University of Groningen

Comparative structural characterization of 7 commercial galacto-oligosaccharide (GOS) products

van Leeuwen, Sander S; Kuipers, Bas J H; Dijkhuizen, Lubbert; Kamerling, Jannis P

Published in:
Carbohydrate Research

DOI:
[10.1016/j.carres.2016.03.006](https://doi.org/10.1016/j.carres.2016.03.006)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2016

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

van Leeuwen, S. S., Kuipers, B. J. H., Dijkhuizen, L., & Kamerling, J. P. (2016). Comparative structural characterization of 7 commercial galacto-oligosaccharide (GOS) products. *Carbohydrate Research*, 425, 48-58. <https://doi.org/10.1016/j.carres.2016.03.006>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Comparative structural characterization of 7 commercial galacto-oligosaccharide (GOS) products

Sander S. van Leeuwen^a, Bas J. H. Kuipers^b, Lubbert Dijkhuizen^{a,*}, Johannis P. Kamerling^a

^a Microbial Physiology, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, Nijenborgh 7, NL-9747 AG Groningen, The Netherlands

^b FrieslandCampina, Stationsplein 4, NL-3818 LE Amersfoort, The Netherlands

ARTICLE INFO

Article history:

Received 4 February 2016

Accepted 11 March 2016

Available online 16 March 2016

Keywords:

Galacto-oligosaccharides

NMR spectroscopy

Prebiotics

High-pH anion-exchange chromatography profiling

ABSTRACT

Many β -galactosidase enzymes convert lactose into a mixture of galacto-oligosaccharides (GOS) when incubated under the right conditions. Recently, the composition of commercial Vivinal GOS produced by *Bacillus circulans* β -galactosidase was studied in much detail in another study by van Leeuwen et al. As a spin-off of this study, we used the developed analytical strategy for the evaluation of 6 anonymous commercial GOS products, in comparison with Vivinal GOS. These GOS products were first subjected to HPLC-SEC, calibrated HPAEC-PAD profiling (glucose units in relation to a malto-oligosaccharide ladder), and 1D ^1H NMR spectroscopy. For a more detailed analysis and support of the conclusions based on the initial analysis, the GOS products were separated into DP-pure subpools on Bio-Gel P-2 (MALDI-TOF-MS analysis), which were subjected to calibrated HPAEC-PAD profiling and ^1H NMR analysis. Unidentified peaks from different GOS products, not present in Vivinal GOS, were isolated for detailed structural characterization. In this way, the differences between the various GOS products in terms of DP distribution and type of glycosidic linkages were established. A total of 13 new GOS structures were characterized, adding structural-reporter-group signals and HPAEC-PAD based glucose unit G.U. values to the analytical toolbox. The newly characterized products enhance the quality of the database with GOS structures up to DP4. The combined data provide a firm basis for the rapid profiling of the GOS products of microbial β -galactosidase enzymes.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Galacto-oligosaccharides (GOS) are mixtures of oligosaccharides with a degree of polymerization (DP) up to 9, produced by the *trans*-galactosylation activity of β -galactosidase enzymes (EC 3.2.1.23), using lactose as both donor and acceptor substrate.¹

Following the discovery of the *trans*-glycosylation potential of β -galactosidases, many of these enzymes have been studied,² especially those of *Lactobacillus reuteri*,^{3,4} *Bacillus circulans*,^{5–9} *Aspergillus oryzae*,^{10,11} *Kluyveromyces lactis*,¹² *Sporobolomyces singularis*,¹³ *Kluyveromyces fragilis*,¹⁴ and *Escherichia coli*.¹⁰ With respect to the characterization of GOS components, the ensemble produced by the *B. circulans* enzyme has been studied in most detail and up to the DP5 level.^{6–8,15–17} Structural studies of GOS mixtures produced by other β -galactosidases (mixtures) have focused in general only on the main DP2 and DP3 structures.^{7,18}

Previously, we have reported the detailed structural characterization of commercial Vivinal GOS produced by *B. circulans* β -galactosidase, making use of our analytical toolbox, which includes HPLC-SEC, HPAEC-PAD, LPLC-SEC, NMR spectroscopy, and MALDI-TOF-MS.^{8,17} Here, we describe the structural analysis of 6 other commercial GOS products, in comparison with Vivinal GOS. The detected differences between the various GOS products in terms of DP distribution and type of glycosidic linkages [(1→2), (1→3), (1→4), (1→6)], provide a firm basis for comparison of the structure/function relationships of (commercial) β -galactosidase enzymes. Knowing these structural differences in GOS products will also provide strong support in studies aiming to understand the prebiotic potential of their different structural elements.

2. Results and discussion

2.1. DP analysis of GOS I–VI

The commercial GOS **I–VI** products were analyzed by HPLC size-exclusion chromatography on a Rezex RSO-oligosaccharide Ag⁺ column, and the percentages of DP1–DP6 are summarized in Table 1. The DP1–DP6 data of the previously studied Vivinal GOS have been

* Corresponding author. Microbial Physiology, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, Nijenborgh 7, NL-9747 AG Groningen, The Netherlands. Tel.: +31 503632150; Fax: +31 503632154.

E-mail address: l.dijkhuizen@rug.nl (L. Dijkhuizen).

Table 1

Percentages of Gal, Glc and GOS DP2–DP6 in Vivinal GOS and in GOS samples **I–VI**, as determined by high-pressure liquid chromatography–size-exclusion chromatography (Rezex RSO-01), monitored by refractive-index detection. Values are averages of triplicate measurements

	Gal	Glc	DP2	DP3	DP4	DP5	DP6
Vivinal GOS	1.70 ± 0.08	18.52 ± 0.12	42.55 ± 0.22	23.63 ± 0.20	10.21 ± 0.23	3.00 ± 0.17	0.39 ± 0.17
GOS I	5.70 ± 0.11	18.04 ± 0.15	32.70 ± 0.24	34.98 ± 0.17	7.37 ± 0.23	1.21 ± 0.17	0
GOS II	0.69 ± 0.04	6.03 ± 0.07	39.30 ± 0.14	32.35 ± 0.13	14.34 ± 0.11	5.49 ± 0.11	1.80 ± 0.11
GOS III	3.30 ± 0.07	5.73 ± 0.12	45.69 ± 0.13	32.34 ± 0.13	11.40 ± 0.11	1.54 ± 0.11	0
GOS IV	0.12 ± 0.02	0.17 ± 0.05	20.04 ± 0.13	46.26 ± 0.16	21.94 ± 0.11	8.48 ± 0.15	2.99 ± 0.08
GOS V	12.48 ± 0.14	25.88 ± 0.11	36.24 ± 0.12	16.56 ± 0.11	6.70 ± 0.15	2.14 ± 0.17	0
GOS VI	1.16 ± 0.07	1.62 ± 0.04	47.21 ± 0.09	36.15 ± 0.08	12.40 ± 0.04	1.46 ± 0.02	0

included for comparison.⁸ It should be noted that data about the production protocols were not available to us. All products showed monosaccharides Gal and Glc, together with DP2 up to DP5 oligosaccharides in detectable levels. Vivinal GOS, GOS **II** and GOS **IV** showed also clear amounts of DP6 oligosaccharides. Except for GOS **IV**, the DP2–DP4 fractions of the various products contain >90% of all GOS components. Most products contain a majority of DP2 components, except for GOS **I** and **IV**, where DP3 is the predominant fraction. GOS **IV** stands out, since it has the highest DP4, the highest DP5, and the highest DP6 content of all GOS products; here the very low amounts of Gal and Glc suggest a DP-based purification step. Also GOS **VI** has minor amounts of Gal and Glc, which may reflect a DP-based purification step.

2.2. Structural analysis of GOS I–VI

The HPAEC CarboPac PA-1 profiles and 1D ¹H NMR spectra of Vivinal GOS and GOS **I–VI** are collected in Figs. 1–3. The GOS **I–V** products were separated into DP subpools by preparative size-exclusion chromatography on Bio-Gel P-2. The isolated fractions were screened by MALDI-TOF-MS, and DP-pure fractions of DP2, DP3, DP4, and DP5 were pooled for further analysis. The HPAEC CarboPac PA-1 profiles and 1D ¹H NMR spectra of these subpools, together with those of Vivinal GOS for comparison,⁸ are collected in Supplementary Information Figs. S3–S14. Scheme 1 summarizes the structures of the different GOS components, assigned in the various HPAEC profiles. Of these components, compounds **1–24** and **xa–zc** were earlier isolated and identified from Vivinal GOS.^{8,17} Compounds **25–27**, **28–33**, and **34–37**, isolated from relevant GOS **I**, GOS **II**, and GOS **III** DP subpools, respectively, were characterized in the present study and their NMR data are presented in Table 2. Descriptions of their structural characterizations and their 1D ¹H NMR spectra (Figs. S1 and S2) are provided as Supplementary Information. GOS **VI** was only investigated as total GOS. Table 3 gives an overview of the components found in Vivinal GOS and GOS **I–VI**.

2.2.1. GOS I preparation

HPAEC-PAD analysis of GOS **I** (Fig. 1B) yielded a profile of which peaks **1–11**, **13**, **15**, **17**, **18**, and **22**, but not **25**, **26**, and **27**, could be annotated by comparison with Vivinal GOS G.U. (glucose units) values (Fig. 1A and Scheme 1).^{8,17} Comparison of the relative percentages of Gal (**1**) and Glc (**2**) with those of Vivinal GOS showed a higher percentage for Gal. Applying the earlier developed structural-reporter-group concept for GOS components,^{8,17} evaluation of the 1D ¹H NMR spectrum of GOS **I** (Fig. 2B) gave insight into most basic structural elements of the components present (Table 4). In the α -anomeric region, these are the H-1 signals **a** of $\rightarrow$ 2)- α -D-Glcp, **b** of $\rightarrow$ 2)- α -D-Glcp of the $\rightarrow$ 4)- β -D-Galp-(1 $\rightarrow$ 2)- α -D-Glcp sequence, **h** of free α -D-Galp, **c** of $\rightarrow$ 3)- α -D-Glcp, and **d** of $\rightarrow$ 4, 6 and/or 1)- α -D-Glcp and of free α -D-Glcp. The β -anomeric region shows a quite heterogeneous signal pattern. Then, in the region 4.25–4.00 ppm the H-6a **e** and **f** signals of $\rightarrow$ 6)-D-Glcp ($\beta\alpha$), the H-4 signal **g** of $\rightarrow$ 3 and/or 4)- β -D-Galp, and the H-6a signal **i** of $\rightarrow$ 6)- β -D-Galp are present.

Signal **b** is very minor. The clear presence of signal **h** in GOS **I** reflects the higher amounts of free Gal, when compared with Vivinal GOS (Fig. 2A, B and Table 1). To support the various assignments based on G.U. values and NMR data, the five GOS **I** DP-pure subpools were further analyzed by HPAEC-PAD and ¹H NMR spectroscopy.

The HPAEC-PAD profile of GOS **I** DP2 (Supplementary Information Fig. S4A and Scheme 1) revealed peaks at G.U. values corresponding with β -D-Galp-(1 $\rightarrow$ 6)-D-Galp (**3**), β -D-Galp-(1 $\rightarrow$ 6)-D-Glcp (**4**; allolactose), β -D-Galp-(1 $\rightarrow$ 4)-D-Glcp (**5**; lactose), β -D-Galp-(1 $\rightarrow$ 4)-D-Galp (**7**), β -D-Galp-(1 $\rightarrow$ 2)-D-Glcp (**8a**), and β -D-Galp-(1 $\rightarrow$ 3)-D-Glcp (**8b**).⁸ Comparison of the profile with that of Vivinal GOS DP2 (Supplementary Information Fig. S3A) shows that peak **4** has a higher relative intensity and peak **8** a lower relative intensity. The various fractions were isolated by semi-preparative HPAEC on CarboPac PA-1 and the identity of the compounds was verified by 1D ¹H NMR spectroscopy. Inspection of the 1D ¹H NMR spectrum of GOS **I** DP2 (Supplementary Information Fig. S10B) revealed the different structural-reporter-group signals expected for the different disaccharides,^{8,17} as summarized in Table 4.

The most significant peaks in the HPAEC-PAD profile of GOS **I** DP3 (Supplementary Information Fig. S4B and Scheme 1) are **6** corresponding with 4,6-digalactosyl-glucose/ β -D-Galp-(1 $\rightarrow$ 4)-[β -D-Galp-(1 $\rightarrow$ 6)-]D-Glcp (**6a**) and 6'-galactosyl-lactose/ β -D-Galp-(1 $\rightarrow$ 6)- β -D-Galp-(1 $\rightarrow$ 4)-D-Glcp (**6b**), **11** corresponding with 4'-galactosyl-lactose/ β -D-Galp-(1 $\rightarrow$ 4)- β -D-Galp-(1 $\rightarrow$ 4)-D-Glcp, and **13**, corresponding with β -D-Galp-(1 $\rightarrow$ 4)- β -D-Galp-(1 $\rightarrow$ 2)-D-Glcp (**13a**) and β -D-Galp-(1 $\rightarrow$ 4)- β -D-Galp-(1 $\rightarrow$ 3)-D-Glcp (**13b**), like previously observed for Vivinal GOS (Supplementary Information Fig. S3B).⁸ Inspection of the 1D ¹H NMR spectrum (Supplementary Information Fig. S10C), however, showed the absence of structural-reporter-group signal **p**, corresponding with $\rightarrow$ 4,6)- β -D-Glcp H-6a, indicating the absence of structure **6a**. The structural-reporter-group signals for both 2- (signal **b**) and 3- (signal **c**) substituted Glc are present, verifying the presence of both **13a** and **13b**. Other assigned HPAEC peaks are **9** (branched trisaccharide), **10** (branched trisaccharides), and **12**; HPAEC peak **25** is absent in Vivinal GOS. The various fractions were isolated by HPAEC and the identity of the compounds, except **25**, was verified by 1D ¹H NMR spectroscopy. Detailed NMR analysis of fraction **25** showed its structure to be β -D-Galp-(1 $\rightarrow$ 6)- β -D-Galp-(1 $\rightarrow$ 3)-D-Glcp/C1 $\rightarrow$ 6B1 $\rightarrow$ 3A (Table 2; see Supplementary Information). Inspection of the 1D ¹H NMR spectrum of GOS **I** DP3 (Supplementary Information Fig. S10C) revealed the structural-reporter-group signals that indicate typical fragments of the identified trisaccharides,^{8,17} as summarized in Table 4. Component **25** can be considered as a β -D-Galp-(1 $\rightarrow$ 6)- elongation of the terminal Gal residue of disaccharide **8b**. The same holds for **6b**, being a similar extension of **5**. No β -D-Galp-(1 $\rightarrow$ 6)- elongations of disaccharides **4** and **8a** were observed. Compounds **11**, **13a**, and **13b** are β -D-Galp-(1 $\rightarrow$ 4)- elongations of the terminal Gal residue of disaccharides **5**, **8a**, and **8b**, respectively, and the minor compound **12** is a β -D-Galp-(1 $\rightarrow$ 3)- elongation of the terminal Gal residue of disaccharide **5**. Analysis by ¹H NMR indicates that HPAEC peak **10** represents both compounds **10a** and **10b** (Scheme 1).

Fig. 1. HPAEC-PAD profiles of A. Vivinal GOS (adapted from van Leeuwen et al.⁸), B. GOS I, C. GOS II, D. GOS III, E. GOS IV, F. GOS V, and G. GOS VI. Assigned peaks are numbered, matching structures presented in Scheme 1. * Lactulose generated from lactose during the industrial preparation of the batches.

The major peaks **26** and **27** in the HPAEC-PAD profile of GOS I DP4 (Supplementary Information Fig. S4C and Scheme 1) could not be correlated with G.U. values reported for Vivinal GOS.⁸ After isolation via HPAEC, they were identified as β -D-Galp-(1 $\rightarrow$ 6)- β -D-Galp-(1 $\rightarrow$ 4)- β -D-Galp-(1 $\rightarrow$ 4)-D-Glcp and β -D-Galp-(1 $\rightarrow$ 6)- β -D-Galp-(1 $\rightarrow$ 4)- β -D-Galp-(1 $\rightarrow$ 3)-D-Glcp, respectively (Table 2; see Supplementary Information). So, structures **26** and **27** represent β -D-Galp-(1 $\rightarrow$ 6)- elongations of two major structures in the DP3 subpool,

i.e. **11** and **13b**, respectively. Other assigned HPAEC peaks are **15**, **17**, and **18**. The presence of both **15a** and **15b** was confirmed by ¹H NMR analysis. The same held for compounds **18a** and **18b**. Inspection of the 1D ¹H NMR spectrum of GOS I DP4 showed structural-reporter-group resonances that support typical fragments of the identified tetrasaccharides^{8,17} as summarized in Table 4. Structures **17**, **18a**, and **18b** are β -D-Galp-(1 $\rightarrow$ 4)- elongations of **11**, **13a**, and **13b**, respectively. Based on the structural analysis of compound **34** in GOS III (see below), the small HPAEC peak at 1.74 is suggested to correlate with β -D-Galp-(1 $\rightarrow$ 6)- β -D-Galp-(1 $\rightarrow$ 6)- β -D-Galp-(1 $\rightarrow$ 4)-D-Glcp, i.e. a β -D-Galp-(1 $\rightarrow$ 6)- elongation of **6b** (Δ G.U. +0.04).

The GOS I DP5 subpool, of which the HPAEC-PAD profile (1.2% of GOS I) is presented in Supplementary Information Fig. S4D, contained too low quantities for isolation and NMR identification of detected fractions (Table 1). Based on the identified structures, when going from DP3 to DP4, it is hypothesized that the main elongations with one β -D-Galp unit of DP4 structures should comprise the addition of β -D-Galp-(1 $\rightarrow$ 4)- and β -D-Galp-(1 $\rightarrow$ 6)- residues. Inspection of the 1D ¹H NMR spectrum of GOS I DP5 (Supplementary Information Fig. S10E) supports this hypothesis. The signals **i** in the range δ 4.08–4.05 can be attributed to H-6a of $\rightarrow$ 6)- β -D-Galp residues; the signals **g** in the range δ 4.21–4.17, attributed to Gal H-4 atoms, fit with β -D-Galp-(1 $\rightarrow$ 4)- elongations. The HPAEC peaks at 3.50 and 3.85 G.U. reflect the addition of one β -D-Galp-(1 $\rightarrow$ 4)- unit to compounds **17** and **18ab**, yielding compounds **22** (+0.64 G.U.) and **23ab** (+0.72 G.U.),⁸ respectively (Scheme 1). Taking into account the Δ G.U. shifts for **5** to **6b** (+0.20), for **11** to **26** (+0.17), and for **6b** to **34** (+0.04), meaning the addition of one β -D-Galp-(1 $\rightarrow$ 6)- residue, it is suggested that the DP5 HPAEC peak with 2.40 G.U. originates from the addition of a β -D-Galp-(1 $\rightarrow$ 6)- residue to structure **26** (Δ G.U. +0.15), resulting in β -D-Galp-(1 $\rightarrow$ 6)- β -D-Galp-(1 $\rightarrow$ 6)- β -D-Galp-(1 $\rightarrow$ 4)- β -D-Galp-(1 $\rightarrow$ 4)-D-Glcp. Following a similar reasoning, it is hypothesized that one of the two peaks at 2.67 and 2.71 G.U. could correlate with a β -D-Galp-(1 $\rightarrow$ 6)- elongation of structure **27** (Δ G.U. +0.08 and +0.12) [β -D-Galp-(1 $\rightarrow$ 6)- β -D-Galp-(1 $\rightarrow$ 6)- β -D-Galp-(1 $\rightarrow$ 4)- β -D-Galp-(1 $\rightarrow$ 3)-D-Glcp]. Based on the Δ G.U. shifts mentioned above, it is also hypothesized that the peak at 2.88 G.U. could correlate with a β -D-Galp-(1 $\rightarrow$ 4)- extension of structure **26** (Δ G.U. +0.63) and/or a β -D-Galp-(1 $\rightarrow$ 6)- extension of branched structures **15**, e.g. **15a** (Δ G.U. +0.19).

2.2.2. GOS II preparation

The HPAEC-PAD profile of GOS II (Fig. 1C) shows a series of major and minor peaks, of which **1**, **2**, **4**–**13**, and **17** could be annotated by comparison with Vivinal GOS G.U. values (Scheme 1).^{8,17} Peaks **28**–**33** did not fit G.U. values annotated for Vivinal GOS and GOS I. The ratio Gal (**1**):Glc(**2**) is comparable with the ratio found in Vivinal GOS (Fig. 1A). Peak **12** (3'-galactosyl-lactose) is of much higher intensity in GOS II than in Vivinal GOS and GOS I. Inspection of the 1D ¹H NMR spectrum of GOS II (Fig. 2C) gave evidence for the following basic structural elements (Table 4): H-1 signal **a**, $\rightarrow$ 2)- α -D-Glcp; H-1 signal **c**, $\rightarrow$ 3)- α -D-Glcp; H-1 signal **d**, $\rightarrow$ 4, 6 and/or 1)- α -D-Glcp and free α -D-Glcp; H-6a signals **e** and **f**, $\rightarrow$ 6)-D-Glcp (β); H-4 signal **g**, $\rightarrow$ 3 and/or 4)- β -D-Galp. The H-1 signal **m** is in agreement with an internal β -D-Galp unit in the sequence β -D-Galp-(1 $\rightarrow$ 3)- β -D-Galp-(1 $\rightarrow$ 4)-D-Glcp. To support the various assignments based on G.U. values and NMR data and to identify the new fractions **28**–**33**, the five DP-pure subpools, prepared from GOS II, were further studied by HPAEC-PAD and ¹H NMR spectroscopy.

The HPAEC-PAD profile of GOS II DP2 (Supplementary Information Fig. S5A) showed peaks at G.U. values that correspond with allolactose (**4**), lactose (**5**), β -D-Galp-(1 $\rightarrow$ 4)-D-Glcp (**7**), β -D-Galp-(1 $\rightarrow$ 2)-D-Glcp (**8a**), and β -D-Galp-(1 $\rightarrow$ 3)-D-Glcp (**8b**), previously described in Vivinal GOS⁸ and also present in GOS I DP2. Here, the shoulder on the right side of peak **8**, i.e. **8b**, is larger than in Vivinal GOS and GOS I DP2. Like in the case of GOS I DP2, the

Fig. 2. One-dimensional 600-MHz ^1H NMR spectra of A. Vivinal GOS (adapted from van Leeuwen et al.⁸), B. GOS I, C. GOS II, and D. GOS III. For assignments of structural-reporter-group signals, see Table 4.

various fractions were isolated, followed by ^1H NMR verification. Furthermore, the 1D ^1H NMR spectrum of GOS II DP2 (Supplementary Information Fig. S11B) contained the different structural-reporter-group signals belonging to the different disaccharides (Table 4). The low-intensity signals **e** and **f** indicate a relatively low amount of allolactose (**4**). The very low amount of **7** is not visible in the ^1H NMR spectrum (signal **g**).

The structures corresponding with the peaks **6**, **9**, **10**, **11**, **12**, and **13** in the HPAEC-PAD profile of GOS II DP3 (Supplementary Information Fig. S5B) are presented in Scheme 1. In contrast to the profiles for Vivinal GOS and GOS I DP3, here the chromatogram is dominated by peak **12**, reflecting the presence of relatively high amounts of 3'-galactosyl-lactose/ β -D-Galp-(1 $\rightarrow$ 3)- β -D-Galp-(1 $\rightarrow$ 4)-D-Glcp, instead of **11** (4'-galactosyl-lactose). Peaks **28**, **29**, and **30** in GOS II showed up in the DP3 subpool. All fractions were isolated, followed by ^1H NMR verification of the already assigned ones and structural analysis of the unknowns, which were identified as β -D-Galp-(1 $\rightarrow$ 3)- β -D-Galp-(1 $\rightarrow$ 6)-D-Glcp (**28**), β -D-Galp-(1 $\rightarrow$ 3)- β -D-Galp-(1 $\rightarrow$ 3)-D-Glcp (**29**), and β -D-Galp-(1 $\rightarrow$ 3)- β -D-Galp-(1 $\rightarrow$ 2)-D-Glcp (**30**) (Table 2; see Supplementary Information), being β -D-Galp-(1 $\rightarrow$ 3)- elongations of disaccharides **4**, **8b**, and **8a**, respectively (Scheme 1). Typical fragments of the identified trisaccharides can be found back directly via structural-reporter-group signals in the 1D ^1H NMR spectrum of GOS II DP3 (Supplementary

Information Fig. S11C; Table 4).^{8,17} The absence of the H-6a resonance of $\rightarrow$ 4,6)- β -D-Glcp at δ 4.29 (signal **p**) indicates that peak **6** consists exclusively of **6b**. The clear H-1 signal **m** stems from the internal β -D-Galp unit in the sequence β -D-Galp-(1 $\rightarrow$ 3)- β -D-Galp-(1 $\rightarrow$ 4)-D-Glcp (**12**). The significant H-1 signal at δ 4.677 supports the presence of a 3-substituted β -D-Galp-(1 $\rightarrow$ 3)- unit in the β -D-Galp-(1 $\rightarrow$ 3)- β -D-Galp-(1 $\rightarrow$ 3)- sequence (**29**).¹⁷ The presence of both **10a** and **10b** was confirmed by ^1H NMR analysis. The absence of structural-reporter-group signal **b** means that HPAEC peak **13** reflects only compound **13b**.

The peaks **31**, **32**, and **33** in the HPAEC-PAD profile of GOS II DP4 (Supplementary Information Fig. S5C) were isolated and characterized as β -D-Galp-(1 $\rightarrow$ 3)- β -D-Galp-(1 $\rightarrow$ 3)- β -D-Galp-(1 $\rightarrow$ 4)-D-Glcp, β -D-Galp-(1 $\rightarrow$ 3)- β -D-Galp-(1 $\rightarrow$ 3)- β -D-Galp-(1 $\rightarrow$ 3)-D-Glcp, β -D-Galp-(1 $\rightarrow$ 3)- β -D-Galp-(1 $\rightarrow$ 3)- β -D-Galp-(1 $\rightarrow$ 2)-D-Glcp, respectively (Table 2, see Supplementary Information), being β -D-Galp-(1 $\rightarrow$ 3)- elongations of compounds **12** (Δ G.U. +0.75), **29** (Δ G.U. +0.63), and **30** (Δ G.U. +0.67), respectively. Furthermore, compound **17** could be assigned from its G.U. value (Scheme 1). The 1D ^1H NMR spectrum of GOS II DP4 (Supplementary Information Fig. S11D) showed the H-1 signals **a** (**33**), **c** (**32**), and **d** (**17**, **31**) for the substitution patterns of α -D-Glcp, as well as **m** (**31**) and **g** (**17**, **31**, **32**, **33**) (Table 4), together with the resonance at δ 4.677. Especially, the latter one is of interest, supporting the $\rightarrow$ 3)- β -D-Galp-(1 $\rightarrow$ 3)- residues in **31**–**33**.

Fig. 3. One-dimensional 600-MHz ^1H NMR spectra of A. Vivinal GOS (adapted from van Leeuwen et al.⁸), B. GOS IV, C. GOS V, and D. GOS VI. For assignments of structural-reporter-group signals, see Table 4.

The HPAEC-PAD profile of GOS II DP5 (5.5% of GOS II) is presented in Supplementary Information Fig. S5D. Because of the low amount of material no component isolations could be carried out. However, its 1D ^1H NMR spectrum (Supplementary Information Fig. S11E) showed all structural-reporter-group signals for further β -D-Galp-(1 $\rightarrow$ 3)- elongations of the DP4 components (the signal at δ 4.677 has a higher relative intensity than observed for GOS II DP4). So, it is hypothesized that in any case, HPAEC peaks with G.U. values of 3.62–3.68, 4.06, and 4.23 are β -D-Galp-(1 $\rightarrow$ 3)- elongations of **31**, **32**, and **33**, respectively.

2.2.3. GOS III preparation

Inspection of the HPAEC-PAD profile of GOS III (Fig. 1D) revealed a series of major and minor peaks, of which the relatively high concentration of the donor/acceptor substrate lactose (peak **5**) is remarkable. The peaks **1–12** could be annotated by comparison with Vivinal GOS G.U. values,^{8,17} and the peaks **25**, **28–31** by comparison with G.U. data of GOS I and GOS II, obtained in this study (Scheme 1). The peaks **34–37** represented new products, as will be shown below [note that **28** (G.U. 2.00) was identified in the DP3 subpool and **35** (G.U. 1.99) in the DP4 subpool]. Evaluation of the 1D ^1H NMR spectrum of GOS III (Fig. 2D, Table 4) already gives a good picture of most basic structural elements of the components present. In the α -anomeric region, the major H-1 signal is **d**, which

reflects the presence of 4-, 6-, and/or 1-substituted α -D-Glcp residues. Relatively low-intense H-1 resonances are seen for the $\rightarrow$ 2)- α -D-Glcp (signal **a**) and $\rightarrow$ 3)- α -D-Glcp (signal **c**) residues. Three additional low-intense signals are clearly observed, of which signal **h** (H-1 of free α -D-Galp) was also present in the ^1H NMR spectrum of GOS I, whereas the signals **j** [$\rightarrow$ 4 and/or 6)- α -D-Galp] and **k** [$\rightarrow$ 3)- α -D-Galp] reflect a significant presence of saccharides with a reducing Gal unit.¹⁷ Compared with the intensities of the **g** signals [H-4 of $\rightarrow$ 3 and/or 4)- β -D-Galp] in the ^1H NMR spectra of Vivinal GOS, GOS I, and GOS II (Fig. 2A–C), here the intensity is rather low. Furthermore, besides the usual **e** and **f** signals [H-6a of $\rightarrow$ 6)-D-Glcp ($\beta\alpha$)], signal **n** corresponds with the H-4 resonance of the bold β -D-Galp-(1 $\rightarrow$ 3)- β -D-Galp-(1 $\rightarrow$ 4)-D-Glcp residue, whereby either the β -D-Galp-(1 $\rightarrow$ 4) residue itself or the β -D-Galp-(1 $\rightarrow$ 3)- unit is 6-substituted (see below), and signal **i** with the H-6a resonance of a 6-substituted (reducing) β -D-Galp unit. So far, the origin of the minor α -anomeric peak at δ 5.482, not observed in Vivinal GOS, GOS I, and GOS II, is unclear (note that this peak is also present in the various GOS III DP-pure subpools). To support the various assignments in the HPAEC chromatogram and the ^1H NMR spectrum of GOS III, the five GOS III DP-pure subpools were further investigated by HPAEC-PAD and 1D ^1H NMR spectroscopy.

Profiling of GOS III DP2 by HPAEC-PAD (Supplementary Information Fig. S6A) showed peaks corresponding with the G.U.

Scheme 1. Galacto-oligosaccharide structures identified in a previous study (1–24 and x, y, z)⁸ or isolated in this study from GOS I (25–27), GOS II (28–33), and GOS III (34–37). Glucose units (G.U.) values are based on HPAEC-PAD (CarboPac PA-1) elution times in relation to an external malto-oligosaccharide ladder.

Table 2
Chemical shift values of GOS compounds **25–37**, as determined by 1D and 2D $^1\text{H}/^{13}\text{C}$ NMR spectroscopy (for structural characterizations, see Supplementary Information)

	25	26		27	28	29	30	31
	^1H	^1H	^{13}C	^1H	^1H	^1H	^1H	^1H
A α -1	5.248	5.220	93.9	5.234	5.227	5.238	5.449	5.227
A α -2	3.73	3.57	72.3	3.72	3.537	3.73	3.66	3.58
A α -3	3.91	3.83	72.5	3.91	3.71	3.939	3.85	3.84
A α -4	3.53	3.65	79.5	3.52	3.50		3.46	3.67
A α -5	3.89	3.94	71.1		4.00		3.82	3.95
A α -6a	3.84	3.86	60.7		4.168		3.84	3.86
A α -6b	3.79	3.83			3.88		3.75	3.83
A β -1	4.687	4.662	96.8	4.675	4.656	4.677	4.733	4.668
A β -2	3.438	3.277	74.9	3.440	3.254	3.445	3.53	3.289
A β -3	3.75	3.64	75.4	3.72	3.49	3.76	3.70	3.64
A β -4	3.52	3.65	79.5	3.50	3.47		3.43	3.67
A β -5	3.55	3.59	75.9	3.49	3.65		3.46	3.61
A β -6a	3.89	3.96	60.9	3.90	4.223		3.88	3.964
A β -6b	3.76	3.81		3.74	3.84		3.70	3.81
B-1	4.632/618	4.483	103.8	4.687/702	4.487/506	4.725/712	4.634/733	4.515
B-2	3.62	3.62	72.3	3.67	3.72	3.76	3.76	3.72
B-3	3.70	3.77	74.0	3.80	3.82	3.86	3.83	3.85
B-4	3.96	4.206	78.2	4.205	4.194	4.198	4.202	4.199
B-5	3.95	3.77	75.5	3.77				3.69
B-6a	4.085	3.87	61.8					3.80
B-6b	3.91	3.82						3.75
C-1	4.466/455	4.614	105.2	4.622	4.613	4.617	4.620	4.683
C-2	3.53	3.57	72.2	3.58	3.61	3.61	3.59	3.78
C-3	3.65	3.67	73.8	3.67	3.66	3.67	3.66	3.84
C-4	3.92	3.942	69.8	3.947	3.921	3.922	3.924	4.203
C-5	3.69	3.88	74.9	3.88		3.74		3.75
C-6a		4.045	70.4	4.048				3.80
C-6b		3.91		3.91				3.74
D-1		4.445	104.5	4.447				4.623
D-2		3.51	71.2	3.54				3.61
D-3		3.64	73.8	3.65				3.68
D-4		3.921	69.8	3.923				3.926
D-5		3.69	76.2	3.70				3.69
D-6a		3.81	62.0					3.80
D-6b		3.76						3.74

	32	33	34	35	36		37	
	^1H	^1H	^1H	^1H	^1H	^{13}C	^1H	^{13}C
A α -1	5.240	5.449	5.223	5.223	5.222	93.7	5.225	93.6
A α -2		3.67	3.59	3.57	3.59	72.1	3.58	72.0
A α -3	3.937	3.86	3.84	3.84	3.85	72.3	3.84	72.5
A α -4		3.47	3.65	3.67	3.65	80.2	3.67	79.8
A α -5			3.92	3.95	3.95	70.9	3.97	71.1
A α -6a				3.85	3.88	61.0	3.88	60.7
A α -6b				3.79	3.80		3.80	
A β -1	4.679	4.730	4.668	4.665	4.668	96.7	4.666	96.8
A β -2	3.447	3.56	3.292	3.291	3.294	74.7	3.288	74.7
A β -3		3.70	3.65	3.64	3.65	75.5	3.65	75.4
A β -4		3.45	3.68	3.67	3.68	80.2	3.68	79.8
A β -5		3.46	3.63	3.60	3.61	75.6	3.60	75.7
A β -6a			3.95	3.96	3.95	61.4	3.96	61.2
A β -6b			3.81	3.81	3.80		3.80	
B-1	4.730/712	4.633/730	4.462	4.459	4.516	103.6	4.510	103.5
B-2		3.77	3.56	3.55	3.72	71.1	3.70	70.9
B-3		3.85	3.69	3.67	3.85	82.6	3.84	82.9
B-4	4.205	4.198	3.97	3.968	4.234	69.1	4.224	69.3
B-5				3.88	3.96	74.6	3.76	75.6
B-6a			4.064	4.068	4.060	69.8	3.76	62.0
B-6b				3.93	3.93		3.72	
C-1	4.685	4.683	4.499	4.551	4.488	104.1	4.621	105.2
C-2		3.78	3.58	3.71	3.54	71.8	3.61	71.9
C-3		3.85	3.67	3.84	3.67	73.3	3.67	73.5
C-4	4.205	4.204	3.96	4.196	3.92	69.5	3.958	69.6
C-5					3.70	76.5	3.88	74.9
C-6a			4.064		3.76	62.0	4.040	70.1
C-6b					3.72		3.90	
D-1	4.622	4.620	4.446	4.620	4.612	105.3	4.431	104.3
D-2		3.62	3.52	3.60	3.61	72.0	3.518	71.8
D-3		3.68	3.65	3.67	3.66	73.3	3.63	73.5
D-4	3.925	3.925	3.93	3.919	3.92	69.5	3.918	69.6
D-5					3.70	76.5	3.69	75.9
D-6a					3.76	62.0	3.76	62.0
D-6b					3.73		3.72	

Table 3

Overview of structures up to DP4 that are present in Vivinal GOS and GOS samples I–VI. Numbers correspond with structures in Scheme 1. ● signifies structures are present; ○ shows they are absent; and ♦ signifies structures that are likely present in the GOS preparations

Structure	Vivinal GOS	GOS I	GOS II	GOS III	GOS IV	GOS V	GOS VI
1	●	●	●	●	●	●	●
2	●	●	●	●	●	●	●
3	●	●	○	●	○	●	●
4	●	●	●	●	●	●	●
5	●	●	●	●	●	●	●
6a	●	○	○	○	●	●	○
6b	●	●	●	●	●	●	●
7	●	●	●	●	●	●	●
8a	●	●	●	●	●	●	●
8b	●	●	●	●	●	●	●
9	●	●	●	●	●	♦	●
10a	●	●	●	●	●	♦	●
10b	●	●	●	●	●	♦	●
11	●	●	●	●	●	●	●
12	●	●	●	●	●	●	●
13a	●	●	○	○	●	○	○
13b	●	●	●	○	●	○	○
14a	●	○	○	○	●	○	○
14b	●	○	○	○	●	○	○
15a	●	●	○	○	●	○	○
15b	●	●	○	●	○	○	●
16a	●	○	○	○	●	○	○
16b	●	○	○	●	○	○	●
16c	●	○	○	♦	○	○	♦
17	●	●	●	●	○	○	●
18a	●	●	○	○	●	○	○
18b	●	●	○	○	●	○	○
xa	●	○	○	○	○	○	○
xb	●	○	○	○	●	○	○
xc	●	○	○	○	●	○	○
ya	●	○	○	○	●	○	○
yb	●	○	○	○	●	○	○
25	○	●	○	●	○	○	●
26	○	●	○	○	○	●	○
27	○	●	○	○	○	○	○
28	○	○	●	●	○	○	●
29	○	○	●	●	○	○	●
30	○	○	●	●	○	○	●
31	○	○	●	●	○	●	●
32	○	○	●	○	○	○	○
33	○	○	●	○	○	○	○
34	○	○	○	●	○	●	●
35	○	○	○	●	○	●	●
36	○	○	○	●	○	●	●
37	○	○	○	●	○	●	●

values of disaccharides **3**, **4**, **5**, **7**, **8a**, and **8b** (Scheme 1). The molar ratios of these compounds differ from those observed for Vivinal GOS and GOS II, but are quite similar to that of GOS I (Supplementary Information Fig. S3A, S4A, and S5A). The HPAEC assignments are supported by the 1D ¹H NMR spectrum of GOS III DP2 (Supplementary Information Fig. S12B), showing the signals **a**, **c**, **d**, **e**, **f**, and **j**, belonging to the different α -D-Glcp and α -D-Galp substitution patterns (Table 4). Although the ¹H NMR spectrum also indicates a signal **k**, the corresponding β -D-Galp-(1→3)- β -D-Galp disaccharide could not be traced in the HPAEC chromatogram.

The HPAEC-PAD profile of the GOS III DP3 subpool (Supplementary Information Fig. S6B) showed peaks that correspond with the G.U. values of **6**, **9**, **10**, **11**, **12**, **25**, **28**, **29**, and **30** (Scheme 1). Their structures, after isolation, were separately verified by ¹H NMR analysis. The absence of ¹H NMR structural-reporter-group signal **p** indicates the absence of structure **6a**. Here, structures **6b** (6'-galactosyl-lactose) and **12** (3'-galactosyl-lactose) are the major components. Evaluation of the structures of the various detected compounds shows especially β -D-Galp-(1→6)- (**6b**, **25**) and β -D-Galp-(1→3)- (**12**, **28**, **29**, **30**) extensions of the DP2 structures **4**, **5**,

8a, and **8b**. These conclusions are supported by the structural-reporter-group signals **a**, **c**, **d**, **e**, **f**, **g**, **i**, and **m**, in the 1D ¹H NMR spectrum of GOS III DP3 (Table 4 and Supplementary Information Fig. S12C). By using ¹H NMR analysis of peak **10**, the presence of **10a** and **10b** was revealed.

The HPAEC-PAD profile of GOS III DP4 is presented in Supplementary Information Fig. S6C, and the peaks coded **15**, **16**, **17**, and **31**, were assigned according to Scheme 1. The peaks **34**, **35**, **36**, and **37** could not be correlated with available G.U. values so far. All fractions were isolated followed by ¹H NMR verification of the earlier reported ones and structural analysis of the new components (Table 2; see Supplementary Information). They show a further elongation of DP3 structures with a β -D-Galp-(1→3)- or a β -D-Galp-(1→6)- unit. So, the major DP3 structure **6b** is elongated with a β -D-Galp-(1→6)- residue, resulting in structure **34** [β -D-Galp-(1→6)- β -D-Galp-(1→6)- β -D-Galp-(1→4)-D-Glcp], and with a β -D-Galp-(1→3)- residue, resulting in structure **35** [β -D-Galp-(1→3)- β -D-Galp-(1→6)- β -D-Galp-(1→4)-D-Glcp] (Scheme 1). And the second major DP3 structure **12** is extended with a β -D-Galp-(1→6)- unit to form structure **37** [β -D-Galp-(1→6)- β -D-Galp-(1→3)- β -D-Galp-(1→4)-D-Glcp] and with a β -D-Galp-(1→3)- unit to form structure **31** [β -D-Galp-(1→3)- β -D-Galp-(1→3)- β -D-Galp-(1→4)-D-Glcp] (Scheme 1). The branched structure **36** [β -D-Galp-(1→3)-[β -D-Galp-(1→6)-] β -D-Galp-(1→4)-D-Glcp] can be formed from **6b** and/or **12**. The minor HPAEC peaks **15**, **16** (except **16c**), and **17** can be explained as derived from β -D-Galp-(1→4)- elongations of DP3 structures **9**, **10a**, and **11**, respectively. However, the absence of structural-reporter-group peak **b**, for $\rightarrow$ 4)- β -D-Galp-(1→2)- α -D-Glcp H-1, indicates the absence of **15a** and **16a**. Components **15b** and **16b** do occur, as shown by ¹H NMR analysis. Inspection of the 1D ¹H NMR spectrum of GOS III DP4 (Supplementary Information Fig. S12D), showing basic elements of the tetrasaccharides, supports these assignments (signals **a**, **c**, **d**, **m**, **n**, **g**, and **i**; Table 4).

The HPAEC-PAD profile of GOS III DP5 (1.5% of GOS III) (Supplementary Information Fig. S6D) showed a complex pattern of peaks that could not be assigned. However, similar activities as observed going from DP3 to DP4, i.e. extensions of DP4 structures with β -D-Galp-(1→3)- and β -D-Galp-(1→6)- residues could be expected. This is supported by the 1D ¹H NMR spectrum of GOS III DP5 (Supplementary Information Fig. S12E), wherein the peak pattern in the range δ 5.5–4.0 is quite similar to that in the same δ range of the GOS III DP4 spectrum.

2.2.4. GOS IV preparation

The HPAEC-PAD profile of GOS IV (Fig. 1E) showed a similar compound composition (peaks **1**, **2**, **4**–**24**; Scheme 1) as found for Vivinal GOS (Fig. 1A),⁸ but in different ratios. Only trace amounts of Gal (**1**) and Glc (**2**) are present, suggesting a purification step. Typically, peak **10** has a higher intensity and peak **17** a lower one than in Vivinal GOS. Peak **3** is absent. Also the HPAEC profiles of the GOS IV DP2–DP5 subpools show similar compound compositions as Vivinal GOS, but in significant different ratios. The 1D ¹H NMR spectrum of GOS IV (Fig. 3B and Table 4) revealed structural-reporter-group signals **a**, **b**, **c**, **d**, **e**, **f**, and **g**, just like what is seen in Vivinal GOS (Fig. 3A), although in different intensities. The structural elements related to the signals **m** and **p** are also present in Vivinal GOS, but in relative quantities too low for detection in the 1D ¹H NMR spectra of the total mixture. Here, however, these structural elements are present in sufficient quantities for detection in the initial analysis.

Although the set of GOS IV DP2 disaccharides (**4**, **5**, **7**, **8a**, and **8b**; Scheme 1) is identical to that of Vivinal GOS DP2 (compare HPAEC profiles in Supplementary Information Figs. S3A and S7A), the amounts of allolactose (**4**) and β -D-Galp-(1→4)-D-Galp (**7**) of GOS IV DP2 are somewhat higher. The 1D ¹H NMR spectrum of GOS IV DP2 (Supplementary Information Fig. S13B and Table 4) is similar to that of Vivinal GOS DP2 (based on the available knowledge so

Table 4¹H NMR structural-reporter-group signals used for the evaluation of 1D ¹H NMR spectra of GOS preparations (references: van Leeuwen et al. ^{8,17} and this study)

Signal code	Chemical shift	Explanation	Component
a	5.45	H-1 of a 2-substituted α -D-Glcp unit; [β -D-Galp-(1 $\rightarrow$ 2)- α -D-Glcp]	8a, 9, 10a, 15b, 16b, 20a, 21a, 30, 33, yb
b	5.42–5.40	H-1 of a 2-substituted α -D-Glcp unit, which bears at O-2 a β -D-Galp unit, elongated at O-4 with another β -D-Galp unit; [β -D-Galp-(1 $\rightarrow$ 4)- β -D-Galp-(1 $\rightarrow$ 2)- α -D-Glcp]	13a, 15a, 16a, 18a, 20bc, 21bc, 23a, ya
c	5.25–5.23	H-1 of a 3-substituted α -D-Glcp unit; [β -D-Galp-(1 $\rightarrow$ 3)- α -D-Glcp]	8b, 10b, 13b, 18b, 23b, 25, 27, 29, 32
d	5.23–5.22	H-1 of a 4-, 6-, and/or 1-substituted α -D-Glcp unit; [β -D-Galp-(1 $\rightarrow$ 4/6)- α -D-Glcp; α -D-Glcp-(1 $\leftrightarrow$ 1)- β -D-Galp]	4, 5, 6ab, 11, 12, 14ab, 16c, 17, 19a, 22, 24, 26, 28, 31, 34, 35, 36, 37, xabc, yab, zabc
e	4.22–4.21	H-1 of free α -D-Glcp	2
f	4.17–4.16	H-6a of a 6-substituted β -D-Glcp unit; [β -D-Galp-(1 $\rightarrow$ 6)- β -D-Glcp]	4, 10ab, 16abc, 21abc, 28
g	4.21–4.17	H-6a of a 6-substituted α -D-Glcp unit; [β -D-Galp-(1 $\rightarrow$ 6)- α -D-Glcp]	4, 10ab, 16abc, 21abc, 28
h	5.26	H-4 of a 3- and/or 4-substituted (reducing) β -D-Galp unit; [β -D-Galp-(1 $\rightarrow$ 3/4)- β -D-Galp]-(1 $\rightarrow$ 1)]	7, 11, 12, 13ab, 14ab, 15ab, 16abc, 17, 18ab, 19abc, 20abc, 21abc, 22, 23ab, 24, 26, 27, 28, 29, 30, 31, 32, 33, xabc, yab, zabc
i	4.08–4.05	H-1 of free α -D-Galp	1
j	5.27	H-6a of a 6-substituted (reducing) β -D-Galp unit; [β -D-Galp-(1 $\rightarrow$ 6)- β -D-Galp]-(1 $\rightarrow$ 1)]	3, 6b, 25, 26, 27
k	5.28	H-1 of a 4- and/or 6-substituted reducing α -D-Galp unit; [β -D-Galp-(1 $\rightarrow$ 4/6)- α -D-Galp]	3, 7
m	4.51–4.52	H-1 of a 3-substituted reducing α -D-Galp unit; [β -D-Galp-(1 $\rightarrow$ 3)- α -D-Galp]	
n	4.23	H-1 of a 3-substituted β -D-Galp-(1 $\rightarrow$ 4)- unit in a β -D-Galp-(1 $\rightarrow$ 3)- β -D-Galp-(1 $\rightarrow$ 4)-D-Glcp sequence	12, 31
p	4.29–4.28	H-4 of a 3,6-disubstituted β -D-Galp-(1 $\rightarrow$ 4)- unit, H-4 of a 3-substituted β -D-Galp-(1 $\rightarrow$ 4)- unit in a β -D-Galp-(1 $\rightarrow$ 6)- β -D-Galp-(1 $\rightarrow$ 3)- β -D-Galp-(1 $\rightarrow$ 4)- sequence	36, 37
p	4.29–4.28	H-6a of a 4,6-disubstituted β -D-Glcp unit; [β -D-Galp-(1 $\rightarrow$ 4,6)- β -D-Glcp]	6a, 14ab, 19abc

far, the origin of the resonance at δ 5.40 for the DP2 subpool is unclear) (Supplementary Information Fig. S9B).

In the HPAEC-PAD profile of GOS IV DP3 (Supplementary Information Fig. S7B) an intense peak **6** [4,6-digalactosyl-glucose (**6a**) and 6'-galactosyl-lactose (**6b**)] is present. The major peaks **11** (4'-galactosyl-lactose) and **13** (¹H NMR analysis: **13a** and **13b**; linear β -D-Galp-(1 $\rightarrow$ 4)- elongations of **5**, **8a**, and **8b**, respectively) in Vivinal GOS are now of equal intensity as peak **10** (¹H NMR analysis: **10a** and **10b**; branched β -D-Galp-(1 $\rightarrow$ 6)- extensions of **8a** and **8b**). In Vivinal GOS, peak **10** has an intermediate intensity. The 1D ¹H NMR spectrum of GOS IV DP3 is similar to that of Vivinal GOS DP3 (compare Supplementary Information Figs. S13C and S9C and Table 4). The relatively intense structural-reporter-group peak **p** indicates significant amounts of structure **6a**.

The HPAEC-PAD profile of the GOS IV DP4 peaks is quite similar to the profile of Vivinal GOS, but the peak intensities differ strongly (compare Supplementary Information Figs. S3C and S7C). It shows that β -D-Galp-(1 $\rightarrow$ 4)- elongations of the branched DP3 structures (**6a** yields **14a** and **14b**; **9** yields **15a** and **15b**; **10a** yields **16a** and **16b**) and β -D-Galp-(1 $\rightarrow$ 4)- elongations of the linear DP3 structures (**11** yields **17**; **13a** and **13b** yield **18a** and **18b**, respectively) do occur (Scheme 1). Compound **17** [β -D-Galp-(1 $\rightarrow$ 4)- β -D-Galp-(1 $\rightarrow$ 4)- β -D-Galp-(1 $\rightarrow$ 4)-D-Glcp], the major peak in Vivinal GOS DP4, now has a reduced intensity. Interestingly, GOS IV DP4 also contains relatively high concentrations of non-reducing galacto-oligosaccharides (peaks **x** and **y**; Scheme 1), not present in such high amounts in Vivinal GOS.⁸ The 1D ¹H NMR spectrum of GOS IV DP4 (Supplementary Information Fig. S13D) showed the same structural reporters as present in Vivinal GOS DP4 (Table 4 and Supplementary Information Fig. S9D), suggesting that all structural isomers of **14**, **15**, **16**, and **18** are included.

The peaks **19**, **20**, and **21** in the HPAEC-PAD profile of GOS IV DP5 (8.5% of GOS IV) (Supplementary Information Fig. S7D), related to β -D-Galp-(1 $\rightarrow$ 4)- elongations of the branched DP4 structures (**14ab**, **15ab**, and **16ab**), dominate over the peaks **22** and **23ab**, related to β -D-Galp-(1 $\rightarrow$ 4)- elongations of the linear DP4 structures (**17** and **18ab**) (Scheme 1). Furthermore, relatively high amounts of non-reducing galacto-oligosaccharides appear to be present (peaks **z**; Scheme 1).⁸ The 1D ¹H NMR spectrum of GOS IV DP5 is similar to that of Vivinal GOS DP4 (Table 4; Supplementary Information

Figs. S9E and S13E), suggesting that all structural isomers of **19**, **20**, **21**, and **23** are included.

2.2.5. GOS V preparation

The HPAEC-PAD profile of GOS V (Fig. 1F) is quite different from that of Vivinal GOS and GOS I–IV (Fig. 1A–E); the assigned peaks **1**–**8**, **11**, and **12** are part of the ensemble of Vivinal GOS components (Scheme 1).⁸ The peak pattern is dominated by Gal (**1**), Glc (**2**), allolactose (**4**), lactose (**5**), and 4,6-digalactosyl-glucose/6'-galactosyl-lactose (**6ab**). The 1D ¹H NMR spectrum of GOS V (Fig. 3C and Table 4) showed a low-intensity signal **a** [H-1 of $\rightarrow$ 2)- α -D-Glcp] and no signal **b**. The clear presence of signal **h** [H-1 of free α -D-Galp] reflects a high abundance of free Gal. The intense H-1 signal **d** stems partly from free Glc (*i.e.* α -D-Glcp; see also δ 4.623 (H-1) and δ 3.235 (H-2) for free β -D-Glcp) and partly from 4- and/or 6-substituted Glc (*i.e.* α -D-Glcp). Further interesting signals are **i** [H-6a of $\rightarrow$ 6)- β -D-Galp], **p** [H-6a of $\rightarrow$ 4,6)- β -D-Glcp], **e** and **f** [H-6a of $\rightarrow$ 6)-D-Glcp], and **m** [internal β -D-Galp unit in the sequence β -D-Galp-(1 $\rightarrow$ 3)- β -D-Galp-(1 $\rightarrow$ 4)-D-Glcp]. The signal **g** [H-4 of $\rightarrow$ 3 and/or 4)- β -D-Galp] is not clearly observed in total GOS V (compare with GOS V DP3). The relatively simple HPAEC pattern is further worked out in the GOS V DP2–DP5 subpools using HPAEC-PAD and 1D ¹H NMR analysis.

The GOS V DP2 fraction consists mainly of lactose (**5**), with minor levels of **3**, **4**, **7**, **8a**, and **8b**, as observed in its HPAEC-PAD profile (Scheme 1 and Supplementary Information Fig. S8A) and confirmed by its 1D ¹H NMR spectrum, which is mainly that of **5** (Supplementary Information Fig. S14B).

The HPAEC-PAD profile of GOS V DP3 (Supplementary Information Fig. S8B) reveals mainly the presence of peak **6**, with minor amounts of 4'-galactosyl-lactose (**11**), 3'-galactosyl-lactose (**12**), and branched DP3 structures **9** and **10a** and/or **10b** (Scheme 1). The 1D ¹H NMR spectrum of GOS V DP3 (Supplementary Information Fig. S14C) revealed the structural-reporter-group signals **a**, **d**, **m**, **p**, **g**, **e**, **f**, and **i** expected for these components (Table 4). Based on ¹H NMR analysis, compound **6a** is present (signal **p**). The relative intensity of **p** [H-6a of $\rightarrow$ 4,6)- β -D-Glcp] compared with **i** [H-6a of $\rightarrow$ 6)- β -D-Galp] indicates that peak **6** consists mainly of **6b**. The detection of signal **a** supports the presence of compounds **9** and **10a**. The occurrence of **10b** cannot be excluded.

The HPAEC-PAD profile of GOS **V** DP4 (6.7% of GOS **V**) (Supplementary Information Fig. S8C) showed a major peak **34**, which is a linear β -D-Galp-(1 $\rightarrow$ 6)- elongation of 6'-galactosyl-lactose (**6b**), the main component of GOS **V** DP3 (Scheme 1). Because of the low amount of material, the unknown earlier eluting peak (1.70 G.U.) could not be identified, but it is hypothesized that in view of the reduced retention time, seen when a β -D-Galp-(1 $\rightarrow$ 4)- unit is replaced by a β -D-Galp-(1 $\rightarrow$ 6)- unit, this component is built up of two β -D-Galp-(1 $\rightarrow$ 6) linear extensions of allolactose (**4**), i.e. β -D-Galp-(1 $\rightarrow$ 6)- β -D-Galp-(1 $\rightarrow$ 6)- β -D-Galp-(1 $\rightarrow$ 6)-D-Glcp. This means that the main activity is toward sequential β -D-Galp-(1 $\rightarrow$ 6)- elongations of lactose (**5**) and allolactose (**4**). The minor peaks **26**, **31**, **35**, **36**, and **37**, annotated on the basis of G.U. values included in Scheme 1, correlate with β -D-Galp-(1 $\rightarrow$ 6)- and β -D-Galp-(1 $\rightarrow$ 3)- extensions of **6b**, **11**, and **12**. The 1D ^1H NMR spectrum of GOS **V** DP4 (Supplementary Information Fig. S14D) supports the presence of major levels of 6-substituted β -D-Galp units (signal i). As none of the numbered components contains a 6-substituted D-Glcp unit, the detection of the H-6a signals **e** and **f** support the suggestion given for the HPAEC peak with 1.70 G.U. The clear presence of the H-6a signal **p** for 4,6-disubstituted β -D-Glcp units also suggests that extensions of trisaccharide **6a** do occur, supported by the presence of compounds **14a** and **14b** in the HPAEC chromatogram (Supplementary Information Fig. S8C and Table 4).

GOS **V** DP5 was only available in a relatively low quantity (2.1% of GOS **V**). As is evident from Supplementary Information Fig. S8D, the whole HPAEC-PAD peak pattern is only slightly shifted up in the G.U. ladder from **34** (Supplementary Information Fig. S8C), suggesting mainly β -D-Galp-(1 $\rightarrow$ 6)- elongations of GOS **V** DP4 structures. Its 1D ^1H NMR spectrum (Supplementary Information Fig. S14E) is roughly comparable with that of GOS **V** DP4.

2.2.6. GOS VI preparation

The HPAEC-PAD profile of GOS **VI** (Fig. 1G) showed a complex pattern of peaks **1–12**, **25**, **28–31**, **34–37** (Scheme 1) with major peaks for allolactose (**4**), lactose (**5**), 6'-galactosyl-lactose (**6b**), β -D-Galp-(1 $\rightarrow$ 2)-D-Glcp (**8a**), β -D-Galp-(1 $\rightarrow$ 3)-D-Glcp (**8b**), and 3'-galactosyl-lactose (**12**). The assignment of peak **6** to only **6b** is derived from the absence of structural-reporter-group signal **p** [H-6a of $\rightarrow$ 4,6)- β -D-Glcp], corresponding with the absence of structure **6a**. So, DP2 structures and the β -D-Galp-(1 $\rightarrow$ 3)- and β -D-Galp-(1 $\rightarrow$ 6)- elongations of lactose. Comparison of the HPAEC-PAD profiles of GOS **VI** and GOS **III** (Fig. 1D) showed a high similarity. Also the 1D ^1H NMR spectrum of GOS **VI** (Fig. 3D) is a nearly exact match to the spectrum of GOS **III** (Fig. 2D). Notable differences, however, are observed in the levels of free Gal (**1**) and Glc (**2**), which suggests that a similar β -galactosidase enzyme has been used, combined with a purification step to reduce the monosaccharide level.

2.3. Final overview

Comparison of the identified GOS DP2 structures in Vivinal GOS and GOS **I–VI** shows that all samples contain the same set of β -D-Galp-(1 $\rightarrow$ x)-D-Glcp (x = 2; 3; 4; and 6) disaccharides (**8a**, **8b**, **5**, and **4**), although in different molar ratios. Also, based on the collected data, the branched DP3 oligosaccharides β -D-Galp-(1 $\rightarrow$ x)-[β -D-Galp-(1 $\rightarrow$ y)-]D-Glcp (x,y = 2,4; 2,6; and 3,6) (**9**, **10a**, and **10b**) are always present, although in different molar ratios. Therefore, the differentiation in β -D-Galp-(1 $\rightarrow$ 3)-, β -D-Galp-(1 $\rightarrow$ 4)-, and/or β -D-Galp-(1 $\rightarrow$ 6)- extensions, dependent on the used β -galactosidase enzyme, appears to start at the level of these 7 components (Scheme 2). Branched trisaccharide **6a** is absent in case of GOS **I**, GOS **II**, GOS **III**, and GOS **VI**. The disaccharides β -D-Galp-(1 $\rightarrow$ 6)-D-Galp (**3**) and β -D-Galp-(1 $\rightarrow$ 4)-D-Galp (**7**) are detected in varying low amounts, in some cases even virtually absent. Remarkably, clear evidence for the extension of these structures could not be detected

Scheme 2. Overview of the major linkage types found for Vivinal GOS and GOS **I–VI**. Numbers for the structures correspond with those in Scheme 1.

so far in any of the GOS products investigated (although some minor signals in the α -anomeric region could not always be identified).

GOS **IV** has the same structural make-up as Vivinal GOS, whereby the main activity is β -D-Galp-(1 $\rightarrow$ 4)- elongation, but the composition in terms of relative intensities is different. GOS **I** showed a mixed activity of β -D-Galp-(1 $\rightarrow$ 4)- and β -D-Galp-(1 $\rightarrow$ 6)- elongation of structures and is unique in that aspect. Another unique product is GOS **V**, showing mainly β -D-Galp-(1 $\rightarrow$ 6)- elongations and only minor structural elements containing other linkage types. It should be noted, however, that GOS **V** mainly contains lactose and 6'-galactosyl-lactose. All other GOS products show different mixtures of structures with mainly β -D-Galp-(1 $\rightarrow$ 3)- and β -D-Galp-(1 $\rightarrow$ 6)- elongations. The most complex of these are GOS **II**, GOS **III**, and GOS **VI**. Moreover, the comparison of the GOS products has shown that some products contain a broad distribution of DPs, while others contain mainly DP2 and DP3. Interestingly, so far β -D-Galp-(1 $\rightarrow$ 2)- elongations have not been seen in the GOS products studied. Ultimately, these GOS profiles show that different products on the market can contain a very different structural make-up.

This study has shown that HPAEC-PAD chromatography in combination with 1D ^1H NMR analysis is a fruitful approach for initial analysis of GOS products prepared with β -galactosidase enzymes from various biological sources.

3. Experimental

3.1. Materials

Commercial Vivinal GOS (57% GOS components, 21% lactose, 22% glucose + galactose) and anonymous test samples of the commercial GOS products **I–VI** were provided by FrieslandCampina Domo, Amersfoort, The Netherlands. Malto-oligosaccharides DP2–7 were obtained from Sigma-Aldrich (St. Louis, MO), and malto-octaose was purchased from Carbosynth, Ltd. (Compton, UK). D₂O (99.9 atom%) was acquired from Cambridge Isotope Laboratories, Ltd. (Andover, MA).

3.2. Analytical size-exclusion chromatography

Size-distribution analysis of GOS **I–VI** products was performed on a Rezex RSO-01 oligosaccharide Ag⁺ (4%) column (200 $\times$ 10 mm; Phenomenex, Utrecht, The Netherlands) using a Waters 2690XE Alliance HPLC system (Waters, Etten-Leur, The Netherlands), equipped with a Waters 2410 RI detector. Elutions were carried out with Milli-Q water at a flow rate of 0.3 mL/min. The positions of Glc, Gal, and DP fractions were determined with reference monosaccharides and a malto-oligosaccharide ladder up to DP8.

3.3. Preparative size-exclusion chromatography

Aliquots of 2-mL GOS **I-VI** products (380 mg/mL), diluted 1:1 with Milli-Q water, were fractionated on a Bio-Gel P-2 column (60 × 5 cm), eluted with water at a flow rate of 1 mL/min at 60 °C, using an AKTA plus instrument (GE Healthcare, Uppsala, Sweden), equipped with a RI-101 Shodex refractive index detector. After a void-volume of 470 mL, oligosaccharides were generally isolated in 65 fractions of 5 mL. Fractions were screened with MALDI-TOF-MS and pooled according to DP.

3.4. High-pH anion-exchange chromatography

GOS **I-VI** products (25 µL, 0.5 mg/mL) and Bio-Gel P-2 GOS **I-VI** DP subpools were profiled on a Dionex ICS-3000 workstation (Dionex, Amsterdam, The Netherlands), equipped with a CarboPac PA-1 column (250 × 4 mm, Dionex) and an ICS-3000 ED pulsed amperometric detector (PAD), using a complex gradient of A: 100 mM NaOH, B: 600 mM NaOAc in 100 mM NaOH, C: Milli-Q water, and D: 50 mM NaOAc. The fractionations were performed at 1.0 mL/min with 10% A, 85% C, and 5% D in 25 min to 40% A, 10% C, and 50% D, followed by a 35-min gradient to 75% A, 25% B, directly followed by 5 min washing with 100% B and reconditioning for 7 min with 10% A, 85% B, and 5% D. For the determination of G.U. (glucose units) values, samples were mixed with fucose (200 µM) and the elution profiles expressed in reference to an external malto-oligosaccharide ladder (DP1-8), also containing internal fucose as reference point.

Semi-preparative separations of Bio-Gel P-2 GOS DP subpools were performed on a CarboPac PA-1 column (250 × 9 mm) at 4 mL/min, using a Dionex ICS-5000 workstation with an ICS-5000 ED detector and adapted gradients as described previously.⁸ Approximately 1 mg carbohydrate per semi-preparative run was injected in a 10-mg/L concentration. In cases of further purifications, isocratic conditions were applied. Collected fractions were immediately neutralized with 4 M acetic acid, desalted on CarboGraph SPE columns (Grace, Breda, The Netherlands) using acetonitrile : water = 4:6 as eluent, and lyophilized.

3.5. Matrix-assisted laser-desorption ionization time-of-flight mass spectrometry

GOS samples were diluted to 80–180 µg/mL concentrations prior to analysis by MALDI-TOF-MS. Samples (1 µL) were mixed on the target with 1 µL 2,5-dihydroxybenzoic acid (10 mg/mL) in 40% aqueous acetonitrile. After drying of the samples under reduced pressure, the spots were recrystallized with 0.2 µL ethanol (p.a. grade) and dried under atmospheric pressure. MALDI-TOF-MS experiments were performed on an Axima™ Performance mass spectrometer (Shimadzu Kratos Inc., Manchester, UK), equipped with a nitrogen laser (337 nm, 3 ns pulse width). Positive-ion mass spectra were recorded using the reflectron mode at a resolution of at least 5000 full width at half maximum and acquired with software-controlled delayed-extraction optimized for *m/z* 800. Masses were calibrated by an external calibration ladder of malto-oligosaccharides (DP1-8). Cumulative mass spectra of 100–200 shots were typically acquired from 1 to 5000 Da with ion-gate blanking at 300 Da.

3.6. Nuclear magnetic resonance spectroscopy

Samples were exchanged two times with 200 µL 99.9 atom% D₂O with intermediate lyophilization, and finally dissolved in 650 µL D₂O,

containing 25 ppm acetone as internal standard ($\delta^1\text{H}$ 2.225; $\delta^{13}\text{C}$ 31.08). ^1H NMR spectra, including ^1H - ^1H and ^{13}C - ^1H correlation spectra were recorded at a probe temperature of 298K on a Varian Inova 600 spectrometer (NMR Department, University of Groningen, The Netherlands). 1D 600-MHz ^1H NMR spectra were recorded with 5000 Hz spectral width at 16k complex data points, using a WET1D pulse to suppress the HOD signal. 2D ^1H - ^1H COSY spectra were recorded in 256 increments in 4000 complex data points with a spectral width of 5000 Hz. 2D ^1H - ^1H TOCSY spectra were recorded with MLEV17 mixing sequences with 30, 50, and 150 ms spin-lock times. 2D ^{13}C - ^1H gHSQC spectra were recorded with a spectral width of 5000 Hz in t_2 and 10,000 Hz in t_1 direction. 2D ^1H - ^1H ROESY spectra with a mixing time of 300 ms were recorded in 128 increments of 4000 complex data points with a spectral width of 5000 Hz. All spectra were processed using MestReNova 5.3 (Mestrelabs Research SL, Santiago de Compostela, Spain), using Whittaker Smoother baseline correction.

Acknowledgments

This work was financially supported by the European Union/European Regional Development Fund, the Dutch Ministry of Economic Affairs, Agriculture and Innovation, the Dutch innovation program 'Peaks in the Delta', the Municipality of Groningen, the Provinces of Groningen, Fryslân, and Drenthe, and the Dutch Carbohydrate Competence Center (CCC Project WP6a).

Supplementary material

Supplementary data to this article can be found online at doi:10.1016/j.carres.2016.03.006.

References

- Prenosil JE, Stuker E, Bourne JR. *Biotechnol Bioeng* 1987;**30**:1026–31.
- Torres DPM, Gonçalves MPF, Teixeira JA, Rodrigues LR. *Compr Rev Food Sci Food Saf* 2010;**9**:438–54.
- Splechna B, Nguyen T-H, Steinböck M, Kulbe KD, Lorenz W, Haltrich D. *J Agric Food Chem* 2006;**54**:4999–5006.
- Nguyen T-H, Splechna B, Steinböck M, Kneifel W, Lettner HP, Kulbe KD, et al. *J Agric Food Chem* 2006;**54**:4989–98.
- Usui T, Kubota S, Ohi H. *Carbohydr Res* 1993;**244**:315–23.
- Yanahira S, Kobayashi T, Suguri T, Nakakoshi M, Miura S, Ishikawa H, et al. *Biosci Biotechnol Biochem* 1995;**59**:1021–6.
- Rodriguez-Colinas B, Poveda A, Jimenez-Barbero J, Ballesteros AO, Plou FJ. *J Agric Food Chem* 2012;**60**:6391–8.
- van Leeuwen SS, Kuipers BJH, Dijkhuizen L, Kamerling JP. *Carbohydr Res* 2014;**400**:59–73. Corrigendum *Carbohydr Res*, 2016;419:69–70.
- Bultema JB, Kuipers BJH, Dijkhuizen L. *FEBS Open Bio* 2014;**4**:1015–20.
- Reuter S, Rusborg Nygaard S, Zimmermann W. *Enzyme Microb Technol* 1999;**25**:509–16.
- Todorova-Balvay D, Stoilova I, Gargova S, Vijayalakshmi MA. *J Mol Recognit* 2006;**19**:299–304.
- Kim CS, Ji E-S, Oh D-K. *Biotechnol Lett* 2003;**25**:1769–74.
- Ishikawa E, Sakai T, Ikemura H, Matsumoto K, Abe H. *J Biosci Bioeng* 2005;**99**:331–9.
- Ladero M, Santos A, García JL, Carrascosa AV, Pessela BCC, García-Ochoa F. *Enzyme Microb Technol* 2002;**30**:392–405.
- Fransen CTM, van Laere KMJ, van Wijk AAC, Brüll LP, Dignum M, Thomas-Oates JE, et al. *Carbohydr Res* 1998;**314**:101–14.
- Coulter L, Timmermans J, Bas R, Van den Dool R, Haaksman I, Klarenbeek B, et al. *J Agric Food Chem* 2009;**57**:8488–95.
- van Leeuwen SS, Kuipers BJH, Dijkhuizen L, Kamerling JP. *Carbohydr Res* 2014;**400**:54–8.
- Urrutia P, Rodriguez-Colinas B, Fernandez-Arrojo L, Ballesteros AO, Wilson L, Illanes A, et al. *J Agric Food Chem* 2013;**61**:1081–7.