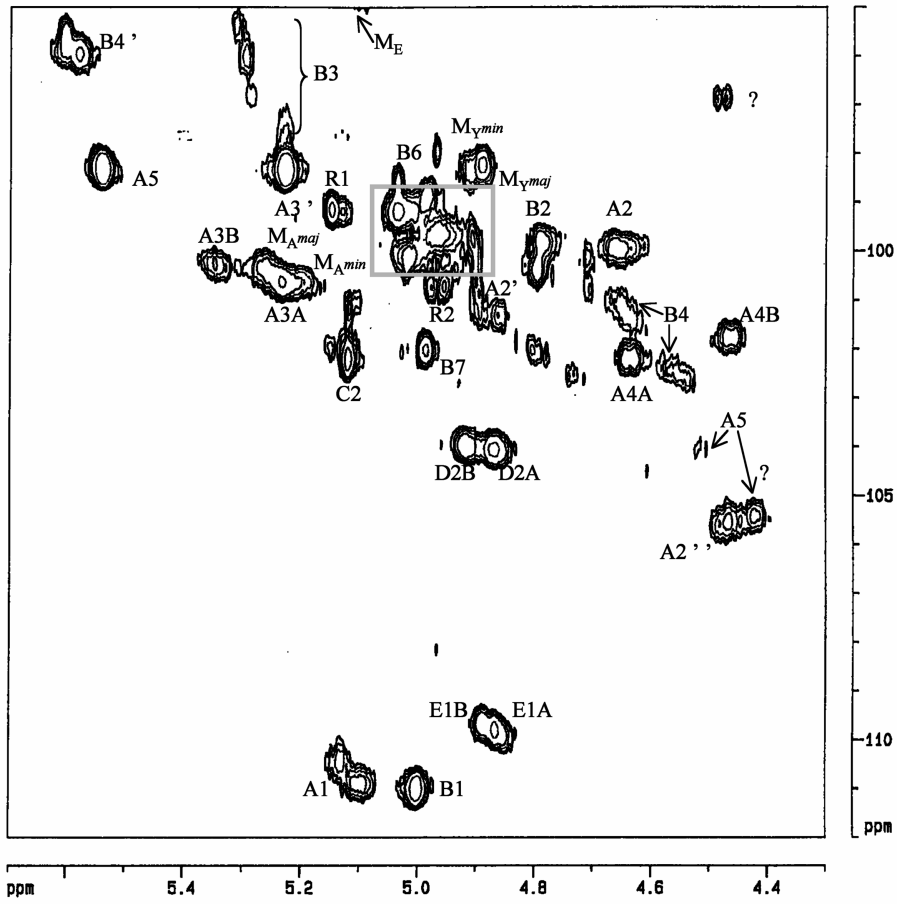


Fig. 1. Anomeric region of the sensitivity-enhanced HSQC spectrum. Crosspeaks have been labeled as in Scheme 1 with the exception of the heavily overlapping α-D-Gal*p*A signals which have been surrounded with a gray rectangle.

2. Results and discussion

The first step in this investigation was to reexamine the ^{1}H and ^{13}C data in light of the new primary structure of mRG-II-ol and the recently acquired sensitivity-enhanced HSQC (Fig. 1) and HSQC-TOCSY spectra. The reassigned NMR chemical shifts of the intact borohydride-reduced RG-II monomer are given in Tables 1 and 2 while the NOE data for the sidechain sugars have been reinterpreted in Table 3. Carbon signals that appear in strongly overlapping regions have been indicated as integer numbers while one decimal has been given for all the chemical shift values of isolated carbon resonances. As discussed below, taking into account the inversion of the anomeric centers of the **B3** and **B4** has resulted in a much more complete set of interresidue NOEs but it also revealed several incorrect assignments of the *galacto* (and related *arabino* and *fuco* spin systems) in our previous study. It should be noted that in spite of the much improved heteronuclear correlation data, some assignments are still solely based on literature data and/or ambiguous correlations in the 2D spectra and they have been given in italics.

2.1. A-chain

In our original NMR study of m-RGII-ol, unusual correlations between low and high field protons (4.38/3.63 and 4.31/3.635 ppm; $J \sim 10$ Hz) in the DQCOSY spectrum were assigned to the H-3′a(A1)/H-3′b(A1) methylene protons. However, crosspeaks in the HSQC-TOCSY spectrum revealed that the low and high field signals belonged to methine and methylene groups, respectively. These H-5(A5)/C-6(A5) (F2/F1) correlations allowed unambiguous assignment of these signals to the H-5–H-6a,b fragment of the **A5** residue. The surprisingly large value of $^{3}J_{5,6}$ indicates that the exocyclic group adopts a unique orientation, presumably the GT one.[14] This new assignment was corroborated by weak (strong) H-4/H-5 crosspeaks in the 750 MHz TOCSY (NOESY) spectrum. The NOESY spectrum also displayed strong crosspeaks between H-5(A5) and H-2 of both **A4** and a *galacto* or *fuco* sugar (for both the major and minor species). Inspection of the optimized molecular models of RG-II presented below indicated that H-2(A3) was the only suitable proton (H-5(A5)/H-2(A3) distance ~ 2.5 Å). Thus, assignment of the former H-5(A5)/H-2(A3) crosspeaks led to the

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

Residue	C-1 H-1	C-2 H-2	C-3 H-3	C-4 H-4	C-5 (C-3′) H-5 (H-3′)	C-6 H-6
A1	109.8	79	78.4	74.3 [a]	71	nap
→3′)-β-D-Api*f*-(1→2	5.09	4.075	nap	4.06, 3.68 [a]	3.945–3.97,	
	nc				3.41–3.45	
A2	100	77.2 and/or 81.4	76.2	76.0 [b]	68.2 [b]	17.6
		br				
2,3,4)-β-L-Rha*p*-(1→3′	4.65–4.67	4.26	3.99	3.69	3.685	1.285
		na	75.7	75.5[b]	3.72	1.30
			4.02	3.765		
A2′	99	71.8	68	71.0	72.7	na
α-D-Gal*p*A-(1→2	4.865	3.59	3.82	4.146	4.485	
A2″	105.0	77.3	73.5	na	na	na
β-D-Gal*p*A-(1→3	4.47	~3.55	3.50			
A3A	99.7	68	74.5	na		15.9
→3,4)-α-L-Fuc*p*-(1→4	5.23	3.91	3.63	4.01		1.43
A3B	99.5	68	74.5	75.6	67.4	na
	5.35	3.95	3.65	4.162	3.96	
A3′	97.55	80.76 [b]	71.8 [b]	70.3 [b]	61.2 [b]	nap
2-*O*-Me-α-D-Xyl*p*-(1→3	5.23	3.122	3.638	3.32	3.525, 3.665	
A4A	101.3	77.1 [b]	77.3 [b]	71.8	74.3 [b]	na
→2)-β-D-Glc*p*A-(1→4	4.64	3.576	3.67	3.495	3.58	
A4B	101.0				74.2	
	4.46	3.555	3.67	3.49	3.58	
A5A	97.5	68.4 [b]	na	68.7	69.8	60.7 [b]
α-D-Gal*p*-(1→2	5.54	3.726	4.005	3.765	4.38 [b]	3.63
A5B			3.87	3.765	4.31 [b]	3.634
B1	110.2	na	nap	74.6 [a]	71	nap
→3′)-β-D-Api*f*-(1→2	5.00			4.11, 3.69 [a]	3.96, 3.41	
B2	100.4	66–67 [b]	75.5	70.4	72	17.1
→3)-β-L-Rha*p*-(1→3′	4.80	4.11	~3.83	3.52	3.83	1.219
		4.135				
	99.8					
	4.795	4.11		3.53	3.65	1.285
		4.13				
		4.14				
B3A-D	94.8	85.8	82.87	78.94	13.8	nap
→2)-β-L-Ace*f*A-(1→3	5.32	4.55, 4.54		4.58	1.05	
	95.2	85.8		79.2		
	5.30	4.505, 4.49		4.605		
	96.0	86.4		79.4		
	5.29			4.62		
	96.6					
	5.29	4.44				
	96.6	86.8				
	5.23	4.325, 4.305,				
		87.4				
		4.26, 4.245				
B4A	100.8–100.6	76.5	73.5	77.0	na	61.6
→2,4)-β-D-Gal*p*-(1→2	4.665	3.57	3.885	3.852		3.47,
						3.775
B4B	4.645	3.57	3.87	nd		3.493,
						3.765
	4.62	3.57	3.86	3.925		
B4C	101.8	77	73.5	78.0		
	4.58	3.685	3.875	3.902		
B4D	4.57	3.665	3.87	3.91		

Table 1 *(Continued)*

Residue	C-1 H-1	C-2 H-2	C-3 H-3	C-4 H-4	C-5 (C-3′) H-5 (H-3′)	C-6 H-6
B4E	102.0	77		78.0		
	4.54	3.726	3.852	3.965		
B4A′	95.19	77.25	∼68	na	66.53	15.8
2-*O*-Me-α-L-Fuc*p*-(1→2	5.612	3.36	3.98		∼4.275	1.12
B4B′	95.02				∼4.275	15.8
	5.584	3.355	3.98, 4.055, 4.085			1.11
B5A [c]	104 [c]				66.4	nap
→2)-α-L-Ara*p*-(1→4	4.62	3.49, 3.47			3.51, 3.83	
B5′ [d]	100.0	70.2	69.7	72.0		16.65
α-L-Rha*p*-(1→3	4.975	3.865	3.62	3.353	3.68	1.18
or **B6** [d]	4.95	3.86				
α-L-Rha*p*-(1→2						
B6 [d]	98.5	78.0	69	70.2	70	16.65
→2)α-L-Rha*p*-(1→2	5.04	3.935	3.735	3.328	3.805	1.174
B7	101.2	78.5	72.5 [b]	81.4	60.7	nap
β-L-Ara*f*-(1→2	4.985	4.03, 4.035	4.098, 4.11	3.79	3.69, 3.59; 3.663, 3.598	
C1A			35.2	67.2	73.3	na
→5)-α-Kdo-(2→3	nap	nap	1.82, 2.02	4.01		
C1B			1.744, 2.025	4.03	4.006	
C2A	98.4	70				16.65
α-L-Rha*p*-(1→5	5.12	4.002	3.735	3.328	3.73	1.177
C2B	5.135					
D1			35.5	64.5	78.7	75.7
→5)-β-D-Dha*p*-(2→3	nap	nap	1.744, and 2.155	4.085	4.105, 4.035	4.82 (D1B)
				∼4.05		4.87 (D1A)
D2A	103.24	78.7	71.5	81.0	61.2	nap
β-L-Ara*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	
E1A	108.7	81.0	77.2	81.3	62.35	
α-Ara*f*-(1→	4.888	3.943	3.575	4.025	3.41, 3.74	
E1B	109.0	81.0	77.2	81.3		
	4.865	3.948	3.52	4.022		

[a] Key: s, singlet; d, doublet; t, triplet; w, weak; br, broad, n, narrow; na, not assigned; nc, not correlated; nap, not applicable; nd, not detected. Fine structure (i.e. $^{3}J_{H\text{-}1,H\text{-}2}$, $^{3}J_{H\text{-}2,H\text{-}3}$, $^{3}J_{H\text{-}3,H\text{-}4}$, $^{3}J_{H\text{-}4,H\text{-}5}$, $^{3}J_{H\text{-}5,H\text{-}6a}$, $^{3}J_{H\text{-}5,H\text{-}6b}$, $^{2}J_{H\text{-}6a,H\text{-}6b}$ 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): α-Ace*f*A m,nap,nap,l; α-Api*f*A vs,nap,nap, $^{2}J_{H\text{-}4a,H\text{-}4b}$ vl, $^{2}J_{H\text{-}3'a,H\text{-}3'b}$ vl; β-Ara*f* m,l,l,s,or; α-Ara*p* l,l,s,nd,s,vl; β-Dha*p* $^{2}J_{H\text{-}3ax,H\text{-}3eq}$ vl, $^{3}J_{H\text{-}3ax,H\text{-}4eq}$ l; $^{3}J_{H\text{-}3eq,H\text{-}4}$ m; β-Gal*p* l,l,m,nd; αGal*p*A m,l,m,nd; β-Gal*p*A l,l,m,nd; β-Glc*p*A l,l,l,l; α-Kdo*p* $^{2}J_{H\text{-}3ax,H\text{-}3eq}$ vl, $^{3}J_{H\text{-}3ax,H\text{-}4}$ l; $^{3}J_{H\text{-}3eq,H\text{-}4}$ m; α-Rha*p* vs,s,l,l,l; β-Rha*p* s,s,l,l,l; α-Xyl*p* m,l,l,l and s,vl. 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] HSQC–TOCSY spectrum.

[c] The following multiple partial spin systems with weak anomeric carbon signals resonating in the 104–105 ppm range likely also correspond to the **B5** →2)-α-L-Ara*p*-(1→4 sugar: 4.51, 3.40, 3.555; 4.485, 3.39, 3.64. The following partial coupling graph (H-1/C-1, H-2/C-2, H-3/C-3, H-4/C-4) of an α-Ara*p* or a β-*galacto* sugar (4.425/104.84, 3.53/71.5, 3.535/72.5, 3.81/68) that exhibits a strong NOE with the proton resonating at 3.965 ppm is also detected and has been tentatively assigned to terminal **B5** linked to **B4E**.

[d] 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: 4.99/99.4, 3.885/70.1, 3.735/69.7, 3.328/72.0, 3.805, 1.174/16.65; 5.115, 3.935, 3.69 and 3.665, 3.235, 3.83, 1.12; 4.92, 4.025 and 4.055 and 4.068, 3.795, 3.37, 3.72, 1.174; ?, 3.895, 3.695, 3.318, 3.565, 1.256 w. Several other very weak signals were detected in the anomeric region of the HSQC spectrum: 4.81/102.2, 4.74/102.8, 4.71/100.2 and 100.8, 4.48/96.2 (strong NOEs with the unusual methylene group resonating at 59.15/3.875, 3.81). The signals of the 'RG-I like' →4)-α-D-Gal*p*A-(1→2)-α-L-Rha*p*-(1→ fragment are as follows: α-D-Gal*p*A (major component) 4.89 (97.5), 3.79, 3.99, 4.298 (77.2), 4.525 (71.2); (minor component) 4.91 (97.5), 3.80, 3.99, 4.30 (76.2), 4.50 (71.2); α-L-Rha*p*-5.15/98.4,3.998/70, 3.775, 3.29, 3.69, 1.13/16.65.

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

Residue	C-1 H-1	C-2 H-2	C-3 H-3	C-4 H-4	C-5 H-5	Interresidue NOEs
M_1	nd	nd	nd	na		
→4)-Galactonic acid		4.72	4.465	4.40[a]		
Aldono-1,4-lactone	180	71.0	78.1	70.1		
Free acid		4.205	4.115	3.645	3.74 [b]	
M_2	98.3	68.2	68.9	77.62	71.0	H-1(M_2)/H-2 and H-3(M_1)
	5.04	3.72	3.865	4.325	4.25	
	na	3.73	3.87	4.34	4.26	H-5(M_2)/H-4′ and H-5(M_1)
M_3	98.8	na	na	77.7	71.3	H-1(M_3)/H-4(M_2)
	4.968 br	3.66	3.88	4.325	4.62	H-5(M_3)/H-2(M_2)
M_4 [c,d], **M_B (M_5)**, **M_D (M_6)**, **M_E**	*5.0–5.045* [d]	*na*	*3.94–3.98*		*72.3*	*H-1(M_4)/4.34* [e]
		3.82	*3.96*	*4.18*	*4.68na*	*H-1(M_4/H-1(B2) w* *H-1(M_4)/H-3′(B1) w*
		75	*68.4*	*78.0*	*71.4*	H-1(M_5)/H-1(B2) w H-1(M_5)/H-3′(B1) w
	5.04	na	4.115	4.43	4.76	H-1(M_5)/H-4(M_4)
	5.015		4.13			4.135–4.18 [e]
	99.9		*74.8*	*75.9*	*71.1*	*H-1(M_6)/H-4(M_5)*
	5.00–5.05 [e]	*~3.72*	*~3.87*	*4.43*	*4.60*	*4.43* [e]
	94			77.0		
	~5.10	3.685, 3.73	3.905	4.495, 4.505		
M_{A1}	99.6	73.5		78.8		H-1/4.505
→2,4)-α-D-Gal*p*A-(1→4	5.26 [f]	3.955	3.585na	4.485		H-1/H-1(A2) w H-1/H-3′(A1) H-3/H-4 intra
M_{A2}	100.7	73.5	na	na		H-1/4.43 [g]
→2,4)-α-D-Gal*p*A-(1→4	5.20 [g]	3.945	3.59	4.51		
M_T	98.4	68	68	71.2	71.2	H-1/4.485
α-D-Gal*p*A-(1→4	5.012	3.608	3.822	4.155	4.55	H-1/4.51
	4.94w	3.608				H-5/3.945

Key as in Table 1.

[a] Correlates with 3.42 (dd, l) and 3.34 (dd, l) in the TOCSY spectrum.

[b] $\delta_{C\text{-}6}$ 63.1 ppm; $\delta_{H\text{-}6a,b}$ 3.55 ppm.

[c] According to Vidal and co-workers, 2000 the fragment containing the first six galacturonic acid residues from the reducing end also includes the α-L-Rha*p*-(1→5)-α-Kdo-(2→3) C-chain.

[d] Multiple broad very weak crosspeaks in the 2D homonuclear spectra.

[e] Overlapping region H-1(M_n)/H-4(M_{n-1})—$\delta_{H\text{-}4(Mn\text{-}1)}$ 4.43, 4.34, and 4.12–4.18 ppm.

[f] Major component.

[g] Two sets of H-1/H-2 crosspeaks at 5.20 ppm. The partial spin systems of other α-D-Gal*p*A fragments are as follows: 5.15 (99.2), 3.775, overlapping at 4 ppm, 4.30; H-4—4.495 (overlapping), H-5—5.13 (70.2).

reassignment of the H-1–H-2–H-3–H-4 coupling graph of the **A3** residue.

The RG-II mega-oligosaccharide contains at least four different rhamnose sugars (**A2**, **B2**, **B6** and **C2**) but due to sample heterogeneity an even larger number of these units had been detected in the NMR spectra. Correct assignment of the key **A2** and **B2** sugars was mandatory for conformational analysis as the exocyclic 6-deoxymethyl groups exhibit numerous long-range contacts with neighboring residues. From both HMBC (1.285/75.6 ppm) and HSQC–TOCSY experiments long range 6-CH_3/C-4 correlations (1.285/75.6 and 1.30/75.6 ppm) were detected for the **A2** β-Rha*p* residue. The low-field glycosylation shifts of the C-4 signals confirmed the assignments of the **A2A** (major species) methyl group and indicated that the signal at 1.30 ppm belonged to the 6-CH_3 of the minor component (**A2B**). However, two H-5 signals were observed for the methyl signal at 1.285 ppm in the DQCOSY spectrum. Perusal of the rapidly increasing body of chemical shift data for diversely-substituted rhamnose residues[15–21] has revealed characteristic chemical shift ranges for the H-4 signal of 2- or 3- (3.24–3.53 ppm—mono- and 2,3-disubstituted rhamnosides) as opposed

Table 3 List of assigned NOES obtained from the 750 MHz NOESY-spectrum (mixing time of 200 ms) at 313 K

Residue	H-1 NOEs	Other Reporter Groups	Reporter Group NOEs
A1 →3′)-β-D-Api*f*-(1→2	H-2(A1) S, H-3(M_A) br S, H-4a(A1) w	H-3′a H-3′b	H-3′b(A1) S, H-1(A2) br S H-1(A2) br S, H-1(M_A) vw
A2A+B →2,3,4)-β-L-Rha*p*-(1→3′	H-2(A2) M (br), H-3′a(A1)S, H-3′b(A1) S, H-1(M_A) w	CH_3(A2A) CH_3(A2B)	H-5(A2A) S, H-5(A3) S, 4.87 M, H-2(A2) vw, 4.85 w, H-1(A3) vw H-5(A2B) S, H-5(A3) S, 3.895 S, 4.89 M, H-6(D1B) w, H-1(A3) vw
A2′ α-D-Gal*p*A-(1→2	H-2(A2′) M, H-4(A2) M	H-5(A2′)	H-4(A2′) S, H-3(A2′) S, H-4(A2) M
A2″ β-D-Gal*p*A-(1→3	H-3(A2″) S, H-3(A2A and B) S		
A3A →3,4)-α-L-Fuc*p*-(1→4	H-2(A3A) S, H-4(A2A) S, H-5(A5B) M CH_3(A2A) w, CH_3(A3) vw	CH_3	H-1(A4A) S, H-3(A4A) M, H-5(A4A) M, H-5(A3A) S, H-4(A3A) S, 4.87 w, H-5(A5A) vw
A3B	H-2(A3B) S, H-4(A2B) S, CH_3(A2B) M, H-5(A5B) M		
A3′ 2-O-Me-α-D-Xyl*p*-(1→3	H-3(A3A+B) S, H-2(A3′) S, H-4(A3A+B) w, OMe(A3′) S	OMe	H-1(A3′) S, H-2(A3′) S
A4A →2)-β-D-Glc*p*A-(1→4	H-4(A3B) S, H-3(A4A) S, H-5(A4A) S, CH_3(A3) S, H-1(A5) vw		
A4B	H-4(A3B) S, H-3(A4B) S, H-5(A4B) S; H-1(A5) vw, H-1(A3′) vw		
A5 α-D-Gal*p*-(1→2	H-2(A5) S, H-2(A4A+B) S, H-3(A4A+B) w	H-5A+B H-5A	H-3(A5A+B) M, H-4(A5) S, H-6(A5) S, H-1(A5) w, H-2(A3A+B) S, H-1(A3A+B) w, H-2(A4A+B) M CH_3(A3) w
B1 →3′)-β-D-Api*f*-(1→2	H-2(B1) S, 3.68–3.73 and 3.82 S ⇒ H-2(M_B) S 4.55 w, 4.865 w	H-3′a(B1) H-3′b(B1)	H-3′b(B1) S H-3′a(B1) S, H-1(B2) S, H-1(M_B) vw
B2A-D →3)-β-L-Rha*p*-(1→3′	H-2(B2A-D) S, H-3′b(B1) S, H-3′a(B1) w, H-1(M_B) vw	CH_3	H-5(B2) M, H-4(B2) w
B3A-E →2)-α-L-Ace*f*A-(1→3	H-2(B3) S, H-2(B2) S	CH_3	H-4(B3)
B4A-E →2,4)-β-D-Gal*p*-(1→2	H-3(B4) S, ∼3.65 ⇒ H-5(B4) S, H-2(B3) M, H-5(B4′) S		
B4′ 2-*O*-Me-α-L-Fuc*p*-(1→2	H-2(B4′) S, H-2(B4A-D) M, Me-6 of α-L-Rha*p* M	CH_3 OMe	H-5(B4′) S; H-1(B4′) S, 3.72, 3.53 H-1(B4′)
B5A →2)-α-L-Ara*p*-(1→4	nnd		
B6 α-l-Rha*p*-(1→2+ →2)-α-L-Rha*p*-(1→2	H-2(B6) S; H-2(B7) w		
B7 β-L-Ara*f*-(1→2	H-2(B7) S, H-2(B6) at 3.935 ppm w		
C1A →5)-α-Kdo-(2→3	nap	H-3eq	H-1(C2A) w
C1B		H-3eq	H-1(C2B) M
C2A α-L-Rha*p*-(1→5	H-2(C2A) M		
C2B	H-2(C2B) M, H-3ax S, H-3eq w		
D1 →5)-β-D-Dha-(2→3	nap	H-3ax H-4 H-5(B) H-6(B)	H-4(D1) S; 4.71, 4.485 H-3ax(D1) M, H-3eq(D1) M H-6(D1B) S, CH_3(A2B) M, H-1(D2) S H-5(B) S, CH_3(A2B) M

Table 3 (*Continued*)

Residue	H-1 NOEs	Other Reporter Groups	Reporter Group NOEs
D2 β-L-Ara*f*-(1→5	H-2(D2) S, H-5(D1) S		
E1A α-Ara*f*-(1→	H-2(E1A) S, 3.69		
E1B	H-2(E1A) S, 3.72		

Key: nnd (no NOE data), S (strong), M medium, w (weak) NOEs when compared to the corresponding H-1/H-2 (strong) intraresidue interactions and otherwise as in Tables 1 and 2.

to that of 4-substituted (3.63–3.94 ppm—mono-, di- and trisubstituted rhamnosides) β-Rha*p* residues. This made it possible to distinguish the overlapping coupling graph of the 3-substituted **B2** residue from that of the 2,3,4-trisubstituted **A2A** sugar and the assignments of the signals of the H-2–H-3–H-4–H-5 fragments of **A2** were then completed from the correlation data in the TOCSY spectrum. **A2A** and **A2B** exhibit very similar chemical shift data and fortunately the 6-deoxymethyl signal of **A2B** was an isolated resonance and thus the corresponding long-range interactions in Table 3 were not ambiguous.

Unassigned crosspeaks between protons resonating at 3.40 and 3.96 ppm in the DQCOSY spectrum were assigned to the H-3′ methylene protons of both the apiose sugars (A1 and B1) on the basis of the strong interresidue H-1(A2)/H-3′(A1) and H-1(B2)/H-3′(B1) through-space interactions and suitable HSQC crosspeaks. Only one of the anomeric protons ($\delta_{H\text{-}1}$ 5.002 ppm) of the apiose residues displayed interglycosidic NOEs with the H-2 of a mainchain α-Gal*p*A residue and it had previously been assigned to H-1(A1). However, the other one ($\delta_{H\text{-}1}$ 5.09 ppm) exhibited strong NOEs with the H-3 of a mainchain α-Gal*p*A residue ($\delta_{H\text{-}1}$ 5.26 ppm) as predicted by the theoretical models for H-1(A1) when the weak H-1(M_A)/CH_3(A2) and H-1(M_A)/H-1(A2) contacts were used in the conformational search (vide infra). Accordingly, this assignment has been reversed and the low-field apiosyl anomeric signal has been assigned to H-1(A1).

Several questions concerning the A-chain subsist such as what type of interaction constrains the **A5** exocyclic group to adopt a unique orientation about the C-5–C-6 bond and what distinguishes the major component from the minor one. Two sets of broad signals are observed for the H-1(A2″)/H-3(A2) interglycosidic crosspeaks of the **A2**″ residues in the NOESY spectrum. In contrast, only one set of sharp signals has been assigned to the other A-chain branching units, **A2**′ and **A3**′. Finally, it should be noted that only one of the two **A4** residues exhibits strong interglycosidic contacts with the CH_3(A3) moiety in agreement with the optimized molecular models from the two lowest energy regions of the **A4**→**A3** disaccharide map that display short and borderline (~ 5 Å) A4/CH_3(A3) distances, respectively.

2.2. B-chain

In the course of the re-examination of the NMR data for the β-rhamnose sugars, two coupling graphs were established for **B2** stemming from 6-CH_3 signals at 1.215 and 1.285 ppm, which included the expected H-1 (4.79–4.80 ppm), H-2 (4.11–4.14 ppm) and H-4 (< 3.53 ppm) chemical shifts. Assignment of the H-1 and H-2 signals of the **B3** α-Ace*f*A sugar was straightforward due to the characteristic C-2(B3) chemical shifts (> 85 ppm) that were unambiguously detected in the sensitivity-enhanced HSQC spectrum in spite of sample heterogeneity (five H-1/H-2 pairs were observed for **B3** in the DQCOSY spectrum). The NOESY spectrum displayed strong H-1(B3)/H-2(B2) crosspeaks corroborating the assignments. These data are very similar to those reported for the related dodecasaccharide that contains the **B1**–**B2**–**B3**–**B4** fragment of the B-chain.[8]

Five distinct H-1–H-2–H-3–H-4 coupling graphs with the anomeric protons resonating in the 4.5–4.7 ppm region have been reassigned to the **B4** β-Gal*p* residue (initially reported as the α-anomer). Here also, a complete set of H-1(B4)/H-2(B3) NOESY crosspeaks was observed. Due to heavy overlapping it was not possible to unambiguously determine the H-5(B4) and H-6(B4) chemical shifts and the only group of unassigned methylene signals with typical carbon chemical shifts (C-6 resonating at 61.8 ppm) has been tentatively attributed to **B4**.

The two coupling graphs, previously assigned to the **B5** α-Ara*p* sugar, displayed multiple overlapping resonances in the DQCOSY spectrum masking the fine structure of the crosspeaks. Due to the much-improved signal-to-noise ratio in the sensitivity-enhanced HSQC

spectrum, complete carbon chemical shifts were obtained and the data indicated the presence of an α-Ara*f* sugar (hereafter referred to as the **E** sidechain). Several H-1/H-2 DQCOSY crosspeaks with the correct fine structure ($^3J_{1,2} = 7$ Hz $< {}^3J_{2,3}$) and chemical shift range for H-1(B5) ($\delta_{H\text{-}1}$ 4.7 ppm)[22] were detected. Moreover, certain HSQC crosspeaks (H-1/C-1, 4.62/104 ppm; H-6a,b/C-6, 3.51 and 3.83/66.4 ppm, two well-resolved doublets, $^2J_{H\text{-}6a,b} = -12.7$ Hz) corroborate the presence of the α-Ara*p* **B5** sugar[23] but the signal-to-noise ratio is too low for further analysis.

The numerous spin systems that were detected for the α-Rha*p* sugars further demonstrate that the primary structure in Scheme 1 is still incomplete. The anomeric region in the HSQC spectrum can be divided into four distinct zones of α-Rha*p* crosspeaks (**R1**, **B6**, **C2**, and **R2**). The low-field shift of the signals of the **R2** anomeric carbons suggests that the corresponding rhamnose units are terminal sugars. A sequential NOE (H-1(B7)/H-2(B6) 4.99/3.935 ppm) allowed the unambiguous assignment one of the expected **B7**–**B6** fragments while the other cross-peaks in the **B6** and **R2** region undoubtedly stem from incomplete substitution by **B7** and additional substitution at the 3-position of **B5** (i.e. **B5′**). Heavy overlapping of the **B6** (**B5′**) and **B7** anomeric protons obscures the sequential NOEs precluding further analysis.

2.3. C-chain

The signals of the H-3a,e(C1) Kdo spins are very broad and prior to the molecular mechanics simulation an explanation for this observation such as chemical exchange or restricted mobility was lacking. The ^{1}H and ^{13}C chemical shifts of the C-chain disaccharide were reported by Ishii and Kaneko[7] and by analogy with their data (4.04/73.8 ppm) strong crosspeaks in the HSQC spectrum (4.005/73.3 ppm) were assigned to H-5(C1)/C-5(C1). The 3-deoxymethylene protons exhibit H-3a,e(C1)/H-1(C2) crosspeaks with the cluster of anomeric proton α-Rha*p* signals resonating near 5.135 ppm (midway between the **R1** and **C2** regions).

The **R1** crosspeaks had been previously identified as those of the **C2** sugar because the corresponding anomeric proton presented a very strong interaction with both an α-Gal*p*A H-1 (4.89 and 4.91 ppm for the major and minor species, referred to as $\mathbf{M_Y}$ in Ref. 13) and a proton resonating at 4.30 ppm (which was assumed to be the H-4′ across the glycosidic linkage of the former α-Gal*p*A). A sequential HMBC correlation (5.15/77.0 ppm) indicates that **R1** is linked to the 4-position of an α-Gal*p*A residue rather than the 5-position of Kdo ($\delta_{C\text{-}5}$ 73.5 ppm). Moreover, strong H-1(α-Gal*p*A)/H-2(R1) NOEs are also observed corroborating the close proximity of this α-Gal*p*A to the **R1** α-rhamnose sugar. However, attempts to model the folding of the C-chain rhamnose onto the galacturonan backbone did not result in low-energy conformers (vide infra). The folding found in the previous paper[12] arises from an erroneous linkage in this chain. All the **R1**/α-Gal*p*A contacts are reminiscent of the RG-I repeating unit, →4)-α-D-Gal*p*A-(1→2)-α-L-Rha*p*-(1→, and the ^{1}H and ^{13}C chemical shift data (see footnote to Table 1) for these two sugars are identical to those reported for RG-I from soybean.[24] Similar intensities are observed for the NMR signals of this 'RG-I like' fragment when compared to those of the other residues suggesting that the former fragment is also a component of the primary structure of RG-II.

The only HSQC-TOCSY long-range correlation concerning the α-rhamnose sugars were crosspeaks at 5.12/70.3 ppm (H-1/C-2) which identified this sugar as a terminal α-Rha*p* residue (**C2**). This sugar displayed the anomeric carbon resonance at lowest field (C-1 at 101.5 ppm) in keeping with the absence of a β-effect due to substitution at C-2. Close inspection of the NOESY spectrum showed that indeed a very weak H-3ax,eq(C1)/5.12 ppm correlation (much weaker than the aforementioned H-3ax,eq(C1)/5.135 ppm one) corroborated this assignment. Only very small variations are observed when one compares the chemical shifts of the mRG-II-ol C-chain ($\Delta\delta_H < 0.04$ ppm; $\Delta\delta_C < 0.3$ ppm with the exception of C-3, $\Delta\delta_{C\text{-}3} + 1.5$ ppm) with those of the corresponding disaccharide reported by Ishii and Kaneko.[7]

2.4. D-chain

The characteristic H-6/C-6 chemical shifts unambiguously assigned in comparison with Vidal and coworkers[8] allowed identification of the key H-5(D1B) signal through the H-6/H-5 crosspeaks in the 750 MHz TOCSY spectrum. Interglycosidic H-5(D1B)/H-1(D2B) NOESY crosspeaks were observed whereas the through-space contacts for H-5(D1A) were overlapping with those of the anomeric protons of **A2′** and one of the E-chain α-Ara*f* sugars. Analysis of the 6-deoxymethyl region of the NOESY spectrum reveals distinct long-range contact between CH_3(A2) and H-6(D1) (1.285/4.865 and 1.30/4.82 ppm for the A and B components, respectively). On the whole the chemical shift variations for the D-chain residues between mRG-II-ol and the related oligosaccharide obtained by cleavage of RG-II are very small ($\Delta\delta_H < 0.04$ ppm with the exception of H-4, $\Delta\delta_{H\text{-}4} - 0.1$ ppm; $\Delta\delta_C < 0.3$ ppm with the exception of C-1(D2), $\Delta\delta_{C\text{-}1} - 1.7$ ppm).

2.5. E-chain

The only difference between the intraresidue spectral data of the A and B components of this residue is the H-3(E1) chemical shift (3.52 and 3.57 ppm). Strong

interresidue NOEs of the anomeric protons with spins resonating at 3.69 (major component) and 3.72 (minor component) ppm, reminiscent of $\delta_{H\text{-}2}$ of α-D-Gal*p*A, are detected but heavy overlapping in the 4.85–4.92 ppm region makes assignments hazardous.

2.6. Homogalacturonan mainchain

Due to severe overlapping in the regions containing both the H-1 and C-1 signals of the α-Gal*p*A units, many of the assignments in Table 2 can only be considered as tentative (given in italics). The carbon and proton chemical shifts for the open chain galactonic acid residue (major form of borohydride-reduced RG-II) **M$_1$**, have been totally elucidated and numerous contacts between H-1(M_2) (or H-5(M_2)) and the galactonic acid protons are detected in the NOESY spectrum. Clear H-1(M_3)/H-4(M_2) interresidue contacts are also observed but the following sequential NOE, H-1(M_4)/H-4(M_3), could not be assigned. The proton and carbon chemical shifts of the **M$_3$**–**M$_2$**–**M$_1$** fragment of m-RG-II-ol are almost identical ($\Delta\delta_H < 0.04$ ppm) to those reported by Vidal and co-workers[8] for the **M$_3$**–**M$_2$**–**M$_1$** fragment of the dodecasaccharide containing the B and D sidechains (this oligomer is composed of a hexasaccharide galacturonan backbone). Prior to the NMR investigation, the C-chain had been selectively removed from this latter oligosaccharide by hydrolysis. Substitution by Kdo is generally accompanied by a substantial shift of the signal of the methine proton α to the branchpoint carbon.[25,26] Consequently, the strong analogy between the proton chemical shifts of the **M$_3$**–**M$_2$**–**M$_1$** moiety in the dodecasaccharide (lacking Kdo) and in m-RG-II-ol (containing Kdo) suggests that the C-sidechain is not linked to this fragment. As it has been shown that the four A–D sidechains are all attached to different α-Gal*p*A residues,[5,27] it appears likely that the C-sidechain is linked to **M$_4$**.

The remaining anomeric protons of the galacturonan backbone sugars gave rise to weak broad crosspeaks in the 2D spectra. Sequential NOEs between the anomeric protons of **A1** and **B1** and spins belonging to **M$_A$** and **M$_B$** were expected to allow identification of these spin systems whereas the glycosidic linkages at **M$_C$** and **M$_D$** involved quaternary carbons at **C1** and **D1** precluding efficient NOE pathways. Appropriate sequential NOESY crosspeaks were indeed detected for the apiosyl sugars (F1/F2 singlet/doublet—5.09/3.58 and 5.00/3.72–3.68 ppm) but their interpretation was not straightforward. The low-field NOESY crosspeaks (5.09/3.58 ppm) did not involve the H-2 signal of an α-Gal*p*A unit and the high-field ones (5.00/3.72) corresponded to the H-2 resonances of several α-Gal*p*A residues (>4 spin systems). Closer inspection of the NOESY spectrum revealed unexpected long-range interactions between mainchain anomeric protons and the spins located at the second glycosidic linkage of both the A- and B-sidechains (i.e. $\delta_{H\text{-}1}$ 5.26 ppm with H-1(A2) and H-3′b(A1); $\delta_{H\text{-}1}$ 5.015 and 5.035 ppm with H-1(B2) and H-3′b(B1)). Assignment of such weak long-range NOEs could only be made on the basis of predicted Overhauser effects for the theoretical models (H-1(M_A)/H-1(A2) and H-1(M_A)/H-3′(A1); H-1(M_C)/H-1(B2) and H-1(M_C)/H-3′(B1)). However, the optimized models revealed the experimentally-observed dichotomy in the apiosyl/mainchain orientations of the A and B sidechains with the closest contacts between H-3(M_A) and H-1(A1) and between H-2(M_B) and H-1(B1), respectively. This pattern would appear to constitute the signature of the RG-II A- and B-sidechain conformations. At present, H-1(M_n)/H-4(M_{n-1}) interresidue contacts have been demonstrated for most of the α-Gal*p*A linkages. However, most of the mainchain sugars also exist as both major and minor components (with different sequential NOEs) and many of the H-1 and H-4 signals for the α-Gal*p*A units are overlapping. As a result, assignment of the backbone residues has still not been feasible on the basis of NMR data alone.

Three major obstacles will need to be overcome to complete the NMR analysis of m-RG-II-ol. In the first place, assignment of the **A2′** and **M$_T$** residues relies on the correct interpretation of the sequential NOEs at 4.485/3.685 and 5.01/4.485 ppm (4.935/4.51 ppm, minor component), respectively (H-5(A2′)/H-4(A2) and H-1(M_T)/H-4(M_A) in Tables 2 and 3. Unfortunately, an alternate explanation for these NOEs is equally probable on the basis of the chemical shift data and it corresponds to reversal of these assignments (namely H-5(M_T)/H-2($M_{T\text{-}1}$) and H-1(A2′)/H-2(A2)). The second difficulty resides in the location of the E-chain. The results of linkage composition analysis (i.e. identification of a 2,3,4-trisubstituted α-Gal*p*A) suggest that **E1** is located at the 3-position of **M$_A$** or **M$_B$**. Both H-1(B1) and H-1(E1A and B) exhibit strong sequential NOEs with protons resonating at the same frequencies (3.68–3.72 ppm—undoubtedly H-2(M_B)) pointing to the latter solution but the high-field shift of the H-3(M_A) signal ($\delta_{H\text{-}3}$ 3.58 ppm) argues for 3-substitution of **M$_A$**.[8,24] Certain unassigned NOEs support alternate solutions (i.e. crosspeaks at 4.55/3.94 ppm can be interpreted as an H-5(M_T)/H-2(M_E) contact suggesting that **M$_E$** = **M$_{T\text{-}1}$**). Finally, the NMR evidence for the position of the 'RG-I like' fragment, with respect to the primary structure in Scheme 1, points to several possibilities (extension of the mainchain at the terminal end, extension of either the A- or E-sidechains at the 3-positions of **A5** or **E1**, respectively). Although the theoretical models have been an enormous help in interpreting the NOE data of mRG-II-ol, only a more complete picture of its primary structure (through structural analysis of oligomeric fragments from the terminal end) will afford the answers to these remaining questions.

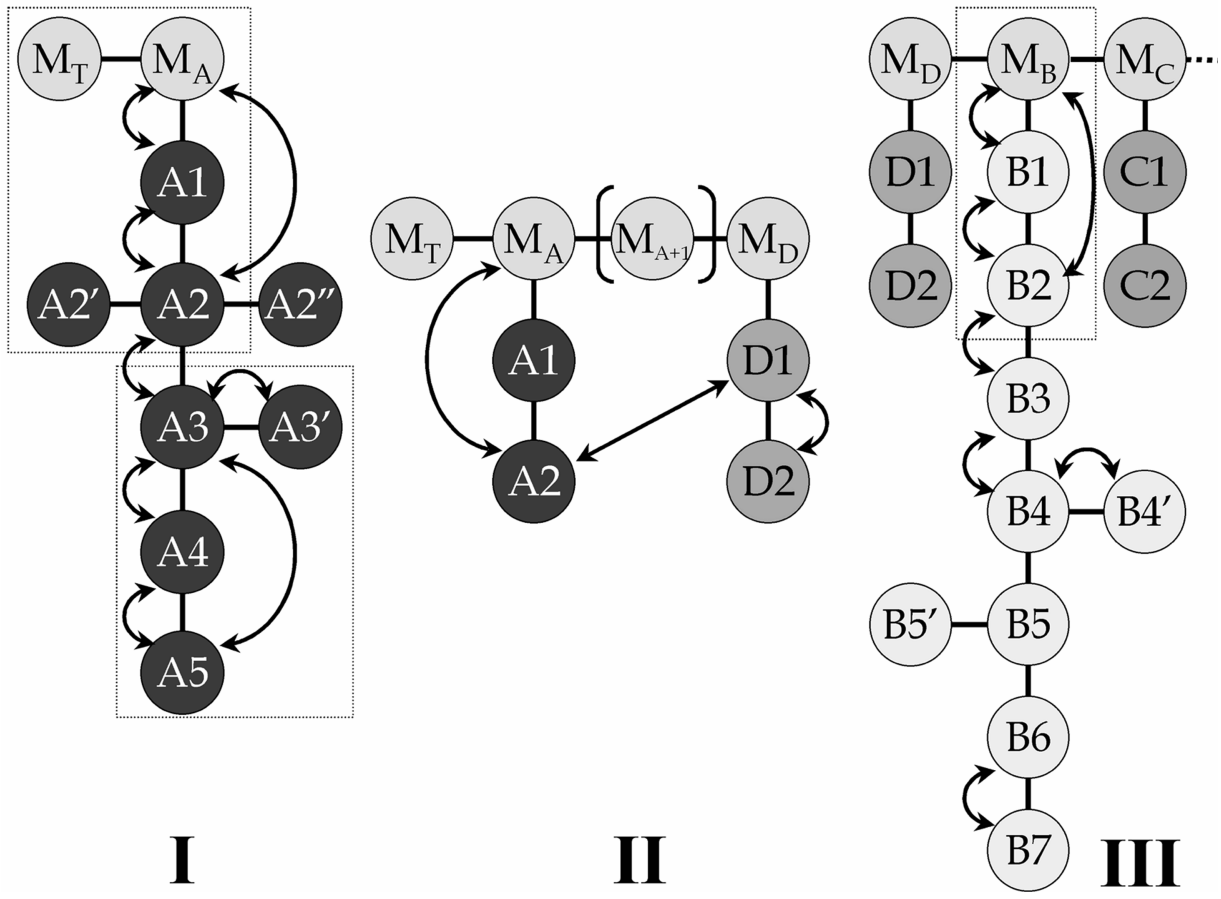


Fig. 2. Fragments studied by Systematic Search. Arrows indicate the inter-residue NOEs.

2.7. Molecular modeling

The conformational study was based on the determination of the preferred orientations of each glycosidic bond as well as the flexibility of the furanose rings in the apiose, aceric acid and arabinose residues. All the low energy conformations of each of the constituting disaccharide segments of RG-II had been established in previous work[12] except for the disaccharides whose configuration had been demonstrated to be incorrect (α-L-Ace*f*A-(1→3)-β-L-Rha*p* (**B3→B2**) and β-D-Gal*p*-(1→2)-α-L-Ace*f*A (**B4→B3**))[8] and the disaccharides α-L-Rha*p*-(1→3)-α-L-Ara*p* (**B5′→B5**) and α-L-Ara*f*-(1→3)-α-D-Gal*p*A (**E1→M$_E$**), not considered previously. Accordingly, the corresponding relaxed maps have been calculated for these segments. The combination of the various minima obtained for all of these fragments generates several millions of possible conformers. Fortunately, the experimental data greatly reduced the number of conformers that had to be considered. Thus, once the NMR spectra of RG-II were reassigned, a list of NOEs was collected including the interactions of the anomeric protons along with those of various reporter groups such as methyl and methylene groups. These data were used to filter the local conformers or building blocks that were used to establish the three-dimensional structure of RG-II.

The are several approaches that can be used to perform conformational search; among them, Monte Carlo, molecular dynamics, or simulated annealing are widely employed. Initially, this latter method was tested on RG-II, but the high degree of ramification of this molecule makes very difficult the exploration of different minima. The method finally applied was the Systematic Search, integrated in the program SYBYL[28] as it is a fast and flexible method that allows enumeration of all of the different torsional conformers which are consistent with some set of distances. In this approach, all the torsion angles (Φ and Ψ) at each linkage were systematically modified, but no minimization was performed. The ranges of values adopted by these torsion angles were governed by the energy maps previously computed. From the generated ensemble of conformers, only those that fulfilled the set of observed interproton distances were selected and their energies were calculated (essentially taking into account the steric interac-

tions) using the internal Tripos force field.[29] Finally, the low-energy conformations were grouped into families according to their glycosidic torsion angles.

Because of the large amount of possible conformers obtained if the whole molecule is submitted to a Systematic Search, RG-II was split into fragments, which were studied separately: fragments I–III (presented in Fig. 2) include A-, D-, and B-chains, and the last one corresponds to the C-chain.

2.8. Fragment I

The first fragment studied corresponds to the A-chain. Due to the large number of interresidue NOEs detected for this fragment (Tables 3 and 4) the conformation adopted by the A-chain seems to be very well defined. Moreover, most of the NOEs implying consecutive residues are compatible with, at least, one of the minima found in the corresponding energy maps.

Table 4
Experimental NOEs vs. calculated distances taken from the five lowest in energy conformers for families IIIa and IIIb. Energies are also indicated

		IIIa					IIIb				
		1	2	3	4	5	1	2	3	4	5
H-1(A1)/H-3(M_A)	S	3.5	3.5	3.5	3.5	3.5	3.4	3.5	3.5	3.5	3.5
H-3′B(A1)/H-1(M_A)	VW	3.7	3.7	3.6	3.7	3.7	3.7	3.8	3.8	3.7	3.7
H-3′A(A1)/H-1(A2)	S	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6
H-3′B(A1)/H-1(A2)	S	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3
H-1(A2)/H-1(M_A)	W	6.4	6.5	6.4	6.4	6.5	6.5	6.6	6.6	6.5	6.6
H-3(A2)/H-1(A2″)	S	2.3	2.3	2.4	2.4	2.4	2.3	2.3	2.3	2.3	2.3
H-4(A2)/H-1(A3)	S	2.8	2.8	2.8	2.8	2.8	2.9	3.0	2.8	2.8	2.8
Me(A2)/H-1(M_A)	VW	5.8	5.9	5.7	5.7	5.9	6.0	6.1	6.0	5.8	6.0
Me(A2)/H-1(A3)	VW	5.0	5.1	5.0	5.0	5.0	5.1	5.1	5.0	5.0	5.1
Me(A2)/H-5(A3)	S	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7
Me(A2)/H-6(D1)	W	6.2	6.1	6.1	6.1	6.0	6.1	6.0	6.0	6.0	6.0
H-3(A3)/H-1(A3′)	S	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
H-4(A3)/H-1(A3′)	W	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4
H-4(A3)/H-1(A4)	S	3.4	3.3	3.4	3.3	3.3	3.3	3.4	3.4	3.4	3.3
Me(A3)/H-1(A4)	S	3.9	3.9	3.9	3.8	3.9	3.9	3.8	3.9	3.9	3.9
Me(A3)/H-3(A4)	M	5.1	5.2	5.2	5.2	5.2	5.2	5.1	5.2	5.1	5.2
Me(A3)/H-5(A4)	M	3.5	3.6	3.5	3.5	3.6	3.6	3.5	3.5	3.5	3.5
H-1(A3)/H-5(A5)	W-M	2.8	2.9	2.8	2.8	2.8	2.9	3.0	2.8	2.8	2.8
H-2(A3)/H-5(A5)	S	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
Me(A3)/H-5(A5)	W	4.3	4.2	4.2	4.2	4.2	4.2	4.3	4.3	4.3	4.2
H-1(A3′)/H-1(A4)	VW	5.5	5.4	5.5	5.4	5.4	5.4	5.5	5.5	5.5	5.4
H-1(A4)/H-1(A5)	VW	3.6	3.6	3.6	3.6	3.6	3.6	3.5	3.6	3.6	3.6
H-2(A4)/H-1(A5)	S	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6
H-2(A4)/H-5(A5)	M	4.4	4.3	4.3	4.3	4.3	4.3	4.4	4.3	4.3	4.3
H-3(A4)/H-1(A5)	W	2.3	2.3	2.3	2.4	2.4	2.3	2.3	2.4	2.4	2.4
H-1(B1)/H-2(M_B)	S	3.4	3.5	3.5	3.5	3.5	3.4	2.3	3.4	3.4	3.4
H-3′b(B1)/H-1(M_B)	VW	3.3	3.2	3.3	3.2	3.0	3.3	3.8	3.3	3.4	3.3
H-3′a(B1)/H-1(B2)	W	3.0	3.0	3.0	2.9	3.0	3.1	2.6	3.1	3.1	3.1
H-3′b(B1)/H-1(B2)	S	2.5	2.5	2.5	2.6	2.5	2.5	2.5	2.5	2.6	2.5
H-1(B2)/H-1(M_B)	W	4.7	4.5	4.7	4.6	4.3	3.8	5.8	4.2	4.3	4.1
H-2(B2)/H-1(B3)	S	2.6	2.4	2.6	2.6	2.5	2.6	2.8	2.8	2.7	2.8
H-2(B3)/H-1(B4)	M	2.9	3.5	2.8	2.7	3.4	3.5	2.7	2.9	2.9	2.7
H-2(B4)/H-1(B4′)	M	3.0	2.4	3.0	2.8	2.4	2.4	3.1	3.0	3.0	2.7
H-1(B6)/H-2(B7)	W	4.6	4.6	4.7	4.6	4.4	4.6	4.6	4.6	4.6	4.5
H-2(B6)/H-1(B7)	W	2.4	3.3	2.5	3.3	3.0	3.2	3.2	2.5	2.5	2.5
H-3ax(C1)/H-1(C2)	S	2.4	2.5	2.4	2.4	2.4	2.5	2.4	2.5	2.5	2.4
H-3eq(C1)/H-1(C2)	W-M	3.7	3.7	3.7	3.7	3.7	3.7	3.7	3.7	3.7	3.7
H-5(D1)/H-1(D2)	S	2.5	2.5	2.5	2.5	2.7	2.5	2.5	2.4	2.4	2.4
Energy (kcal)		**112.9**	**113.6**	**114.6**	**116.8**	**118.9**	**107.9**	**109.2**	**113.6**	**115.5**	**116.3**

When the distance implied a methyl group, the distance was pondered among the three protons: $\langle r^{-3} \rangle^{-1/3}$.

The A-chain was, in turn, split into two fragments: a first fragment (Ia) containing the residues $\mathbf{M_T}$, $\mathbf{M_A}$ (both α-D-Gal*p*A), **A1** (β-D-Api*f*), **A2** (β-L-Rha*p*), and **A2′** (α-D-Gal*p*A), and a second fragment (Ib) containing the residues **A3** (α-L-Fuc*p*), **A3′** (2-*O*-Me-α-D-Xyl*p*), **A4** (β-D-Glc*p*A), and **A5** (α-D-Gal*p*). **A2″** (β-D-Gal*p*A) was further added to complete the chain, as initially no long range NOEs were found for this residue. Furthermore, according to the $^3J_{1,2}$ (1–2 Hz) of apiose, the furanose ring presents a conformation in the northern region of the pseudorotation wheel ($^3T_2$).

There are two long-range NOEs that could define the global orientation of the Ia fragment: NOEs between the anomeric proton of $\mathbf{M_A}$ and both H-1 and the methyl group of **A2**. Actually, Table 3 includes five interresidue NOEs between the monosaccharides composing fragment Ia, but some of them correspond to short distances (e.g. H-3′a(A1)/H-1(A2)) for all of the low energy regions, and others (e.g. H-1(A1)/H-3(M_A)) had not been initially assigned: these latter NOEs served as a test for the resulting structures.

A conformational search was done by modifying the torsion angles for **A1**(Φ, Ψ, ω) and **A2**(Φ, Ψ), while applying the aforementioned distances as the criterion of selection. A large number of conformations was obtained, which were grouped into families including low energy families having **A1**(Φ, Ψ) values close to (300, 160) and **A1**(ω) either 60 or $-60°$. The **A2**(Φ, Ψ)parameters were close to (80, 180) for **A1**(ω) = 60°, and (150, 180) in the second case, that is, there is some adjusting but essentially the same area is explored. The relative orientation of **A1** and $\mathbf{M_A}$ implies a close approach of protons H-1(A1) and H-3(M_A). A broad intense signal in the NOESY spectrum for the apiosyl anomeric proton resonating at 5.09 ppm with a spin resonating at 3.58 ppm could be assigned to H-1(A1)/H-3(M_A), allowing the identification of H-1(A1). The orientation of **A2′** was set to that of the global minimum in the energy map, as the assignment of this unit still can only be considered as tentative.

In the case of the Ib fragment, the NOE data allowed us to determine which of the minima found in the disaccharide energy maps were compatible with the conformation in solution (Fig. 3A). Thus, NOEs between the methyl group of **A3** and H-1, H-3, and H-5 of **A4** indicated that this group was located over the pyranose ring of this latter. This orientation is compatible with the global minimum (Φ, Ψ) = (280, 120) found for the **A4**→**A3** segment. In the same way, the intense NOE between H-1(A3′) and H-3(A3) indicated an orientation near that of the global minimum (80, 100), although the weak H-1(A3′)/H-4(A3) contact suggested values of Φ and Ψ corresponding to a shoulder in the lowest energy domain (140, 160). This latter minimum gives distances of about 5.0–5.5 Å between H-1(A3′) and H-1(A4) (long-range) which could explain the very weak NOE observed for this pair of protons. In the case of **A5**, on one hand the strong H-1(A5)/H-2(A4) NOE and the weak H-1(A5)/H-1(A4) and H-1(A5)/H-3(A4) interactions are compatible with both the two lowest energy minima of the **A5**→**A4** segment ((80, 120) and (100, 160), respectively). However, long-range NOEs between H-5(A5) and several protons of **A3** (especially the intense NOE with H-2(A3)), indicate the presence of another conformer having (Φ, Ψ) values of (60, 270) close to the third minimum for this residue (100, 300) in solution. This folding of **A5** towards **A3** also implies a modification of the **A4**(320, 180) orientation. These new values are located in low energy regions of the Φ, Ψ-map, although not at the global minimum, as in the previous case. No explanation has been found for this preference, since apparently other more stable orientations are accessible. Nevertheless, there is some indication that **A5** could be substituted at C-3, and the presence of an additional unit might shed some light on this subject. To reduce the number of possible conformations for the complete molecule, only this latter conformer was included in the calculations, since it is quite well defined by these interresidue NOEs. Nevertheless, an equilibrium between the three minima may well exist.

2.9. Linking fragments Ia and Ib

The intense NOE between the methyl group of A2 and H-5(A3), as well as the strong H-1(A3)/H-4(A2) NOE is compatible with an orientation corresponding to the global minimum found in the related disaccharide (280, 280), for which the calculated distance is 2.7 Å in both cases. Once the conformation of the A-chain was established, the **A2″** residue was added. Initially, the NOEs found for this residue were considered to be ambiguous and could not be used to define the orientation of this residue. Thus, the torsion angles of **A2″** were tested for each of the minima found for the disaccharide **A2″**→**A2**. Values of (Φ, Ψ) near to the minima (280, 60) and (60, 120) were the most favorable energetically and this former conformation is compatible with the strong H-1(A2″)/H-3(A2) NOE (calculated distance 2.4 Å) that has been assigned since.

The preferred conformation of the A-chain, obtained when **A3** and **A5** are not close, is very similar to the one obtained in the previous study of Mazeau and Pérez,[12] with the exception of the **A1** and **A2** residues as the interactions with the main chain were not considered then. However, the folding of **A5** towards **A3** is difficult to explain with the present primary structure.

2.10. Fragment II

A second group of long-range NOEs suggests the relative orientation of the A- and D-chains, Fig. 2. Thus,

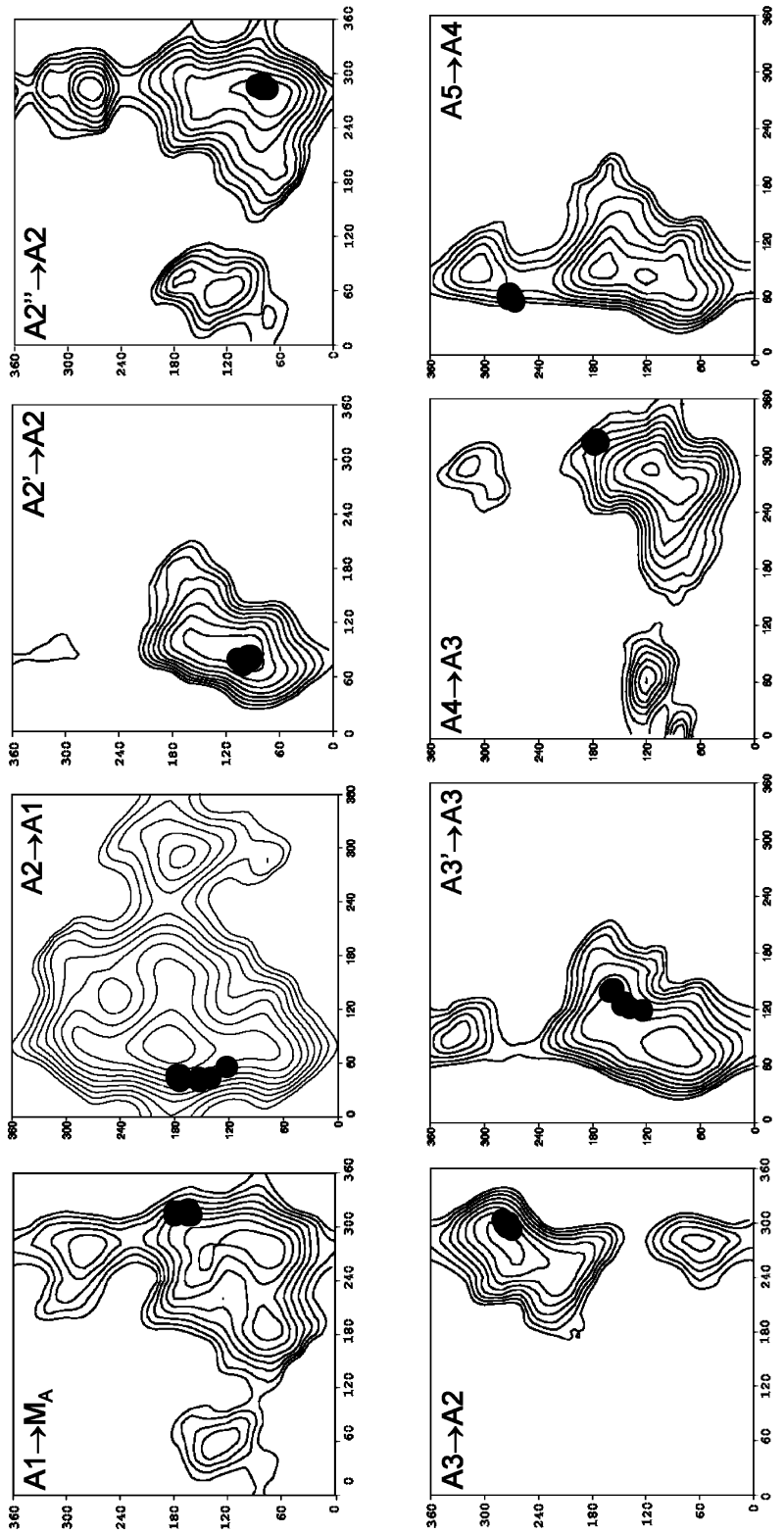


Fig. 3. Potential energy surfaces of all the disaccharide segments studied in the A- and D-sidechains of RG-II as a function of the Φ and Ψ torsion angles. The isoenergy contours are drawn with interpolation of 1 kcal/mol above the minimum of each map. The dots correspond to the 28 conformations found in the systematic search, after minimization of the complete molecule.

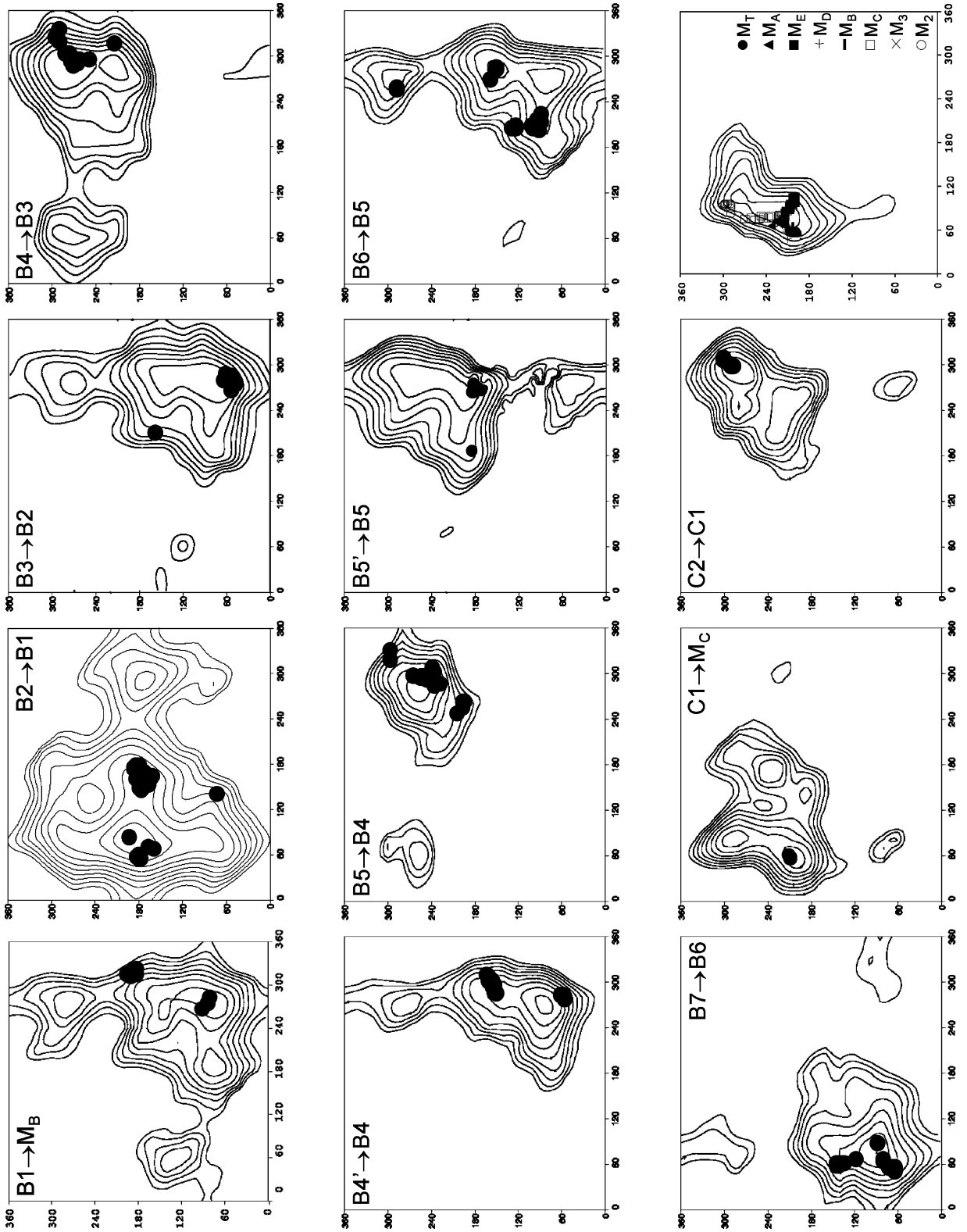


Fig. 3. (*Continued*)

weak CH_3(A2)/H-6(D1) NOEs can be explained if this latter residue exposes the bottom side of its pyranose ring towards the methyl group. To determine whether this orientation is compatible with the rest of the molecule, fragments containing **D1** (β-D-Dha*p*), **D2** (β-L-Ara*f*), **A1**, **A2**, and the Gal*p*A residues belonging to the main chain were built. As it is not known whether there is a mainchain residue separating the branchpoint sugars, M_A and M_D, the two cases where were considered (Scheme 1) and zero or one Gal*p*A unit between the branching points of these chains was added.

Distance criterions were used to select conformations so that the methyl group of **A2** was close to both H-6(D1) and H-1(M_A) and this latter proton was not more than 5 Å away from H-1(A2). The torsion angles of **A1** (Φ, Ψ, and ω) and **A2** (Φ, Ψ), together with those of the Gal*p*A units of the main chain and **D1** (Φ, Ψ) were used in the search.

In both cases (with or without additional Gal*p*A), structures in agreement with the above requirements were found. **D1** presented an orientation near to the second and third minima in energy ((60, 240) and (80, 280)). The resulting structures adopt conformations close to that found for the fragment Ia as regards the torsion angles of **A2** (60, 120) and **A1** (330, 150, 60) associated with $\mathbf{M_A}$ (60, 120), $\mathbf{M_E}$ (90, 90) and **D1** (80, 280) orientations that correspond to the minima found in the corresponding disaccharide energy maps. Nevertheless, when the structure was minimized, the distance between the methyl group and **D1** became slightly longer than in the initial geometry (about 6 Å, see Table 4). Finally, the strong H-1(D2)/H-5(D1) NOE is compatible with the global minimum found in the energy map for the **D2**→**D1** segment (80, 240). The residue **D2** was then added in this orientation without further search.

2.11. Fragment III

The third fragment studied by Systematic Search corresponds to the B-chain, Fig. 2. In order to avoid conformations that would be sterically unfavorable in the complete molecule, the C- ($M_C = M_4$) and D-chains ($M_D = M_6$) were added to a hexasaccharide (M_1 up to M_6) mainchain with the orientation obtained for the corresponding fragments. A fragment containing D1, D2, B1 (β-D-Api*f*), B2 (β-L-Rha*p*), C1 (α-Kdo*p*), C2 (α-L-Rha*p*), and six Gal*p*A units (M_D–M_B–M_C–M_3–M_2–M_1) was built and submitted to a conformational search.

The A- and B-chains possess the same disaccharide connection to the main chain: β-L-Rha*p*-(1→3′)-β-D-Api*f*-(1→2). It is not surprising to notice that these two identical fragments share some common NOEs: interactions between linked fragments (H-3′b(Api*f*)/H-1(Gal*p*A)) and (H-3′a,b(Api*f*)/H-1(Rha*p*)), as well as some long range NOEs (H-1(Gal*p*A/H-1(Rha*p*)). However, one cannot expect these two fragments to have the same folding, because some of the other observed NOEs are specific for each fragment: the orientation of the glycosidic linkage of apiosyl residue with respect to the Gal*p*A of the main chain involves the proximity of H-1(A1) and H-3(M_A) for the A-chain, and H-1(B1) and H-2(M_B) for the B-chain. Furthermore, the orientation of β-Rha*p* with respect to β-Api*f* is defined by the proximity of the methylenic protons (H-3′a and H-3′b) and the anomeric proton of Rha*p*, but the pattern of intensities is different in both cases. Finally, the distant NOE between Me(A2) and H-1(M_A) is not echoed in fragment III. Thus, although an orientation related to that obtained for A-chain is expected for the B-chain significant differences are also anticipated.

Although there is only one assigned long-range NOE detected for fragment III, (H-1(B2)/(H-1)M_B), a set of interresidue NOE data allowed us to better define the orientation of M_B with respect to B2: a weak H-3′b(B1)/H-1(M_B) NOE, the strong H-1(B1)/H-2(M_B) interaction, and the relative intensities of the H-1(B2)/H-3′a(B1) and H-1(B2)/H-3′b(B1) crosspeaks (weak and strong, respectively). Unfortunately, the protons of the hydroxymethyl group of B1 (β-D-Api*f*) have not been stereospecifically assigned, and the two possible combinations had to be considered.

Only the torsion angles of **B1**, **B2**, $\mathbf{M_D}$, $\mathbf{M_B}$ and $\mathbf{M_C}$ were considered in the search, since the conformations of C- and D-chains were taken from the results of the studies of other fragments. In the first place, H-3′a(B1) was considered as pro-*S* and H-3′b(B1) as pro-*R* (distribution IIIa). Only 28 conformations were obtained that were compatible with the mentioned distances owing to the great number of residues close to units **B1** and **B2**. When these structures are grouped into families, a main global family is obtained, that encompasses 86% of all the conformations. The torsion angles for **B1** (300, 210, 30) are near to those found for **A1** (300, 160, 60), although for **B2** these angles do not correspond to the global minimum (Fig. 3B), as in the case of **A2**, but to a shoulder close to (180, 180).

When H-3′a(B1) is considered as pro-*R* and H-3′b(B1) as pro-*S* (distribution IIIb) 43 possible conformations were obtained and grouped into three main families. Values for the glycosidic linkage of **B1** at (300, 210) and (240, 120) were found, with the **B1**(ω) orientation close to 180°. However, the **B2** glycosidic torsion angles adopted values close to the global minimum in the first two families (30, 180) and (60, 150) and close to a shoulder at (120, 90) for the third one.

For the rest of the B-chain, the interresidue NOEs imply consecutive residues, most of them corresponding to the protons at the linkage point. The NOE data do not allow discrimination of the low energy regions, and a great number of possibilities had to be taken into

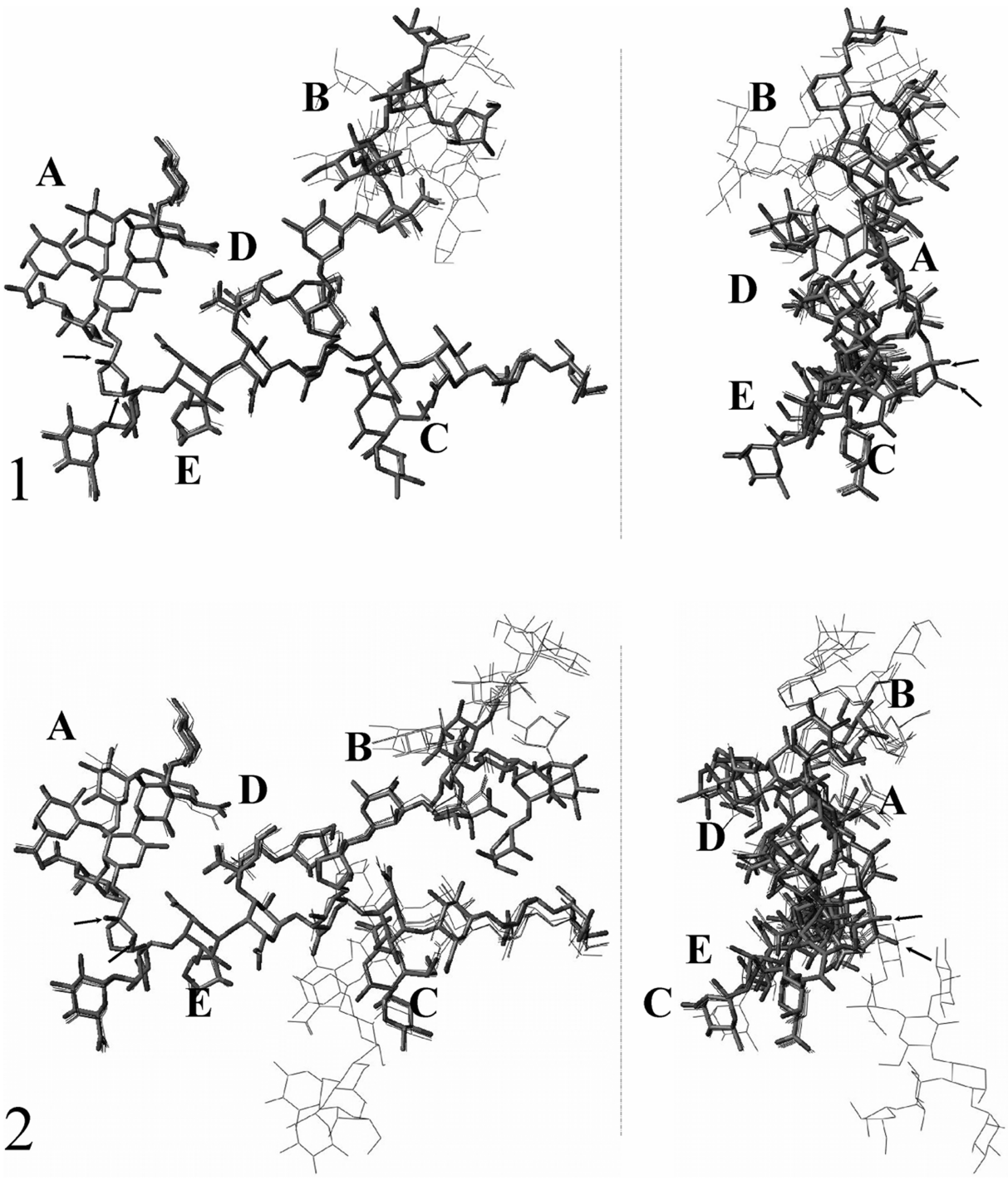


Fig. 4. Frontal (left) and side (right) views of the superposition of the five lowest in energy conformations obtained from (1) a distribution IIIa (H-3′b(B1) pro-*R*) and (2) a distribution IIIb (H-3′b(B1) pro-*S*). Arrows indicate the linkage points of boron in dimer RG-II.

account. Strictly-speaking, two possible low energy conformations ($^3T_2$ and $^2T_3$) adopted by the furanose ring of **B3** (α-L-Ace*f*A) would have to be considered as the $^3J_{1,2}$ (approx 5 Hz) does not allow discrimination between the two conformational families of the five-membered ring. However, in order to simplify the study, a $^2T_3$ conformation was assumed. Thus, the rest of the B chain (units **B3**–**B7**) was added to each fragment coming from the above families and the resulting chain was studied by a conformational search. A large number of possible conformations was obtained: in the case of the IIIa distribution and the first family of the IIIb distribution, about 35,000 conformations were found that were grouped into more than 40

families after having applied an energy window of 50 kcal/mol. The first eight families were retained in both cases, involving more than 96% of all the selected conformations. For the other two families of the IIIb distribution, the number of possible conformations was lower (roughly 10,000 and 20,000, respectively), and only the most populated families were selected. Thus, the study of the III fragment resulted in 20 possible conformations for the B chain.

2.12. C-chain

The branching position of the C-chain onto the main chain remains uncertain, although from the study of the dodecasaccharide related to RG-II[8] it can be deduced that is located near the reducing end of the molecule. As discussed above, the NMR data points to M_4 as the linkage position of the C-chain and the corresponding primary structure (see Scheme 1) was chosen for the conformational study.

From the NOEs between H-1(C2) (α-L-Rha*p*) and H-3ax(C1) (α-Kdo*p*) the conformation adopted by the **C2**→**C1** segment is expected to be near to the global minimum (300, 280) found for the corresponding disaccharide. As regards **C1**, there were no assigned NOEs to indicate its orientation. All the low energy minima are compatible with the **C2**–**C1**–$\mathbf{M_C}$ fragment, including the global minimum (60, 200). As stated previously (NMR discussion), although a strong NOE had been found between the H-1 signal of a Rha*p* previously assigned to H-1(C2) and protons of the main chain, we have not found any energetically favorable structure that could explain this effect. The position of **C2** is always far from the main chain, and it only bends towards the galacturonic acid chain when **C1** adopts orientations corresponding to minima of higher energy. Moreover, M_C can adopt the same orientation as the other 4-linked α-D-Gal*p*A residues and the unusual NOE pattern (i.e. absence of the H-1(M_n)/H-4(M_{n-1} contact) reported previously could not be explained on the basis of the model structures.

2.13. The complete molecule

Once the fragments were studied separately, they were assembled to build the complete molecule with a little adjusting to relax steric interactions between fragments. Then, the E-chain was linked to fragment II (with the $\mathbf{M_A}\rightarrow\mathbf{M_E}\rightarrow\mathbf{M_D}$ distribution in Scheme 1). The orientation of this new fragment (240, 210) was obtained from a systematic search of the α-L-Ara*f*-(1→3)-α-D-Gal*p*A fragment that explored the lowest energy conformation compatible with the rest of the molecule. In the same way, B- and C-chains were added. When the selected conformations for fragment III were incorporated to give the complete molecule, 20 conformers were obtained and subsequently submitted to exhaustive minimization in Sybyl. The most favorable ones are represented in Fig. 4, and their geometries indicated in Fig. 3 and Table 4. In this table, the distances calculated in the minimized structures are checked against the NOEs found in NOESY spectrum. The minimization modified the structures slightly, since there are some deviations from the initial distances. Nevertheless, the resulting distances can explain most of the assigned NOEs, at least qualitatively (in some cases the predicted intensities are different from the experimental ones). Fig. 4 presents the superposition of the five conformations of lowest energy obtained from distributions IIIa and IIIb. In both cases, quite flat structures are obtained, with a length of about 37 Å for the main chain (horizontal diameter) and roughly the same length between the ends of the side chains A and B (vertical diameter). The general shape is a disk, with a thickness of about 17 Å, which is in agreement with the results obtained by AFM (V. Morris, personal communication) and light-scattering (G. Bradbrook, personal communication). The translational self-diffusion coefficient of mRG-II-ol ($D_t = 0.97 \pm 0.08$ m^2/s at 298 K in D_2O), has been established with the pulsed-field gradient spin-echo technique.[30] When this parameter is interpreted according to hydrodynamic theory for a simple model (a spherical object) its molecular volume corresponds to a radius of roughly 20 Å. The conformations arising from the IIIb distribution are more packed than their counterparts in the IIIa distribution, where the chain B is very extended far from the main chain. As regard as the main chain, it presents a very extended orientation, adopting a shape close to that of a helix (between two and threefold).

Ishii and Kaneko[7] determined the location of the 1:2 borate-diol ester cross-link in the dimer to be the apiofuranosyl residue of the A-chain. Thus, it is reasonable to think that the OH-2 and OH-3 hydroxyls in this residue must be free from steric interactions, so that the second monomer can easily approach the first one to form the dimer. This condition is fulfilled in both cases as the hydroxyls (Fig. 4) are well exposed and the side chains are directed towards the opposite side of the molecule. Regarding the energies (Table 4), the values obtained for the IIIa distribution are lower than those of the IIIb distribution, although the difference is small considering the size of the molecule.

3. Experimental

3.1. Sample preparation

As described previously,[5,13] native RG-II was saponified, reduced with sodium borohydride, and then purified by size exclusion chromatography on a ACA

202 (IBF; 1–15 kDa, 400 × 16 mm^2, 13 mL/h).[31] These steps, which remove the acetyl and methylester substituents, also reduce the anomeric center of the first galacturonic acid residue of the main chain (M_1) to afford 3-linked galactonic acid. Twenty milligrams of pure RG-II monomer (according to analytical HPLC) were 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.

3.2. NMR spectroscopy

The homonuclear data have been described previously[13] and were processed with the program NMRPipe.[32] Sensitivity-enhanced HSQC[33–35] and HSQC–TOCSY[36] experiments were recorded at 500 MHz using standard pulse sequences from the Bruker library. Quadrature detection in the non-acquisition dimension was achieved with Echo-Antiecho-TPPI gradient selection. For the 800 MHz gradient-enhanced and TROSY[37,38] HSQC spectra, quadrature detection in the non-acquisition dimension was achieved with Echo-Antiecho-States gradient selection. All experiments were carried out at 40 °C. The 2D spectra were recorded with the following acquisition times (F2) or sweep widths (F1) and actual total number of data points: sensitivity-enhanced HSQC, F_1 15,094.3 Hz and 512, and F_2, 682 ms and 4096 (256 transients); HSQC–TOCSY, F_1 13,834.4 Hz and 1024, and F_2, 341 ms and 2048 (120 transients); gHSQC, F_1 24,140 Hz and 800, and F_2, 512 ms and 8192 (40 transients), TROSY, F_1 24,140 Hz and 800, and F_2, 512 ms and 8192 (40 transients). The mixing time for the HSQC–TOCSY (MLEV17) spectra was 65 ms. Phase-shifted squared sinebell apodization functions were applied to the 2D data in both dimensions and the first points were scaled. Data were zero-filled to 4096 (8192 for the 800 MHz spectra) and 2048 points in F2 and F1, respectively. A polynomial baseline correction was applied in F2 after Fourier transformation. The heteronuclear 500 and 800 MHz 2D data were processed with XWINNMR and NMRPipe, respectively.

3.3. Molecular modeling

The starting geometry and conformation of each of the monomeric units composing RG-II was taken from the MONOBANK database or from the previous conformational study.[12] In the case of aceric acid (α-L-Ace*f*A) the conformation of the furanose ring was obtained by complete optimization with MM3(92)[39,40] of the two minima found for the β-anomer after inversion of the configuration at the anomeric center.

3.4. Relaxed maps

The global shape of a disaccharide is mainly governed by the rotations about the glycosidic linkage.[41] The relative orientations of saccharide units are expressed in terms of the glycosidic linkage torsion angles Φ and Ψ which have the definition $\Phi = $ O-5–C-1–O1–C′-x and $\Psi = $ C-1–O-1–C′-x–C′-($x-1$) for a $1 \rightarrow x$ linkage. An additional rotatable bond ($\omega = $ C-4–C-3–C-3′–O-3′, in apiofuranoses) occurs in particular in the $1 \rightarrow 3'$ glycosidic linkage. Each linkage is denoted according to the corresponding non-reducing residue: e.g. A2(Φ) means the torsion angle Φ for the disaccharidic fragment A2 → A1. The conformational space of each of the three new disaccharide segments of RG-II was explored in a systematic fashion by stepping the glycosidic Φ and Ψ torsion angles in 10° increments over the whole angular range. At each conformational microstate a geometry optimization through MM3 was performed by allowing coordinates of each atom to vary except those defining the Φ and Ψ torsion angles. Relaxed maps were thus obtained, indicating roughly the geometry of the different minima for the new three segments.

3.5. Oligomeric fragments

Starting geometries of the oligomers and the complete molecule were constructed with the POLYS program.[42]

3.6. Systematic search

Systematic searches were performed with the SYBYL program.[28] Since this method does not apply any kind of energy minimization to the resulting conformers, strong steric interactions can arise that cause the conformation to become extremely unfavorable. In an attempt to reduce this kind of interaction, hydroxylic protons were removed, as their orientation could be easily modified in a further geometry optimization. In a first step, the torsion angles to be modified were defined. The range of values to be considered was taken from the energy maps and the step set to 30°, starting from the initial geometry. All of the conformations in which the distance between atoms was less than the sum of their van der Waals radii should in principle be rejected, although a softening multiplicator factor of 0.8–0.7 was applied to the van der Waals radii so that more conformations met this criterion. Finally, the distance constraints were defined. In most cases, this distance was selected so that only those conformers that provide a distance in the range of 0–5 Å were taken into account, without considering the intensity of the corresponding NOE. If the NOE included a methyl group, the distance implied the carbon atom and the range was increased to 0–6 Å.

3.7. Force fields

Geometry optimization of low molecular weight oligosaccharides and monosaccharides was performed using the molecular mechanics program MM3.[39,40] This force field contains a correction for the anomeric effect and has been shown to be especially well adapted for the study of carbohydrates. The block-diagonal minimization method, with the default energy convergence criterion of $0.00003 \cdot n$ kcal/mol per five iterations (n being the number of atoms), was used for grid point optimizations. To mimic a hydrated environment of the molecules, the dielectric constant of water ($\varepsilon = 80.0$) was used in all the calculations. Geometry optimization of high molecular weight fragments and the complete molecule was performed using the Tripos force field[29] included in the SYBYL program.[28] This force field has shown suitable for the study of carbohydrates. The charges were generated according to the algorithm of Gasteiger–Marsili,[43] and a dielectric constant of 80.0 was used to emulate the shielding of the solvent. Fifteen thousand iterations of Powell's algorithm[44] were applied, followed by 1000 iterations of BFGS,[45–48] with appropriate torsional constraints applied to the furanose rings of apiose and aceric acid in order to preserve their geometries. All calculations were carried out on a network of Silicon Graphics Indigo workstations.

4. Conclusions

The first study of the three-dimensional structure of monomer RG-II has been presented in this report. It began with the reassignment of the NMR spectra, owing to the modification of the primary structure indicated by Vidal and co-workers.[8] New ^{13}C NMR data have been presented that allowed us to assign most of the signals corresponding to the primary structure in Scheme 1. Unfortunately, several assignments are ambiguous and several of the spin systems, that have been clearly identified, have never been demonstrated in studies of RG-II. This study has revealed the presence of a RG-I backbone repeating unit in mRG-II-ol from red wine. The determination of the glycosylation site of this fragment would shed light on the relative locations of the RG-I and RG-II components in the pectic network.

In spite of this, a large body of NMR spectral data has been collected, especially concerning the assigned NOEs, and it has been applied to the study of the preferred conformation of RG-II in solution. The conformational analysis of RG-II started with the publication of Mazeau and Pérez[12] in which adiabatic energy surfaces, along with the locations of the local energy minima for all disaccharide components of RG-II were established. These data, along with the NOEs, were combined to determine possible time-averaged conformations of monomer RG-II in solution.

Two families of conformers, stemming from the two possible diastereospecific assignments of H-3′b(B1), have been obtained. Both of them are in agreement with other experimental data like the general shape of the molecule and the accessibility of apiose in A-chain. When calculated distances are compared with the NOEs (Table 4) good agreement is obtained, at least qualitatively in most cases; the intensity of the expected NOEs deviates from the experimental ones only for a very few interactions. Considering on one hand that the NOEs correspond to a time-averaged or virtual structure, and, on the other hand, that the isolated-spin approximation has been applied to systems with nearby protons, we consider the agreement to be very good.

This study has been an iterative process between the computational study and the NMR assignments. Thus, initial conformations for RG-II or its fragments have led to the revision of known signals and the assignment of new ones that, in turn, result in a new set of NOEs constraints or the modification of the previous ones. The resulting structures give slightly longer distances after minimization than expected in the case of most long-range NOEs. Nevertheless, the optimized conformers can be considered as good starting geometries for further studies of the molecular dynamics of the RG-II monomer and the conformational analysis of the RG-II dimer, which are being carried out in our group.

Acknowledgements

This study made use of the Avance 500 NMR spectrometer at the CGRM in Grenoble, the Varian INOVA-800 NMR spectrometer at the IBS 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). This research has been supported by a Marie Curie Fellowship of the European Community programme 'Improving the Human Research Potential and the Socio-Economic knowledge base' under the contract number HPMF-CT-2000-00717. Thanks are extended to B. Brutscher for help with the 800 MHz TROSY and HSQC experiments. The authors would like to acknowledge fruitful discussions with A. Imberty. Finally, we would also like to thank C. Monteiro for communicating her most recent experimental value for translational diffusion coefficient of mRG-II-ol.

References

1. Darvill, A. G.; McNeill, M.; Albersheim, P. *Plant Physiol*. **1978**, *62*, 418–422.
2. Blevins, D. G.; Lukaszewski, K. M. *Annu. Rev. Plant Physiol. Plant Mol. Biol*. **1998**, *49*, 481–500.
3. Ishii, T.; Matsunaga, T.; Pellerin, P.; McNeill, M.; Darvill, A. G.; Albersheim, P. *J. Biol. Chem*. **1999**, *274*, 13 098–13 104.
4. Ishii, T.; Matsunaga, T.; Hayashi, N. *Plant Physiol*. **2001**, *126*, 1698–1705.
5. Pellerin, P.; Doco, T.; Vidal, S.; Williams, P.; Brillouet, J.-M.; O'Neill, M. A. *Carbohydr. Res*. **1996**, *290*, 183–197.
6. Whitcombe, A.; O'Neill, M. A.; Steffan, W.; Albersheim, P.; Darvill, A. G. *Carbohydr. Res*. **1995**, *271*, 15–29.
7. Ishii, T.; Kaneko, S. *Phytochemistry* **1998**, *49*, 1195–1202.
8. Vidal, S.; Doco, T.; Williams, P.; Pellerin, P.; York, W. S.; O'Neill, M. A.; Glushka, J.; Darvill, A. G.; Albersheim, P. *Carbohydr. Res*. **2000**, *326*, 277–294.
9. Doco, T.; Williams, P.; Vidal, S.; Pellerin, P. *Carbohydr. Res*. **1997**, *297*, 181–186.
10. Shin, K.-S.; Kiyohara, H.; Matsumoto, T.; Yamada, H. *Carbohydr. Res*. **1997**, *300*, 239–249.
11. Stevenson, T. T.; Darvill, A. G.; Albersheim, P. *Carbohydr. Res*. **1988**, *182*, 207–226.
12. Mazeau, K.; Pérez, S. *Carbohydr. Res*. **1998**, *311*, 203–217.
13. Hervé du Penhoat, C.; Gey, C.; Pellerin, P.; Perez, S. *J. Biomol. NMR* **1999**, *14*, 253–271.
14. Bock, K.; Duus, J. O. *J. Carbohydr. Chem*. **1994**, *13*, 513–543.
15. Karlsson, C.; Jansson, P. E.; Widmalm, G.; Sørensen, U. B. S. *Carbohydr. Res*. **1997**, *304*, 165–172.
16. Molinaro, A.; Evidente, A.; Fiore, S.; Iacobellis, N. S.; Lanzetta, R.; Parrili, M. *Carbohydr. Res*. **2000**, *325*, 222–229.
17. Molinaro, A.; Evidente, A.; Lanzetta, R.; Parrilli, M.; Zoina, A. *Carbohydr. Res*. **2000**, *328*, 203–217.
18. Eichler, E.; Jennings, H. J.; Gilbert, M.; Whitfield, D. M. *Carbohydr. Res*. **1999**, *319*, 1–16.
19. Toukach, F. V.; Bartodziejska, B.; Senchenkova, S. N.; Wykrota, M.; Shashkov, A. S.; Rozalski, A.; Knirel, Y. A. *Carbohydr. Res*. **1998**, *318*, 146–153.
20. Shashkov, A. S.; Senchenkova, S.; Laux, P.; Ahohuendo, B. A.; Kecskés, M. L.; Rudolph, K.; Knirel, Y. A. *Carbohydr. Res*. **2000**, *323*, 235–239.
21. Jones, C.; Whitley, C.; Lemercinier, X. *Carbohydr. Res*. **2000**, *325*, 192–201.
22. Spellman, M. W.; McNeil, M.; Darvill, A. G.; Albersheim, P. *Carbohydr. Res*. **1983**, *122*, 131–153.
23. Castro, V. H.; Ramirez, E.; Mora, G. A.; Iwase, Y.; Nagao, T.; Okabe, H.; Matsunaga, H.; Katano, M.; Mori, M. *Chem. Pharm. Bull*. **1997**, *45*, 349–358.
24. Huisman, M. M. H.; Fransen, C. T. M.; Kamerling, J. P.; Vliegenthart, J. F. G.; Schols, H. A.; Voragen, A. G. *J. Biopolymers* **2001**, *58*, 279–294.
25. Gil-Serrano, A. M.; Rodríguez-Carvajal, M. A.; Tejero-Mateo, P.; Espartero, J. L.; Thomas-Oates, J.; Ruiz-Sainz, J. E.; Buendía-Clavería, A. M. *Biochem. J*. **1998**, *334*, 585–594.
26. Gil-Serrano, A. M.; Rodríguez-Carvajal, M. A.; Tejero-Mateo, P.; Espartero, J. L.; Menendez, M.; Corzo, J.; Ruiz-Sainz, J. E.; Buendía-Clavería, A. M. *Biochem. J*. **1999**, *342*, 527–535.
27. O'Neill, M. A.; Warrenfeltz, D.; Kates, K.; Pellerin, P.; Doco, T.; Darvill, A. G.; Albersheim, P. *J. Biol. Chem*. **1996**, *271*, 22 923–22 930.
28. Tripos Associates, S.H.R., Suite 303 *Sybyl*; 6.7 ed.: St. Louis, MO 63144, 2000.
29. Clark, M.; Cramer, R. D., III; Van Opdenbosh, N. *J. Comp. Chem*. **1989**, *10*, 982–1012.
30. Monteiro, C.; Hervé du Penhoat, C. *J. Phys. Chem*. **2001**, *105*, 9827–9833.
31. Bourlard, T.; Pellerin, P.; Morvan, C. *Plant Physiol. Biochem*. **1997**, *35*, 623–629.
32. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277–293.
33. Palmer, A. G., III; Cavanagh, J.; Wright, P. E.; Rance, M. *J. Magn. Reson*. **1991**, *93*, 151–170.
34. Kay, L. E.; Keifer, P.; Saarinen, T. *J. Am. Chem. Soc*. **1992**, *114*, 10 663–10 665.
35. Schleucher, J.; Schwendinger, M.; Sattler, M.; Schmidt, P.; Schedletsky, O.; Glaser, S. J.; Sorensen, O. W.; Griesinger, C. *J. Biomol. NMR* **1994**, *4*, 301–306.
36. Köver, K. E.; Hruby, V. J.; Uhrin, D. *J. Magn. Reson*. **1997**, *129*, 125–129.
37. Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12 366–12 371.
38. Brutscher, B.; Simorre, J. P. *J. Biomol. NMR* **2001**, *21*, 367–372.
39. Allinger, N. L.; Rahman, M.; Lii, J. H. *J. Am. Chem. Soc*. **1990**, *112*, 8293–8307.
40. Allinger, N. L.; Yuh, Y. H.; Lii, J. H. *J. Am. Chem. Soc*. **1989**, *111*, 8551–8566.
41. Tvaroska, I.; Bleha, T. *Adv. Carbohydr. Chem. Biochem*. **1989**, *47*, 45–123.
42. Engelsen, S. B.; Cros, S.; Mackie, W.; Pérez, S. *Biopolymers* **1996**, *39*, 417–433.
43. Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219–3228.
44. Powell, M. J. D. *Math. Program*. **1977**, *12*, 241–254.
45. Broyden, C. *J. Inst. Math. Appl*. **1970**, *6*, 76–90.
46. Fletcher, R. *Comp. J*. **1970**, *13*, 317–322.
47. Goldfarb, D. *Math. Comp*. **1970**, *24*, 23–26.
48. Shanno, D. *Math. Comp*. **1970**, *24*, 647–657.