Supporting information material to

Tackling the chemical diversity of microbial nonulosonic acids – a universal large-scale survey approach

Hugo B. Kleikamp, Yue Mei Lin, Duncan G.G. McMillan, Jeanine S. Geelhoed , Suzanne N.H. Naus-Wiezer, Peter van Baarlen, Chinmoy Saha, Rogier Louwen, Dimitry Y. Sorokin, Mark C.M. van Loosdrecht and Martin Pabst*

Contact: m.pabst@tudelft.nl

Experimental procedures

Additions to results and discussions

Experimental Procedures

A. Cell lysis, acid hydrolysis and alpha-keto acid specific labelling. Liquid cultures were pelleted at 4200 rpm for 10 minutes, and supernatants were removed and washed with PBS. Further, cell pellets and all other solid materials were freeze-dried, followed by physical disruption and homogenisation. Then, 1 mL of 2 M acetic acid solution was added to 2.5 mg of freeze-dried biomass. The resolubilised material was hydrolysed at 80 °C for 2 hours, after which samples were centrifuged for 5 min at 14K rpm. From every sample hydrolysate, 10 µL were dried at 45 °C under reduced pressure using a speed vac concentrator. Labelling was performed by adding 20 µL labelling solution at 50 °C for 2.5 hours. The labelling solution consisted of 1.4 M acetic acid, 0.75 M 2-mercaptoethanol (beta-mercaptoethanol), 18 mM of sodium dithionate and 7 mM of DMB. Reference sialic acid standards were labelled directly and were prepared to reach a final concentration of 2.5 pmol per µL injection solution. MS-grade water was added to the reference panel to reach a concentration of 10 pmol/µl. The labelling procedure follows the original work published by Hara et al.,[1] where 1,2-diamino-4,5-methylene dioxybenzene was introduced as the selective label for alpha-keto acids, forming a fluorescence-active quinoxaline derivative. Here, we decided to select the well-established DMB label (otherwise used for fluorescence detection) because it has a high degree of saturation and therefore the mass defect of labelled ulosonic acids distinguishes significantly from non-labelled metabolites or background signals. Furthermore, the large quinoxaline core guided highly unique ulosonic fragmentation pattern. The reaction is outlined in Figure S2.

B. Reverse-phase-Orbitrap-MS and segmented very small mass window scanning. An M-Class HSS T3 300µm x150mm C18 was mounted to an Acquity M-Class UPLC (Waters) using 97% H_2O plus 3% acetonitrile as solvent A and 97% acetonitrile plus 3% H_2O as solvent B (both 0.1% formic acid). A gradient from 12% B to 25% B was maintained at a flow-rate of 9 µL/min over 20 minutes, followed by a washing-step. Samples were injected in duplicates followed by blanks. Continuous fragmentation of very small mass segments was performed using a QE plus bench top Orbitrap mass spectrometer, operated in ES+ mode, in 2.5 Da steps from 380–520 Da (2.75 Da width). Fragmentation was performed at a NCE of 28. Alternating MS1 and MS2 scans at a loop count of 51 and a resolution of 70K with an AGC target of 5e5 for MS1, and 17K with an AGC target of 5e4 for MS2, were acquired in centroid mode. Confirmatory high-resolution and targeted experiments were performed at 140K resolution in HRMS, or PRM mode at 1 Da isolation, respectively. Calibration was carried out within every 24 hours.

C. Channel hit detection and NulO parent assignment by chemical filtering and structural scoring. Nonulosonic acid parent candidate identification was performed using Matlab (R2019b). Briefly, raw files were converted to mzxml ('.mzXML') format using the msConvert software tool. Peak lists were imported into Matlab using the 'mzxmlread' function, followed by 'mzxml2peaks' to extract MS1 and MS2 mass lists. MS1 data were deisotoped (within 5 ppm) by removing up to three ^{13}C isotopes, provided that intensity ratios were close to natural abundance ratios ($1^{13}C$=1, $2^{13}C$<0.25, $3^{13}C$<0.05, $4^{13}C$<0.01). Furthermore, MS2 peaks with less than 10^3 counts were excluded. For each fragmentation channel (2.75 Da window) a cell array was created, containing scan index, mass channel range, retention time, fragment peaks and peak intensities. MS2 scans containing core, class or reporter fragments were extracted (allowing a max mass error of 15ppm, see SI-table, sheet 9, script box 1), and only MS2 scans containing both, class and core fragments, were considered for further processing. For each scan, the class of the ulosonic acid was determined by the carbon length of the largest class fragment (see SI-table, sheet 9, script box 2). Neighbouring MS1 scans were then analysed for potential precursors (see SI-table, sheet 9, script box 3), where only precursor masses with a mass defect between 0.075–0.225 m/z were considered. For this, an empirical (NulO) chemical composition space was constructed considering min/max element counts ($C_{15\text{-}30}H_{16\text{-}35}N_{2\text{-}5}O_{7\text{-}15}P_{0\text{-}1}S_{0\text{-}1}$) and constrained ratios between elements and degree of unsaturation (RDBE >7.75, C/H ratio = 0.65–0.9, C/O ratio = 1.25–2.5, which was further corrected for amidation, phosphorylation and sulphation) (see SI-table, sheet 9, script box 4-5).[3] P/O and S/O ratios were set to <0.09, which elements also required the presence of in-source loss fragments of -79.96 Da or -79.95 Da, respectively. All precursors matching the constructed (NulO) chemical composition space, thereby allowing no more than 5 ppm mass error, received a chemical composition and were considered for further processing (see SI-table, sheet 9, script box 6). Precursor candidates occurring in >75% of all scans were regarded as static background and not further considered.

The passed MS1 parent mass candidates were finally subjected to a structural evidence scoring as described in the following (=decision tree). Parent mass candidates were scored based on the occurrence of water loss peaks, either observed as a neutral loss peak in the MS2 scan or as an in-source fragment in the MS1 scan (=Water score, +1 score for MS1/MS2 scans with water loss, $[M+H^+]^+$-H_2O, -18.0105). Further, scores were added according to the max. number of observed ulosonic acid (UA) fragments (2. # Fragments, +X scores equal to X = #core + #class-fragments (one per oxidation state)), and for the uniqueness of a parent mass candidate within a MS1 fragmentation window (=Window score, +1 score if parent mass candidate is the only 'realistic candidate'). Additional scores were granted for the fraction of co-elution of a parent mass candidate with the MS2 scans containing UA fragments (=Overlap, expressed as fraction= aligned/total). Parent mass candidates belonging to the same class and chemical composition were combined (intensity, scores, retention times and fragment signatures). In order to provide a comparative measure (to compare to random matches, as described below) a total score was established (=sum of Water score, Window score and # Fragments, multiplied by the fraction of Overlap).

To verify the significance over purely random matches, a 'total score-cut-off, was defined for every sample. For this, every dataset was subjected to semi-randomisation of the MS scan mass peaks and processed through the same pipeline as described above. Randomisation was performed using Matlabs 'rand' function, with the constrain, that randomised masses fall within the window of the lowest and highest masses of the original data (see SI-table, sheet 9, script box 7). The largest total score observed for the randomised data set defined the 'total score-cut-off'. Only parent mass candidates (from the correct dataset) with a total score greater than the 'score-cut-off' were considered for further analysis.
Finally, the minimum threshold for a valid ulosonic parent mass candidate (to be reported in SI-table, sheet 1) required a minimum MS1 intensity of 10^4 counts, a MS1 water loss in-source peak (a water loss peak (MS1 or MS2, during filtering), being matched twice in 2 consecutive runs, a minimum of 3 fragments (at least one class), and in addition either an MS2 water score or a window score. Hits showing strong mono-linker marker peaks as well as degradation makers were rejected. The hits were finally exported using the 'writetable' function. Peaks from the LC gradient wash region were excluded. The complete output of channel-hits and assigned NulO parents for all species/references as analysed above is presented in the **SI-table, sheets 1-3.**

D. Genome-level analysis of ulosonic acid biosynthesis (UAB) pathways. Analysis for homologues of oct- and non-ulosonic acid biosynthetic routes in genomes of species highlighted in Figure 2B was established as described in Lewis et al., 2009.[2] The following protein sequences were used for protein homology search using the NCBI Blast tool: Kdn Q8A710 (NAB1) and Q8A711 (NAB2), Neu P13266/Q8NFW8 (NAB1) and A0A0H3MPX1/Q9NR45 (NAB2); Pse A0A3K5CFB7 (NAB1) and A0A3X8VGM4 (NAB2), Leg Q0P8S7 (NAB1) and Q0P8T1 (NAB2). The Matlab function '[RID1, ROTE]=blastncbi(seq,'blastp','Entrez',species,'expect',1e-3)' was used to obtain the homology search request ID ('RID1') and estimated search time ('ROTE') for every fasta template (seq) and selected species ('species'). The homology search report was collected using 'report1=getblast(RID1,'WaitTime',ROTE,'ToFile','1CIV_report.xml')'. Results are summarised in **SI-table, sheet 7**.

E. Graphical representation of Channel-hit maps, mass binning graphs and fragment statistics.
The theoretical chemical space for NulOs was established by considering the three currently known core compositions (**Figure 1, A**) and by performing a combinatorial addition of up to 4 ('functional') modifications out of 10 frequently found modifications, plus one small mass-shift such as oxidation, dehydration and reduction. To further expand the chemical space and mass defect for yet undescribed diversifications we allowed also for one additional single amino acid and one additional sugar modification (see SI-table, sheet 8, script boxes 1-3). The formula weight for every composition was calculated and masses were binned in 5 Da channels using the 'histogram' function of Matlab. The difference to the nominal mass unit of every composition was binned and visualised in a mass defect histogram (**Figure 1, B**). The established combinatorial space described above is outlined in detail in the **SI-table, sheet 8.**
Initial mining for realistic ulosonic acid species in the large scale data was performed using a Matlab script. Thereby, all acquired spectra were searched for the highly conserved core ulosonic acid DMB fragment features 205.06, 217.06, 229.06 and 231.06. Spectra containing those features were extracted and fragments were binned according to their chemical compositions. Continuous carbon chain fragments with chain lengths of 7, 8, 9 and potential higher, allowing different degrees of oxidation and saturation (=realistic backbone compositions), were extracted and translated into structures (considering continuous and steady fragmentation trees) and visualised using the 'histogram' function of Matlab (SI-table, sheet 4, including script boxes 1-4, **Figure 2 A-C**).
The channel-hit map (**Figure 3**) showing all references/species analysed was prepared using the 'pcolor' function in Matlab. Limits were set so that every valid channel hit (containing a potential parent NulO) was represented by a black/red channel. In-between sample run blanks are shown first, followed by the analysed samples/references. The order of samples shown in the map (from left to right) follows the order listed in the **SI-table, sheet 3**, except that duplicates were not combined, but represented separately. The channel hit-map was once filtered for Neu/Kdn-like signals and once for bacterial-type NulOs across samples analysed. The binning of identified NulOs from every channel was performed using the 'histogram' function of Matlab. Thereby, masses calculated from chemical compositions (of individual runs) were combined within 0.1 Da bins. Selected species from the survey (**Figure 4**) are presented as a combined graph using Matlabs 'bar' function (stacked). For this, the intensity of the most intense parent ion was used to present the proportion of every channel hit. Bars are annotated by the channel mass (rounded to unit mass). Boxes representing the NAB hits were included manually between the relative and absolute proportions. For sample 13, NulOs are annotated with their chemical composition, mass error and class (Kdn derivatives) and scores.

Additions to results and discussions

F. NulO core compositions. The to-date three distinct core NulO compositions are known, which are outlined in Figure S1. The only composition without any amino functionality is keto-deoxy-neuraminic acid (Kdn, Figure S1, middle). The second (Figure S1, right) refers to N-acetyl-neuraminic acid (Neu5Ac, 5-acetamido-2-keto-3,5-dideoxy-D-glycero-D-galactonononic acid) commonly found in the D-glycero-D-galacto configuration. The third class is a 5,7-diamino variant with an additional deoxy on the 9 position. This composition is commonly found for pseudaminic acid (Pse, 5,7-Diamino-3,5,7,9-tetradeoxy-L-glycero-L-manno-non-2-ulopyranosonic) found in L-glycero-L-manno configuration or its stereoisomers legionaminic acid (Leg, D-glycero-D-galacto configuration, including a 3/8-epilegionaminic acid referred to as 3/8eLeg) as well as acinetaminic (Aci, D-glycero-L-altro configuration 8 epimer, D-glycero-L-altro and a

8 epimer referred as to 8eAci). Diversification can be found on any of the amino and hydroxyl positions. A recent comprehensive summary of a large collection of discovered sialic acids and other NulOs has been collected by Schauer et al.[3]

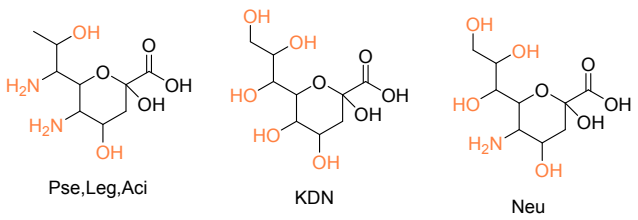

Figure S1. Three most commonly found classes of non-ulosonic acids, including groups undergoing frequent diversification (orange). Common isomerisation points for Pse, Leg and Aci (also referred to as to P.L.A) are not further differentiated in this study.

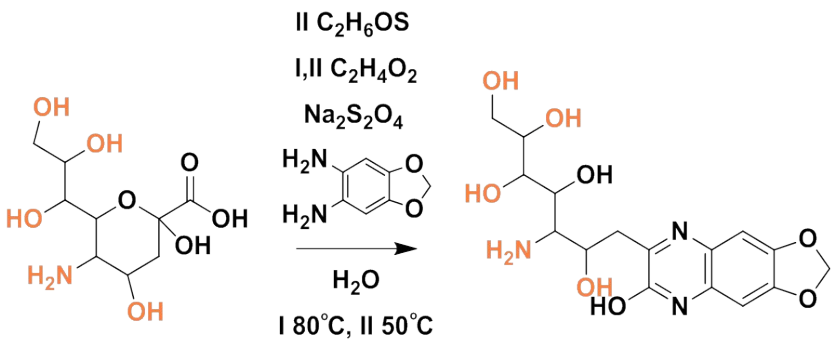

Figure S2. DMB labelling reaction scheme, first described by Hara et al.[1] I) first step includes incubation with 2M acetic acid at 80°C for 2 hours to release glycosidically as well as nucleotide activated sialic acids or other NulOs. II) Alpha keto acid specific labelling using 1,2-diamino-4,5-methylene dioxybenzene at 50°C for 2.5 hours.

G. Sialic acid reference standard spiking experiments. Initial spiking experiments were performed using a mix of 6 commercial sialic acid standards Neu5Ac, NeuGc, $Neu(AC)_2$, $Neu(Ac)_3$, NeuGcAc (Ludger, CatNo CM-SRP-01) and Kdn (Sigma Aldrich, CatNo 60714). All 6 sialic acid standards could be successfully recovered when spiking into E-coli K12 lysate at concentrations of 1:1, 1:5 and 1:1). It was also possible to distinguish from *E. coli*'s octulosonic acid derivatives.

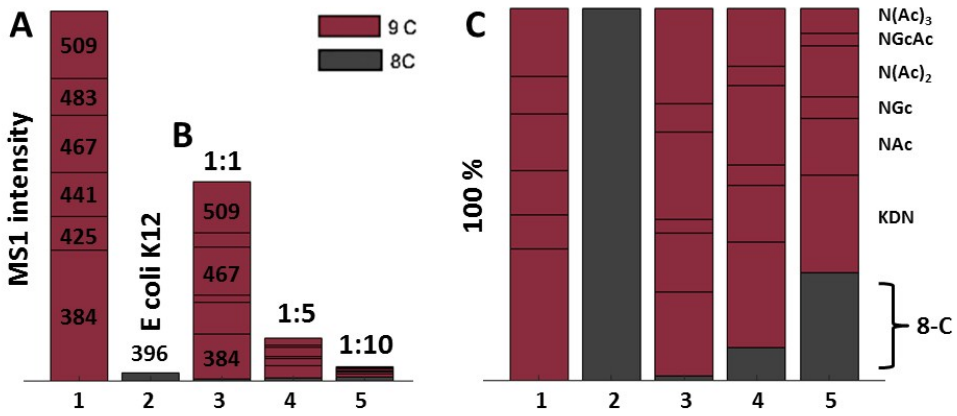

Figure S3. Stacked bar graphs show the observed peak intensities/proportions for reference sialic acids Kdn, NeuAc, NeuGc, $NeuAc_2$, NeuGcAc and $NeuAc_3$ and an 8-Carbon ulosonic acid naturally present in E-coli (presumably Kdo-Ac, the mass analysis did not include the native Kdo mass). A) Bars 1 to 2 represent the (absolute) summed MS1 peak intensities of the 6 sialic acid reference standards and the *E. coli* octulosonic acid derivate measured separately. B) Bars 3-4 show the (absolute) summed MS1 peak intensities of the 6 reference standards, after spiking in ratios 1:1 (3) 1:5 (4) and 1:10 (5) into the *E. coli* extract. The extract was prepared with protocols and biomass quantities, exactly as used for all other samples in this survey. B) shows the same order of samples normalised to 100%.

H. Towards establishing universal fragmentation features. We systematically established a core fragmentation framework for ulosonic acids with different carbon chain lengths ranging from C-8 to theoretical C-10 sugars (C-9+), deoxygenation states and degrees of saturation. Ulosonic acids are identified for which class fragments determine the carbon length, core fragments show attachment to the DMB-label and reporter fragments show side products from DMB-labelling. After an initial water loss peak (-H_2O), the neutral losses will be all –N and –O modified side groups. This leaves behind the backbone carbon chain, with many unmodified hydroxyl groups, varying from 0 to 3 -OHs that remain attached to the backbone. Depending on which –O positions are unmodified, fragmentation of the backbone can occur before all hydroxyl groups are fragmented, such as in Kdn. The fragments with the largest complete intact carbon backbone can be used to allocate the carbon length and corresponding class of the ulosonic acid (C-8–C-10). Due to the higher degree of modification, C-9 backbone with 1 –OHs Kdn/Neu core (295.07 m/z) can be differentiated from Pses with a less-saturated backbone (297.09 m/z). Extrapolating the fragmentation route can also be used to include octulosonic acids Kdo, Kdo8N (283.07 m/z), as well as more theoretical chemical space of deoxy Pse/Leg 299.10, deoxy Kdo (285.09 m/z) and larger C-10 sugars (311.09). Almost exclusively, the methylenedioxybenzene proximate C1-hydroxyl remained stabilised and contributed to the 3-carbon core fragments as described further below. Fractionation of the product alkene and alkyne chains led then to losses of 12.00 Da and 27.99 Da corresponding to the loss of C or CO, respectively. The smallest, but unique ulosonic fragments observed (core fragments), were a C-3 (205.061 m/z), C-4 (217.061) and C-5 (229.061 m/z) respectively, and an alternative C-5 fragment for 4-deoxy ulosonic acids (231.076 m/z).

Finally, reporter fragments were determined for detection of mono-labelled species (223.071, 343.114, 283.093, 297.108, 313.103) and DMB label degradation products (201.066, 189.066), which shows chemical instability at the 4,5-methylenedioxy end. Lower mass fragments at 177.07 m/z displayed additional DMB label-related fragmentation events.

Core	205.06	217.06	229.06	231.08

	C-6			C-7			C-8			C-9			Higher carbon (C-10)		
O6	**295.09**	**297.11**	**299.12**	**307.09**	**309.11**	**311.12**	**319.09**	**321.11**	**323.12**	**331.09**	**333.11**	**335.12**	**343.09**	**345.11**	**347.12****
O5	**277.08**	**279.10**	**281.11**	**289.08**	**291.10**	**293.11**	**301.08**	**303.10**	**305.11**	**313.08**	**315.10**	**317.11**	**325.08**	**327.10**	**329.11***
O4	**259.07**	**261.09**	**263.10**	**271.07**	**273.09**	**275.10**	**283.07**	**285.09**	**287.10**	**295.07**	**297.09**	**299.10**	**307.07**	**309.09**	**311.10**
O3	**241.06**	**243.08**	**245.09**	**253.06**	**255.08**	**257.09**	**265.06**	**267.08**	**269.09**	**277.06**	**279.08**	**281.09**	**289.06**	**291.08**	**293.09**

Table S1. Empirical conserved fragmentation trees for different kinds of ulosonic acid classes considering varying degrees of oxidation, saturation and carbon backbone length. For every backbone class, we considered a range of 12 indicative fragments. Following the same rationale, potential higher carbon ulosonic acid derivatives e.g. (10-carbon) where considered using 289.06, 291.08, 293.09, 307.07, 309.09, 311.10, 325.08, 327.10, 329.11, 343.09, 345.11 and 347.12. Out of these derivatives, 293 and 311 were most prominently observed in our study. This mass table was used for the screening study to discriminate between different carbon chain length ulosonic acids.

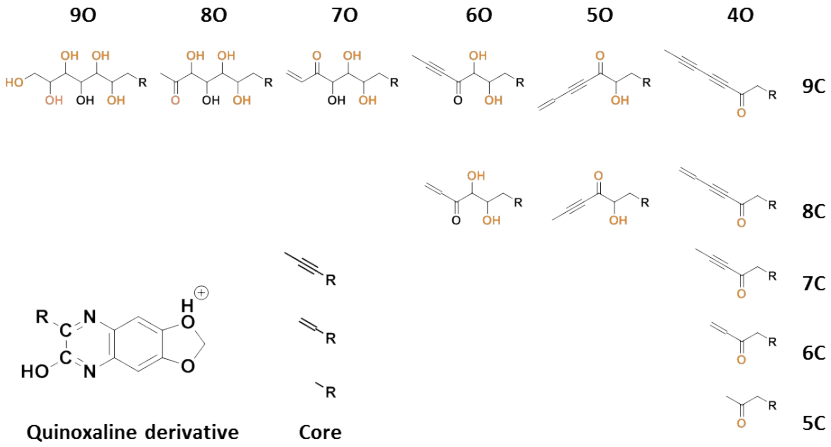

Figure S4. CID fragmentation tree for Kdn including quinoxaline (label) core.

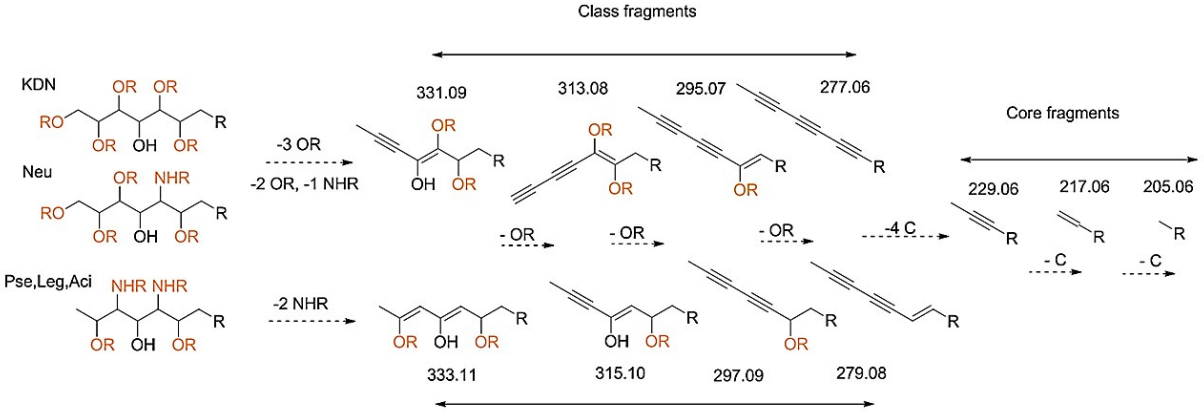

Figure S5. Comparative fragmentation trees of Kdn/Neu and Pse/Leg/Aci, where R represents the quinoxaline label core. The here shown fragmentation tree could also be extended towards other C-9 derivatives, showing highly comparable fragmentation behaviour. The main difference between Pse-type and Neu/Kdn is the 2 Da difference, which is not caused by the number of substitutions reducing the saturation. This difference is seen for the 'Class fragments' because the Pse-type C-9 sugars have an exposed primary carbon. However, this difference did not appear to be influenced by the number of amine versus hydroxyl groups; amine groups were not further considered during fragment mining in automatic data processing. Apart from core fragment and class fragments, reporter fragments were included to assess false positives due to side reactions of the DMB-label, or artefacts generated upon fragmentation of the DMB-labelled compounds.

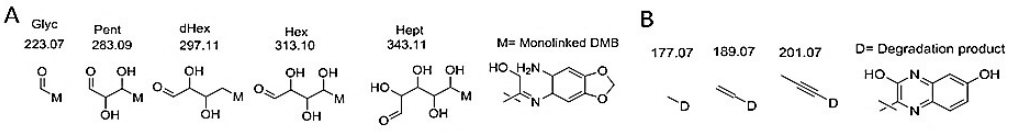

Figure S6. Proposed fragmentation for 'mono-linked' species (A), and proposed instability of the DMB label itself (B).

I. Verification of the empirical fragmentation features. The empirical fragmentation tree was verified with known ulosonic acid standards for Kdo (E-Coli), Kdn (commercial), Neu5Ac (commercial) and Pseudaminic acid (Pse) as obtained from Campylobacter.

	C-6			C-7			C-8			C-9			Higher carbon (C-10)		
O6	**295.09**	**297.11**	**299.12**	**307.09**	**309.11**	**311.12**	**319.09**	**321.11**	**323.12**	**331.09**	**333.11**	**335.12**	**343.09**	**345.11**	**347.12****
O5	**277.08**	**279.10**	**281.11**	**289.08**	**291.10**	**293.11**	**301.08**	**303.10**	**305.11**	**313.08**	**315.10**	**317.11**	**325.08**	**327.10**	**329.11***
O4	**259.07**	**261.09**	**263.10**	**271.07**	**273.09**	**275.10**	**283.07**	**285.09**	**287.10**	**295.07**	**297.09**	**299.10**	**307.07**	**309.09**	**311.10**
O3	**241.06**	**243.08**	**245.09**	**253.06**	**255.08**	**257.09**	**265.06**	**267.08**	**269.09**	**277.06**	**279.08**	**281.09**	**289.06**	**291.08**	**293.09**

Table S2. Kdo C-8 fragmentation markers were as predicted 283.07, 301.08 and 319.09; Neu/Kdn C-9 markers were 295.07, 313.08 and 331.09. For Pse, the C-9 markers were 297.09, 315.10 and 333.11. Pse showed compared to Neu an additional lower oxidation peak of 279.08. The identified peaks for the ulosonic acid form Gryphiswaldense (GW) were 331.10 and 328.13 (amine instead of H_2O*). GW showed in addition (similar to Pse) the lower oxidised species 293.09. Due to high levels of acetylation, the further oxygen was only added after addition of 2 acetylations (**).

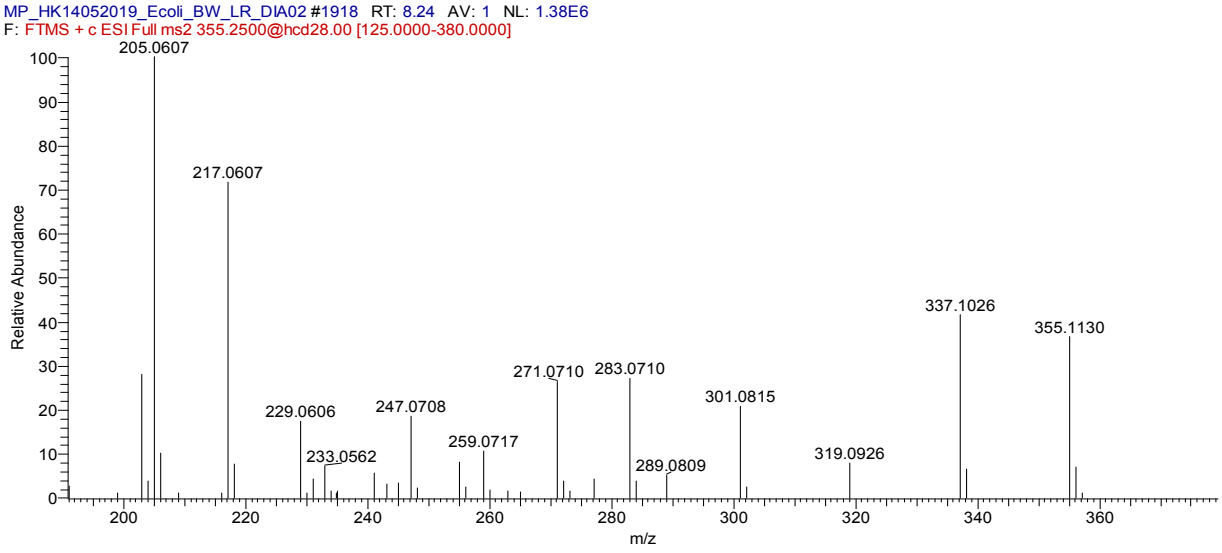

Figure S7a. DMB-Kdo fragmentation profile.

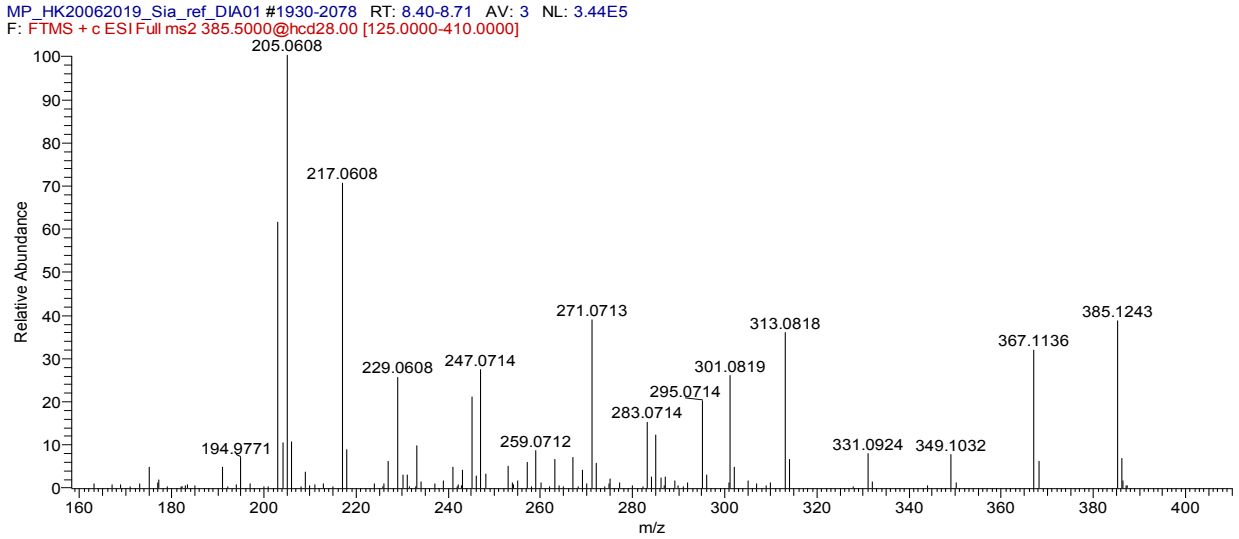

Figure S7b. DMB-Kdn fragmentation profile.

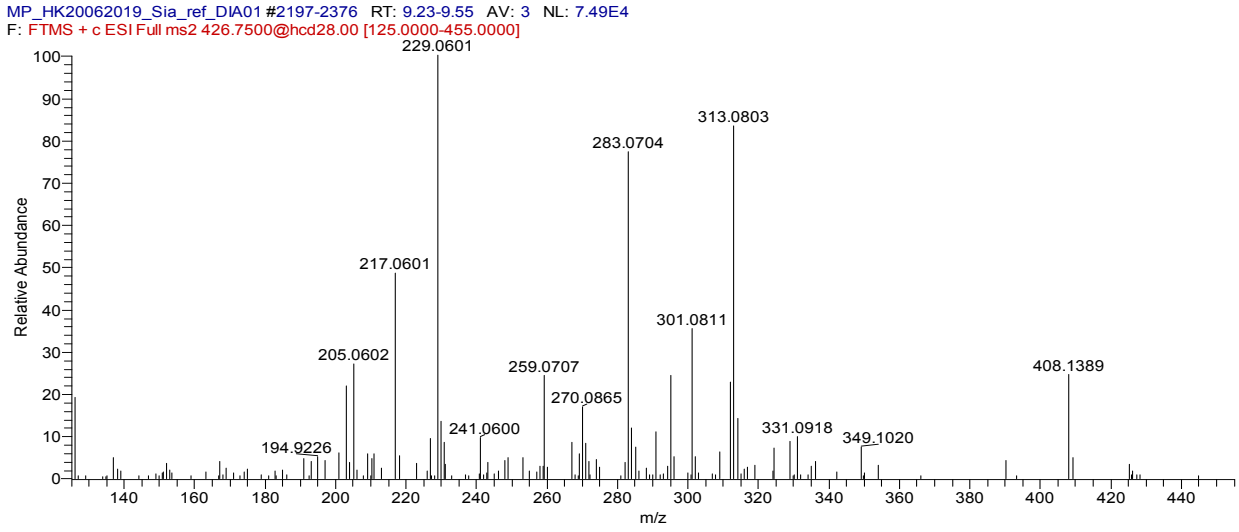

Figure S7c. DMB-NeuAc fragmentation profile.

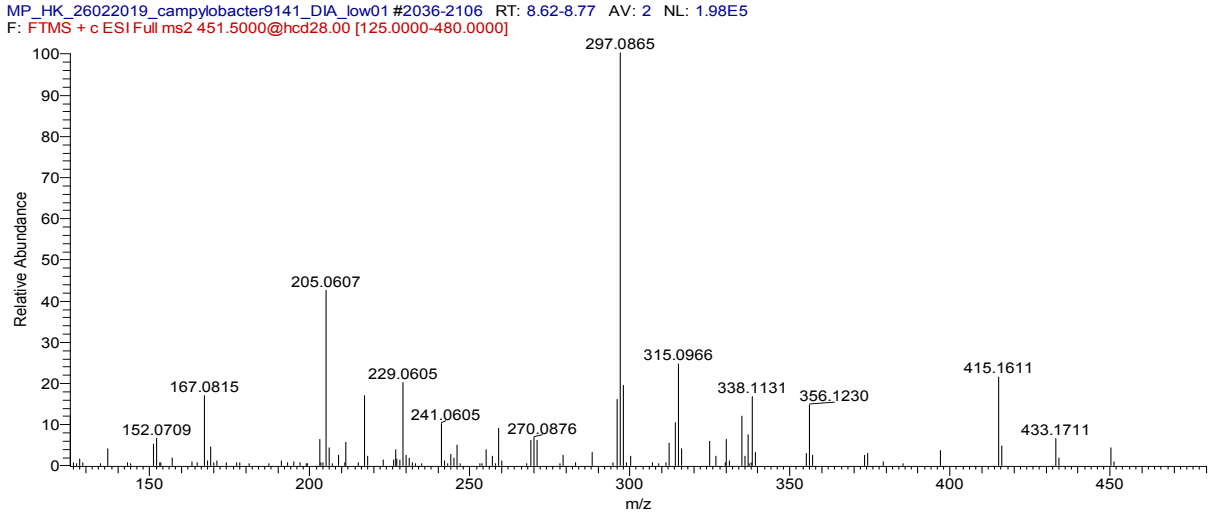

Figure S7d. DMB-Pse fragmentation profile.

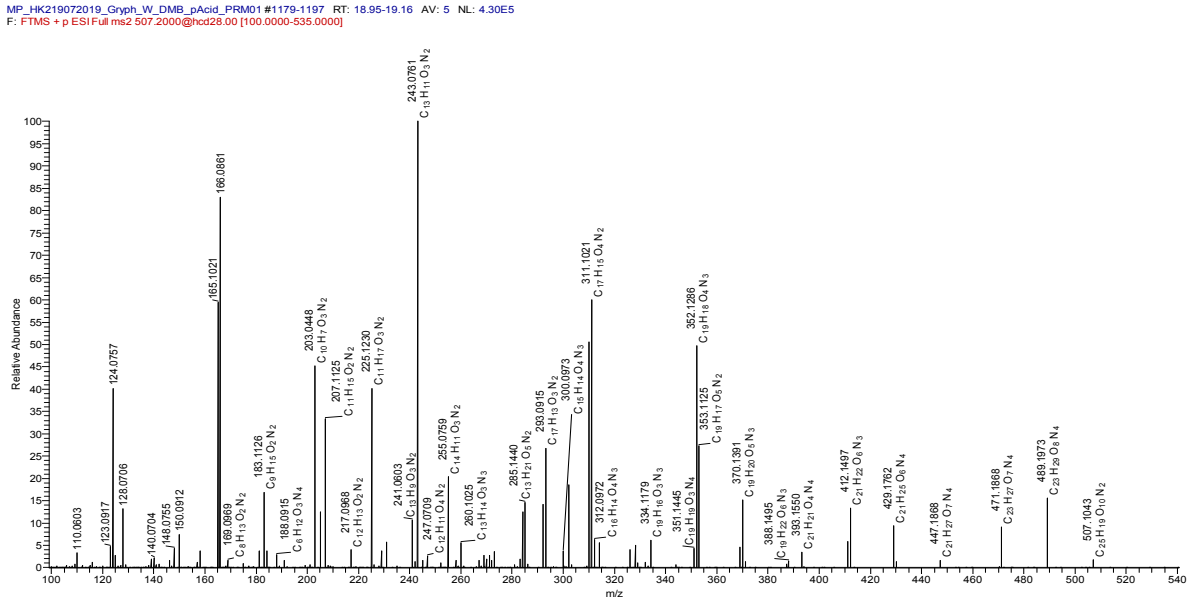

Figure S7e. DMB-UA GW fragmentation spectrum (of 507 m/z) annotated with sum formulae as determined by the Xcalibur chemical composition annotation tool (max mass deviation 10 ppm, DMB core set as min elemental composition).

J. Potential higher carbon ulosonic acids. The ulosonic acid peaks observed for *M. Gryphiswaldense* matched a fragmentation tree of a potential 10 carbon backbone (or any other structure indistinguishable by means of MS). More precisely, three ulosonic acid peaks were observed that matched the UA backbone peaks 293.09 and 311.10 $[M+H]^+$, which correspond to 3 and 4 oxygen-containing species (2 oxygens derive from the DMB label, and one from a hydroxyl group which is part of the unique core quinoxaline fragment). The mass peak with the sum formula $C_{23}H_{30}N_4O_9$ (m_i = 506.2013, confirmed by high-resolution mass spectrometry) revealed following fragmentation the loss of three acetyl groups. Other losses derived from the loss of water, ammonia and -CH_2 groups. No other type of backbone modification was observed that could have increased the carbon count compared to the known ulosonic acid fragmentation features. The other mass peaks observed derived from a water loss of 506.2013 mass peak and from a species with only 2 acetyl groups instead of 3. The latter (twofold acetylated) species was further also observed in the methanogenic bacterium sample (**SI-table, sheet 1-3**). The same sample (*M. Gryphiswaldense*) was also measured without prior DMB labelling (plus/minus acid hydrolysis) which did not show any of the abovementioned ulosonic acid peaks. Further, the same sample was measured without acid hydrolysis, but with DMB labelling, which resulted in only trace quantities of the above mentioned ulosonic acid peaks, only observable after manual investigations of the mass traces.

To confirm the said peaks as ulosonic acids and to obtain additional fragmentation spectra, we analysed the acid hydrolysed and non-derivatised (non DMB labelled) material for the theoretical free ulosonic acid mass peak (predicted sum formula from the AiRM experiment: $C_{16}H_{26}N_2O_9$ after subtraction of the label). Thereby, we indeed observed an abundant mass peak corresponding to the predicted accurate mass of the free ulosonic acid and fragmentation profiles similar to those observed for the nonulosonic acid standards. To verify the carbon backbone length, we investigated the fragmentation tree closer and compared it with Neu5Ac (commercial standard) and Pse (Campylobacter jejuni). Since the fragmentation tree of the triply acetylated ulosonic acid from M. Gryphiswaldense (GW) appeared somewhat more complex, we undertook a chemical deacetylation using 20 mM sodium hydroxide. To do so, a microscale quantity of the peak (giving the parent ion 391 $[M+H^+]^+$) was fractionated manually from the capillary HPLC, speed-vacuum dried and redissolved in 20 mM aqueous sodium hydroxide. The mixture was incubated at 30°C for 12 hours and (re)injected to the LC-MS system. We performed a targeted analysis for masses of species with 3 (starting material), 2, 1 and no acetylation(s). The fractionated ctrl, as expected, showed only a peak for the threefold-acetylated species, where the (mild) base treated sample showed only a peak for a twofold-acetylated species (loss of O-Ac, but not N-Ac). In the following, the fragmentation tree for the twofold (N) acetylated species was compared to Neu5Ac (commercial, single-NAc, 9 carbon) and Pse (Campylobacter, di-NAc, 9 carbon).

Neu5Ac shows major fragments for a threefold water loss followed by the fragmentation of the N-acetyl group (-42.01 Da -C_2H_2O or -59.03Da -C_2H_5ON). From the later peak (197.04 m/z), the backbone fragmentation occurred from either loss of carboxylic acid (-46.005 Da, -H_2CO_2) or the cleavage of the C-9 carbon methanol group (-30.01, -CH_2O). The fractionation from the resulting peak at 176.03 Da was followed by the carboxylic acid loss (-46.005 Da, -H_2CO_2) to give a peak at 121.02 m/z. Pse shows the same major fragments for a threefold water loss followed by the fragmentation of one acetyl group (major -42.01 Da, -C_2H_2O) and a second N-acetyl group (major -59.03Da, -C_2H_5ON). The resulting C-9 backbone fragment (180.06 m/z) showed further the loss of carboxylic acid (-46.005 Da, -H_2CO_2) or the cleavage of the C-9 carbon methanol group (-30.01, -CH_2O). The fractionation from the resulting peak at 176.03 Da was followed by the carboxylic acid loss (-46.005 Da, -H_2CO_2) to give a peak at 121.02 m/z.

The (doubly N-acetylated) ulosonic acid from GW showed a comparable (simple) fragmentation tree compared to Neu5Ac and Pse. First, 2 water loss peaks were observed followed by twice a loss of N-Acetyl groups (major -59.03 Da -C_2H_5ON) to 254.1 m/z and to the 10-carbon fragment 195.06 m/z ($C_{10}H_{11}O_4$). From here, we observed the backbone fragmentation by the loss of the carboxylic acid (-46.005 Da, -H_2CO_2) to give a 9-carbon peak at 149.05 m/z. A second lower abundant route branching from 2 water loss peaks to the loss of one N-Acetyl group (major -59.03 Da, -C_2H_5ON) to the peak 254.1 m/z. From there, an (early) CO_2 loss to 210.1 m/z (including a minor carboxylic acid loss to 208.09 m/z) was found taking place. This loss was followed by the loss of acetyl to the C-9 168.1 m/z (-42.01 Da, -C_2H_2O), where the loss of -59.03 Da (-C_2H_5ON) was much weaker. The (early) loss of CO_2 (-43.98 Da) in parallel to the carboxylic acid loss was also observed in Neu5Ac and Pse but was significantly less pronounced.

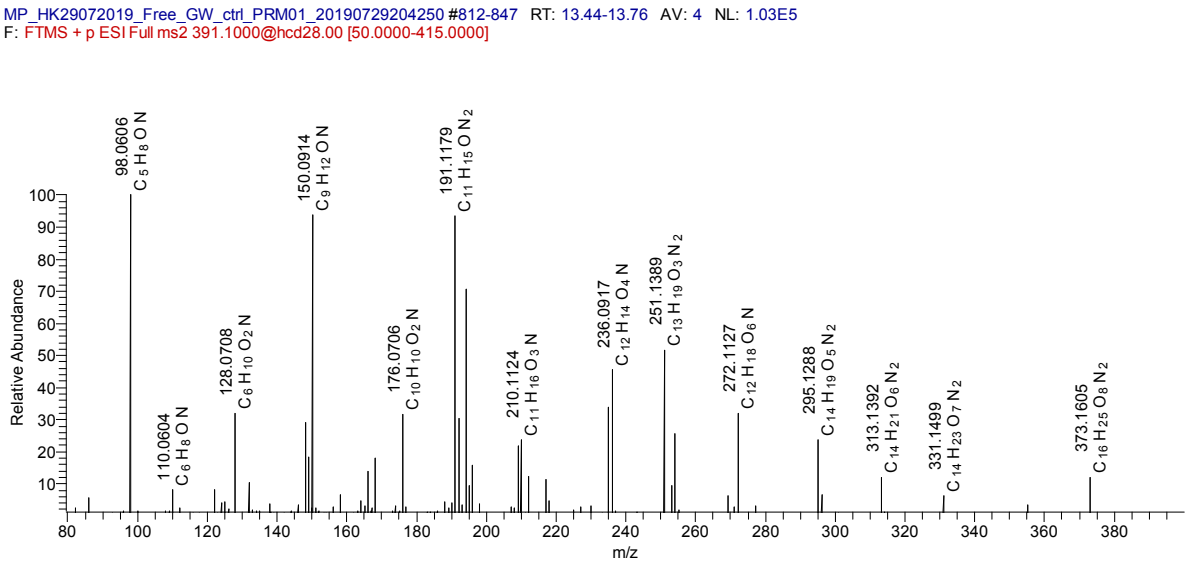

Figure S8a. HCD fragmentation profile of unlabelled ulosonic acid from GW with the proposed sum formula $C_{16}H_{26}N_2O_9$ and the theoretical $[M+H]^+$ of 391.1711 Da.

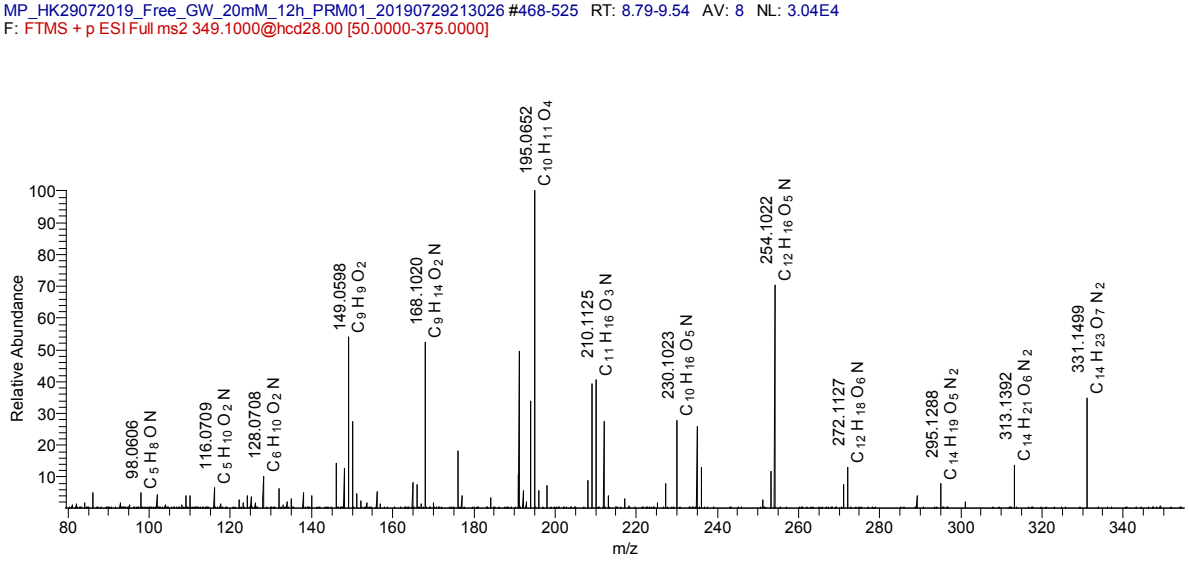

Figure S8b. HCD fragmentation profile of unlabelled and de O-acetylated ulosonic acid from GW.

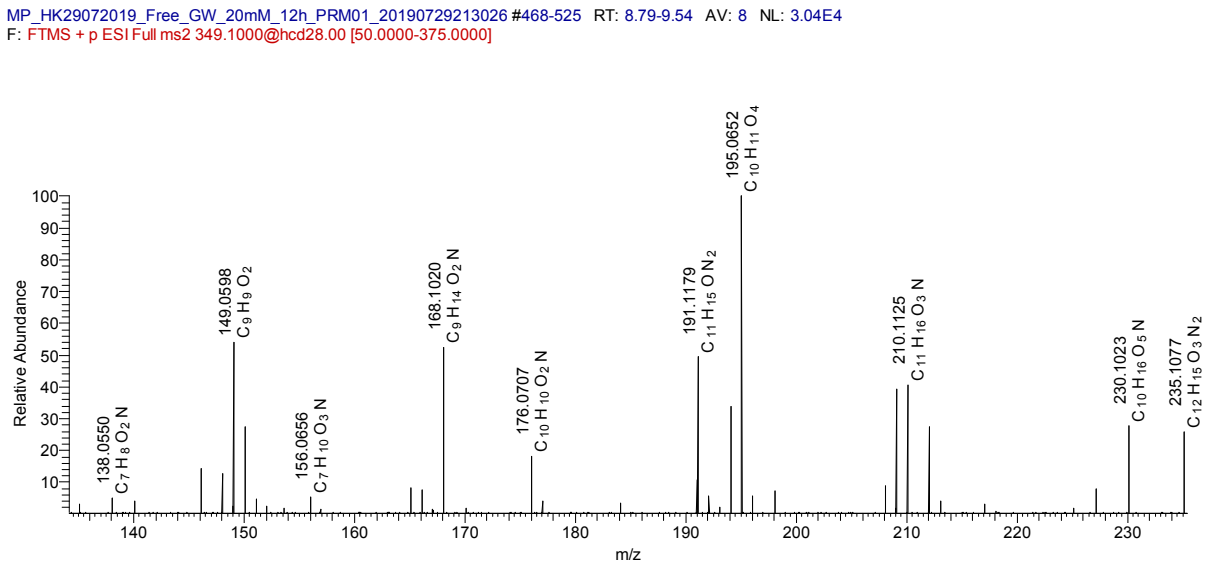

Figure S8c. HCD fragmentation profile of unlabelled and de O-acetylated ulosonic acid from GW, with zoom to the C-10 peak which fragments to a C-9 following the loss of carboxylic acid.

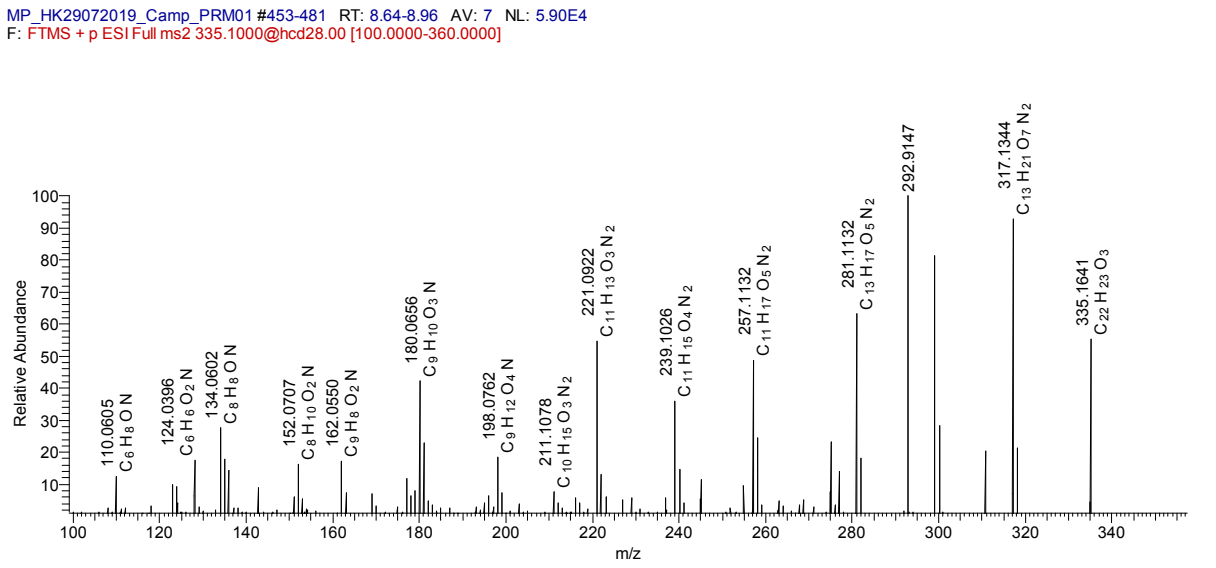

Figure S8d. HCD fragmentation profile of unlabelled Pse from Campylobacter jejuni.

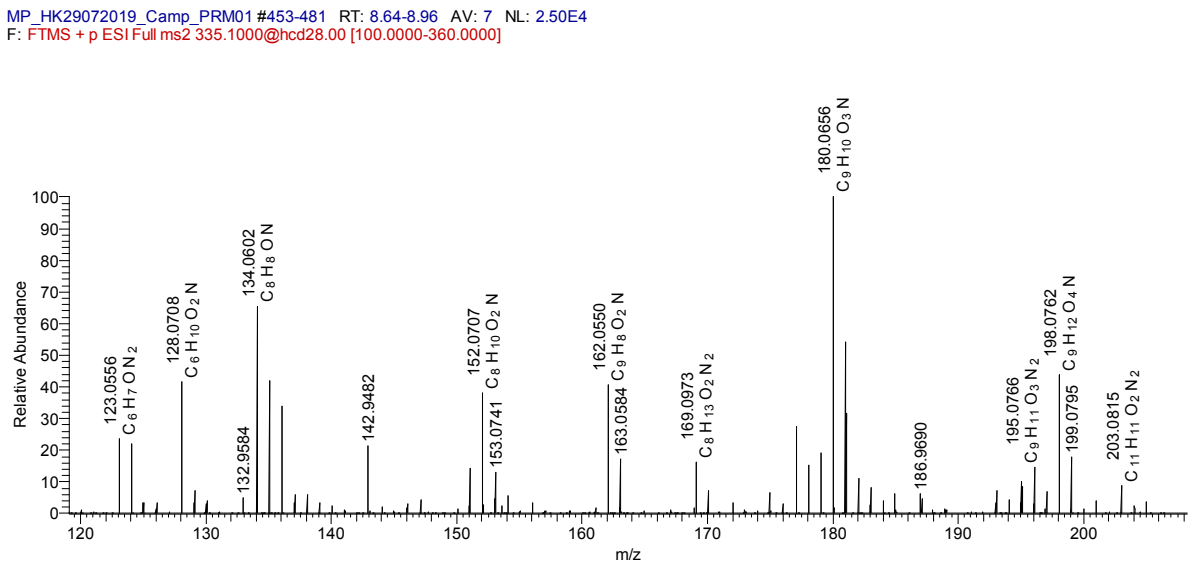

Figure S8e. HCD fragmentation profile of Pse from Campylobacter, with zoom to the C-9 peak which fragments to a C-8 by carboxylic acid loss.

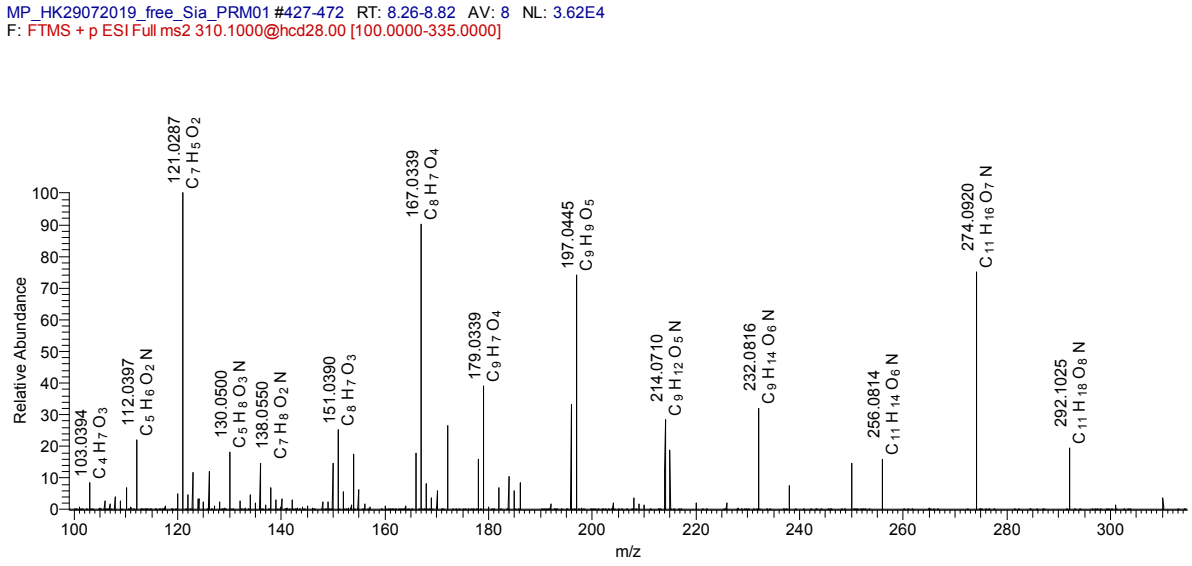

Figure S8f. HCD fragmentation profile of the commercial, unlabelled Neu5Ac standard.

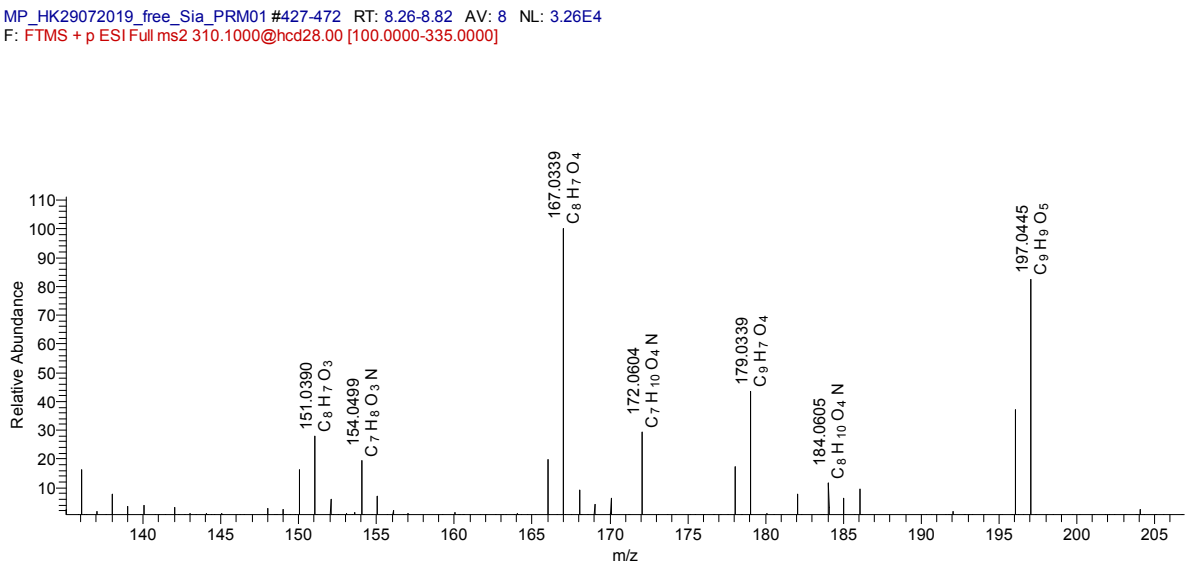

Figure S8g. HCD fragmentation profile of Neu5Ac, with zoom to the C-9 peak which fragments to a C-8 by carboxylic acid loss.

K. Degradation and non-specific labelling side products. Most lysates showed additional artefacts deriving from nonspecific labelling side reactions with other sugars and that unfortunately showed upon fragmentation the ulosonic acid core fragments 205.061, 217.061 and 229.061 $[M+H^+]^+$. However, to distinguish from actual hept-, oct- and nonulosonic acids, we included the 'Reporter ions 1' that were established through initial validation experiments using common bulk monosaccharides. In addition, we included in-source fragments of common bulk monosaccharides such as for 162.053, 146.058 and 132.042. Spectra showing reporter ions and in-source fragments at high frequency were rejected, or marked as possible labelling artefacts. Furthermore, chemical instability of the DMB derivative may produce low abundant artefacts peaks. A presumable hydrolysis of the methylenedioxy group followed by elimination would result in a mass decrease by 28 Da. Unfortunately, higher fragments may mimic the intact quinoxaline core fragments. For example degraded DMB-Neu5Ac would appear with the mass of potential 'N-methyl-neuraminic acid' and could only be distinguished by fragmentation tree analysis. Therefore, we included in our fragmentation trees the 'Reporter ions 2' 177.066, 189.066 and 201.066 $[M+H^+]^+$. Spectra were rejected, or marked as degradation products when all 3 reporter ions were present at high frequency.

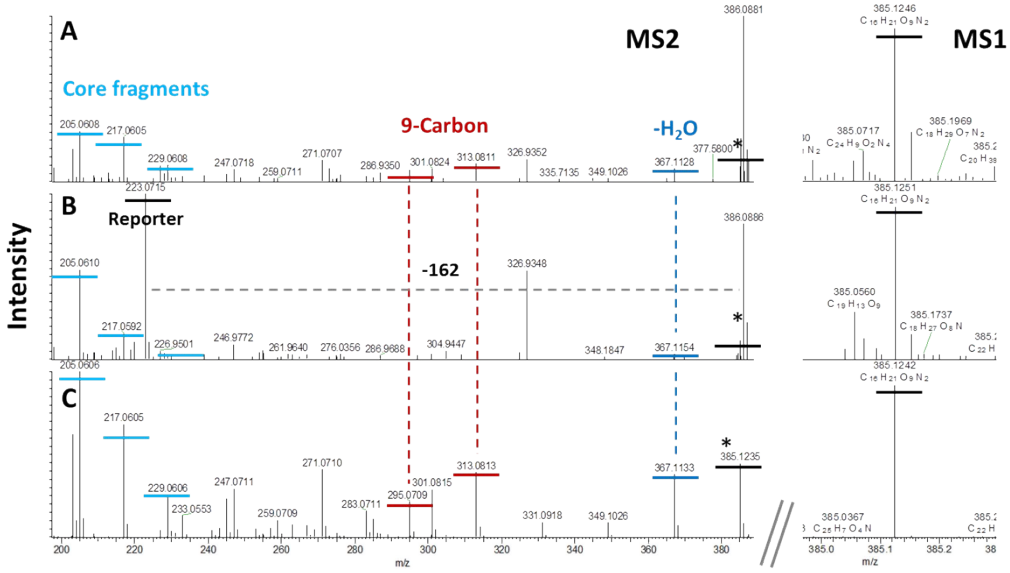

Figure S9. Non-specific derivatisation (artefacts) mimic ulosonic acid mass peaks, including core ulosonic acid fragment ions, as shown for a "pseudo-Kdn" peak C16H21O9N2. Spectra could be differentiated from genuine peaks by the carbon number fragments and reporter ions. A) Genuine Kdn from Cricosphaera carterae, B) 'pseudo-Kdn' peak observed in Nicotiana benthamiana and C) commercial Kdn standard.

References

[1] S. Hara, M. Yamaguchi, Y. Takemori, M. Nakamura, Y. Ohkura, *J Chromatogr* **1986**, *377*, 111-119.

[2] A. L. Lewis, N. Desa, E. E. Hansen, Y. A. Knirel, J. I. Gordon, P. Gagneux, V. Nizet, A. Varki, *Proc Natl Acad Sci U S A* **2009**, *106*, 13552-13557.

[3] R. Schauer, J. P. Kamerling, *Adv Carbohydr Chem Biochem* **2018**, *75*, 1-213.