



Structural studies of the O-antigen polysaccharide from *Escherichia coli* TD2158 having O18 serogroup specificity and aspects of its interaction with the tailspike endoglycosidase of the infecting bacteriophage HK620

Mona V. Zaccheus^a, Nina K. Broeker^b, Magnus Lundborg^a, Charlotte Uetrecht^{b,†}, Stefanie Barbirz^{b,*}, Göran Widmalm^{a,*}

^a Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, S-106 91 Stockholm, Sweden

^b Physical Biochemistry, Potsdam University, Karl-Liebknecht-Str. 24-25, D-14476 Potsdam-Golm, Germany

ARTICLE INFO

Article history:

Received 5 April 2012

Received in revised form 18 May 2012

Accepted 19 May 2012

Available online 28 May 2012

Keywords:

Escherichia coli

Tailspike

Endoglycosidase

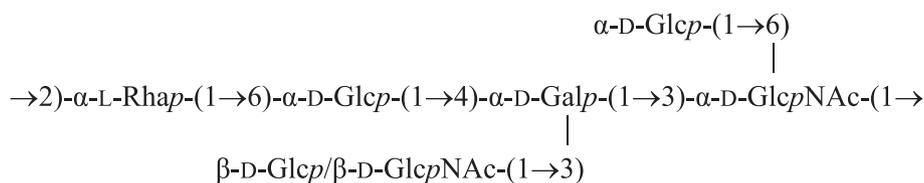
Lipopolysaccharide

NMR

Mass spectrometry

ABSTRACT

We have analyzed the O-antigen polysaccharide of the previously uncharacterized *Escherichia coli* strain TD2158 which is a host of bacteriophage HK620. This bacteriophage recognizes and cleaves the polysaccharide with its tailspike protein (TSP). The polysaccharide preparation as well as oligosaccharides obtained from HK620TSP endoglycosidase digests were analyzed with NMR spectroscopy. Additionally, sugar analysis was performed on the O-antigen polysaccharide and MALDI-TOF MS was used in oligosaccharide analysis. The present study revealed a heterogeneous polysaccharide with a hexasaccharide repeating unit of the following structure:



A repeating unit with a D-GlcNAc substitution of D-Gal has been described earlier as characteristic for serogroup O18A1. Accordingly, we termed repeating units with D-Glc substitution at D-Gal as O18A2. NMR analyses of the polysaccharide confirmed that O18A1- and O18A2-type repeats were present in a 1:1 ratio. However, HK620TSP preferentially bound the D-GlcNAc -substituted O18A1-type repeating units in its high affinity binding pocket with a dissociation constant of $140\ \mu\text{M}$ and disfavored the O18A2-type having a $\beta\text{-D-Glcp-(1}\rightarrow\text{3)}$ -linked group. As a result, in hexasaccharide preparations, O18A1 and O18A2 repeats were present in a 9:1 ratio stressing the clear preference of O18A1-type repeats to be cleaved by HK620TSP.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Bacterial cell wall surfaces contain many different carbohydrate structures.¹ Moreover, they can also produce capsular

* Corresponding authors.

E-mail addresses: barbirz@uni-potsdam.de (S. Barbirz), gw@organ.su.se (G. Widmalm).

[†] Present addresses: Sample Environment Group, European XFEL GmbH, Notkestr. 85, D-22607 Hamburg, Germany; Laboratory for Molecular Biophysics, Uppsala University, Box 596, S-75124 Uppsala, Sweden.

polysaccharides and exopolysaccharides which can be involved in the biofilm formation.² In Gram-negative bacteria high diversity in their O-polysaccharide part of lipopolysaccharide has been found.^{3,4} This part is also termed O-antigen because it is a serotyping feature of bacteria and defines their O-serospecificity.⁵ The O-antigen polysaccharide has multiple functions, for example, it prevents the bacterium from the host's immune response or from bacteriophage attack. On the other hand, phages also exploit the structural diversity of the O-antigen polysaccharide for initial attachment and infection.^{6,7} Bacteriophage HK620 is a temperate podovirus, which means that it possesses a dsDNA genome in an

icosahedral capsid and has a short, non-contractile tail.^{8,9} The phage uses its tailspike proteins (TSP) to specifically recognize and cleave the O-polysaccharide from *Escherichia coli* TD2158 LPS.¹⁰ TSP are large trimeric fibrous proteins with molecular weights of about 200 kDa.^{10–12} They have long shallow binding grooves and in case of *Salmonella* phage P22TSP micromolar dissociation constants have been determined for specific oligosaccharides.¹³ As a host phage HK620 uses the so far uncharacterized *E. coli* strain TD2158 isolated from sewage.⁸ When lipopolysaccharide preparations of this strain were incubated with HK620TSP, hexasaccharide fragments were produced. These could be purified and co-crystallized, showing that HK620TSP is an *endo*-N-acetylglucosaminidase.¹⁰ The O-antigen hexasaccharide fragment found in complex with the TSP was of serogroup O18A1.¹⁴ In this work we have analyzed the O-polysaccharide of *E. coli* TD2158 by NMR spectroscopy. Moreover, we analyzed the composition of oligosaccharide mixtures obtained by enzymatic action of HK620TSP. Our results show that O-antigen of *E. coli* TD2158 has a structure which differs from the classical O18A1-type and that HK620TSP can discriminate between the two serological types.

2. Results

2.1. Structural analysis of the O-antigen polysaccharide

Hydrolysis with aq trifluoroacetic acid and subsequent sugar analysis of the *E. coli* TD2158 polysaccharide (PS) showed Rha:Glc:Gal:GlcN in relative ratio of 10:58:20:12. The determination of the absolute configuration of sugar components revealed L-Rha, D-Glc, D-Gal, and D-GlcN, that is, sugars that are present in *E. coli* O18 serogroups.^{14,15} The ¹H NMR spectrum of the *E. coli* TD2158 PS (Fig. 1) was complex with several resonances discernible in the spectral region for anomeric protons, indicating that the polymer was not composed of regular repeating units. The structural determination was carried out using, in particular, 2D NMR spectroscopy.

The ¹H,¹³C-HSQC NMR spectrum of the O-antigen PS from *E. coli* TD2158 showed ten cross-peaks in the region of anomeric resonances (Fig. 2a). In both ¹H and 1D ¹³C NMR spectra, resonances of similar intensity were observed for resolved resonances, indicating that the suggested heterogeneity was due to approximately equal partial substitution, that is, in a 1:1 relative ratio. This is consistent with resonances from L-Rha residues at δ_{H6} 1.34 and 1.35. Furthermore, resonances from N-acetylated D-GlcN residues were present at δ_H 2.03 and 2.09, but in a 1:2 relative ratio. The sugar residues present in the PS are denoted A–L (12 in total) starting from the anomeric proton having the highest ¹H chemical shift of 5.49 ppm and decreasing to 4.65 ppm, respectively. The ¹³C NMR chemical shifts showed that the hexose sugar residues are present in the pyranoid ring form. Therefore, the anomeric configuration

can be deduced from the ¹J_{C-1,H-1} coupling constants,¹⁶ revealing that residues A–J have the α -anomeric configuration whereas residues K and L have the β -anomeric configuration (Table 1). Analysis of ¹H,¹H-TOCSY spectra revealed that residues A and B have the galacto-configuration and that residues I and J derived from rhamnose residues. Subsequent analysis identified three sugar residues with δ_C 56.71, 52.69, and 52.67 (Fig. 2b) for residues K, C, and D, respectively, originating from N-acetyl-D-glucosamine residues. The remaining sugar residues E–H and L consequently are glucose residues. The ¹³C NMR glycosylation shifts (Table 1) were used to identify the substitution positions.¹⁷ Thus, residues A and B are \rightarrow 3,4)- α -D-Galp-(1 \rightarrow , residues C and D are \rightarrow 3,6)- α -D-GlcpNac-(1 \rightarrow , residues E and H are \rightarrow 6)- α -D-Glcp-(1 \rightarrow , residues F and G are α -D-Glcp-(1 \rightarrow , residues I and J are \rightarrow 2)- α -L-Rhap-(1 \rightarrow , residue K is β -D-GlcpNac-(1 \rightarrow , and residue L is β -D-Glcp-(1 \rightarrow . For residues C and D as well as for residues E, F, and G severe spectral overlap occurred for the resonances from the respective anomeric protons (Table 1). ¹H,¹H-BASDH-TOCSY experiments¹⁸ were indispensable in elucidation of the ¹H,¹H-spin systems because a narrow spectral region can be selected in the F₁-dimension accompanied by homonuclear decoupling along the evolution dimension during t₁. The ¹H and ¹³C resonances from the N-acetyl groups of the D-glucosamines were readily assigned using the ¹H,¹³C-BS-CT-HMBC experiment (Fig. 3a and b). Hence, it was possible to assign and compile the ¹H and ¹³C chemical shifts of the 12 sugar residues (Table 1).

The information about the sequence of sugar residues was obtained from ¹H,¹H-NOESY (Fig. 3c) and ¹H,¹³C-HMBC experiments, in which informative correlations were obtained for all residues (Table 1). The particular connectivities from residues F and G to residues C and D or vice versa still remained unresolved due to the chemical shift degeneracy of the H6 and C6 resonances of the latter residues. However, from a primary structural point of view, this does not pose any problem because the same structural element, α -D-Glcp-(1 \rightarrow 6)- α -D-GlcpNac-(1 \rightarrow , is identified in both cases. Importantly, the sequential information identifies β -D-GlcpNac-(1 \rightarrow 3)- α -D-Galp-(1 \rightarrow for K–B, and β -D-Glcp-(1 \rightarrow 3)- α -D-Galp-(1 \rightarrow for L–A, consistent with heterogeneity due to approximately equal partial substitution in a 1:1 relative ratio. For the remaining sugar residues small chemical shift differences were present as a consequence of the structural heterogeneity. Sequential information was additionally observed for A–C, B–D, C–I, D–J, E–A, H–B, I–H, and J–E. The structure of the O-antigen PS from *E. coli* TD2158 can then be described by hexasaccharide repeating units in which one of the side-chains carries either β -D-GlcpNac or β -D-Glcp as terminal groups (Fig. 4).

When exclusively the β -D-GlcpNac-(1 \rightarrow 3)-linked group substitutes the backbone of the O-antigen polysaccharide it has been denoted O18A1,¹⁴ but its ¹H and ¹³C NMR chemical shifts were not reported. In the following, we denote the hexasaccharide

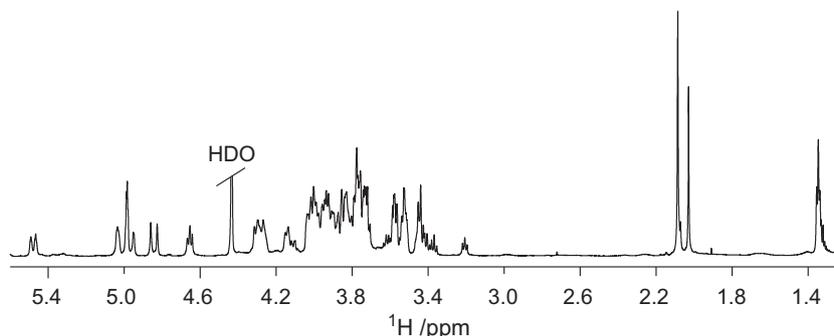


Figure 1. The ¹H NMR spectrum of the O-antigen PS from *E. coli* TD2158.

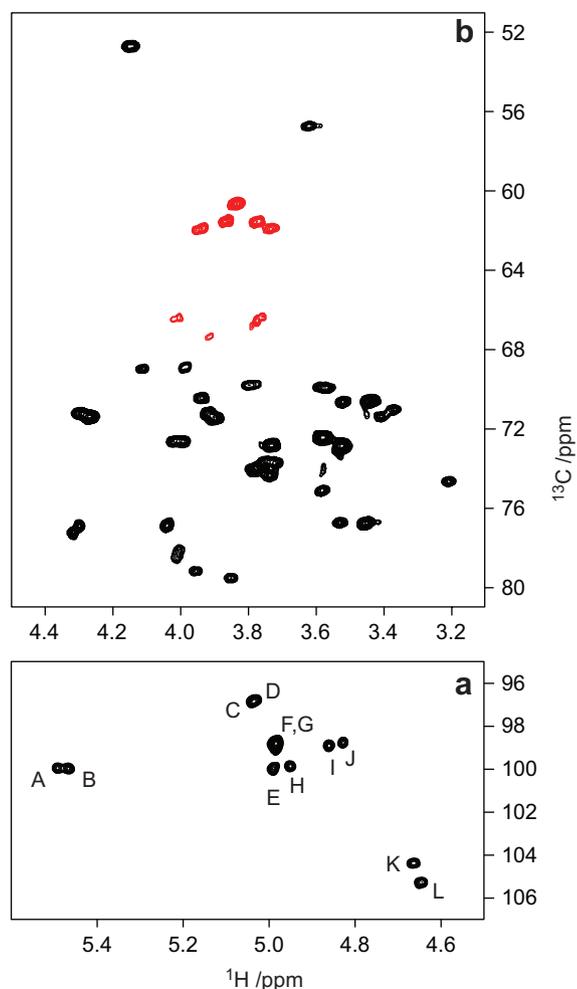


Figure 2. Selected regions of the multiplicity-edited ^1H , ^{13}C -HSQC NMR spectrum of the O-antigen PS from *E. coli* TD2158. The sugar residues are denoted A–L and cross-peaks in the anomeric region (a) are annotated accordingly; (b) the region for ring and hydroxymethyl atoms in which the cross-peaks from methylene groups are colored in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

repeat with the β -D-Glcp group as O18A2 although no specific antibody for serotyping has been isolated yet. It is important to note that the heterogeneity of the O-antigen polysaccharide found in the NMR analysis of polysaccharide preparations indeed resided on single polysaccharide chains and was not due to two different polymer populations. This was shown by endoglycosidase treatment of the polysaccharide and is described in the next section.

2.2. HK620TSP endoglycosidase produces a heterogeneous mixture of oligosaccharides

Polysaccharide of *E. coli* TD2158 was incubated with HK620TSP endoglycosidase and the reaction products were separated by size exclusion chromatography (SEC) as described previously.¹⁰ Oligosaccharides containing one to four repeats were isolated and analyzed with MALDI-TOF mass spectrometry^{19,20} (Table 2). All oligosaccharides isolated were mixtures of different compounds having different substituents on the backbone chain D-Gal residue, that is, SEC could not resolve whether D-Gal was substituted by D-GlcNAc (O18A1) or D-Glc (O18A2). Moreover, for preparations built of two, three, and four repeating units (RUs) MALDI-MS peaks occurred at m/z values that corresponded to oligosaccharides lacking a hexose residue. However, no further fragmentation

experiments were performed to obtain information whether the lacking hexose was the β -(1 \rightarrow 3)-linked side-chain branched D-Glc.

For the hexasaccharide corresponding to 1RU, D-GlcNAc- and D-Glc-containing oligosaccharides were found. Their ratio estimated from mass peak intensities was about 3:1, and when quantified with ^1H NMR spectroscopy about 9:1 (see below). In oligosaccharides of higher molecular mass the different types of RUs were found combined; for example, dodecasaccharides composed from two A2-type RUs or mixed A1A2-types (Table 2). In general, the relative intensity of homogenous A1-type oligosaccharides decreased whereas the intensity of mixed types increased with increasing oligosaccharide length. As a result, tetracosasaccharides composed of four A1-type RUs were not observed.

2.3. The hexasaccharide cleavage product has a β -D-GlcpNAc residue as a side-chain

We analyzed the smallest cleavage product, that is, the hexasaccharide, by ^1H NMR spectroscopy and found a major and a minor component in a \sim 9:1 relative ratio. In the ^1H , ^{13}C -HSQC spectrum of the major component eight cross-peaks were resolved in the region for anomeric resonances (Fig. 5a). In order to determine which glycosidic linkage in the O-antigen was cleaved by endoglycosidase, ^1H and ^{13}C NMR chemical shifts for different oligosaccharide structures were predicted using the computer program CASPER^{21,22} to aid the structural determination. ^{13}C chemical shifts in the spectral region 91–97 ppm are typical for reducing pyranoid sugar residues.²³ The cross-peak intensity of the α -anomeric form of the reducing end of the oligosaccharide was stronger than that of the β -anomeric form (Fig. 5a). For the monosaccharides in D_2O solution this is the case for D-GlcNAc and L-Rha but not for D-Glc and D-Gal where the β -anomeric form is favored. This suggests that one of the two former sugar residues is present at the reducing end. Further analysis of the ^1H , ^{13}C -HSQC spectrum showed three cross-peaks in the region for nitrogen-carrying carbon atoms (Fig. 5b); consistent with the above MS-analysis the oligosaccharide is a hexasaccharide containing two D-GlcNAc residues. Again, the α - to β -anomeric ratio favors one of the two former monosaccharides but the chemical shifts of these H-2 and C-2 atoms support D-GlcNAc as the reducing end residue. The additional cross-peak in this region should then come from the terminal β -D-GlcpNAc-(1 \rightarrow 3)-linked residue, the chemical shifts of which not being particularly perturbed by the two anomeric forms at the reducing end. Furthermore, the anomeric resonances at >5.4 ppm from the α -D-Galp residue (confirmed by ^1H , ^1H -TOCSY experiments) differed slightly (Fig. 5a), which is reasonable if this residue is then α -(1 \rightarrow 3)-linked to the reducing end D-GlcpNAc residue. Thus, the endoglycosidase cleaved the α -D-GlcpNAc-(1 \rightarrow 2)- α -L-Rhap-linkage of the O-antigen PS from *E. coli* TD2158.

2.4. Architecture of the hexasaccharide binding site determines the enzymatic specificity of HK620TSP

The accumulation of polysaccharide digestion products of A2-type may be explained by the architecture of the binding and cleavage site on HK620TSP (Fig. 6). In the hexasaccharide–protein complex the terminal β -D-GlcpNAc-(1 \rightarrow 3)-linked residue establishes a large hydrogen bond network to the protein, either directly or via ordered water molecules. Some contacts would not be possible if D-Glc was present instead of D-GlcNAc. In the latter, the carbonyl oxygen of the acetamido group accepts a hydrogen bond from O-4 as the donor of the backbone α -D-Glc residue (3.3 Å) and a water-mediated contact (2.6 Å) from Asn315. Moreover, N-2 donates a hydrogen bond to Glu400 (3.0 Å) and accepts an intramolecular hydrogen bond from O-2 of α -D-Gal (3.5 Å). As a consequence we hypothesize that D-Glc-containing

Table 1¹H and ¹³C NMR chemical shifts (ppm) of the resonances from the O-antigen polysaccharide of *E. coli* TD2158 and inter-residue correlations from ¹H,¹H-NOESY and ¹H,¹³C-HMBC spectra^a

Sugar residue		¹ H/ ¹³ C					Correlation to atom (from anomeric atom)		
		1	2	3	4	5	NOE	HMBC	
→3,4)-α-D-Galp-(1→	A	5.49	4.11	3.95	4.31	4.02	3.83, 3.83	H3, C	C3, C
		(0.27)	(0.33)	(0.14)	(0.36)	(−0.01)			
		99.99 [173]	68.92	79.10	77.23	72.60	60.65		
		(6.81)	(−0.41)	(8.97)	(6.95)	(1.30)	(−1.39)		
→3,4)-α-D-Galp-(1→	B	5.47	3.98	3.85	4.30	3.99	3.83, 3.83	H3, D	C3, D
		(0.25)	(0.20)	(0.04)	(0.35)	(−0.04)			
		99.99 [174]	68.91	79.45	76.85	72.60	60.59		
		(6.81)	(−0.44)	(9.32)	(6.57)	(1.30)	(−1.45)		
→3,6)-α-D-GlcpNAc-(1→ ^b	C	5.04	4.15	4.00	3.90	4.26	3.76, 4.01	H2, I H1, I	C2, I
		(−0.17)	(0.27)	(0.25)	(0.41)	(0.40)			
		96.92 [172]	52.69	78.51	71.38	71.36	66.37		
		(5.15)	(−2.31)	(6.77)	(0.12)	(−1.15)	(4.59)		
→3,6)-α-D-GlcpNAc-(1→ ^b	D	5.03	4.14	4.00	3.89	4.25	3.76, 4.01	H2, J H1, J	C2, J
		(−0.18)	(0.26)	(0.25)	(0.40)	(0.39)			
		96.78 [172]	52.67	78.04	71.48	71.36	66.37		
		(5.01)	(−2.33)	(6.30)	(0.22)	(−1.15)	(4.59)		
→6)-α-D-Glcp-(1→	E	4.99	3.53	3.75	3.58	4.28	3.80, 3.91	H4, A H6, A	C4, A
		(−0.24)	(−0.01)	(0.03)	(0.16)	(0.44)			
		99.99 [171]	72.71	73.60	69.90	71.36	67.21		
		(7.00)	(0.24)	(−0.18)	(0.81)	(−1.01)	(5.37)		
α-D-Glcp-(1→	F	4.985	3.57	3.78	3.44	3.73	3.78, 3.87	H6, C/D	C6, C/D
		(−0.25)	(0.03)	(0.06)	(0.02)	(−0.11)			
		98.88 [171]	72.40	73.98	70.55	72.77	61.54		
		(5.89)	(−0.07)	(0.20)	(−0.16)	(0.40)	(−0.30)		
α-D-Glcp-(1→	G	4.982	3.57	3.78	3.44	3.73	3.77, 3.86	H6, C/D	C6, C/D
		(−0.25)	(0.03)	(0.06)	(0.02)	(−0.11)			
		98.88 [172]	72.40	73.98	70.55	72.77	61.50		
		(5.89)	(−0.07)	(0.20)	(−0.16)	(0.40)	(−0.34)		
→6)-α-D-Glcp-(1→	H	4.95	3.52	3.72	3.56	4.30	3.92, 3.96	H4, B H6, B	C4, B
		(−0.28)	(−0.02)	(0.00)	(0.16)	(0.46)			
		99.87 [171]	72.77	73.68	69.90	71.11	67.36		
		(6.88)	(0.30)	(−0.10)	(0.81)	(−1.26)	(5.52)		
→2)-α-L-Rhap-(1→	I	4.86	4.04	3.94	3.53	3.80	1.35	H6a, H H6b, H	C6, H
		(−0.26)	(0.12)	(0.13)	(0.08)	(−0.06)	(0.07)		
		98.92 [169]	76.91	70.39	73.05	69.82	17.90		
		(4.08)	(5.10)	(−0.61)	(−0.14)	(0.70)	(0.23)		
→2)-α-L-Rhap-(1→	J	4.83	4.03	3.93	3.53	3.78	1.34	H6a, E H6b, E	C6, E
		(−0.29)	(0.11)	(0.12)	(0.08)	(−0.08)	(0.06)		
		98.77 [170]	76.74	70.39	73.05	69.77	17.85		
		(3.93)	(4.93)	(−0.61)	(−0.14)	(0.65)	(0.18)		
β-D-GlcpNAc-(1→ ^c	K	4.66	3.62	3.58	3.41	3.45	3.74, 3.95	H3, B	C3, B
		(−0.06)	(−0.03)	(0.02)	(−0.05)	(−0.01)			
		104.34 [161]	56.71	75.05	71.36	76.63	61.91		
		(8.49)	(−1.15)	(0.24)	(0.30)	(−0.19)	(0.06)		
β-D-Glcp-(1→	L	4.65	3.21	3.53	3.37	3.46	3.72, 3.93	H3, A	C3, A
		(0.01)	(−0.04)	(0.03)	(−0.05)	(0.00)			
		105.26 [161]	74.60	76.68	70.98	76.77	61.89		
		(8.42)	(−0.60)	(−0.08)	(0.27)	(0.01)	(0.05)		

^a $J_{H-1,C-1}$ values are given in Hz in square brackets. Chemical shift differences as compared to corresponding monosaccharides are given in parentheses.^b Chemical shifts for NAc are δ_H 2.09; δ_C 23.07 and 174.85.^c Chemical shifts for NAc are δ_H 2.03; δ_C 23.17 and 175.27.

oligosaccharides of A2-type have a reduced affinity to the HK620TSP binding pocket. At the C-terminus of the binding pocket, catalytic acids E372 and D339 are located¹⁰ and they would therefore preferentially cleave polysaccharide and oligosaccharides at A1-type repeating units.

Crystal structure analysis showed a binding pocket for a hexasaccharide (Fig. 6). HK620TSP contains seven tryptophane residues and two of them, Trp308 and Trp314, are located in the binding site where they may become restricted by the bound carbohydrate molecule (Fig. 6a). Fluorescence titration of HK620TSP with hexasaccharide showed a tryptophane fluorescence increase (Fig. 7a). This may be interpreted as a less efficient static quench

by the bound carbohydrate compared to a more efficient dynamic solvent quench in the apo protein. From the binding isotherm a dissociation constant of about 140 μ M was obtained for the hexasaccharide. To investigate whether another high affinity binding site existed we analyzed the kinetics of the fluorescence signal upon incubation with longer carbohydrate fragments, that is, dodecasaccharides (2RU) (Fig. 7b). Upon mixing an immediate initial signal increase to 129% occurred. If dodecasaccharides had a higher affinity to HK620TSP than hexasaccharides a signal decrease should occur because HK620TSP cleaves dodecasaccharides into hexasaccharides. However, the signal continued to rise which is in agreement with an increase in absolute saccharide concentra-

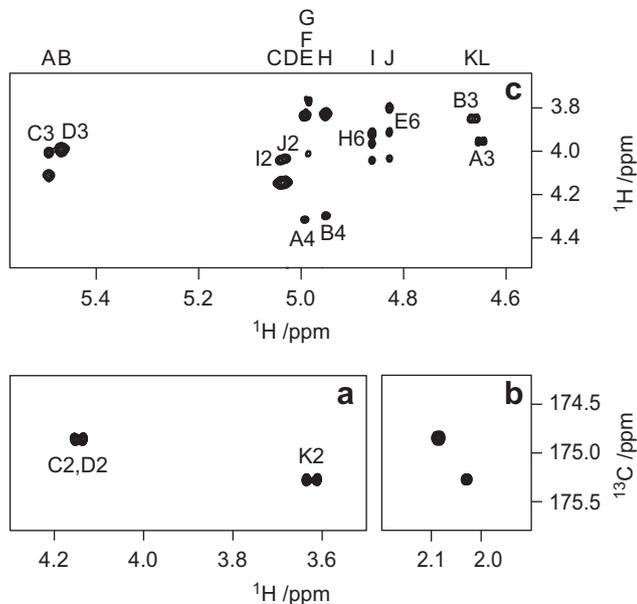


Figure 3. (a, b) Selected regions of the band-selective constant-time ^1H , ^{13}C -HMBC spectrum of the O-antigen PS from *E. coli* TD2158 with pertinent annotations; (c) part of the ^1H , ^1H -NOESY NMR spectrum with a mixing time of 100 ms of the O-antigen PS from *E. coli* TD2158. Residues are indicated by capital letters and ^1H or ^{13}C atom annotations by numbers.

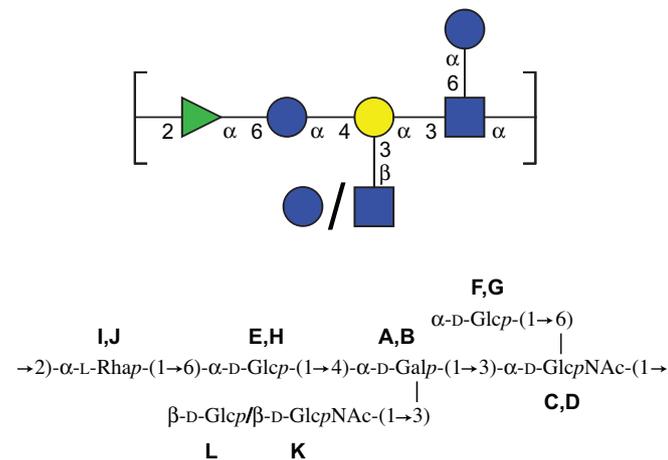


Figure 4. Structure of *E. coli* TD2158 O-antigen polysaccharide in CFG-notation (top) and standard nomenclature (bottom).

tion and an indication of similar affinities for both, hexa- and dodecasaccharides. We therefore conclude that there is no additional high affinity binding site. Thus, the sole binding site shows a structural preference for hexasaccharides of the O18A1-type, whereas O18A2-type RUs are disfavored. As a consequence, reduced cleavage rates will occur at O18A2 RUs, and likewise if the β -(1 \rightarrow 3)-linked residue is absent, in agreement with the observed masses of cleavage products in MALDI-MS and the 9:1 ratio of O18A1 to O18A2 hexasaccharides accumulated during TSP polysaccharide cleavage.

3. Discussion and conclusions

Our results showed that the O-antigen polysaccharide of *E. coli* TD2158 had an O18A-type backbone. However, the data revealed a heterogeneous β -(1 \rightarrow 3)-linked substitution at the D -Gal, which was either D -GlcNAc as in serogroup O18A1 or D -Glc termed

Table 2

MALDI-TOF MS analysis of oligosaccharide fractions obtained from gel permeation chromatography of *E. coli* TD2158 polysaccharide digests with HK620TSP

[M+Na] ⁺ _{exptl} / Da ^a	[M+Na] ⁺ _{calcd} / Da ^a	Relative intensity/%	O-Antigen composition ^b
1 RU			
1038.5	1038.5	25	O18A2
1079.6	1079.6	75	O18A1
2 RU			
1914.9	1915.0	7	O18A1 O18A1*
2035.9	2036.0	15	[O18A2] ₂
2076.9	2077.1	52	O18A1 O18A2
2117.9	2118.1	26	[O18A1] ₂
3 RU			
2912.8	2912.5	4	O18A1 O18A2 O18A1*
2953.8	2953.6	4	[O18A1] ₂ O18A1*
3034.8	3033.6	4	[O18A2] ₃
3074.8	3074.6	23	O18A1 [O18A2] ₂
3115.8	3115.7	37	[O18A1] ₂ O18A2
3156.8	3156.7	27	[O18A1] ₃
4 RU			
3909.5	3910.1	9	O18A1 [O18A2] ₂
3951.8	3951.1	10	O18A1* [O18A1] ₂ O18A2 O18A1*
3993.6	3992.2	10	[O18A1] ₃ O18A1*
4071.6	4072.1	13	O18A1 [O18A2] ₃
4113.7	4113.2	25	[O18A1] ₂ [O18A2] ₂
4154.6	4154.2	33	[O18A1] ₃ O18A2

^a Monoisotopic mass.

^b An asterisk denotes O18A1 fragments presumably lacking β -(1 \rightarrow 3)-substitution.

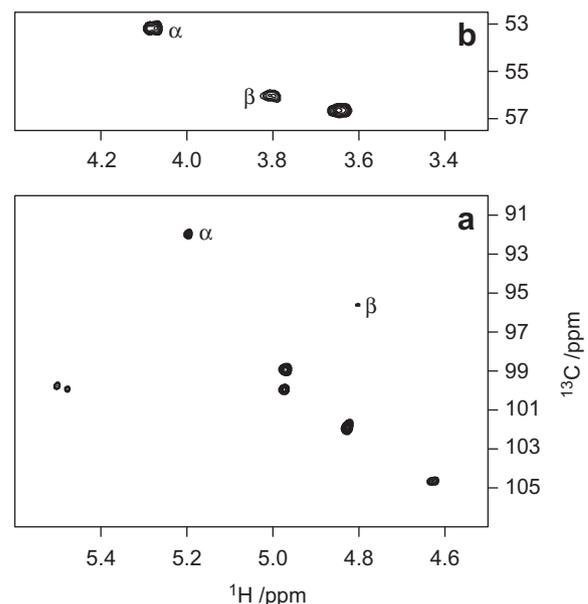


Figure 5. Selected regions of the ^1H , ^{13}C -HSQC NMR spectrum of a hexasaccharide produced from the *E. coli* TD2158 O-antigen polysaccharide by the cleavage action of the tailspike endoglycosidase of *E. coli* phage HK620. (a) The anomeric region in which the cross-peaks from the reducing end residue are annotated according to anomeric configuration, that is, α and β ; (b) the spectral region for C-2 resonances of D -GlcNAc of the hexasaccharide with the corresponding annotations.

O18A2. The latter is not a true serogroup because no defined antibody has been isolated so far that would be suitable for typing and distinguishing O18A2 unambiguously from other O18 serogroups. Moreover, in the present work we combined polysaccharide NMR spectroscopy with the analysis of TSP endoglycosidase-derived

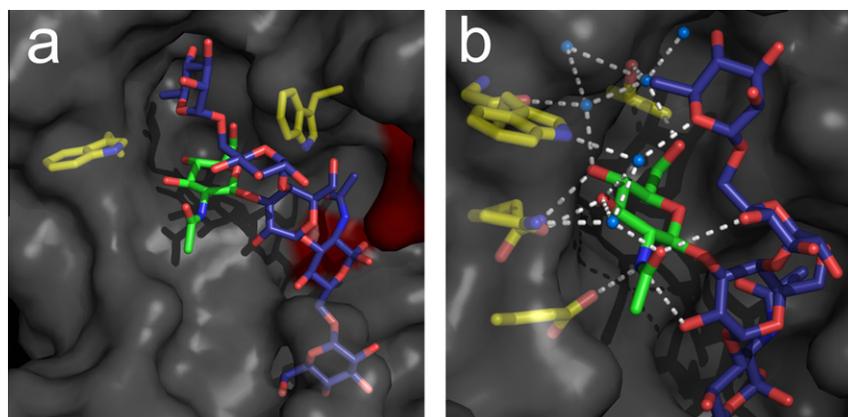


Figure 6. *E. coli* TD2158 hexasaccharide binding site on HK620TSP. (a) View of the full binding site with the active site residues E372 and D339 as red surfaces, and the β -D-GlcNAc-(1 \rightarrow 3)-linked residue in green. Amino acids W308 and W314 (yellow) show hyperfluorescence upon carbohydrate binding (cf. Fig. 7); (b) hydrogen bonding network in the D-GlcNAc binding pocket with interacting water molecules and amino acid side-chains (T311, G313, N314, N315, N346, E400) in yellow. Figure created with PyMOL (Schrödinger, LLC) from PDB entry 2vjj. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

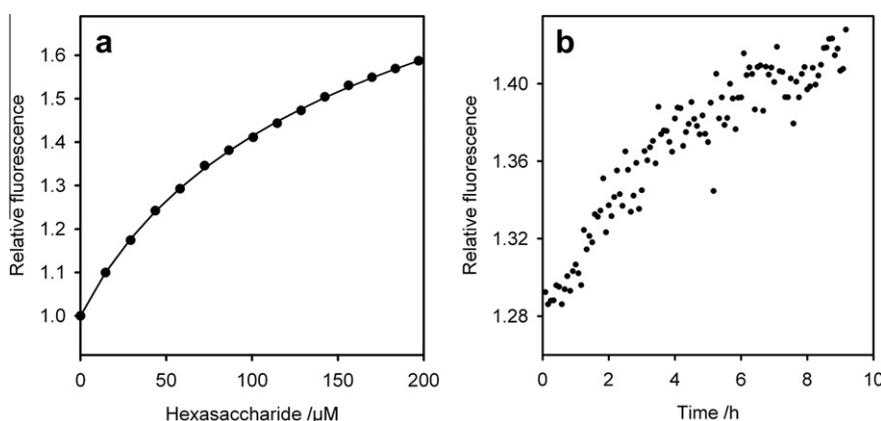


Figure 7. (a) Fluorescence titration of 178 nM HK620TSP with *E. coli* TD2158 hexasaccharide at 20 °C. Fitting of the data to a single independent site model yielded $K_D = 140 \mu\text{M}$; (b) kinetics of relative fluorescence change during incubation of 175 nM HK620TSP with 104 μM *E. coli* TD2158 dodecasaccharide at 10 °C.

oligosaccharides by NMR and MALDI MS. Whereas NMR analysis of the polysaccharide showed an approximate 1:1 ratio of D-GlcNAc and D-Glc substitutions at the D-Gal residue, oligosaccharide cleavage products had a different distribution, that is, $\sim 9:1$, for the hexasaccharides. This is in agreement with a preference of the HK620TSP binding site for O18A1-type hexasaccharide. Accordingly, as a result of incomplete polysaccharide hydrolysis, mixtures of larger oligosaccharides were obtained that were enriched in O18A2 repeating units. The distribution pattern of D-Glc or D-GlcNAc substitutions was irregular according to MALDI MS analysis; however, no quantification of single species was possible from the MALDI MS experiment because of the non-linear dependencies of ionization efficiencies.²⁴ Moreover, an accurate size distribution could not be obtained from size exclusion chromatography, which was unable to separate longer oligosaccharide fragments varying only by one or two monosaccharide residues. In principle, these could be separated by ion exchange albeit with very low yields of oligosaccharides (data not shown). As a consequence, it was not possible to obtain pure O18A2 hexasaccharide fragments in sufficient quantities to determine their affinity toward HK620TSP. Furthermore, given hexasaccharide preparations that contain a $\sim 9:1$ ratio of O18A1 to O18A2 structures, the dissociation constant obtained from fluorescence binding titrations can only be regarded as an estimate. Nevertheless, the dissociation constant of about 140 μM for HK620TSP hexasaccharide complexes was rather high when compared to the P22TSP octasaccharide complex having

$K_D = 3.9 \mu\text{M}$ at 25 °C.¹³ The reason for this difference is unclear and possibly related to the large amount of solvent liberated from the HK620TSP hexasaccharide binding site upon complex formation. Future studies could therefore investigate thermodynamics of the binding process. However, a prerequisite would be the use of oligosaccharides obtained from strains with a homogenous O18A1 polysaccharide. In addition, it is important to note that the heterogeneity of the polysaccharide is not an artifact of the delipidation procedure with acetic acid. MALDI MS analysis of oligosaccharides purified from *E. coli* strain IHE3042¹⁴ by the same procedure showed a homogenous polysaccharide preparation in agreement with the published O18A1 O-antigen polysaccharide structure (data not shown).

Heterogeneity in polysaccharides may be due to the presence of two populations of polysaccharides as found for the O-antigen polysaccharides from *Salmonella boecker*²⁵ and *Salmonella thompson* serogroup C1 (6,7).²⁶ Structural heterogeneity is caused either by partial^{27,28} or variable sugar substitution of the polymer backbone and can occur randomly or in blocks. The random substitution with either L-Rha and L-Man side-chain groups was found in the extracellular polysaccharide S-130, also known as Welan gum, produced by an *Alcaligenes* strain.^{29,30} In general, heterogeneity of the O-antigen polysaccharide is intimately linked to factors that influence its synthesis at all stages.^{1,5} Random distribution by a monosaccharide leading to branching of the polymer backbone as found in the present study points to a biosynthetic event that

occurs only after the oligosaccharide precursor of the repeating unit has been assembled and transferred to the periplasm.^{1,31} In *Salmonella* serogroups B and D introduction of an additional D-Glc was found to be a post-polymerization event with an unusual α -glucosylmonophosphorylundecaprenol donor.¹ Although fully plasmid-encoded O-antigen polysaccharide synthesis has been reported, it is more likely that alterations of single genes are due to serogroup conversion by temperate bacteriophages. No information exists on genes responsible for O-antigen polysaccharide synthesis in *E. coli* strain TD2158 used in the present study. Originally, this strain was isolated as a host for the temperate bacteriophage HK620,⁸ but the infection mechanism of this P22-like lipopolysaccharide recognizing phage is unknown. For phage P22, polysaccharide binding and hydrolysis by its TSP³² is a prerequisite for infection.⁶ The present study revealed a heterogeneous polysaccharide and a clear preference of HK620TSP for cleavage at O18A1-type over O18A2-type polysaccharide repeating units. However, if a random distribution of O18A2 repeating units occurs over the whole polysaccharide chain this will not prevent the phage from O-antigen polysaccharide hydrolysis with its TSP endoglycosidase and will not limit phage infection.

4. Experimental procedures

4.1. Materials

Escherichia coli TD2158 was obtained from A. J. Clark (University of Arizona, Tucson, AZ). All experiments were carried out with HK620TSP lacking the N-terminal domain as described previously.¹⁰

4.2. Preparation of polysaccharide

E. coli TD2158 cells⁸ were grown overnight at 18 °C in LB medium and harvested by centrifugation. Cells were washed with water and resuspended with 10% acetic acid and refluxed for 1.5 h.³³ After centrifugation the supernatant contained the polysaccharide; the pellet was treated once again as described. The combined supernatants were dialyzed against flowing water, cleared by centrifugation and lyophilized.

To remove DNA the crude material was subsequently treated batchwise with DE52 anion exchange cellulose (Whatman, Springfield Mill, UK). The polysaccharide (PS) was precipitated from aqueous solution with 80% ethanol three times and lyophilized. It was free of nucleic acids and proteins, as shown by the absence of DNA and protein absorbance.

4.3. Preparation of oligosaccharides via enzymatic digestion

The PS (10 mg) was dissolved in 200 μ L of 50 mM TrisHCl pH 6.0 and digested with 50 μ g mL⁻¹ of HK620TSP overnight at 45 °C. Remaining PS was precipitated with 80% ethanol. The supernatant contained the oligosaccharides which were separated by size exclusion chromatography (SEC) on a column of Superdex 30 26/60 (GE Healthcare, Freiburg, Germany) as described before.¹⁰ Collected oligosaccharides were rechromatographed twice by SEC and analyzed with MALDI-MS.

4.4. Fluorescence titrations

Protein fluorescence was measured on a Spex-FluoroLog-3 spectrofluorometer (HORIBA Jobin Yvon GmbH, Unterhaching, Germany) in stirred poly(methyl methacrylate) cuvettes (Carl Roth, Karlsruhe, Germany). Samples were excited at 295 nm and emission was recorded at 347 nm. Sample volumes did not change more than 5% upon oligosaccharide titration. Data were corrected

for the dilution and fitted to a binding isotherm with single independent sites as described.³⁴ In the kinetic measurements samples were mixed and protein fluorescence was recorded at 5 min intervals.

4.5. Component analyses

The PS was hydrolyzed with 2 M trifluoroacetic acid at 120 °C for 30 min. The sample was then reduced with NaBH₄ and acetylated, after which it was analyzed by gas liquid chromatography (GLC). The absolute configurations of the sugar components, save for rhamnose, were determined by GLC analysis of their acetylated (+)-2-butyl glycoside derivatives ((+)-2-butanol, AcCl, 80 °C, overnight) essentially as described³⁵ using also racemic 2-butanol. The absolute configuration of rhamnose was determined by GLC analysis of the acetylated (+)-2-octyl glycoside derivatives prepared in the corresponding manner employing also racemic 2-octanol.

4.6. GLC analyses

The alditol acetates were separated using a temperature program of 190 °C for 2 min, 4 °C min⁻¹ up to 220 °C and then 10 min at 220 °C. The injector and detector temperatures were set to 200 °C and 250 °C, respectively. The acetylated butyl glycosides were separated using a temperature program of 170 °C for 12 min, 10 °C min⁻¹ up to 220 °C, and then 2 min at 220 °C. The injector and detector temperatures were set to 180 °C and 250 °C, respectively. The octyl glycoside derivatives of rhamnose were separated using a temperature program of 180 °C for 2 min, 1 °C min⁻¹ up to 220 °C followed by 30 min at 220 °C. The injector and detector temperatures were set to 220 °C and 250 °C, respectively. All separations, with the exception of the octyl glycoside derivatives of rhamnose, were carried out on a PerkinElmer Elite-5 column with hydrogen as the carrier gas (25 psi/1724 hPa). The octyl rhamnose derivatives were separated on a PerkinElmer Elite-225 column. The columns were fitted to a PerkinElmer Clarus 400 Gas Chromatograph equipped with flame ionization detectors.

4.7. Mass spectrometry

MS analyses were carried out on a Bruker REFLEX II MALDI-TOF mass spectrometer (Bruker Daltonik, Bremen, Germany) after calibration with 100 pmol μ L⁻¹ Angiotensin II in α -cyano-4-hydroxycinnamic acid as a matrix. Carbohydrate samples were crystallized with 2,5-dihydroxybenzoic acid (15 mg mL⁻¹ in 20% (v/v) methanol) and measured in reflective mode at 25 kV according to the manufacturer's instructions.

4.8. NMR spectroscopy

NMR spectra of the PS (8.5 mg) in D₂O solution (0.55 mL) were recorded at 60 °C on a Bruker AVANCE III 700 MHz spectrometer equipped with a 5 mm TCI Z-Gradient Cryoprobe. Data processing was performed using vendor-supplied software. Chemical shifts are reported in ppm using internal sodium 3-trimethylsilyl-(2,2,3,3-²H₄)-propanoate (δ _H 0.00) or external 1,4-dioxane in D₂O (δ _C 67.40) as references.

The assignments of the ¹H and ¹³C resonances of the PS were obtained by analysis of 1D ¹H and ¹³C NMR spectra together with 2D NMR spectra from ¹³C, ¹H-HETCOR experiments, multiplicity-edited ¹H, ¹³C-HSQC experiments,³⁶ ¹H, ¹H-TOCSY experiments³⁷ with mixing times 10, 30, 60, 90, and 120 ms, a ¹H, ¹H-NOESY experiment³⁸ with a mixing time of 100 ms as well as a ¹H, ¹³C-H2BC experiment³⁹ with a constant time delay of 33 ms for the *J*_{H,H} evolution.

To assign ^1H spin-systems of the constituent monosaccharides that showed severe spectral overlap, band-selective homonuclear decoupled TOCSY experiments¹⁸ with mixing times of 50 and 150 ms were used. The selective excitation of anomeric protons was performed with a double pulsed field gradient spin echo (DPFGSE) sequence⁴⁰ using 11–17 Hz wide and 50–75 ms long Gaussian 180° pulses; 60 kHz broad 30 ms constant adiabatic WURST sweep pulses⁴¹ were used during zero-quantum suppression⁴² filters which were placed before and after the TOCSY mixing sequence. ^1H , ^{13}C -HSQC- ^1H , ^1H -TOCSY experiments^{43,44} with mixing times of 50, 100, and 200 ms were also performed.

The inter-residue correlations were assigned using the ^1H , ^1H -NOESY experiment and a ^1H , ^{13}C -HMBC experiment⁴⁵ with a 45 ms delay for evolution of long-range couplings. The band-selective constant-time ^1H , ^{13}C -HMBC experiment⁴⁶ with a 60 ms delay for evolution of long-range couplings was used to correlate the *N*-acetyl groups to corresponding sugar residues. The chemical shifts were compared to those of the corresponding monosaccharides.²³

Acknowledgements

We thank Sophie Haebel (Interdisciplinary Centre for Mass Spectrometry of Biomolecules, Universität Potsdam) for help with the MS analyses. This work was supported by grants from the Swedish Research Council, The Knut and Alice Wallenberg Foundation and the Deutsche Forschungsgemeinschaft [grant number BA 4046/1-1].

References

- Raetz, C. R.; Whitfield, C. *Annu. Rev. Biochem.* **2002**, *71*, 635–700.
- Whitfield, C. *Annu. Rev. Biochem.* **2006**, *75*, 39–68.
- Herget, S.; Toukach, P. V.; Ranzinger, R.; Hull, W. E.; Knirel, Y. A.; von der Lieth, C. W. *BMC Struct. Biol.* **2008**, *8*, 35.
- Jansson, P.-E.; Stenutz, R.; Widmalm, G. *Carbohydr. Res.* **2006**, *341*, 1003–1010.
- Wang, L.; Wang, Q.; Reeves, P. R. The Variation of O Antigens in Gram-Negative Bacteria. In *Endotoxins: Structure, Function and Recognition*; Wang, X., Quinn, P. J., Eds.; Springer: Dordrecht, 2010; pp 123–152.
- Andres, D.; Hanke, C.; Baxa, U.; Seul, A.; Barbirz, S.; Seckler, R. *J. Biol. Chem.* **2010**, *285*, 36768–36775.
- Andres, D.; Roske, Y.; Doering, C.; Heinemann, U.; Seckler, R.; Barbirz, S. *Mol. Microbiol.* **2012**, *83*, 1244–1253.
- Dhillon, T. S.; Poon, A. P. W.; Chan, D.; Clark, A. J. *FEMS Microbiol. Lett.* **1998**, *161*, 129–133.
- Clark, A. J.; Inwood, W.; Cloutier, T.; Dhillon, T. S. *J. Mol. Biol.* **2001**, *311*, 657–679.
- Barbirz, S.; Müller, J. J.; Uetrecht, C.; Clark, A. J.; Heinemann, U.; Seckler, R. *Mol. Microbiol.* **2008**, *69*, 303–316.
- Steinbacher, S.; Baxa, U.; Miller, S.; Weintraub, A.; Seckler, R.; Huber, R. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10584–10588.
- Müller, J. J.; Barbirz, S.; Heinle, K.; Freiberg, A.; Seckler, R.; Heinemann, U. *Structure* **2008**, *16*, 766–775.
- Baxa, U.; Cooper, A.; Weintraub, A.; Pfeil, W.; Seckler, R. *Biochemistry* **2001**, *40*, 5144–5150.
- Jann, B.; Shashkov, A. S.; Gupta, D. S.; Jann, K. *Eur. J. Biochem.* **1992**, *210*, 241–248.
- Jansson, P.-E.; Kenne, L.; Widmalm, G. *Carbohydr. Res.* **1989**, *193*, 322–325.
- Bundle, D. R.; Lemieux, R. U. *Methods Carbohydr. Chem.* **1976**, *7*, 79–86.
- Söderman, P.; Jansson, P.-E.; Widmalm, G. *J. Chem. Soc., Perkin Trans. 2* **1998**, 639–648.
- Krishnamurthy, V. V. *Magn. Reson. Chem.* **1997**, *35*, 9–12.
- Linnerborg, M.; Weintraub, A.; Widmalm, G. *Eur. J. Biochem.* **1999**, *266*, 246–251.
- Harvey, D. J. *Mass Spectrom. Rev.* **2012**, *31*, 183–311.
- Lundborg, M.; Widmalm, G. *Anal. Chem.* **2011**, *83*, 1514–1517.
- Lundborg, M.; Fontana, C.; Widmalm, G. *Biomacromolecules* **2011**, *12*, 3851–3855.
- Jansson, P.-E.; Kenne, L.; Widmalm, G. *Carbohydr. Res.* **1989**, *188*, 169–191.
- Hsu, N.-Y.; Yang, W.-B.; Wong, C.-H.; Lee, Y.-C.; Lee, R. T.; Wang, Y.-S.; Chen, C.-H. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 2137–2146.
- Brisson, J.-R.; Perry, M. B. *Biochem. Cell Biol.* **1988**, *66*, 1066–1077.
- Lindberg, B.; Leontein, K.; Lindquist, U.; Svenson, S. B.; Wrangsell, G.; Dell, A.; Rogers, M. *Carbohydr. Res.* **1988**, *174*, 313–322.
- Staa, M.; Urbina, F.; Weintraub, A.; Widmalm, G. *Eur. J. Biochem.* **1999**, *262*, 56–62.
- Säwén, E.; Östervall, J.; Landersjö, C.; Edblad, M.; Weintraub, A.; Ansaruzzaman, M.; Widmalm, G. *Carbohydr. Res.* **2012**, *348*, 99–103.
- Jansson, P.-E.; Widmalm, G. *Carbohydr. Res.* **1994**, *256*, 327–330.
- Kumar, N. S.; Ratnayake, R. M. S. K.; Widmalm, G.; Jansson, P.-E. *Carbohydr. Res.* **1996**, *291*, 109–114.
- Guan, S.; Bastin, D. A.; Verma, N. K. *Microbiology* **1999**, *145*, 1263–1273.
- Landström, J.; Nordmark, E.-L.; Eklund, R.; Weintraub, A.; Seckler, R.; Widmalm, G. *Glycoconjugate J.* **2008**, *25*, 137–143.
- Freeman, G. G.; Philpot, J. St. L. *Biochem. J.* **1942**, *36*, 340–356.
- Baxa, U.; Steinbacher, S.; Miller, S.; Weintraub, A.; Huber, R.; Seckler, R. *Biophys. J.* **1996**, *71*, 2040–2048.
- Leontein, K.; Lönngrén, J. *Methods Carbohydr. Chem.* **1993**, *9*, 87–89.
- Parella, T.; Sánchez-Ferrando, F.; Virgili, A. J. *Magn. Reson.* **1997**, *126*, 274–277.
- Braunschweiler, L.; Ernst, R. R. *J. Magn. Reson.* **1983**, *53*, 521–528.
- Kumar, A.; Ernst, R. R.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1–6.
- Nyberg, N. T.; Duus, J. Ø.; Sørensen, O. W. *J. Am. Chem. Soc.* **2005**, *127*, 6154–6155.
- Stott, K.; Stonehouse, J.; Keeler, J.; Hwang, T. L.; Shaka, A. J. *J. Am. Chem. Soc.* **1995**, *117*, 4199–4200.
- Kupče, Ě. *Methods Enzymol.* **2002**, *338*, 82–111.
- Thrippleton, M. J.; Keeler, J. *Angew. Chem., Int. Ed.* **2003**, *42*, 3938–3941.
- Domke, T. *J. Magn. Reson.* **1991**, *95*, 174–177.
- De Beer, T.; Van Zuylen, C. W. E. M.; Hård, K.; Boelens, R.; Kaptein, R.; Kamerling, J. P.; Vliegthart, J. F. G. *FEBS Lett.* **1994**, *348*, 1–6.
- Bax, A.; Summers, M. F. *J. Am. Chem. Soc.* **1986**, *108*, 2093–2094.
- Claridge, T. D. W.; Pérez-Victoria, I. *Org. Biomol. Chem.* **2003**, *1*, 3632–3634.