nature communications

Article

https://doi.org/10.1038/s41467-024-52683-x

The lipooligosaccharide of the gut symbiont *Akkermansia muciniphila* exhibits a remarkable structure and TLR signaling capacity

Received: 2 May 2024

Accepted: 16 September 2024

Published online: 27 September 2024

Check for updates

Pilar Garcia-Vello[1,11], Hanne L. P. Tytgat[2,10,11], Janneke Elzinga[2], Matthias Van Hul[3,4], Hubert Plovier[3,4], Marta Tiemblo-Martin[1], Patrice D. Cani[3,4,5], Simone Nicolardi[6], Marco Fragai[7], Cristina De Castro[1], Flaviana Di Lorenzo[1], Alba Silipo[1,8], Antonio Molinaro[1,8,12] ✉ & Willem M. de Vos[2,9,12] ✉

The cell-envelope of Gram-negative bacteria contains endotoxic lipopolysaccharides (LPS) that are recognized by the innate immune system via Toll-Like Receptors (TLRs). The intestinal mucosal symbiont *Akkermansia muciniphila* is known to confer beneficial effects on the host and has a Gram-negative architecture. Here we show that *A. muciniphila* LPS lacks the O-polysaccharide repeating unit, with the resulting lipooligosaccharide (LOS) having unprecedented structural and signaling properties. The LOS consists of a complex glycan chain bearing two distinct undeca- and hexadecasaccharide units each containing three 2-keto-3-deoxy-D-*manno*-octulosonic acid (Kdo) residues. The lipid A moiety appears as a mixture of differently phosphorylated and acylated species and carries either linear or branched acyl moieties. Peritoneal injection of the LOS in mice increased higher gene expression of liver TLR2 than TLR4 (100-fold) and induced high IL-10 gene expression. *A. muciniphila* LOS was found to signal both through TLR4 and TLR2, whereas lipid A only induced TLR2 in a human cell line. We propose that the unique structure of the *A. muciniphila* LOS allows interaction with TLR2, thus generating an anti-inflammatory response as to compensate for the canonical inflammatory signaling associated with LOS and TLR4, rationalizing its beneficial host interaction.

Intestinal bacteria are in continuous dialog with each other and the host where they shape the immune system since early life. Many intestinal bacteria contain a Gram-negative cell envelope that is decorated by lipopolysaccharides (LPS), the leading glycoconjugate recognized by the innate immune system, specifically via Toll-Like Receptors (TLRs)[1,2]. The intestinal tract has a tight barrier that prevents LPS from reaching the circulation since, as a powerful endotoxin, LPS can induce metabolic endotoxemia, resulting in local inflammation and the onset of metabolic disease such as obesity and type 2 diabetes[3,4]. The mucosal symbiont *Akkermansia muciniphila*[5,6] was found to play an important role in maintaining this intestinal barrier function and has developed into one of the few intestinal microbes with evidence of generating a beneficial host response in mice and human[7]. Administration of *A. muciniphila* cells was found to increase barrier function, decrease serum LPS levels, and protect from weight gain and glucose intolerance in diet-induced obese mice[8]. This was

A full list of affiliations appears at the end of the paper. ✉e-mail: molinaro@unina.it; willem.devos@wur.nl

confirmed in various other studies that also revealed increased immune signaling by *A. muciniphila* via the expansion of regulatory T cells (Tregs)[7,9,10].

Pasteurized *A. muciniphila* cells were found to be as active, or even more, as live cells in mouse models. This was attributed to the signaling of a heat-stable outer membrane protein, Amuc_1100 that is thought to signal through the TLR2 receptor[7,11,12]. These results were partly replicated in a human study with subjects at risk for type 2 diabetes, in which pasteurized and live *A. muciniphila* cells increased gut barrier function and decreased serum LPS levels[13]. However, *A. muciniphila*, as a Gram-negative bacterium belonging to the *Verrucomicrobia*, also produces heat-stable LPSs[14]. This generates a paradox since a beneficial symbiont harbors a potentially inflammatory component. This provided the basis for the present study that shows the *A. muciniphila* strain Muc^T to produce a lipooligosaccharide (LOS, Fig. S1), i.e., an LPS devoid of the O-antigen polysaccharide, with an unprecedented chemical structure, made up of an exceptionally large non-repetitive carbohydrate backbone together with a unique lipid A portion, signaling through both TLR4 and TLR2. This peculiar chemistry contributes to explaining the anti-inflammatory effect of *A. muciniphila* in the host by interaction with TLR2 receptor to compensate for the canonical inflammatory signaling associated with TLR4 recognition.

Results

The chemical structure of the A. muciniphila LOS carbohydrate domain reveals an unprecedented complexity

To determine the full structure of *A. muciniphila* Muc^T LOS, including its carbohydrate (core oligosaccharide) and glycolipid (lipid A) moieties, we employed a combination of organic chemistry, high-resolution mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy. LOS was isolated, purified (Supplementary material and Fig. S2) and fully de-lipidated to obtain the core oligosaccharide **OS** (composed of OS_{1deAc} and OS_{2deAc}) covalently connected to lipid A sugar skeleton (Figs. 1–3). Due to its high complexity, the NMR assignment of core **OS** required the use of an ultra-high magnetic field at 1.2 GHz, which allowed to attribute all spin systems and to establish the saccharide sequence (Fig. 1, supplementary information and Fig. S3). NMR spectra unveiled the occurrence of 26 sugar units (including the lipid A sugar backbone), all present as pyranose rings (Table S1). Their glycosylated positions were identified by downfield shift of carbon resonances; *inter*-residual NOE contacts, together with the long-range correlations derived from the HMBC spectrum, allowed to determine the **OS** structure as sketched in Fig. 1. The sugar units were distributed along two glycan chains, the longest composed of a hexadecasaccharide, the shortest of an undecasaccharide chain ($OS1_{deAc}$ and $OS2_{deAc}$, Table S1 and Figs. 1–2), both starting with a Kdo (2-keto-3-deoxy-D-*manno*-octulosonic acid) disaccharide unit, in turn connected to an additional, highly branched Kdo residue carrying the whole glycan chain. To further elucidate and corroborate the structure of *A. muciniphila* **OS**, an in-depth structural characterization was performed by ultrahigh-resolution ESI FT-ICR MS and collision-induced dissociation (CID) (Fig. 2, supporting information and Figs. S4–S14). The **OS** MS showed a high heterogeneity resulting from the presence of both the undeca- and hexadecasaccharides and their variants generated from incomplete deacylation/deacetylation and/or in-source fragmentation. The two oligosaccharides were detected at *m/z* 817.9286 and *m/z* 812.5114 and with a mass measurement error of 0.37 and 0.12 ppm, respectively ($OS1_{deAc}$ and $OS2_{deAc}$, Fig. 2, Figs. S4–S14). CID fragmentation analysis on selected precursor ions confirmed the final resulting structures, with the presence of three Kdo residues per oligosaccharide and a specific fucose acetylation. In parallel, a mild acid hydrolysis on the LOS allowed to split the lipid A from the core portion; the resulting mixture was further purified and, additionally to the lipid A, two isolated oligosaccharides were indeed identified by NMR as the expected tetradecasaccharide and nonasaccharide (**OS_1** and **OS_2**), thus fully confirming the structural assignment (Supporting Information for complete composition, spectroscopic and spectrometric analyzes and full structural discussion; Figures S15–S20, Tables S2–S3). The structure of *A. muciniphila* lipid A was defined by merging data from fatty acid compositional analysis, MALDI-TOF MS and MS/MS investigation

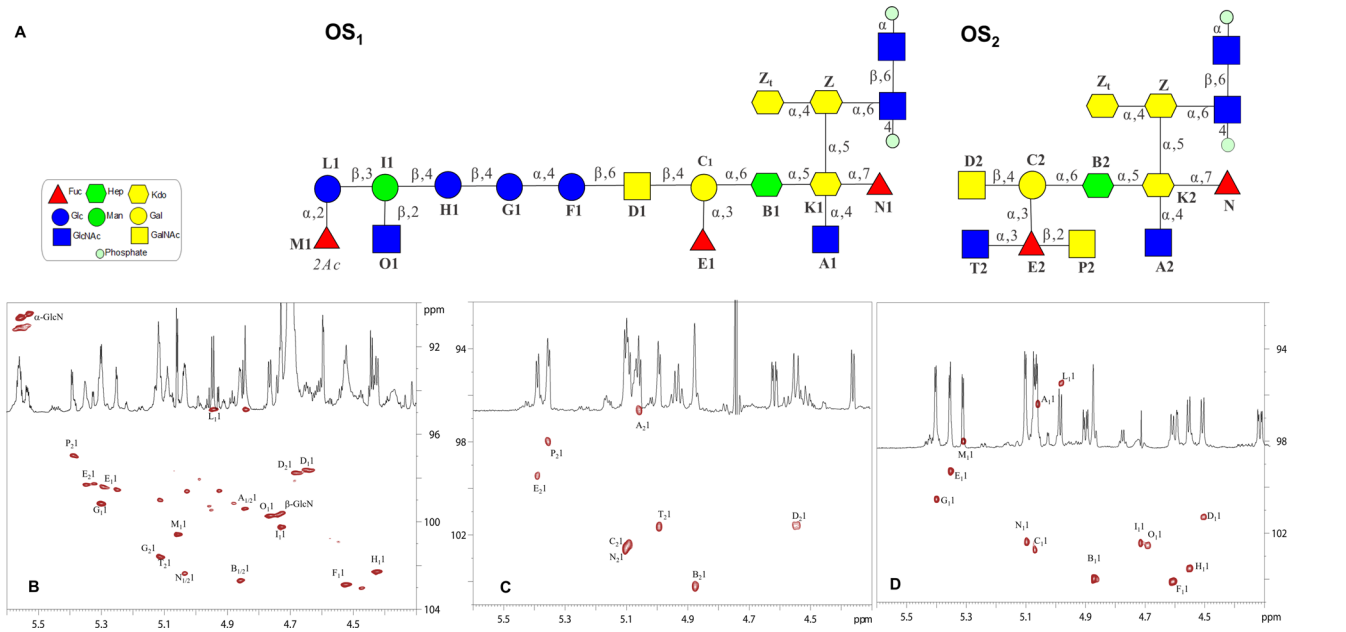

Fig. 1 | *A. muciniphila* LOS chemical structure. A The glycan components of the LOS from *A. muciniphila* sketched according to the symbology of monosaccharide residues (in the inset); the structure of the core region has been assigned by HR MS and NMR analyzes of the acetic acid and full deacylated products; **B**, **C**, **D** multiplicity edited HSQC NMR spectra of (from left to right) *A. muciniphila* **OS** (full deacylated LOS, 1.2 GHz), **OS_1** and **OS_2** (from acetic acid hydrolysis, at 950 MHz): zoom of the anomeric region. Letters refers to the spin systems and are as reported in Supplementary Tables S1, S2 and S3. The glycosidic linkages of **Z_t** and **K** residues are putative and can be interchanged.

https://doi.org/10.1038/s41467-024-52683-x

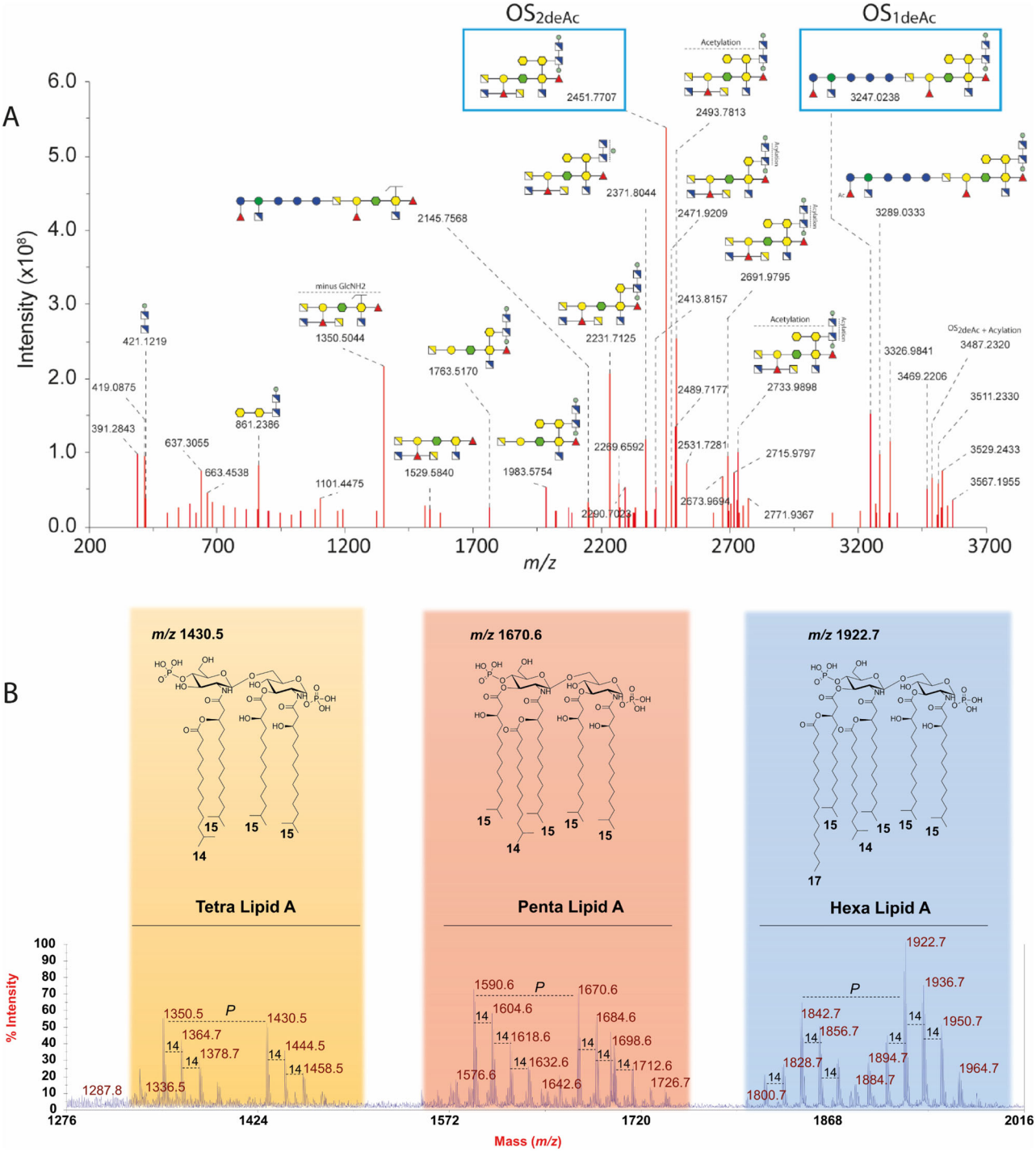

Fig. 2 | Mass spectrometry of LOS and lipid A. A HR and sub-ppm mass measurement of Nona- and Tetradeca- oligosaccharides from the deacylated LOS performed by ESI-FT-ICR MS (mass spectrum deconvoluted to $[M + H]^+$). **B *A. muciniphila* lipid A**. Negative ion MALDI-TOF mass spectrum of the lipid A from *A. muciniphila*. The families of lipid A species that differ in the degree of acetylation are indicated as "Hexa", "Penta", and "Tetra" Lipid A (*P*: phosphate group). The proposed interpretation of relevant ion peaks is reported in Table S3. The structure of the main *bis*-phosphorylated lipid A species for each family (Tetra, Penta and Hexa) has been sketched in each colored panel. The C17:0 moiety has been depicted in its linear form for description purposes only.

(Fig. 2, Supplementary Fig. S21, and Tables S4-S5). A heterogenous blend of lipid A species was identified, differing in acyl chain length and phosphorylation pattern. Three main families of peaks were identified as tetra-, penta- and hexa-acylated species with representative peaks at *m/z* 1922.7, 1670.6, and 1430.5 respectively, in approximately comparable area similarly to the corresponding *mono*-phosphorylated species. The main peak at *m/z* 1922.7 matched with a *bis*-phosphorylated lipid A carrying *iso* (i) C15:0 (3-OH) as primary directly *O*- and *N*-linked fatty acids, and i-C14:0 and C17:0 (or i-C17:0 or a-C17:0) as secondary acyl substituents. The peak at *m/z* 1670.6 was assigned to a *bis*-phosphorylated penta-acylated lipid A species lacking the C17:0 unit, whereas the *bis*-phosphorylated tetra-acylated lipid A species at *m/z* 1430.5 was devoid of both C17:0 and the primary *O*-linked i-15:0 (3-OH) decorating the non-reducing glucosamine unit (Fig. 2, S21 and Tables S4–S5). In summary, *A. muciniphila* lipid A is made up of a complex blend of *mono*- and *bis*-phosphorylated, tetra-

https://doi.org/10.1038/s41467-024-52683-x

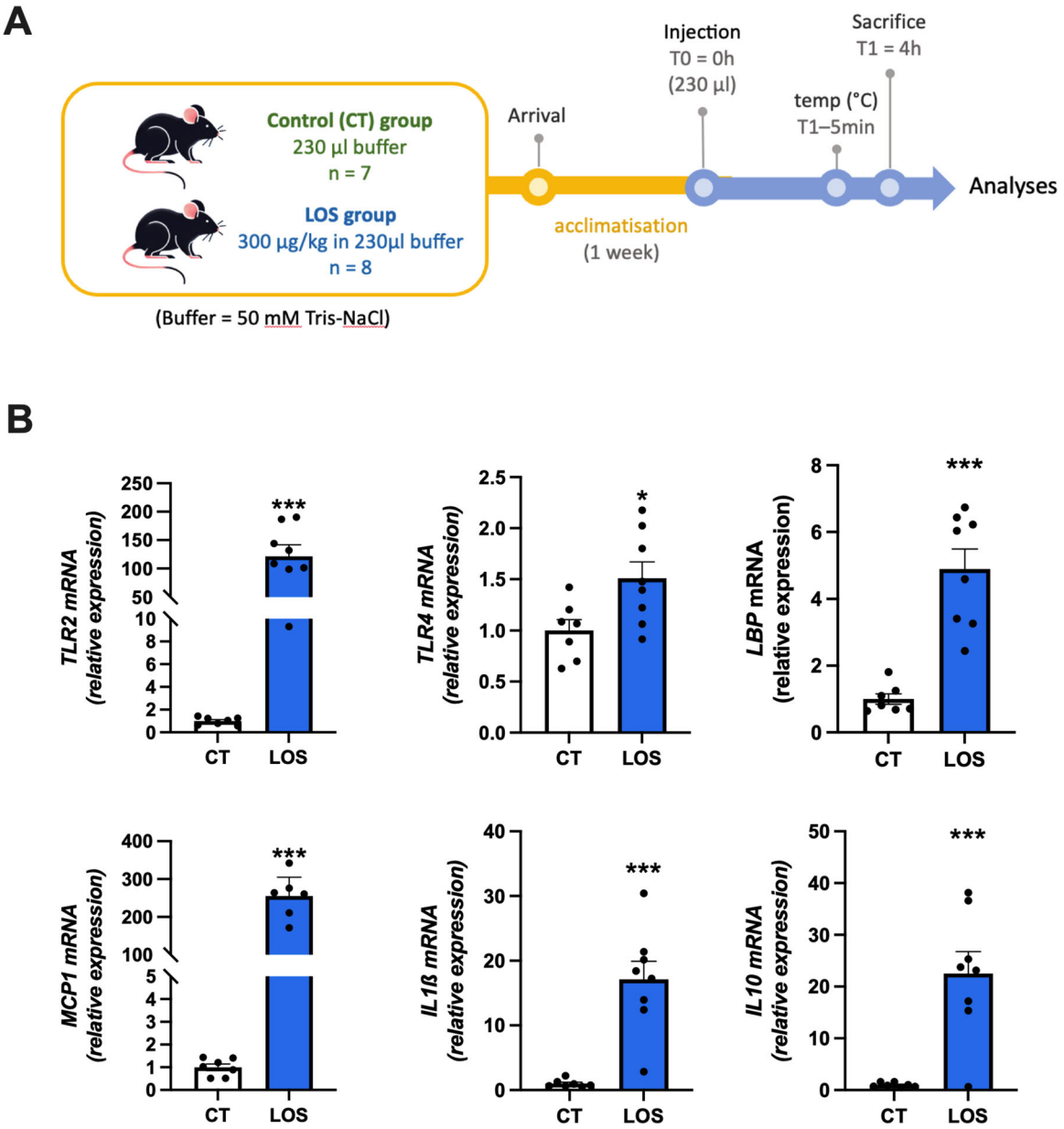

Fig. 3 | Injection of *A. muciniphila* LOS in mice triggers mRNA expression in the liver. A Outline of the experimental setup. Two groups of mice were injected interperitoneally with 230 µl of LOS solution (generating a final concentration of 300 µg/kg body weight; LOS group, $n = 8$) or control group (CT; 50 mM Tris-NaCl buffer pH 7; CT group, $n = 7$). **B** Expression of relevant genes such as Toll like receptor 2 and 4 (TLR2, TLR4), monocyte chemoattractant protein 1 (MCP1), Lipopolysaccharides binding protein (LBP), interleukin 1beta and 10 (IL1β and IL10) were determined in the liver by qPCR, RPL19 (ribosomal protein L19) was chosen as housekeeping genes. All the data were compared to the CT group as fold changed (CT group set at 1). Data are expressed as mean ± SEM and compared using two-tailed Mann Whitney test (*p-value < 0.05; ***p-value < 0.0005). Source data are provided as a Source Data file.

to hexa-acylated lipid A species, whose forms displaying the highest acylation degree has a 4 + 2 symmetry of the acyl chains with the respect to the glucosamine backbone (Fig. 2).

Notably, this remarkable structural heterogeneity was also reflected in the observed large set of expressed genes and proteins predicted to be involved in the LOS biosynthesis (Table S6, see discussion).

Injection of *A. muciniphila* LOS in mice triggers TLR2 mRNA and interleukin 10 expression

To examine its functional importance, the purified LOS of *A. muciniphila* and a vehicle (buffer) were injected intraperitoneally in two groups of mice ($n = 7$–8) which were sacrificed 4 h later (Fig. 3A). LOS injection showed no effect on body weight or weight of organs, such as liver, cecum or cecal content, although a small increase in spleen weight was reported compared to the vehicle injection, possibly indicating an effect on the immune system (Fig. S22). Subsequently, liver samples were analyzed for expression of genes involved in immune signaling in response to the *A. muciniphila* LOS injection as compared to that of the vehicle control (Fig. 3B). The expression of several genes involved in the attenuation of inflammation was increased, including LPS binding protein (LBP), the MCP1 complement regulatory protein and interleukins IL1ß (mediator of inflammatory response) and IL10 (anti-inflammatory cytokine). As expected, the LOS triggered the expression of TLR4 mRNA, the canonical receptor of all bacterial LOS and LPS. However, most noteworthy was the very strong (over 100-fold) surge in relative TLR2 mRNA expression, suggesting that *A. muciniphila* LOS also can interact via this receptor.

https://doi.org/10.1038/s41467-024-52683-x

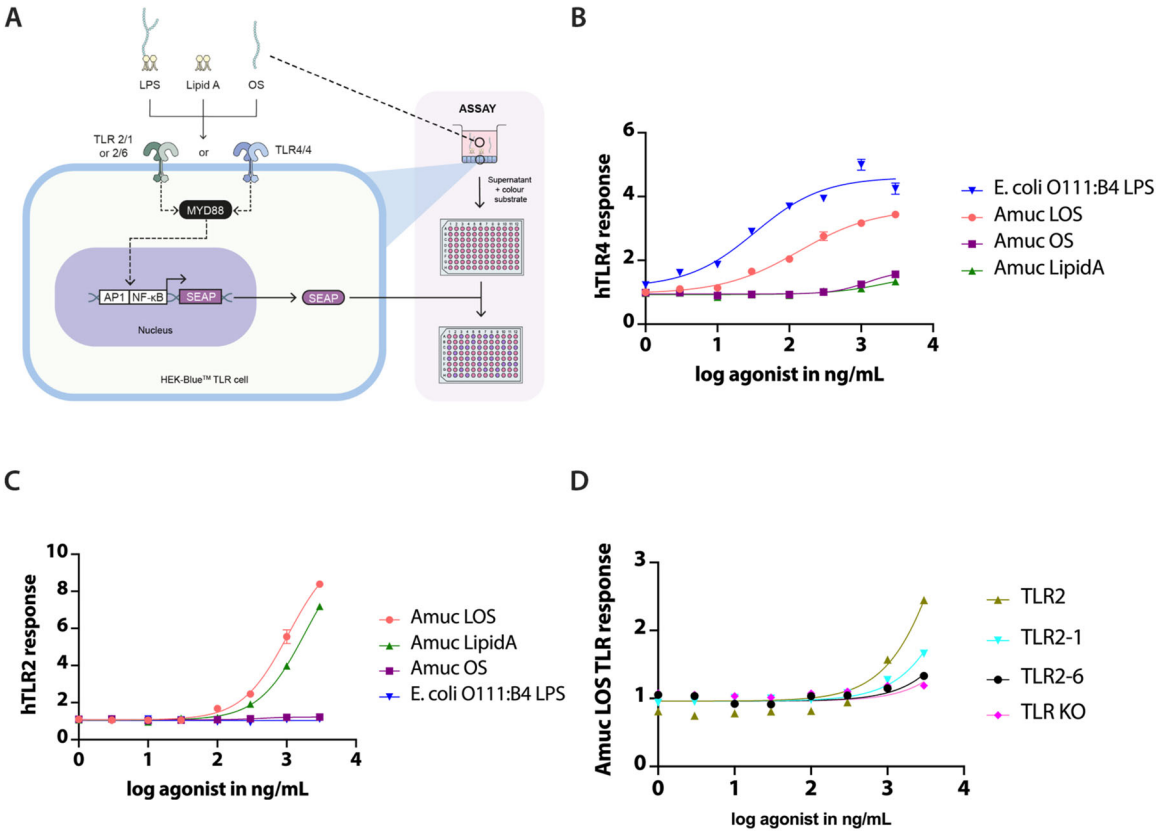

Fig. 4 | Activation of TLR4 and TLR2 by *A. muciniphila* LOS and its structural components. The signaling capacity of *A. muciniphila* LOS, OS and lipid A was evaluated by incubating them in HEK-Blue reporter cell lines (**A**), expressing hTLR4 (**B**) and hTLR2 (**C**). A comparative analysis of LOS signaling to the different hTLR2-1, hTLR2-6, and hTLR2 KO1/6 cell lines was made, as the hTLR2 reporter cell line expresses both TLR1 and TLR6, allowing hetero- and homodimerization of TLR2 with TLR1, TLR6 and TLR2 (**D**). All experiments were performed at least in three biological replicates and representative experiments of three technical replicates are shown. The dose-response curves depict normalized mean TLR responses of three replicates +- SD, and log of the agonist in ng/mL. (Abbreviations used: LOS = lipooligosaccharide, OS = oligosaccharide, hTLR = human Toll-Like Receptor, TLR4 = TLR4 expressing cell line, TLR2 = TLR2 expressing cell line also harboring TLR1 and TLR6, TLR2-1 = cell line only harboring TLR2 and TLR1 heterodimers, TLR2-6 = cell line in which TLR2 and TLR6 are expressed and can form heterodimers, TLR KO = cell line in which all TLR receptors are knocked out, except TLR2). Source data are provided as a Source Data file, and more details are recorded in the "Methods" section.

A. muciniphila LOS signals through both TLR4 and TLR2, with a key role for lipid A

The capacity of *A. muciniphila* LOS and of its components, the isolated core oligosaccharide (OS) and lipid A to signal via TLR2 and TLR4 receptors was further investigated using HEK-Blue hTLR2 and hTLR4 reporter cell lines, respectively (Fig. 4, Fig. S24). *A. muciniphila* LOS was found to activate the TLR4 receptor only slightly less than the LPS of *Escherichia coli* O111:B4 (a Gram-negative bacterium used as positive control for TLR4 activation) (Fig. 4B). Remarkably, in contrast to the LPS of *E. coli* O111:B4, the *A. muciniphila* LOS was also capable of efficient signal transduction via the TLR2 receptor (Fig. 4C, and Supplementary Figs. S23, S24).

Next, we determined the signaling activity of the *A. muciniphila* LOS components, OS and lipid A, to elucidate which part of the molecule contributes to the recorded TLR4 and TLR2 activation. Unexpectedly, both lipid A and OS showed little to no signaling activity with hTLR4-expressing cells, indicating the need for intact LOS to mediate signaling through TLR4 (Fig. 4B). However, Lipid A showed comparable signaling through the hTLR2 cell line as the complete *A. muciniphila* LOS, while OS showed no activity (Fig. 4C). Hence, we conclude that lipid A is the structural *A. muciniphila* LOS component involved in TLR2 signaling.

As the TLR2 receptor functions as a heterodimer formed with either TLR1 or TLR6 and both are present in the hTLR2 reporter cell line (TLR2, TLR1 and TLR6 expression), we tested the LOS in a comparative analysis using the hTLR2-1 (only TLR2 and TLR1 present), hTLR2-6 (only TLR2 and TLR6) and hTLR2 KO1/6 cell lines (only TLR2 present). A slight preference for signaling of *A. muciniphila* LOS via the TLR2-1 heterodimer was observed compared to the TLR2-6 heterodimer, whereas virtually no activation of the hTLR2KO1/6 cell line was detected, indicating that a potential TLR2 homodimer does not recognize LOS (Fig. 4D).

In conclusion, these results confirm and extend the structural and mice experimental data and show that the *A. muciniphila* LOS is a weak agonist of TLR4, with the lipid A moiety to be virtually silent whilst showing an unusual strong activation of the TLR2 receptor, likely via TLR2-6 heterodimer.

Discussion

A. muciniphila and its health benefits have attracted considerable attention in the last decades offering the possibility of not only using it as a next-generation therapeutic approach but also as a diagnostic and prognostic tool for a plethora of different inflammatory and metabolic disorders[7]. In addition to metabolites produced by *A. muciniphila*, particular attention has been given to cell envelope components, such as the exposed outer membrane protein Amuc_1100 that induces the production of the anti-inflammatory cytokine IL-10 through the TLR2 signaling pathway, likely playing a critical role in maintaining the host immune homeostasis and intestinal mucosa[7,11,12].

Here we describe the structural and functional elucidation of the LPS, the main constituents of the Gram-negative outer membrane that *A. muciniphila* presents as LOS, thus lacking the O-antigen portion. Strikingly, the *A. muciniphila* LOS contains both core oligosaccharide and lipid moieties of a unique chemical architecture and structure with no chemical or biosynthetic correlation/analogy with any other LOS/LPS[1,2]. *A. muciniphila* LOS lipid A contains a complex blend of similar amounts of tetra-, penta- and hexa-acylated lipid A species. Notably, hexa-acylated lipid A species have been associated with a strong activation of the TLR4-dependent response[2,15]. The observation that the purified lipid A fraction of *A. muciniphila* is virtually silent in TLR4 signaling may be attributed to the unusual presence of a significant amount of *mono*-phosphorylated and hypo-acylated lipid A species as well as to the chemical nature and length of the acyl moieties. While *A. muciniphila* OS fraction was also silent in TLR4 signaling, the entire LOS shows some TLR4 activation but much less than the inflammatory LPS of *E. coli* O111:B4.

A special finding was the observation that the *A. muciniphila* LOS not only signals to TLR4 but activates the TLR2 receptor even much more efficiently. This is highly relevant since the latter signaling has been recognized as the main facilitator of the anti-inflammatory effects of *A. muciniphila*[7]. Detailed analysis showed that the LOS signal transduction pathway to TLR2 signaling can be mainly ascribed to the TLR2 activation by its lipid A moiety that consists of branched fatty acids with 15–17 carbon atoms. Of note, several studies reported that lipid A species of intestinal *Bacteroides* spp. have unique properties[16] and also contain such branched aliphatic chains and contribute to the TLR2 activation of the LPS together with glycine lipids[17,18]. The latter are made up of the same or very similar acyl components that in *Porphyromonas gingivalis* were found to activate TLR2[18–20]. Glycine lipids as detected in these inflammatory bacteria have not been reported in *A. muciniphila* cells. In contrast, *A. muciniphila* was found to be a major source of ornithin lipids that dampen the *E. coli* LPS-induced inflammation by upregulating the expression of ATF3, a master regulator of metabolic and immune homeostasis[21]. This is another example of how specific cell-envelope located molecules of *A. muciniphila* affect the inflammatory response.

It has been well established that the TLR2 and TLR4 signaling pathways shape the development and function of the gut and enteric nervous system, providing a link between the gut-brain axis[22]. Signaling between intestine and hypothalamus was recently demonstrated to take place in a mouse model by the action of pasteurized *A. muciniphila*[23]. As these pasteurized cells are metabolically inert, this obviates the involvement of in situ-produced metabolites but is likely to include the activity of heat-stable components such as the LOS described here.

An unexpected finding was that the LOS oligosaccharide moiety is a unique chemical structure possessing a mixture of two complex glycan units with no repeating motif and including a distinctive fucosylation pattern. Interestingly, the highly branched fucose unit (residue E_2 Fig. 1) present in the shorter undecasaccharide chain (OS_2) might act stopping a further elongation of the saccharide moiety. Indeed, the same fucose (residue E_1) is terminal in the longer hexadecasaccharide chain (OS_1). Scrutinizing the *A. muciniphila* genome and its observed transcriptome and proteome during growth on mucin, revealed a large set of expressed genes and proteins predicted to be involved in the LOS synthesis (Table S6). It is tempting to hypothesize that the occurrence of fucose as well as the large number of GalNAc and GlcNAc residues, might be a way for the *A. muciniphila* LOS to be recognized by host receptors, such as by specific lectins, as *self*. In this way *A. muciniphila* would adopt the same strategy as some human pathogens that decorate their LPS with glycan motifs mimicking host structures to escape immune surveillance. This stealth protection mechanism may tune down the inflammatory signaling capacity of the cell-envelope located LOS, as proposed for other mutualistic gut-associated bacteria[24].

The complex signaling unraveled here further addresses the paradox that *A. muciniphila* is an abundant mucosal symbiont while being a Gram-negative bacterium with the potential to activate an inflammatory signal via TLR4. Using in vivo mice experiments, we show that when LOS is introduced in the circulation via intraperitoneal injection, the liver expression of the gene encoding TLR2 is hundred-fold more increased than that encoding TLR4. We propose that this leads to an amplification of the beneficial LOS signaling via the TLR2 rather than the TLR4 receptor. This would explain the observed large increase of IL10, which represents the hallmark of anti-inflammatory activity attenuating inflammation. It is worth noting that IL10 is well known to inhibit the production of proinflammatory cytokines including IL1ß[25]. A recent study using various mouse models showed that *A. muciniphila* strain Muc^T signals via TLR2-dependent pathway when administered intraperitoneally[26]. Altogether, our in vitro data and these in vivo *data* support the hypothesis that *A. muciniphila* membrane components including LOS are activating TLR2-dependent pathways.

Our study has some limitations as we have used the highly purified and characterized *A. muciniphila* LOS components but not synthesized these de novo as their chemical complexity is very high. Indeed, although we strongly deem that such a structural complexity and heterogeneity are responsible for the peculiar immunological behavior of *A. muciniphila* LOS, the lack of synthetic derivative(s) also precludes a structure activity relationship study of the LOS moieties. Furthermore, this also impaired addressing intricate details of the immune signaling, like the elucidation of in vivo inflammatory response or lack thereof upon LOS exposure or their comparison to other LOS structures for microbiota-relevant bacteria like *B. thetaiotaomicron*[27].

This study paves the way to a new concept of LPS, not anymore only as a foe but instead a friendly glycolipid. Indeed, we here clearly show that *A. muciniphila* produces an intricate LOS that, due to its uncommon and complex chemical structure with an exceptionally large non-repetitive carbohydrate backbone, might provide a decisive contribution to the health-promoting effects of this gut symbiont by uniquely modulating the immune and potentially the gut-brain response via TLR2 and TLR4. Future structure to function studies will be pivotal for identifying the direct role of LOS in the crosstalk between *A. muciniphila* and the human host. In this context, studies aimed to fully dissect the chemical features and the epitopes responsible of immune recognition will be crucial in expanding the potential use of *A. muciniphila* and its constituents to conceive new health-promoting therapeutic avenues.

Methods

Growth conditions and production of biomass

Akkermansia muciniphila Muc^T (ATCC BAA-835) cells were grown in anaerobic serum bottles sealed with butyl-rubber stoppers at 37 °C with N_2:CO_2 (80:20 ratio) in the headspace at 1.5 atm[1]. Cells were cultured anaerobically in a basal mucin-based medium as previously described[1] or in a synthetic medium where mucin was replaced by 16 g/l soy-peptone (Organotechnie, France), 4 g/l threonine, and a mix of glucose and N-acetylglucosamine (25 mM each). The latter medium was used for large-scale production and bacterial biomass was harvested, washed with PBS and stored at −20 C. Upon thawing, the cell suspension was dialyzed against milliQ water and subsequently lyophilized. The dried cells were kept at 4 °C prior to further use.

Extraction and purification of the LOS

A. muciniphila dried cells (8 g) were treated through a combination of the phenol/chloroform/light petroleum (PCP) and the hot phenol/water extractions to isolate the lipooligosaccharide (LOS)[2,3]. Followed by a dialysis (membrane cut-off 12-14 kDa) against distilled water until

the last traces of phenol were removed. The 10 % sodium dodecyl sulfate–polyacrylamide electrophoresis gel (SDS-PAGE) with silver staining[4] evidenced the presence of LOS in both phases resulting from the PCP extraction and the water phase of the hot phenol/water extraction. Subsequently, the sample underwent enzymatic digestion: 4 mg DNAse (Merck), 4 mg RNAse (Merck), 4 mg amylase (Merck) and 100 mg pullulanase at 37 °C (Merck) overnight and 4 mg proteinase K (Merck) at 56 °C, for 4 h[3], which was followed by a dialysis (membrane cut-off 12–14 kDa) against distilled water for 48 h. Afterwards, the samples were freeze-dried. The resulting material was re-suspended in distilled water to then undergo a separation by centrifugation (5000 × *g*, 4 °C, 30 min) and a subsequent ultra-centrifugation of the supernatant (108,000 × *g*, 4 °C, 4 h). The 13 % SDS-PAGE with silver staining evidenced the presence of LOS in both the centrifugation and ultra-centrifugation precipitates (Fig. S1) (yield 26.2 mg LOS from PCP extraction/ per gram dried cells; 80 mg LOS from hot phenol-water extraction per gram of dried cells). All LOS containing fractions underwent the by Hirschfeld et al[5]. protocol to remove any traces of lipoproteins possibly contaminating the isolated material. SDS-PAGE followed by Coomassie Brilliant Blue (Sigma-Aldrich) gel staining was performed to evaluate the presence of protein/lipoprotein contaminants. Both silver and Coomassie Brilliant Blue staining of the gels was executed according to the standard protocols[4]. Then, compositional analyzes were performed to define the sugar and fatty acid content.

Characterization of the LOS and chromatographic purifications

Mild acid hydrolysis (10 mg LOS / 1 mL of acetic acid 1%, 100 °C, 2 h) was performed on the purified LOS to separate the polysaccharidic chain from the lipid A. The resulting product was then centrifuged (8000 × *g*, RT, 15 min). In the precipitate, the lipid A was successfully isolated for its study through MALDI (5.6 mg, 35% of LOS). In the supernatant, however, a mixture of different polysaccharides was obtained. These polysaccharides were separated through a High-Performance Liquid Chromatography (HPLC) system Agilent 1100 Phenomenex C18 Reversed Phase in isocratic conditions (flow 0.8 mL/min) and monitoring the eluate with a refractive index detector (206 nm). The HPLC resulted in a chromatogram with three differentiated peaks named from A to C, (**OS1, OS2** and their mixture). ^{1}H-NMR experiments were performed for all the three fractions.

An aliquot of lipid A obtained as above was also treated with ammonium hydroxide as previously reported[5,6], dried and analyzed by Matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry (MS). For O-deacylation, the LOS (106 mg) were dissolved in anhydrous methylhydrazine (5 mL), stirred at 37 °C for 90 min, cooled, poured into ice-cold acetone (50 mL) until precipitation. The precipitate was then centrifuged (5000 × *g*, 40 min), washed twice with ice-cold acetone, dried, and then dissolved in water and lyophilized (78 mg, 73% of LOS)[7]. The sample was subsequently *N*-deacylated with KOH 4 M (7.8 mL, 120 °C, 18 h)[8]. It was then neutralized with HCl and subsequently, desalted using a Sephadex G-10 column (166 cm^3; Amersham Biosciences, Inc.). The resulting oligosaccharide fraction (32 mg, 30.2 % of the LOS). Following, it was further purified by Bio-Gel P-10 chromatography (166 cm^3). The full deacylated LOS was obtained with a total yield of 19.8 %.

LOS compositional analysis

The monosaccharide content was established by the acetylated *O*-methyl glycoside derivatives method[9]. For this purpose, the LOS (0.5 mg) was treated with HCl/MeOH (1.25 M, 80 °C, 16 h) followed by an acetylation step with acetic anhydride in pyridine (80 °C, 30 min). The fatty acid content was established by a hexane extraction after treating the LOS with HCl/MeOH[10]. The absolute configuration of each monosaccharide was determined by the *O*-octylglycoside derivatives method[11]. The sugar linkage pattern was defined by the partially methylated alditol acetated method. The sample (0.5 mg) was solved in DMSO, methylated with CH_3I and powdered NaOH[12], hydrolyzed with trifuoroacetic acid (2 M, 100 °C, 1 h), carbonyl reduced with $NaBD_4$, and acetylated with pyridine and acetic anhydride. All chemical derivatives were analyzed by using a Gas chromatography–mass spectrometry (GC–MS) Agilent Technologies 7820 A (Santa Clara, CA, USA) equipped with a mass selective detector 5977B, equipped with the automatic injector 7693 A and HP-5ms capillary column Agilent, Italy (30 m × 0.25 mm i.d., 0.25 mm as film thickness, flow rate 1 mL/min, He as carrier gas). Electron impact mass spectra were recorded with ionization energy of 70 eV and an ionizing current of 0.2 mA. The temperature program used was: 150 °C for 5 min, 150 up to 300 °C at 10 °C/min, 300 °C for 12 min.

NMR spectroscopy

For structural assignments of the saccharide part, ^{1}H NMR and 2D NMR spectra were recorded using Bruker 600, 800, 950 MHz and 1.2 GHz spectrometers equipped with a reverse cryo-probe with gradients along the *z* axis. The samples were dissolved at a concentration of 1 mg in 550 μL of D_2O and the spectra were calibrated with internal acetone (δ_H = 2.225 ppm; δ_C = 31.45 ppm). The conditions were as follows: for the **OS1** was recorded 550 μL D_2O, 32 °C at 600 MHz and the **OS2** 550 μL D_2O, 27 °C at 950 MHz. Total Correlation Spectroscopy (TOCSY) were performed with spinlock times of 100 ms using data sets (t1 × t2) of 4096 × 1024 points. Nuclear Overhauser Enhancement Spectroscopy (NOESY) experiments were performed using data sets (t1 × t2) of 4096 × 1024 points with mixing times of 200–400 ms. Heteronuclear Single-Quantum Coherence (HSQC), and Heteronuclear Multiple-Bond Correlation (HMBC) experiments were performed in the ^{1}H-detection mode by single-quantum coherence with proton decoupling in the ^{13}C domain using data sets of 2048 × 900 points[13]. HSQC was performed using sensitivity improvement and the phase-sensitive mode using echo/antiecho gradient selection, with multiplicity editing during the selection step[14]. HMBC was optimized on long-range coupling constants, with a low-pass *J* filter to suppress one-bond correlations, using gradient pulses for selection. HMBC spectra were optimized for 6–15 Hz coupling constants. The data matrix in all the heteronuclear experiments was extended to 4092 × 2048 points and transformed by applying a qsine or a sine window function[15].

MALDI-TOF mass spectrometry and MS/MS

MALDI-TOF MS and MS/MS analysis of lipid A were performed on an ABSCIEX TOF/TOF™ 5800 Applied Biosystems mass spectrometer equipped with a Nd:YLF laser with a λ of 345 nm, a 3 ns pulse width and a repetition rate of up to 1000 Hz. The lipid A was dissolved in $CHCl_3$/MeOH (1:1, v/v), as previously described[16]. The matrix was the trihydroxyacetophenone dissolved in methanol/0.1%trifluoroacetic acid/acetonitrile (7:2:1, v/v/v) at a concentration of 75 mg/mL. As for the ammonium hydroxide-treated lipid A this was dissolved in $CHCl_3$-trifluoroethanol (4:1, v/v) and the matrix used was 2,5-dihydroxybenzoic acid (DHB) in acetonitrile 0.2% trifluoroacetic acid (7:3, v/v)[6]. The lipid A solution (0.5 μL) and the matrix solution (0.5 μL) were deposited on the MALDI plate and dried at room temperature. All spectra were a result of the accumulation of 2000 laser shots, whereas 6000–7000 shots were summed for the MS/MS data acquisitions as described previously[5]. Each experiment was performed in triplicate.

High-resolution electrospray ionization (ESI) Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR MS)

Measurements were performed on a 15 T solariX XR FT-ICR mass spectrometer (Bruker Daltonics) equipped with a CombiSource and a ParaCell. The system was operated using ftmsControl software (Bruker Daltonics). Direct infusion experiments of a 20 μg/mL solution of the deacylated LOS in water/ACN/formic acid (v/v 49.95%:50%:0.05%) were performed at a 2 μL/min using a syringe pump. The ESI dry gas

flow rate, drying temperature, nebulizer gas pressure and capillary voltage were 4.0 L/min, 200 °C, 1.0 bar, and -4500 V respectively. Mass spectra were acquired in the *m/z*-range 92–5000 with 2 M data points. For collision-induced dissociation (CID) experiments, precursor ions were selected in the quadrupole and fragmented in the hexapole cell. Collision energies were optimized for each precursor ion to improve the fragmentation efficiency. Interpretation of the MS and MS/MS spectra was visually performed in DataAnalysis 5.0 SR1 (Bruker Daltonics) and complemented by software-based fragment assignment using GlycoWorkbench 2. Fragment ions were assigned within 10 ppm error and allowing a maximum of up to 3 glycosidic cleavages. For each fragment ion, only one isomeric structure out of those possible was illustrated in the figures[17,18]. Deconvolution to $[M+H]^+$ ions was performed in DataAnalysis. Deconvoluted *m/z*-values and intensities (both absolute and relative) were exported in Microsoft Excel. Ions with a relative intensity higher than 3% were then plotted to generate a deconvoluted spectrum.

Structural assignment of the fully deacylated LOS

NMR-based structure determination of OS. An aliquot of pure LOS was fully de-acylated to obtain the core oligosaccharide connected to the lipid A sugar skeleton (**OS**, Figs. 1–2), in turn further purified and analyzed by means of NMR and MS techniques. In particular, due to its high complexity (Fig. 1 and supporting text and figures), the full assignment of **OS** required the use of ultra-high magnetic field at 1.2 GHz and allowed to assign all spin systems and to establish the challenging saccharide sequence (^{1}H NMR in Fig. 1). The anomeric configuration of each monosaccharide unit was determined based on the $^3J_{H1,H2}$ coupling constant values, whereas those of vicinal $^3J_{H,H}$ ring coupling constants led to the identification of the relative configuration of each sugar residue. Finally, combined information derived from *inter*-residue NOE connectivity and heteronuclear long-range correlations, together with compositional analysis, allowed the complete characterization of core **OS** from *A. muciniphila*.

The analysis of the NMR spectra revealed the presence of 26 sugar units, all present as pyranose rings, according to both ^{13}C chemical shift values and long-range correlations between C-1/H-1 and H-5/C-5 in the ^{1}H,^{13}C HMBC spectrum (for the Kdo residues between C-2 and H-6) (Table S1, Fig. 1). These sugar units were distributed along two saccharide chains, the longest composed of a tetradecasaccharide (OS_{1deAc} residues from $\mathbf{A_1}$ to $\mathbf{O_1}$, Table S1 and Fig. 1), the shortest of a nonasaccharide (OS_{2deAc} residues from $\mathbf{A_2}$ to $\mathbf{T_2}$), both starting with a Kdo residue (**K** and **Z**), in turn connecting these oligosaccharides to a further branched Kdo unit. Ten spin systems were identified as *galacto* configured sugar units, as indicated by the $^3J_{H3,H4}$ and $^3J_{H4,H5}$ values (3 Hz and <1 Hz, respectively). Residue $\mathbf{C_1}$ and $\mathbf{C_2}$ were identified as a-Gal residues, with $^3J_{H1,H2}$ and $^1J_{CH}$ coupling constant values, together with *intra*-residual NOE contact of H-1 with H-2, diagnostic of their α-configurations and $^4C_1$ ring conformation. $\mathbf{D_1}$, $\mathbf{D_2}$ and $\mathbf{P_2}$ were identified as a GalNAc residues, as attested by the correlation in the HSQC spectrum of H-2 resonance to a nitrogen bearing carbon signals (Fig. 1, and Table S1). The *intra*-residual NOE contact of H-1 with H-3 and H-5, $^3J_{H1,H2}$ and $^1J_{CH}$ values were indicative of their β–anomeric configuration. Residues $\mathbf{E_1}$, $\mathbf{M_1}$, $\mathbf{N_1}$, $\mathbf{E_2}$ and $\mathbf{M_2}$ were identified as a-Fucose units, as attested by the scalar correlations of the ring protons with the methyl signal at position 6, visible in the TOCSY spectra.

Spin systems $\mathbf{I_1}$ was identified as a b-Mannose, i.e., the *manno* configuration was established by $^3J_{H-1,H-2}$ and $^3J_{H-2,H-3}$ values, both below 2 Hz and diagnostic of the H-2 equatorial orientation; similarly, $\mathbf{B_1}$ and $\mathbf{B_2}$ were identified as α-Heptose residues. Eight residues were *gluco*-configured, as indicated by the ring $^3J_{H,H}$ coupling constants (values about 8-10 Hz); the anomeric configuration was established as above. In detail, residues $\mathbf{A_1}$, $\mathbf{O_1}$, $\mathbf{A_2}$ and $\mathbf{T_2}$ were identified as GlcNAc residues, all a-configured but $\mathbf{O_1}$; residues $\mathbf{F_1}$, $\mathbf{H_1}$, and $\mathbf{L_1}$ were Glucose units, all b-configured whereas $\mathbf{G_1}$ unit was in a-configuration.

Spin system $\mathbf{K_1}$ and $\mathbf{K_2}$ were attributed to 4,5,7-Kdo residues on the basis of the diastereotopic methylene proton signals resonating in the high-field region of the ^{1}H NMR spectrum; on the basis of the H-3 proton chemical shift it was possible to assign them an α-anomeric configuration. Interestingly, spin systems were identified and attributed to Kdo residue ($\mathbf{Z_t}$ and **Z**) whose location was definitively assessed by the punctual analysis of HR-MS spectra (see below). The glycosylated positions of the sugar units were identified with the downfield shift of carbon resonances (Table S1). The *inter*-residual NOE contacts (Fig. S3) together with the long-range correlations derived from the HMBC spectrum (Fig. 1) allowed to determine the **OS** sequence as depicted in Fig. 1.

Mass spectrometry-based structure determination of OS. To further confirm the NMR-based structural information, ultrahigh-resolution ESI FT-ICR MS experiments were recorded on *A. muciniphila* **OS**. This MS investigation of the OS allowed to confirm the presence of three Kdo moieties per oligosaccharide, a specific fucose decoration and the final resulting structure (Fig. 1). Of note, the compositional analysis revealed the presence of t-Kdo, 4,5-Kdo and 4,5,7-Kdo.

The deacylated LOS sample was analyzed by direct infusion positive ion mode ESI FT-ICR MS (Fig. S4a). Among the most abundant ion species that were detected, two major oligosaccharide structures, namely $\mathbf{OS_{2deAc}}$, which included the nonasaccharide unit, and $\mathbf{OS_{1deAc}}$, which included the tetradecasaccharide unit, and some variants with residual acylation (+ 240.2089 Da) and/or acetylation (+ 42.0106 Da) were identified (Fig. S4b). Smaller and truncated oligosaccharide structures, which could have been generated by in-source fragmentation, were also detected. Identifications were based on accurate mass measurement in MS (i.e., sub-ppm errors) and MS/MS analysis using collision-induced dissociation (CID) on selected precursor ions. The high heterogeneity of the ion species and their charge state distribution limited MS/MS analysis since the overlaps between different isotopic distributions hampered the isolation of a single ion species for further fragmentation experiments. Therefore, MS/MS analysis was solely performed on some of the most intense ion species detected. Examples of assigned MS/MS spectra are reported in Figs. S6-S14. Interpretation of the CID MS spectra was performed considering neutral losses of H_2O, HPO_3, H_3PO_4, C_2H_2O (loss of acetylation), $C_{15}H_{28}O_2$ (loss of acylation C15:0 3OH), possible fucose rearrangement, and the presence of residual acetylation on $HexNH_2$ residues. Cross-ring fragmentation was not evaluated. Since isomeric fragment ions cannot be distinguished in the mass spectra, only one of all possible structures is reported per fragment ion in the figures. The fragmentation scheme and the description of the symbols are reported in Fig. S5. In all FT-ICR mass spectra, the *m/z*-values of monoisotopic peaks are indicated.

$\mathbf{OS_{2deAc}}$ was detected as $[M+5H]^{5+}$, $[M+4H]^{4+}$ and $[M+3H]^{3+}$ at *m/z* 491.1601, *m/z* 613.6982 and *m/z* 817.9285, respectively, and at *m/z* 2451.7707 in the deconvoluted mass spectrum. $\mathbf{OS_{2deAc}}$ variants included an acetylated form (*m/z* 2493.7813), an acylated form (*m/z* 2471.8044), an acetylated and acylated form (*m/z* 2733.9898), and different truncated forms (i.e., *m/z* 2231.7125, *m/z* 1983.5754, *m/z* 1763.5170, *m/z* 1529.5840 and *m/z* 1350.5044) which were detected with or without residual acetylation and acylation. Variants generated from the loss of HPO_3 were also detected (e.g., *m/z* 2371.8044). The structure of $\mathbf{OS_{2deAc}}$ was corroborated by CID analysis (Figs. S6-S7). Additional structural information was obtained from the MS/MS analysis of the $\mathbf{OS_{2deAc}}$ lacking one Kdo residue (Fig. S8).

$\mathbf{OS_{1deAc}}$ was detected as $[M+4H]^{4+}$ and $[M+3H]^{3+}$ at *m/z* 812.5113 and *m/z* 1083.0128, respectively and at m/z 3247.0238 in the deconvoluted mass spectrum. As for $\mathbf{OS_{2deAc}}$, acetylated and acylated variants were also observed for $\mathbf{OS_{1deAc}}$ (*m/z* 3289.0333 and *m/z*

3487.2320). Other **OS$_{1deAc}$** variants, lacking one or more monosaccharide residues or that generated from the loss of HPO_3, were also detected but at a lower intensity compared to what was observed for **OS$_{2deAc}$**. Informative MS/MS spectra were obtained for **OS$_{1deAc}$** carrying one acetylation (Fig. S9) and **OS$_{1deAc}$** carrying one acetylation and one acylation (i.e., Figs. S10–S14).

To be noted, the MS approach does not allow the distinction between isomeric monosaccharide residues (e.g., galactose and glucose) therefore the reported monosaccharide compositions were based on the results obtained by NMR and compositional analysis.

Structural assignment of the LOS after acetic acid treatment

Structural characterization of the lipid A by MALDI-TOF MS and MS/MS. The structure of the lipid A was established by merging data from chemical analysis (Table S5) and MALDI-TOF MS and MS/MS investigation of the product derived from a mild acid hydrolysis (Fig. S21, Supplementary data 1) executed on an aliquot of *A. muciniphila* MucT LOS. The reflectron MALDI-TOF MS spectrum recorded in negative-ion polarity of the lipid A is reported in Fig. 2, which showed extensive heterogeneity of the lipid A under investigation. The MS spectrum shows three intense groups of ions around *m/z* 1922.7, 1670.6, and 1430.5 identified and assigned as tetra-, penta, and hexa-acylated lipid A; each of these groups is characterized by the presence of mass differences of 14 amu (−CH_2− unit) indicative of lipid A species differing in the length of the fatty acid chains. In addition, for each of these clusters of peaks, *mono*-phosphorylated forms were also identified at around *m/z* 1842.7, 1590.6, and 1350.5, respectively (Fig. 2, Table S4–S5). In detail, the most intense peak at *m/z* 1922.7 was attributed to a *bis*-phosphorylated hexa-acylated lipid A species carrying *i*C15:0 (3-OH) as primary *O*-linked and *N*-linked fatty acids, whereas *i*C14:0 and C17:0 [or *i*C17:0, or *a*C17:0] residues were the secondary acyl substituents. The peak at *m/z* 1670.6 was assigned to a *bis*-phosphorylated penta-acylated lipid A species lacking the C17:0 unit, whereas the peak at *m/z* 1430.5 was attributed to a *bis*-phosphorylated tetra-acylated lipid A species devoid of both C17:0 and of one primary *O*-linked *i*15:0 (3-OH).

A negative-ion MS/MS analysis was conducted to unveil the location of the lipid A acyl moieties with respect to the glucosamine disaccharide backbone as well as the position of the phosphate in the *mono*-phosphorylated lipid A forms. The MS/MS spectrum of the precursor ion at *m/z* 1922.7 (Fig. S21a) showed two intense peaks at *m/z* 1824.7 and *m/z* 1664.7 attributed to ions derived from the loss of one phosphate and a *i*15:0 (3-OH), respectively. A less intense peak at *m/z* 1652.7 was matched with a lipid A fragment originated by the loss of the secondary C17:0 [or *i*C17:0, or *a*C17:0]. The observation of the ion peak at *m/z* 1314.6, attributed to a fragment derived from the loss of a whole unit of *i*15:0 (3-OH) carrying C17:0 [or *i*C17:0, or *a*C17:0], was fundamental for the structural characterization. The occurrence of this peak in fact gave a first indication that the C17:0 was bound as a secondary substituent of a primary ester linked *i*15:0 (3-OH). In parallel, the presence of the ion at *m/z* 738.1, originated from the cleavage of the glycosidic linkage of the glucosamine backbone (Y_1)[19], in turn suggested the absence of any secondary acyl chain on the reducing glucosamine, which carried the sole primary *i*15:0 (3-OH). Likewise, the structures of the *bis*-phosphorylated lipid A species detected at *m/z* 1670.6 and 1430.5 were also defined (Figs S21b and S21c, Table S5), revealing a penta-acylated lipid A carrying four primary *i*15:0 (3-OH), and one *i*14:0 as secondary acyl substituent on the amide-linked *i*15:0 (3-OH) of the nonreducing glucosamine for the species detected at *m/z* 1670.6, and a tetra-acylated species bearing two *i*15:0 (3-OH) on the reducing glucosamine, and one *i*15:0 (3-OH) carrying *i*14:0 on the nonreducing glucosamine for the species at *m/z* 1430.5. This structural deduction was further corroborated by the MALDI-TOF MS analysis of the lipid A after treatment with NH_4OH, as previously described by Silipo et al[6].. This treatment selectively removes acyl and acyloxyacyl esters, while leaving the acyl and acyloxyacyl amides unaffected. The negative-ion MALDI-TOF MS spectrum of NH_4OH treated lipid A displayed a main peak at *m/z* 1189.9 that matched with deprotonated $[M-H]^-$ tri-acylated lipid A made up of the *N*-linked *i*15:0 (3-OH) and one secondary *i*14:0 substituent. This analysis, complemented with data attained from MS/MS analysis, definitively proved that the hexa-acylated lipid A from *A. muciniphila* MucT possesses a 4 + 2 symmetry of the acyl chains with respect to the disaccharide backbone, with C17:0 [or *i*C17:0, or *a*C17:0] and *i*14:0 as ester linked and amide-linked 3-acyloxyacyl residues, respectively for the species at m/z 1922.7 (Fig. 1). Also in this case, lipid A forms differing in 14 amu (-CH2-) and for the absence of one phosphate unit were also identified.

Finally, to gain insight into the position of the phosphate unit in the *mono*-phosphorylated lipid A species, a negative-ion MS/MS analysis was performed on several peaks and agreed with the phosphate carried only on the nonreducing glucosamine. As example, in the MS/MS spectrum recorded on the precursor ion at *m/z* 1350.5 (Fig. S21b), a representative ion peak of the cluster ascribed to *mono*-phosphorylated tetra-acylated lipid A, it was fundamental to observe the peaks at *m/z* 1051.2 and *m/z* 750.8, which derived from the cross-ring fragmentations ^{0,2}A2 and ^{0,4}A2, respectively. The presence of these ion peaks clearly demonstrated that the phosphate is on the nonreducing glucosamine, which in turn is acylated by the *i*15:0 (3-OH) bearing the secondary *i*14:0 unit. Therefore, combining data from the MALDI-TOF MS and MS/MS analysis of the mild acid hydrolysis and NH_4OH products with information acquired from compositional analysis, it was possible to settle that the LOS from *A. muciniphila* MucT is made up of a complex mixture of *mono*- and *bis*-phosphorylated tetra- to hexa-acylated lipid A species mostly decorated by branched primary and secondary acyl chains.

NMR analysis of the OS1 and OS2 of A. muciniphila. The LOS from *A. muciniphila* underwent a mild acid hydrolysis to split the lipid A and the core oligosaccharide portion. After separation and purification two oligosaccharides were isolated, **OS1** and **OS2**, and for both a thorough compositional and methylation analyzes were carried out (Fig. S15). The results fully matched with NMR data (see below, Supplementary data 2) driving to a univocal structural assignment for **OS1** and **OS2**.

OS1-NMR

The NMR spectra of **OS1** (Figs. S16–S18, Table S2) revealed the presence of 14 sugar units. Spin systems **C$_1$**, **D$_1$**, **E$_1$**, **M$_1$** and **N$_1$** were identified as *galacto* configured sugar units, as indicated by the $^3J_{H3,H4}$ and $^3J_{H4,H5}$ values (3 Hz and < 1 Hz, respectively). **C$_1$**, **E$_1$**, **M$_1$** and **N$_1$** had $^3J_{H1,H2}$ and $^1J_{CH}$ coupling constant values, together with *intra*-residual NOE contact of H-1 with H-2, were diagnostic of their α-configurations and $^4C_1$ ring conformation. **D$_1$** was in b-configuration ($^3J_{H1,H2\ value}$ = 8 Hz and *intra*-residue NOE contacts H1-H3-H5) and was identified as a GalN as its C-2 correlated to nitrogen bearing carbon signal. **E$_1$**, **M$_1$** and **N$_1$** were identified as a-Fuc units, as attested by the scalar correlations of the ring protons with the methyl signal at position C6; in particular, residue **M$_1$** was acetylated at position O2 as shown by its H2 downfield shifted signal and methyl signal at 2.15 which, in the HMBC, correlated to a carbonyl signal at 173.0 ppm which in turn correlated to H2 of residue **M$_1$**. Spin systems **A$_1$**, **F$_1$**, **G$_1$**, **H$_1$**, **L$_1$** and **O$_1$** were identified as *gluco* configured sugar units, indicated by the ring $^3J_{H,H}$ coupling constants values (8–10 Hz), the anomeric configurations were identified by $^3J_{H1,H2}$ values. **A$_1$** and **O$_1$** were identified as GlcN due to the correlation of C-2 to nitrogen bearing carbon signal. **B$_1$** and **I$_1$** were identified as *manno* configured rings, because of the lack of signals in the COSY and TOCSY spectra with other protons from the spin system other than H-2; **B$_1$** was in b-configuration, as showed by the intra-residue NOE contacts of H1 with H3 and H5 and by the $^1J_{CH}$ coupling constant of 164 Hz. Spin system **K** was attributed to the Kdo residue on

the basis of the diastereotopic methylene proton signals resonating in the high-field region of the ^{1}H NMR spectrum.

OS2-NMR

The NMR spectra of **OS2** (Figs. S19 and S20, Table S3) revealed the presence of 9 sugar units. Spin systems $\mathbf{C_2}$, $\mathbf{D_2}$, $\mathbf{E_2}$, $\mathbf{N_2}$ and $\mathbf{P_2}$ were identified as *galacto* configured sugar rings, as indicated by the $^3J_{H3,H4}$ and $^3J_{H4,H5}$ values (3 Hz and <1 Hz, respectively). $\mathbf{C_2}$, $\mathbf{E_2}$, $\mathbf{N_2}$ and $\mathbf{P_2}$ had $^3J_{H1,H2}$ and $^1J_{CH}$ coupling constant values, together with *intra*-residual NOE contact of H-1 with H-2, were diagnostic of their α-configurations and $^4C_1$ ring conformation. $\mathbf{D_2}$ is b-configured. $\mathbf{D_2}$ and $\mathbf{P_2}$ were identified as a GalN as their C-2 correlated to nitrogen bearing carbon signal. $\mathbf{E_2}$ and $\mathbf{N_2}$ were identified as a-Fuc units as attested by the scalar correlations of the ring protons with the methyl signal at position 6. Spin systems $\mathbf{A_2}$ and $\mathbf{T_2}$ were identified as *gluco* configured sugar rings, indicated by the ring $^3J_{H,H}$ coupling constants (8–10 Hz), the anomeric configurations were identified as above. $\mathbf{A_2}$ and $\mathbf{T_2}$ were identified as GlcN due to the correlation of C-2 to nitrogen bearing carbon signal.

$\mathbf{B_2}$ was *manno* configured sugar ring, because of the lack of signals in the COSY and TOCSY spectra with other protons from the spin system other than H-2. Spin system $\mathbf{K_2}$ was attributed to the Kdo residue on the basis of the diastereotopic methylene proton signals resonating in the high-field region of the ^{1}H NMR spectrum. Interestingly, this spectrum contains a free, reducing Kdo unit, $\mathbf{X_2}$.

Mice experiments

Sixteen 9-week-old male C57BL/6 J mice were purchased from Janvier (Le Genest-Saint-Isle, France) and co-housed in pairs under Specific and Opportunistic Pathogen Free (SOPF) conditions in a controlled environment (temperature of 22 ± 2 °C, 12 h daylight cycle, and humidity-controlled room) with free access to food and water. Upon arrival, mice underwent a 1-week acclimatization period, during which they were fed a control diet (CT) (AIN93Mi, Research Diet, New Brunswick, NJ, USA). On the day of the experiment, mice were body weight-match assigned to the two experimental groups. One mouse had to be excluded due to an abnormally low body weight, as confirmed by the GRUBBS outlier test.

Subsequently, one set of mice was injected interperitoneally with 230 µl buffer (vehicle) solution (Tris-NaCl 50 mM; CT group, $n = 7$), while the other set was injected with 230 µl of LOS solution (for a final concentration of 300 µg/kg body weight in buffer; LOS group, $n = 8$). Four hours after injection body temperature was measured using a rectal probe for small rodents (model BAT-12, Physitemp, Clifton, NJ, USA) after which mice were anesthetized with isoflurane (Forene®, Abbott, Queenborough, Kent, England) before being killed by cervical dislocation. The liver, spleen and cecum were dissected and weighted. The liver was clamped in liquid nitrogen and stored at −80 °C for further analysis.

Our research complies with all relevant ethical regulations. All mice experiments were approved by the Ethical Committee for Animal Care of the Health Sector of the Université Catholique de Louvain (UCLouvain) headed by Prof. J-P Dehoux, under number 2017/UCL/MD/005, and were performed in accordance with the guidelines of the Local Ethics Committee and in accordance with the Belgian Law of 29 May 2013, regarding the protection of laboratory animals (agreement number LA1230314).

LOS injection showed no effect on body weight or weight of organs, such as liver, cecum or cecal content, although a small increase in spleen weight was reported compared to the vehicle injection, possibly indicating an effect on the immune system (see Supplementary material Fig. S22).

RNA Preparation and gene expression analysis by real-time qPCR analysi

Total RNA was prepared from liver tissue using TriPure reagent (Roche). Quantification and integrity analysis of total RNA was performed by running 1 µL of each sample on an Agilent 2100 Bioanalyzer (Agilent RNA 6000 Nano Kit, Agilent). cDNA was prepared by reverse transcription of 1 µg total RNA using a Reverse Transcription System kit (Promega, Leiden, The Netherlands). Real-time PCRs were performed with the StepOnePlus real-time PCR system and software (Applied Biosystems, Den Ijssel, The Netherlands) using Mesa Fast qPCR sybr green mix (Eurogentec, Seraing, Belgium) and with the CFX Manager 3.1 software (Bio-Rad, Hercules, CA) using Mesa Fast qPCR (GoTaq qPCR Master Mix, Promega, Madison, WI, USA) for detection, according to the manufacturer's instructions. RPL19 was chosen as housekeeping gene. All samples were run in duplicate in a single 96-well reaction plate, and data were analyzed according to the $2^{-\Delta\Delta Ct}$ method. The identity and purity of the amplified product was checked through analysis of the melting curve carried out at the end of amplification.

TLR analysis

HEK-Blue™ hTLR2, hTLR2-TLR1, hTLR2-TLR6, hTLR2 KO-TLR1/6 and hTLR4 cells (resp. hkb-htlr2, hkb-htlr21, hkb-htlr26, hkb-htlr2k16, hkb-htlr4, Invivogen, CA, USA) were used to screen for TLR2, hTLR2/1, TLR2/6 and/or TLR4 activation. In these cell lines, stimulation of TLR2-1, TLR2-6 and TLR4 and subsequent activation of NF-κB and AP-1 induces the production of secreted embryonic alkaline phosphatase (SEAP), which can be quantified spectrophotometrically. These cell lines are widely used in the field (>150 citations for utilization of hTLR2 cell lines alone) and are a validated model system to explore hTLR interactions (Invivogen, CA, USA). No commonly misidentified cell lines were used in the study.

Cell lines were grown and subpassaged according to manufacturer's instructions. Cells were subpassaged at 70–80% of confluency in a maintenance medium of Dulbecco's Modified Eagle Medium (DMEM) supplemented with GlutaMAX™ and 4.5 g/L D-glucose (Gibco), 10% (v/v) of Fetal Calf Serum (Hyclone), 1x MEM non-essential amino acids (Gibco), 100 U/mL penicillin, 100 µg/mL streptomycin (Gibco), 100 µg/mL normocin (Invivogen), and 1x HEK-Blue™ Selection (Invivogen). Cells were maximally maintained until passage 30 for hTLR2 cells and passage 5 for the other cell lines. Receptor activation was tested by seeding cells at 280.000 (all but hTLR4 cells) and 140.000 cells/mL (hTLR4 cells only), respectively, in flat-bottom 96-well plates in maintenance medium without HEK-Blue™ Selection in a total volume of 180 µL/well. After 1 day, cells were stimulated by addition of 20 µL of the stimulant of interest (up to 3 µg/mL diluted in PBS) *in triplo*. Plates were incubated for 14–20 h at 37 °C under 5 % CO_2. The following receptor ligands were used as positive control to test and ensure activity of the cell lines: Recombinant human interleukin 1 bèta (10 ng/µL, Invivogen) for all cells; Pam3CSK4 (100 ng/mL, Invivogen) for TLR2 and TLR2-1 cells; FSL-1 (1 µg/µL, Invivogen) for TLR2-6 cells; and Ultrapure LPS from *E. coli* O111:B4 and K-12 (up to 3 µg/mL, Invivogen) for TLR4 cells. PBS (dilution reagent of the proteins of interest) and culture medium without selection antibiotics were used as a negative control. SEAP activity – and thus NF-kB activation – was detected by measuring the absorbance at 620 nm (SpectraMax® iD3, Molecular Devices, LLC, USA). All signals were normalized to negative control signals (DMEM). All experiments were conducted in three independent biological repeats, each consisting of three technical repeats. Normalized signals were plotted as dose-response curves of data poins +-SD.

Data analysis

All data analyzes were carried out using GraphPad Prism v10 and are non-parametric (tests).

Statistics & reproducibility

All statistical details and reproducibility details are included in the sections on the relevant experiments.

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability
All elements necessary to allow interpretation and replication of results, including full datasets and detailed experimental procedures are provided in the Supplementary Information, as supplementary data files and in the source data file. Source data are provided with this paper.

References

1. Raetz, C. R. & Whitfield, C. Lipopolysaccharide endotoxins. *Annu Rev. Biochem* **71**, 635–700 (2002).
2. Di Lorenzo, F. et al. A journey from structure to function of bacterial lipopolysaccharides. *Chem. Rev.* **122**, 15767–15821 (2022).
3. Cani, P. D. et al. Metabolic endotoxemia initiates obesity and insulin resistance. *Diabetes* **56**, 1761–1772 (2007).
4. Cani, P. D. et al. Changes in gut microbiota control metabolic endotoxemia-induced inflammation in high-fat diet-induced obesity and diabetes in mice. *Diabetes* **57**, 1470–1481 (2008).
5. Derrien, M., Vaughan, E. E., Plugge, C. M. & de Vos, W. M. *Akkermansia muciniphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int J. Syst. Evol. Microbiol.* **54**, 1469–1476 (2004).
6. de Vos, W. M. Microbe profile: *Akkermansia muciniphila*: a conserved intestinal symbiont that acts as the gatekeeper of our mucosa. *Microbiol. (Read.)* **163**, 646–648 (2017).
7. Cani, P. D., Depommier, C., Derrien, M., Everard, A. & de Vos, W. M. *Akkermansia muciniphila*: paradigm for next-generation beneficial microorganisms. *Nat Rev Gastroenterol Hepatol*, https://doi.org/10.1038/s41575-022-00631-9 (2022).
8. Everard, A. et al. Cross-talk between *Akkermansia muciniphila* and intestinal epithelium controls diet-induced obesity. *Proc. Natl Acad. Sci. USA* **110**, 9066–9071 (2013).
9. Shin, N. R. et al. An increase in the *Akkermansia* spp. population induced by metformin treatment improves glucose homeostasis in diet-induced obese mice. *Gut.* **63**, 727–735 (2014).
10. Liu, S. et al. Oral administration of miR-30d from feces of MS patients suppresses MS-like symptoms in mice by expanding *Akkermansia muciniphila*. *Cell Host microbe* **26**, 779–794 e778 (2019).
11. Plovier, H. et al. A purified membrane protein from *Akkermansia muciniphila* or the pasteurized bacterium improves metabolism in obese and diabetic mice. *Nat. Med.* **23**, 107–113 (2017).
12. Wang, L. et al. A purified membrane protein from *Akkermansia muciniphila* or the pasteurised bacterium blunts colitis associated tumourigenesis by modulation of CD8(+) T cells in mice. *Gut.* **69**, 1988–1997 (2020).
13. Depommier, C. et al. Supplementation with *Akkermansia muciniphila* in overweight and obese human volunteers: a proof-of-concept exploratory study. *Nat. Med.* **25**, 1096–1103 (2019).
14. Ottman, N. et al. Pili-like proteins of *Akkermansia muciniphila* modulate host immune responses and gut barrier function. *PloS One* **12**, e0173004 (2017).
15. Molinaro, A. et al. Chemistry of lipid A: at the heart of innate immunity. *Chemistry* **21**, 500–519 (2015).
16. Vatanen, T. et al. Variation in microbiome LPS immunogenicity contributes to autoimmunity in humans. *Cell* **165**, 842–853 (2016).
17. Di Lorenzo, F. et al. Pairing *Bacteroides vulgatus* LPS structure with its immunomodulatory effects on human cellular models. *ACS Cent. Sci.* **6**, 1602–1616 (2020).
18. Pither, M. D. et al. *Bacteroides thetaiotaomicron* rough-type lipopolysaccharide: the chemical structure and the immunological activity. *Carbohydr. Polym.* **297**, 120040 (2022).
19. Lynch, A., Tammireddy, S. R., Doherty, M. K., Whitfield, P. D. & Clarke, D. J. The glycine lipids of Bacte*roides thetaiotaomicron* are important for fitness during growth In Vivo and In Vitro. *Appl Environ Microbiol.* **85**, e02157–18(2019).
20. Nichols, F. C. et al. Glycine lipids of *Porphyromonas gingivalis* are agonists for toll-like receptor 2. *Infect. Immun.* **88**, e00877–19 (2020).
21. Zhang, Q. et al. Genetic mapping of microbial and host traits reveals production of immunomodulatory lipids by *Akkermansia muciniphila* in the murine gut. *Nat. Microbiol.* **8**, 424–440 (2023).
22. Geng, Z. H., Zhu, Y., Li, Q. L., Zhao, C. & Zhou, P. H. Enteric nervous system: the bridge between the gut microbiota and neurological disorders. *Front Aging Neurosci.* **14**, 810483 (2022).
23. Abot, A. et al. Pasteurized *Akkermansia muciniphila* improves glucose metabolism is linked with increased hypothalamic nitric oxide release. *Heliyon* **9**, e18196 (2023).
24. Shimoyama, A. et al. Lipopolysaccharide from gut-associated lymphoid-tissue-resident *Alcaligenes faecalis*: complete structure determination and chemical synthesis of its lipid A. *Angew. Chem.* **60**, 10023–10031 (2021).
25. Manzanillo, P., Eidenschenk, C. & Ouyang, W. Deciphering the crosstalk among IL-1 and IL-10 family cytokines in intestinal immunity. *Trends Immunol.* **36**, 471–478 (2015).
26. Wang, Y. et al. *Akkermansia muciniphila* induces slow extramedullary hematopoiesis via cooperative IL-1R/TLR signals. *EMBO Rep.* **24**, e57485 (2023).
27. Jacobson, A. N., Choudhury, B. P. & Fischbach, M. A. The biosynthesis of lipooligosaccharide from bacte*roides thetaiotaomicron*. *mBio.* **9**, e02289–17 (2018).

Acknowledgements
PGV was supported by the European Commission via the International Training Network Train2Target (721484). MTM was supported by the European Commission via the H2020-MSCA-ITN-2020 GLYTUNES (contract n° 956758). PDC is honorary research director at FRS-FNRS (Fonds de la Recherche Scientifique) and is recipient of grants from FRFS-WELBIO: WELBIO-CR-2022A-02, EOS: program no. 40007505). The authors acknowledge the support and the use of resources of Instruct-ERIC, a landmark ESFRI project, and specifically the CERM/CIRMMP Italy center. AS and AM acknowledge PNRR, Missione 4 – Componente 2 – NextGenerationEU – Partenariato Esteso INF-ACT – One Health Basic and Translational Research Actions Addressing Unmet Needs on Emerging Infectious Diseases MUR: PE00000007. AM and FDL acknowledge PNRR, Missione 4 – Componente 2 – NextGenerationEU – Partenariato Esteso ON Foods-Project code PE00000003, Concession Decree No. 1550 of 11 October 2022 adopted by the Italian Ministry of University and Research, CUP D93C22000890001, Project title "ON Foods"-Research and innovation network on food and nutrition Sustainability, Safety and Security–Workin on Foods". This study was also supported with a grant to AM and FDL from the Italian Ministry of Health - Health Operational Plan Trajectory 5 - Line of action "Creation of an action program for the fight against malnutrition in all its forms and for the dissemination of the principles of the diet Mediterranean" (Mediterranean Diet for Human Health Lab "MeDiHealthLab"; code T5-AN-07). AM acknowledges National Recovery and Resilience Plan (NRRP), CN2 Agritech National Research Center and the European Union Next-Generation EU Missione 4, Componente 2, Investimento 1.4- D.D. 1032 CN00000022. FDL acknowledges Ministry of Education, Universities and Research, PRIN MUR 2022 (2022SHW3KY) and PRIN MUR PNRR 2022 (P202293ZMC), European Research Council (ERC) under the Horizon Europe program under grant agreement No 101039841 (DEBUGGING LPS). This work was also supported in part by the Italian Ministry of Foreign Affairs and International Cooperation (Italy-Germany Science and Technology cooperation–Call for joint research proposals for the years 2023–2025) to FDL. WMdV and his team were supported by

the SIAM Gravitation Grant 024.002.002 of the Netherlands Organization for Scientific Research and the Advanced Research Grant Microbes Inside (AdG 250172) of the European Research Council.

Author contributions

The study was conceived by A.M. and W.Md.V. P.G.V. and M.T.M. isolated the LOS under the supervision of C.D.C. and A.S. S.N. and F.D.L. carried out and assigned MS on the OS and the Lipid A respectively. A.S. and M.F. carried out NMR experiments. A.S. performed NMR assignments of OS, OS_1 and OS_2. H.L.P.T. and J.E. carried out signaling experiments. P.D.C., M.M.V., and H.P. conceived and carried out the mouse experiments. P.G.V., H.L.P.T., A.S., F.D.L., S.N., A.M., W.Md.V. drafted the manuscript.

Competing interests

WMdV and PDC are inventors on patent applications dealing with the use of specific bacteria and components in the treatment of different diseases and have co-founded The Akkermansia Company SA. PDC was cofounder of Enterosys. All the other authors declare that they have no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-52683-x.

Correspondence and requests for materials should be addressed to Antonio Molinaro or Willem M. de Vos.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

[1]Department of Chemical Sciences, University of Naples Federico II, Napoli, Italy. [2]Laboratory of Microbiology, Wageningen University, Wageningen, The Netherlands. [3]Louvain Drug Research Institute, Metabolism and Nutrition Research group, UCLouvain, Université Catholique de Louvain, Brussels, Belgium. [4]Walloon Excellence in Life Sciences and BIOtechnology (WELBIO), WELBIO department, WEL Research Institute, avenue Pasteur, 6, Wavre, Belgium. [5]Institute of Experimental and Clinical Research (IREC), UCLouvain, Université catholique de Louvain, Brussels, Belgium. [6]Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, The Netherlands. [7]Magnetic Resonance Center (CERM), CIRMMP and Department of Chemistry "Ugo Schiff", University of Florence, Florence, Italy. [8]Department of Chemistry, School of Science, Osaka University, Osaka, Japan. [9]Human Microbiome Research Program, Faculty of Medicine, University of Helsinki, Helsinki, Finland. [10]Present address: Nestlé Institute of Health Sciences, Nestlé Research, Société des Produits Nestlé S.A., Lausanne, Switzerland. [11]These authors contributed equally: Pilar Garcia-Vello, Hanne L. P. Tytgat. [12]These authors jointly supervised this work: Antonio Molinaro, Willem M. de Vos. ✉e-mail: molinaro@unina.it; willem.devos@wur.nl