RESEARCH PAPER

Towards an automated analysis of bacterial peptidoglycan structure

Marshall Bern[1] · Richard Beniston[2,4] · Stéphane Mesnage[3]

Abstract Peptidoglycan (PG) is an essential component of the bacterial cell envelope. This macromolecule consists of glycan chains alternating *N*-acetylglucosamine and *N*-acetylmuramic acid, cross-linked by short peptides containing nonstandard amino acids. Structural analysis of PG usually involves enzymatic digestion of glycan strands and separation of disaccharide peptides by reversed-phase HPLC followed by collection of individual peaks for MALDI-TOF and/or tandem mass spectrometry. Here, we report a novel strategy using shotgun proteomics techniques for a systematic and unbiased structural analysis of PG using high-resolution mass spectrometry and automated analysis of HCD and ETD fragmentation spectra with the Byonic software. Using the PG of the nosocomial pathogen *Clostridium difficile* as a proof of concept, we show that this high-throughput approach allows the identification of all PG monomers and dimers previously described, leaving only disambiguation of 3–3 and 4–3 cross-linking as a manual step. Our analysis confirms previous findings that *C. difficile* peptidoglycans include mainly deacetylated *N*-acetylglucosamine residues and 3–3 cross-links. The analysis also revealed a number of low abundance muropeptides with peptide sequences not previously reported.

Keywords Peptidoglycan · Cell wall · Muropeptides · Proteomics · Tandem mass spectrometry · Glycoproteomics · Cross-link

✉ Marshall Bern
bern@proteinmetrics.com

✉ Stéphane Mesnage
s.mesnage@sheffield.ac.uk

[1] Protein Metrics Inc., 1622 San Carlos Ave., Suite C, San Carlos, CA 94070, USA

[2] Biological Mass Spectrometry Facility biOMICS, University of Sheffield, Brook Hill Road, Sheffield S3 7HF, UK

[3] Department of Molecular Biology and Biotechnology, University of Sheffield, Western Bank, Sheffield S10 2TN, UK

[4] Present address: Covance Laboratories Limited, Otley Road, Harrogate, North Yorkshire HG3 1PY, UK

Introduction

The major and essential polymer within the bacterial cell envelope is peptidoglycan (PG), which forms a single bag-shaped macromolecule (or sacculus) around the cell [1]. PG is unique to the bacterial kingdom, and its synthesis is the target of the most clinically important antibiotics ever discovered such as the beta-lactams (penicillin) and glycopeptides (vancomycin). In addition to a role as an exoskeleton that confers cell shape and resistance to the intracellular osmotic pressure, PG is used as a scaffold for the display of a myriad of polymers and proteins at the cell surface.

PG composition is variable amongst the bacteria, but is usually highly conserved within a given species [2]. The macromolecule is made of beta-1,4-linked glycan chains alternating *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc). MurNAc residues are substituted via a lactyl group by pentapeptide stems most frequently made of L-Ala-iso-D-Glu-*m*eso-DAP (or L-Lys)-D-Ala-D-Ala (mDAP, *meso*-diaminopimelic acid). During growth, PG precursors are assembled in the cytoplasm and translocated at the cell surface where they are polymerized. Whilst glycan chain polymerization is

conserved, two types of enzymes can cross-link peptide stems (Fig. 1). D,D-Transpeptidases recognize the C-terminal D-Ala-D-Ala extremity of a donor stem. They form a covalent intermediate with the D-Ala in position 4 and link the COOH group of this residue to the NH₂ group of the mDAP residue in position 3 of an acceptor peptide stem, thereby forming 4–3 cross-links. D,D-Transpeptidases are also called penicillin-binding proteins (PBPs) as they are inhibited irreversibly by beta-lactam antibiotics, which are D-Ala-D-Ala structural analogs. Peptidoglycan cross-links can also result from the activity of L,D-transpeptidases. These enzymes recognize the last two residues of a tetrapeptide donor stem substrate generated by a carboxypeptidase. They form a covalent intermediate with the meso-DAP in position 3 and link the COOH group of this residue to the NH₂ group of the mDAP of an acceptor peptide stem. L,D-Transpeptidases form 3–3 cross-links and are not inhibited by beta-lactam antibiotics. In addition to distinct modes of polymerization, several enzymatic modifications take place during growth, a process called "PG remodeling." Such modifications can occur on glycan strands (e.g., O-acetylation or de-N-acetylation), reviewed by Vollmer [3], or on peptide stems (e.g., amidation of γD-Glu, cleavage of C-terminal residues, or amino acid substitutions mediated by L,D-transpeptidases), reviewed by Vollmer et al. [1].

Surprisingly, the strategy proposed nearly 30 years ago by Bernd Glauner [4] is still used routinely to explore PG structure. It involves the purification of PG sacculi, followed by digestion with lysozyme to generate disaccharide peptides (muropeptides) that are separated by reversed-phase HPLC. Individual fractions are collected, desalted and analysed by MS and MS/MS. Improvement of chromatography equipment has increased the throughput of PG analyses, allowing the separation of several micrograms of muropeptides in less than 30 min [5, 6]. Recent studies have reported PG structural analysis using online LC-ESI-MS [7] or LC-ESI-MS/MS [8–10] for the identification of muropeptides. Although these approaches overcome a major limitation (time-consuming offline fractionation), they still rely on manual analysis of the major ions and, hence, have limited capability to search for new PG components, likely biasing the analysis towards previously identified species. To the best of our knowledge, no automated analysis of peptidoglycan tandem mass spectra has been reported before.

Here, we describe a novel strategy for a systematic, unbiased, structural analysis of PG that combines high mass accuracy and automated analysis of HCD and ETD fragmentation spectra using Byonic and Byologic software (Protein Metrics, San Carlos, CA). Using the PG of the nosocomial pathogen Clostridium difficile as a proof of concept, we show that this
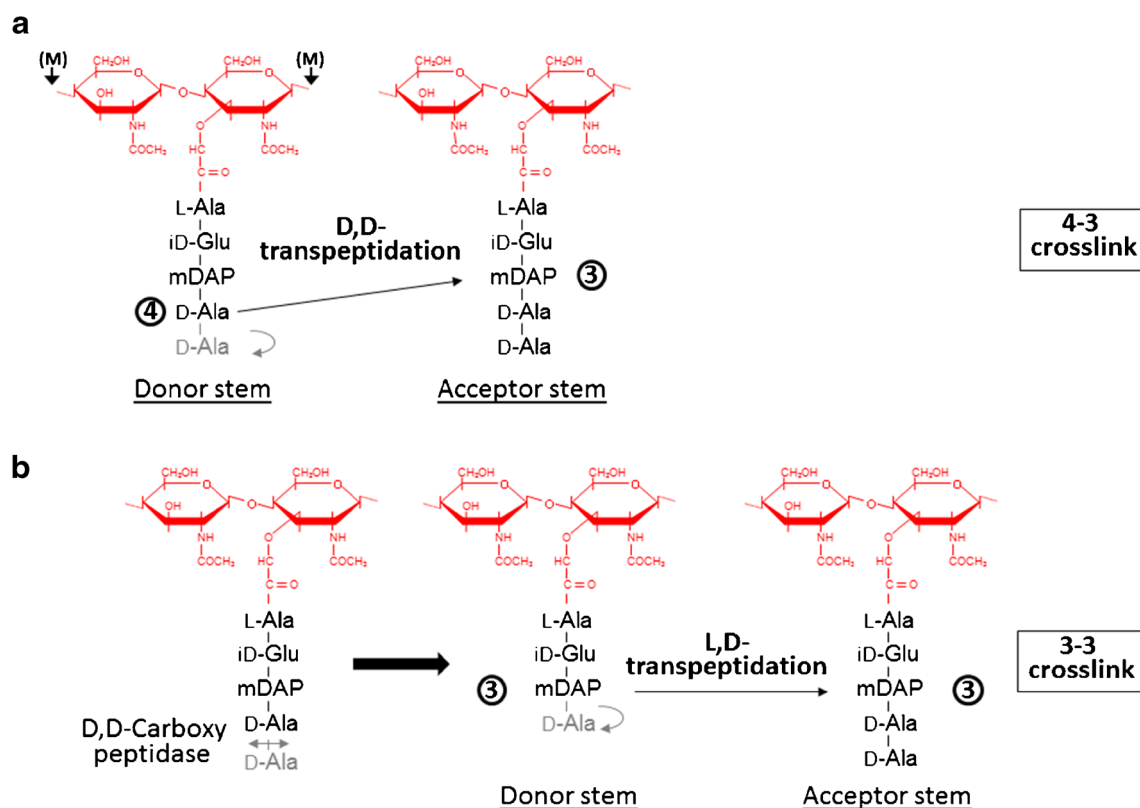
Fig. 1 Composition and polymerization of bacterial peptidoglycan (PG). PG building blocks, cut from larger PG chains with mutanolysin cleavages indicated by (M), correspond to the disaccharide-pentapeptide GlcN-MurNAc-L-Ala-iD-Glu-mDAP-D-Ala-D-Ala. PG components can be cross-linked by two distinct mechanisms: (a) D,D-transpeptidation results in the formation of a 4–3 bond between a donor and an acceptor stem; (b) L,D-transpeptidation follows cleavage of D-Ala in position 5 and results in the formation of a 3–3 bond between a donor and an acceptor stem

systematic, unbiased, "bottom up" approach allows the identification and quantitation of all PG monomers and dimers previously described, leaving only disambiguation of 3–3 and 4–3 cross-linking as a manual step. Our analysis confirms previous findings [11] that *C. difficile* peptidoglycans include mainly deacetylated GlcNAc residues (GlcN) and 3–3 cross-links, and that acceptor peptides often include glycine substituted for the D-Ala in position 4. Byonic's "wildcard" search, which allows any mass delta on any one residue, also finds a number of low abundance muropeptides with peptide sequences not previously reported. Variable residues are located at the C-terminus of acceptor peptides, after the mDAP in position 3.

Experimental section

Extraction and purification of peptidoglycan samples

Fifty milliliters of TY broth was inoculated with an overnight culture of *C. difficile* strain 630 [11] at a starting OD_{600} of 0.05. Exponentially growing cells (OD600 = 0.45) were collected by centrifugation at 4500 rpm for 10 min at 25 °C. The pellet was resuspended in 10 mL of boiling MilliQ water (MQ) before the addition of an equal volume of boiling SDS 8 %. After 30 min at 100 °C, the cells were left to cool down to room temperature. Insoluble cell walls were pelleted at 45,000*g* for 15 min and washed five times using 30 mL of MilliQ water. Proteins covalently bound to peptidoglycan were removed by pronase treatment (final concentration of 2 mg/mL for 4 h at 60 °C). Protease-treated cell walls were washed 6 times with 30 mL of MilliQ water before covalently bound polymers were removed by incubation in 1 M HCl for 5 h at 37 °C. Insoluble pure peptidoglycan was washed 6 times with MilliQ water, freeze-dried and resuspended at a final concentration of 10 mg/mL.

Preparation of soluble muropeptides

One milligram of purified peptidoglycan (HCl-treated cell walls) were digested overnight in 50 mM phosphate buffer (pH 5.5) supplemented with 100 µg of mutanolysin in a final volume of 125 µL. Following heat inactivation of mutanolysin (5 min at 100 °C), soluble disaccharide peptides were then mixed with an equal volume of 250 mM borate buffer (pH 9.25) and reduced with 1 % (m/v) sodium borohydride. After 20 min at room temperature, the pH was adjusted to 4.5 using phosphoric acid. Reduced muropeptides were desalted by rp-HPLC using water–formic acid 0.1 % (v/v) as buffer A and one-step elution with 25 % acetonitrile (v/v)–formic acid 0.1 % (v/v) gradient.

Data acquisition Samples were analysed online using an UltiMate 3000 RSLCnano LC System (Dionex) coupled to an LTQ Orbitrap Elite hybrid mass spectrometer (Thermo) equipped with a nanospray ion source. Desalted muropeptides (300 ng) were separated on a PicoFrit™ Hypersil Gold aQ analytical column (1.9 µm, 75 µm id × 50 cm,) (New Objective). Muropeptides were eluted for 10 min with water + 0.1 % (v/v) formic acid + 0.8 % (v/v) acetonitrile (buffer A) and then with a 50-min acetonitrile linear gradient (0 to 22.5 %) in buffer A at a flow rate of 0.3 µl/min. The mass spectrometer was operated in standard data-dependent acquisition mode controlled by Xcalibur 2.2. The instrument was operated with a cycle of one MS (in the Orbitrap) acquired at a resolution of 120,000 at *m/z* 400 from 150 to 2000 *m/z* and the top 5 most abundant multiply charged (2+ and higher) ions in a given chromatographic window were subjected to either HCD fragmentation (isolation window 3 *m/z*, normalized collision energy = 25, activation time 10 ms) or ETD fragmentation (isolation window 3 *m/z*, normalized collision energy = 35, activation time 300 ms) in the linear ion trap with supplemental activation enabled. An FTMS target value of 1e6, an HCD target value of 50,000 and an ion trap MSn target value of 10,000 were used. Dynamic exclusion was enabled with a repeat duration of 30 s with an exclusion list of 500 and exclusion duration of 30 s. Lock mass of 445.120025 was enabled for all experiments.

Data analysis We used Byonic version 2.8.0 to identify peptidoglycan forms, Byologic version 2.4.21 to compute extracted ion chromatograms (XICs) and relative abundances, and Byomap 2.4.21 to view and annotate peaks in the total ion chromatogram (TIC). We started by searching for PG monomers using a FASTA database containing only the peptide sequences AEM, AEMA and AEMG. We specified a fixed modification of +41.0443 on methionine (M) so that modified M represented mDAP with mass 172.0848 (=131.0405 + 41.0443). We enabled variable modifications of 277.116 (MurNAc), 396.174 (GlcN-deacetyl-MurNAc), 420.174 (GlcN-anhydro-MurNAc), 438.185 (GlcN-MurNAc) and 480.196 (GlcNAc-MurNAc) on the peptide N-terminus. We also enabled variable modifications of –0.984 (amidation) on E and mDAP. In a subsequent search, we turned off amidation, anhydro and deacetylation and used a wildcard modification of mass –130 to +210 Da to find unanticipated modifications and sequence variants. All searches after some initial exploratory searches used a "Manual score cut" of zero, 5 ppm precursor mass tolerance, 20 ppm fragment mass tolerance for HCD spectra and 0.5 Da fragment mass tolerance for ETD spectra.

We searched for PG dimers two different ways: cross-link search on a monomer database and "ordinary" search on a dimer database. Byonic includes automatic cross-link search with user-defined cross-link mass and residue specificity, but with the limitation that the cross-link connects two residues of the same type. We specified −18.0106 as the cross-link mass

and M (meaning mDAP) as the residue type. We ran Byonic on a FASTA database of monomer sequences and enabled the glycans GlcN-MurNAc and GlcNAc-MurNAc on the N-terminus. Byonic then predicts the correct b/y (or c/z) ions for 3–3 linkages from Glycan-A-E-mDAP to any Glycan-A-E-mDAP-X (or –X-X) peptide, but slightly incorrect b/y (or c/z) ions for 4–3 linkages from Glycan-A-E-mDAP-A to Glycan-A-E-mDAP-X peptides, because it fails to predict the b-ion containing Glycan-A-E-mDAP and the complementary y-ion containing A and Glycan-A-E-mDAP-X. We also searched for dimers using a FASTA database containing concatenated sequences such as AEMAEMA and allowing

Table 1 List of detected monomers with peptides (m = mDAP), glycans, precursor masses, apex elution times, XIC integrals and percent of reported XIC integrals

Peptide	Charges	Glycan	Neutral Mass (Daltons)	Elution Time (Minutes)	XIC Integral (Ion Counts)	% of XICs Sum=29.2%
AEmA	2	GlcN-MurNAc	899.3971	9.97	3.81E+10	22.35%
	2	MurNAc	738.3283	9.89	4.78E+08	0.28%
	1	(none)	461.2122	9.89	2.82E+07	0.02%
AEmG	2	GlcN-MurNAc	885.3815	9.12	3.28E+09	1.92%
	2	MurNAc	724.3126	8.94	3.93E+07	0.02%
AEmF	2	GlcN-MurNAc	975.4284	37.37	1.41E+09	0.83%
	2	MurNAc	814.3596	37.37	1.42E+07	0.01%
AEm	2	GlcN-MurNAc	828.3600	8.66	1.13E+09	0.66%
AEmA	2	GlcNAc-MurNAc	941.4077	13.25	1.11E+09	0.65%
AEm(I/L)	2	GlcN-MurNAc	941.4441	32.67	8.76E+08	0.51%
	1	MurNAc	780.3752	32.67	1.46E+08	0.09%
AEmV	2	GlcN-MurNAc	927.4284	27.26	7.55E+08	0.44%
AEmS	2	GlcN-MurNAc	915.3920	8.75	5.53E+08	0.32%
AEmY	2	GlcN-MurNAc	991.4233	30.25	3.28E+08	0.19%
AEm(NH$_2$)A	2	GlcN-MurNAc	898.4131	9.37	2.08E+08	0.12%
AEmM	2	GlcN-MurNAc	959.4005	27.11	1.94E+08	0.11%
AEmR	2, 3	GlcN-MurNAc	984.4611	8.66	1.71E+08	0.10%
	2	MurNAc	823.3923	8.75	7.77E+07	0.05%
AEmAA	2	GlcN-MurNAc	970.4342	14.45	1.23E+08	0.07%
AEm(NH$_2$)	2	GlcNAc-MurNAc	869.3905	9.37	1.02E+08	0.06%
AEmGG	2	GlcN-MurNAc	942.4029	8.48	9.53E+07	0.06%
AEm[A-V]	2	GlcN-MurNAc	998.4655	33.92	9.12E+07	0.05%
AEmK	2	GlcN-MurNAc	956.4550	6.85	8.03E+07	0.05%
	2	MurNAc	795.3861	8.37	2.38E+07	0.01%
AEm[A-G]	2	GlcN-MurNAc	956.4186	11.37	7.77E+07	0.05%
AEmH	2, 3	GlcN-MurNAc	965.4189	8.64	7.17E+07	0.04%
AEmGV	2	GlcN-MurNAc	984.4499	27.26	6.93E+07	0.04%
AEm	2	GlcN-anh-MurNAc	810.3494	10.15	4.40E+07	0.03%
AEmF	2	GlcNAc-MurNAc	1017.4390	40.97	3.29E+07	0.02%
AEm(I/L)	2	GlcNAc-MurNAc	983.4546	36.40	1.66E+07	0.01%
AEmASV	2	GlcN-MurNAc	1085.4975	29.04	1.56E+07	0.01%

The XIC integral is the ion count for the monoisotopic precursor, summed over charge states for the monomers observed in more than one charge state, that is, AEmR and AEmH. Rows with peptides left blank show monomers formed in the electrospray source by loss of monosaccharides, recognizable as in-source decay by exact co-elution. Shaded rows show monomers with the most common glycan (GlcN-MurNAc). Brackets as in [A-G] indicate uncertain order of the amino acid residues; (I/L) indicates either isoleucine or leucine. 77.6 % of all monomers are GlcN-MurNAc-AEmA. Of the peptide sequence variants, only AEmG has been reported before

glycan modifications on both N- and C-termini; this method predicts the correct ions for cleavages of peptide bonds up to the cross-link, that is, correct b1, b2, b3 and y6, y5, y4 for the 3–3 link, but incorrect ions for cleavages after the cross-link. Incorrect ion predictions adversely affect Byonic scores, and this is one of the reasons that we used a manual score cutoff of zero (the other reason is that the usual methods for creating decoy databases, for example, reversing peptide sequences, do not work well for PG analysis, in which accurate precursor mass may be more informative than fragmentation peaks, many of which are common to multiple scored candidates).

Byonic does not currently offer cross-link search for more than two linked peptides, so we searched for trimers using a FASTA database containing sequences such as AEMAEMAEMA and allowing glycan modifications on N- and C-termini as well as on M. As above, this method predicts correct ions for cleavages up to the first cross-link but incorrect ions after that. When searching for dimers and trimers, we did not search for all possible combinations of monomers, but limited attention to those in which all donor peptides are either AEM or AEMA, that is, the peptides acted upon by the L,D and D,D transpeptidases. We tried all the monomers we discovered in the monomer searches as the terminal acceptor peptide for dimers and trimers.

Byologic and Byomap offer complementary ways to view and quantify peptides, by XIC and TIC or UV trace, respectively.

Table 2 List of most abundant dimers with peptides (m = mDAP), glycans, masses, apex elution times and abundances

Peptides	Charges	Glycans	Neutral Mass	Elution Time	XIC Integral	% of XICs Sum=57.4%
AEm, AEmA	2, 3	GlcN-MurNAc, GlcN-MurNAc	1709.747	28.47	4.18E+10	24.52%
	2, 3	GlcN-MurNAc, MurNAc	1548.678	28.47	9.43E+09	5.53%
	2	MurNAc, MurNAc	1387.609	28.47	1.28E+09	0.75%
AEmA, AEmA	2, 3	GlcN-MurNAc, GlcN-MurNAc	1780.784	30.49	1.23E+10	7.21%
	2, 3	GlcN-MurNAc,-MurNAc	1619.715	30.41	1.77E+09	1.04%
	2	MurNAc, MurNAc	1458.646	30.41	2.47E+08	0.14%
AEm, AEmG	2, 3	GlcN-MurNAc, GlcN-MurNAc	1695.731	27.11	6.74E+09	3.95%
	2, 3	GlcN-MurNAc, MurNAc	1534.662	27.11	1.66E+09	0.97%
	2	MurNAc, MurNAc	1373.593	27.11	2.92E+08	0.17%
AEm, AEmA	2, 3	GlcN-MurNAc, GlcNAc-MurNAc	1751.757	31.39	4.84E+09	2.84%
AEm, AEmA	2, 3	GlcN-MurNAc	1271.562	21.17	3.34E+09	1.96%
	2	MurNAc	1110.493	21.17	1.51E+09	0.89%
AEm, AEm	2, 3	GlcN-MurNAc, GlcN-MurNAc	1638.709	27.85	2.46E+09	1.44%
	2	GlcN-MurNAc, MurNAc	1477.641	27.85	7.17E+08	0.42%
AEmA, AEmA	2, 3	GlcN-MurNAc, GlcNAc-MurNAc	1822.794	32.98	1.74E+09	1.02%
AEmA, AEmG	2, 3	GlcN-MurNAc, GlcN-MurNAc	1766.768	28.24	1.57E+09	0.92%
AEm, AEmF	2, 3	GlcN-MurNAc, GlcN-MurNAc	1785.778	43.84	9.47E+08	0.56%
	2, 3	GlcN-MurNAc, MurNAc	1624.709	43.84	2.03E+08	0.12%
AEmA, AEmA	2, 3	GlcN-MurNAc	1342.599	24.70	8.42E+08	0.49%
	2	MurNAc	1181.530	24.79	3.01E+08	0.18%
AEm, AEmG	2, 3	GlcN-MurNAc	1257.546	11.98	6.06E+08	0.36%
AEm, AEm(I/L)	2, 3	GlcN-MurNAc, GlcN-MurNAc	1751.794	41.05	5.20E+08	0.31%
AEm, AEmY	2, 3	GlcN-MurNAc, GlcN-MurNAc	1801.773	38.02	4.87E+08	0.29%
AEm, AEmV	2, 3	GlcN-MurNAc, GlcN-MurNAc	1737.778	36.71	3.38E+08	0.20%
AEm, AEmR	2, 3, 4	GlcN-MurNAc, GlcN-MurNAc	1794.811	26.71	3.21E+08	0.19%
AEm, AEmG	2	GlcN-MurNAc, GlcNAc-MurNAc	1737.741	29.04	3.16E+08	0.19%
AEmA, AEmF	2, 3	GlcN-MurNAc, GlcN-MurNAc	1856.815	45.10	2.52E+08	0.15%
AEm, AEmK	2, 3, 4	GlcN-MurNAc, GlcN-MurNAc	1766.804	25.52	2.46E+08	0.14%
AEm, AEm	2	GlcN-MurNAc, GlcNAc-MurNAc	1680.720	29.70	2.41E+08	0.14%
AEm, AEmGG	2, 3	GlcN-MurNAc, GlcN-MurNAc	1752.752	26.63	2.37E+08	0.14%
AEm, AEm	2, 3	GlcN-MurNAc	1200.525	13.25	2.34E+08	0.14%
AEmA, AEmV	2, 3	GlcN-MurNAc, GlcN-MurNAc	1808.815	40.47	1.19E+08	0.07%

Shaded rows show dimers with the most common donor peptide (AEm) and glycans (GlcN-MurNAc); acceptor peptide prevalences (AEmA, AEmG, AEm, AEmF, etc.) agree well with monomer abundances. About 54 % of the dimers correspond to GlcN-MurNAc-AEm donors cross-linked to GlcN-MurNAc-AEmA acceptors via a 3–3 bond

We used Byologic to validate and quantify PG forms identified by Byonic and Byomap to obtain a TIC trace, inspect MS full scans and map PG forms to TIC elution peaks. We used manual analysis to distinguish 3–3 from 4–3 linkages for the dimers containing one Glycan-A-E-mDAP and one Glycan-A-E-mDAP-A or two Glycan-A-E-mDAP-A's, which is isobaric with a Glycan-A-E-mDAP and Glycan-A-E-mDAP-A-A dimer.

Results and discussion

As shown in Tables 1, 2 and 3, the Byonic searches found 29.2, 57.4 and 13.4 % monomers, dimers and trimers, as percentages of the total of identified XIC quantities. These proportions are consistent with the percentages of 35.1, 56.6 and 8.3 % reported previously [11] based on UV traces quantifications and MALDI-TOF offline analyses rather than LC-ESI quantitation. We find that 77.6 % of monomers are GlcN-MurNAc-AEmA (here m = mDAP), in close agreement with 74.9 % from Peltier et al. These monomers containing tetrapeptide stems result from carboxypeptidase activity (see Electronic Supplementary Material (ESM) Fig. S2). GlcN-MurNAc-AEm is not the most common monomer because it is usually linked rather than free.

We find that 42.7 % of dimers are GlcN-MurNAc-AEm linked to GlcN-MurNAc-AEmA, in close agreement with 43.8 % from Peltier et al. Our number, however, rises to 53.7 % if we include forms missing GlcN that, due to almost exactly correlated co-elution, we judge to be produced within the electrospray source. In-source decay also occurs with MALDI, but is harder to judge without elution profiles. The most common trimer contains two GlcN-MurNAc-AEm and one GlcN-MurNAc-AEmA, accounting for 30.7 % of all trimers (34.9 % including in-source decay) and 42.7 % for Peltier et al. We did not find anhydrous MurNAc as reported by Peltier et al., but rather anhydrous mDAP in the C-terminal position (see spectra on page 18 of the Fig. S3 in ESM), but this may be an in vitro water loss.
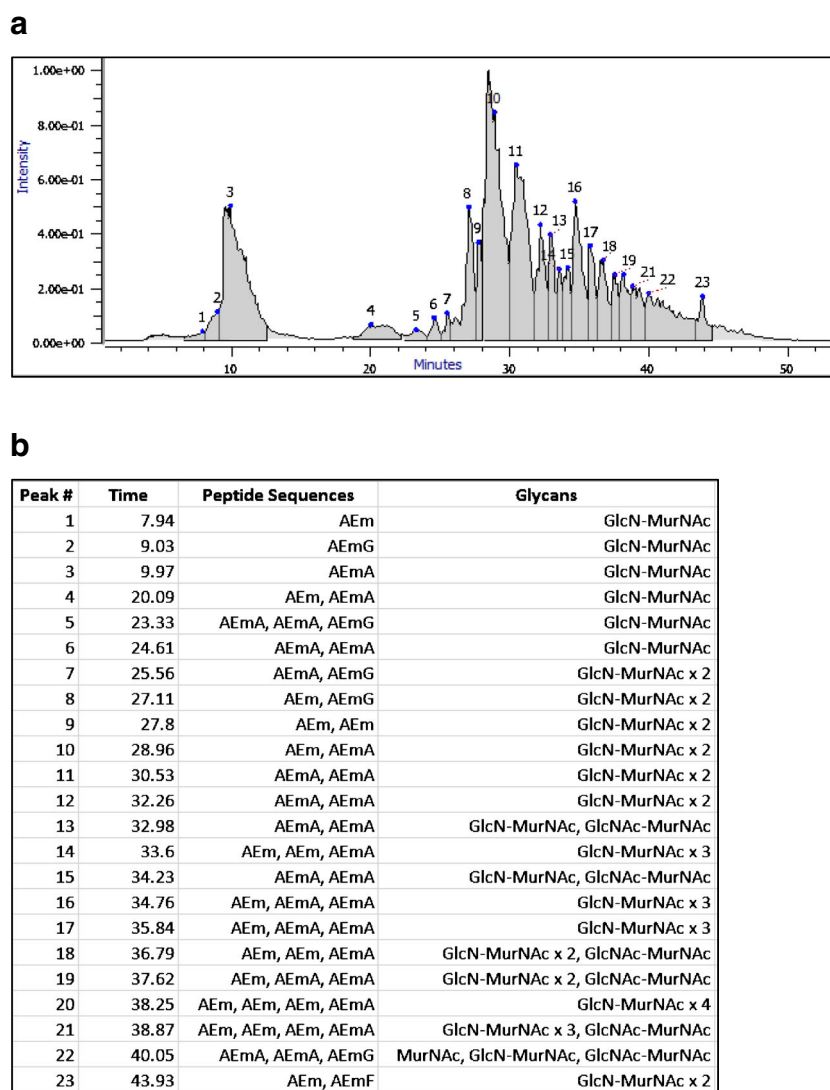
As shown in Fig. 2, we found at least one identification for each distinct peak in the TIC elution profile. Byomap computes a summed MS spectrum for each elution peak, and in almost all cases, Fig. 2 lists the identification matching the most intense peak in the MS spectrum. An exception is elution peak 14, in which the most intense spectral peak has neutral mass 1707.731 and remains unidentified. The MS spectra for elution peaks 21 and 22 contain numerous low abundance forms, most but not all of which match masses from Tables 2 and 3.

Table 3 List of most abundant trimers with glycans, precursor masses, apex elution times and abundances

Peptides	Charges	Glycans	Neutral mass	Elution time	XIC integral	% of XICs sum = 13.4 %
AEm, AEm, AEmA	2, 3, 4	GlcN-MurNAc, GlcN-MurNAc, GlcN-MurNAc	2520.096	34.76	7.03E + 09	4.12 %
	3, 4	GlcN-MurNAc, MurNAc, GlcN-MurNAc	2359.027	34.84	9.39E + 08	0.55 %
AEm, AEmA, AEmA	2, 3, 4	GlcN-MurNAc, GlcN-MurNAc, GlcN-MurNAc	2591.133	35.77	4.48E + 09	2.63 %
AEm, AEm, AEmA	2, 3, 4	GlcN-MurNAc, GlcN-MurNAc	2081.911	31.87	1.50E + 09	0.88 %
AEm, AEm, AEmA	2, 3	GlcN-MurNAc, GlcN-MurNAc, GlcNAc-MurNAc	2562.107	36.79	1.38E + 09	0.81 %
AEm, AEm, AEmG	2, 3, 4	GlcN-MurNAc, GlcN-MurNAc, GlcN-MurNAc	2506.080	33.21	1.19E + 09	0.70 %
AEm, AEmA, AEmA	2, 3	GlcN-MurNAc, GlcN-MurNAc, GlcNAc-MurNAc	2633.144	37.54	1.14E + 09	0.67 %
AEmA, AEmA, AEmA	3, 4	GlcN-MurNAc, GlcN-MurNAc, GlcN-MurNAc	2662.170	36.95	8.39E + 08	0.49 %
AEm, AEmA, AEmA	2, 3	GlcN-MurNAc, MurNAc, GlcN-MurNAc	2430.064	36.87	7.07E + 08	0.41 %
AEm, AEm, AEmA	3, 4	GlcNAc-MurNAc, GlcNAc-MurNAc, GlcNAc-MurNAc	2646.138	35.69	5.05E + 08	0.30 %
AEm, AEmA, AEmG	3	GlcN-MurNAc, GlcN-MurNAc, GlcN-MurNAc	2577.117	34.00	4.77E + 08	0.28 %
AEmA, AEmA, AEmG	3, 4	GlcNAc-MurNAc, MurNAc, GlcNAc-MurNAc	2571.107	40.13	3.88E + 08	0.23 %
AEm, AEmA, AEmG	3	GlcNAc-MurNAc, MurNAc, GlcNAc-MurNAc	2500.070	39.34	3.87E + 08	0.23 %
AEm, AEm, AEm	3, 4	GlcN-MurNAc, GlcN-MurNAc, GlcN-MurNAc	2449.059	33.68	3.45E + 08	0.20 %
AEm, AEm, AEmA	2, 3, 4	GlcN-MurNAc	1643.726	26.29	3.43E + 08	0.20 %
AEmA, AEmA, AEmA	3	GlcN-MurNAc, GlcN-MurNAc, GlcNAc-MurNAc	2704.181	38.57	3.14E + 08	0.18 %
AEm, AEmA, AEmA	3, 4	GlcN-MurNAc, GlcN-MurNAc	2152.948	32.51	2.31E + 08	0.14 %
AEm, AEmA, AEmG	3	GlcN-MurNAc, GlcN-MurNAc, GlcNAc-MurNAc	2619.128	35.54	1.99E + 08	0.12 %
AEm, AEm, AEmG	3	GlcN-MurNAc, GlcNAc-MurNAc, GlcN-MurNAc	2548.091	34.46	1.82E + 08	0.11 %
AEm, AEm, AEmG	3	GlcN-MurNAc, MurNAc, GlcN-MurNAc	2345.012	33.21	1.54E + 08	0.09 %

35 % of the trimers correspond to a GlcN-MurNAc-AEm donors cross-linked to GlcN-MurNAc-AEmA acceptor stems via 3–3 bonds

Fig. 2 (**a**) Total ion chromatogram shows PG components ordered by hydrophobicity, monomers generally eluting before dimers and trimers. (**b**) Major component for each elution peak. Most peaks also contain several minor components. Peaks 11 and 12 and also 16 and 17, contain isomers, possibly epimers, indistinguishable by precursor mass and MS/MS scans


Peak #	Time	Peptide Sequences	Glycans
1	7.94	AEm	GlcN-MurNAc
2	9.03	AEmG	GlcN-MurNAc
3	9.97	AEmA	GlcN-MurNAc
4	20.09	AEm, AEmA	GlcN-MurNAc
5	23.33	AEmA, AEmA, AEmG	GlcN-MurNAc
6	24.61	AEmA, AEmA	GlcN-MurNAc
7	25.56	AEmA, AEmG	GlcN-MurNAc x 2
8	27.11	AEm, AEmG	GlcN-MurNAc x 2
9	27.8	AEm, AEm	GlcN-MurNAc x 2
10	28.96	AEm, AEmA	GlcN-MurNAc x 2
11	30.53	AEmA, AEmA	GlcN-MurNAc x 2
12	32.26	AEmA, AEmA	GlcN-MurNAc x 2
13	32.98	AEmA, AEmA	GlcN-MurNAc, GlcNAc-MurNAc
14	33.6	AEm, AEm, AEmA	GlcN-MurNAc x 3
15	34.23	AEmA, AEmA	GlcN-MurNAc, GlcNAc-MurNAc
16	34.76	AEm, AEmA, AEmA	GlcN-MurNAc x 3
17	35.84	AEm, AEmA, AEmA	GlcN-MurNAc x 3
18	36.79	AEm, AEm, AEmA	GlcN-MurNAc x 2, GlcNAc-MurNAc
19	37.62	AEm, AEmA, AEmA	GlcN-MurNAc x 2, GlcNAc-MurNAc
20	38.25	AEm, AEm, AEm, AEmA	GlcN-MurNAc x 4
21	38.87	AEm, AEm, AEm, AEmA	GlcN-MurNAc x 3, GlcNAc-MurNAc
22	40.05	AEmA, AEmA, AEmG	MurNAc, GlcN-MurNAc, GlcNAc-MurNAc
23	43.93	AEm, AEmF	GlcN-MurNAc x 2

The most obvious departure from previous results is the discovery of a number of peptide sequence variants, including 5- and even 6-residue peptides, which have not been reported before. Their presence in *C. difficile* is likely to be due to the exchange activity of L,D-transpeptidases (Fig. S2 in the ESM), which has been demonstrated in vitro [12] for several amino acids. These sequence variants are evidenced by precursor masses matching within 5 ppm, in most cases within 1 ppm, as well as high-quality fragmentation spectra, one of which is shown Fig. 3 and others in Fig. S3 in the ESM. The elution times of the monomers, with Arg, Lys, His and Ser giving earlier elution times than Gly and Ala, and Met, Val, Tyr, Lys and Phe giving later elution times, matches measurements made for peptide elution time predictors [13]. It is tempting to assume that such unconventional muropeptides are present in other bacterial peptidoglycans but have not been reported so far due to their low abundance. Both high MS sensitivity and automated

MS/MS spectrum analysis will be required to test if similar unconventional muropeptides are found in other PG.

It is worth noting that our nano-LC strategy uses only 300 ng of PG material, about 30 times less material than a recently published "highly sensitive" UPLC strategy [6]. The relatively short run time (about 50 min) is much faster than traditional offline fractionation (140 min for *C. difficile*), and comparable to the UPLC strategy [6], which employs no mass spectrometry and is hence more suited to comparison than to discovery. Nano-LC-ESI-MS/MS thus appears to be the most effect data acquisition strategy yet devised for PG structural analyses.

Complementing the data acquisition strategy, our bioinformatics search strategy gives high sensitivity and acceptably low false-positive rate, even for trimers for which almost 2/3 of the fragment peak predictions are incorrect. The primary reason for the low error rate is the simplicity of the search relative to other glycoproteomics searches. The trimer search

Fig. 3 Collisional dissociation (HCD) spectrum of an unanticipated sequence variant, recognized by high-accuracy (5 ppm precursor and 20 ppm fragment) mass spectrometry. Other fourth residue variants include: Phe, Ile/Leu, Ser, Met, Val, Arg, Lys and His. The small peptide fragmentation diagram on the right side of the tandem mass spectrum shows Byonic's automatic assignment of spectrum peaks
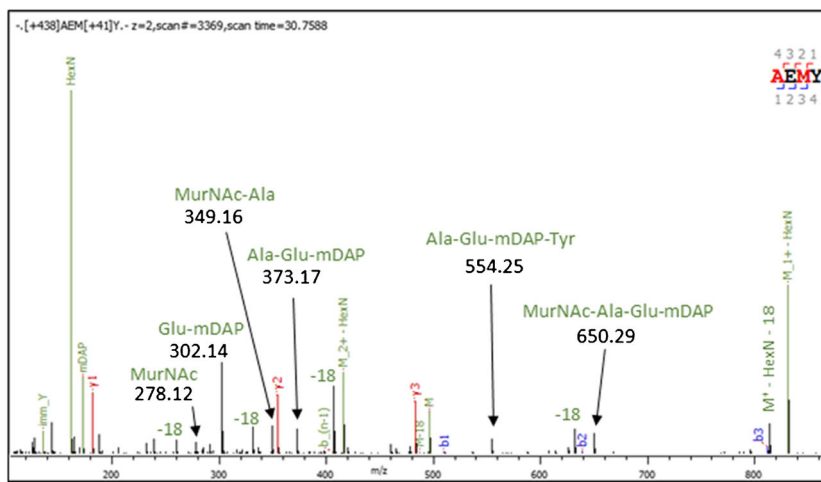
Fig. 4 (**a**) HCD and (**b**) ETD spectra of a dimer with that could be either GlcN-MurNAc-A-E-mDAP 3–3 linked to GlcN-MurNAc-A-E-mDAP-A-A or GlcN-MurNAc-A-E-mDAP-A 4–3 linked to GlcN-MurNAc-A-E-mDAP-A. The HCD spectrum contains no peaks (such as an A-A y-ion for 3–3) that can distinguish the two possibilities, but the ETD spectrum contains peaks at 899.45 and 884.42 matching the theoretical masses (given on the fragmentation diagram) of 899.32 and 884.39 for c- and z-ions splitting donor and acceptor peptides. The ETD spectrum therefore shows that the dimer is 4–3 linked. Byonic gives a fragmentation diagram and annotates c2, y2, y3, z2 and z3 for [+438]AEM[+41][+881]A crosslinked to [+438]AEM[+41]A, where [+438] represents the disaccharide, M[+41] represents mDAP, and [+881] the cross-link "modification" on M[+41], but Byonic does not annotate internal fragments nor the disambiguating peaks at 899.45 and 884.42
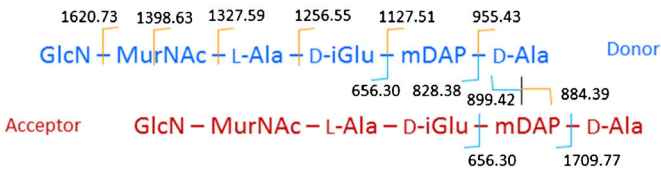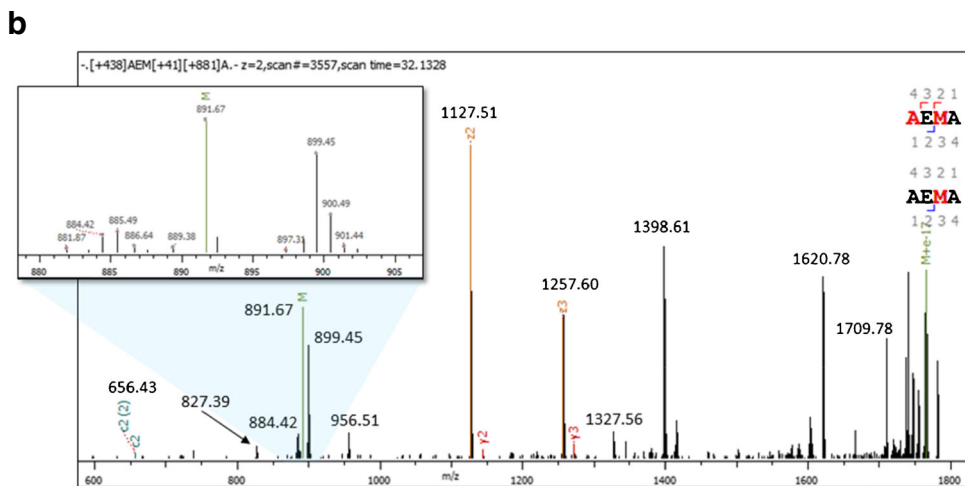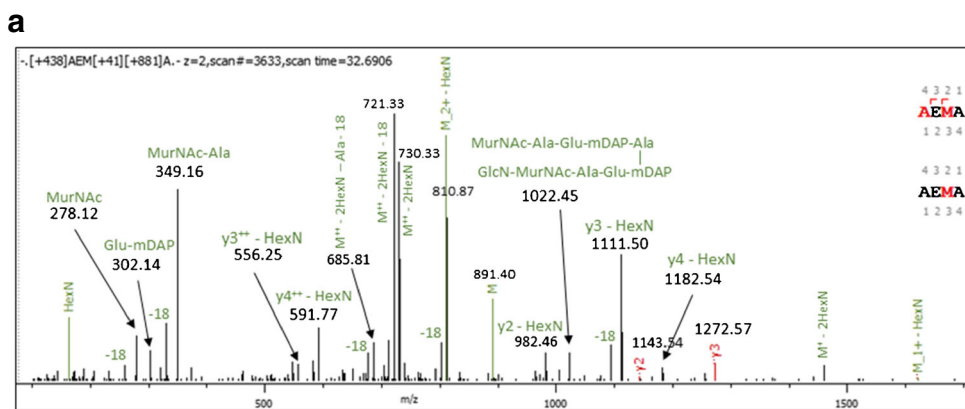
includes about $2 \times 2 \times 20$ peptide sequences (two possibilities for each donor peptide and up to 20 for the terminal acceptor peptide) and $5 \times 5 \times 5$ possibilities for glycosylation (no glycan, MurNAc, GlcN-MurN, GlcN-MurNAc and GlcNAc-MurNAc), for a total of about ~5000 candidates with less than 500 distinct masses. Accurate precursor mass alone, without fragmentation, can resolve almost all precursors in a "search space" of this size; for example, GlcNAc-MurNAc-AEmA and GlcN-MurNAc-AEM(I/L) differ by 38 ppm, almost 8 times the precursor mass tolerance of 5 ppm. We also find that our search strategy gave acceptably low false-negative rate: All the TIC elution peaks in Fig. 2 have identifications, and in all but a few cases mentioned above, these identifications explain the major peaks in the summed MS spectra.

As mentioned above, we relied on manual analysis to distinguish 3–3 from 4–3 cross-links, so we limited attention to the most abundant dimers. All the MS/MS spectra we examined of precursors with mass matching $2 \times$ GlcN-MurNAc-AEmA appeared to contain 4–3 cross-links, rather than a 3–3 link between GlcN-MurNAc-AEm and GlcN-MurNAc-AEmAA. This finding is of course unsurprising because it agrees with monomer abundances. The sample may, however, contain some amount of the 3–3 form co-eluting with the 4–3 form. Figure 4 gives an example manual analysis for both HCD and ETD spectra. The HCD spectrum contains no peaks distinguishing 3–3 and 4–3 cross-links, partly because HCD spectra of PG components are dominated by internal fragments, few of which can be unambiguously linked to one cross-link form. For example, an internal fragment with the mass of two mDAP residues would be unambiguous evidence

of a 3–3 link, but an internal fragment with the mass equal to the sum of one Ala and two mDAP residues could arise from either cross-link form. ETD spectra, however, are much simpler, dominated by c- and z-ions, and we found that ETD spectra of PG dimers generally include strong peaks such as 899.45 and 884.42 in Fig. 4 for the c- and z-ions splitting the monomers. We find that ETD, which to our knowledge has not been applied to PG analysis before, offers a promising alternative or adjunct to collisional fragmentation.

All the MS/MS spectra we examined of dimers containing GlcN-MurNAc-AEm and GlcN-MurNAc-AEmA appeared to contain 3–3 cross-links. Figure 5 gives a manually annotated spectrum. Notice that in both Figs. 4 and 5, the ETD spectra show strong cleavage at the glycosidic bonds between GlcN and MurNAc (thereby localizing the deacetylation), between MurNAc and the lactyl group, and between the lactyl group and the alanine residue in position 1; this cleavage pattern is somewhat unexpected because ETD generally cleaves peptide bonds more readily than glycosidic bonds [14]. Also notice that in Fig. 4, the cleavage between mDAP and D-Ala shows a hydrogen transfer from the c-ion to the z-ion, a common occurrence in ETD spectra of peptides with precursor charge $z = 2+$ [15].

As shown in Figure S1 in the ESM, the 4–3 dimer shows two distinct elution peaks at 30.6 and 32.2 min, with the earlier peak about twice as abundant as the later. The 3–3 dimer has distinct peaks at 28.6 and 31 min in the ratio of about 4:1. In both cases, the MS/MS spectra show no differences between the two elution peaks. We speculate as others have before [11, 16] that these dimers have two forms each, differing only in stereochemistry.

Fig. 5 ETD spectrum of GlcN-MurNAc-A-E-mDAP 3–3 linked to GlcN-MurNAc-A-E-mDAP-A. The peaks at 828.45 and 884.42 match the theoretical masses of 828.38 and 884.39 for c- and z-ions splitting donor and acceptor peptides. ETD fragments PG components at both peptide and glycosidic bonds, including the C–O bond between the lactyl group and MurNAc, shown in the fragmentation diagram with a line across MurNAc
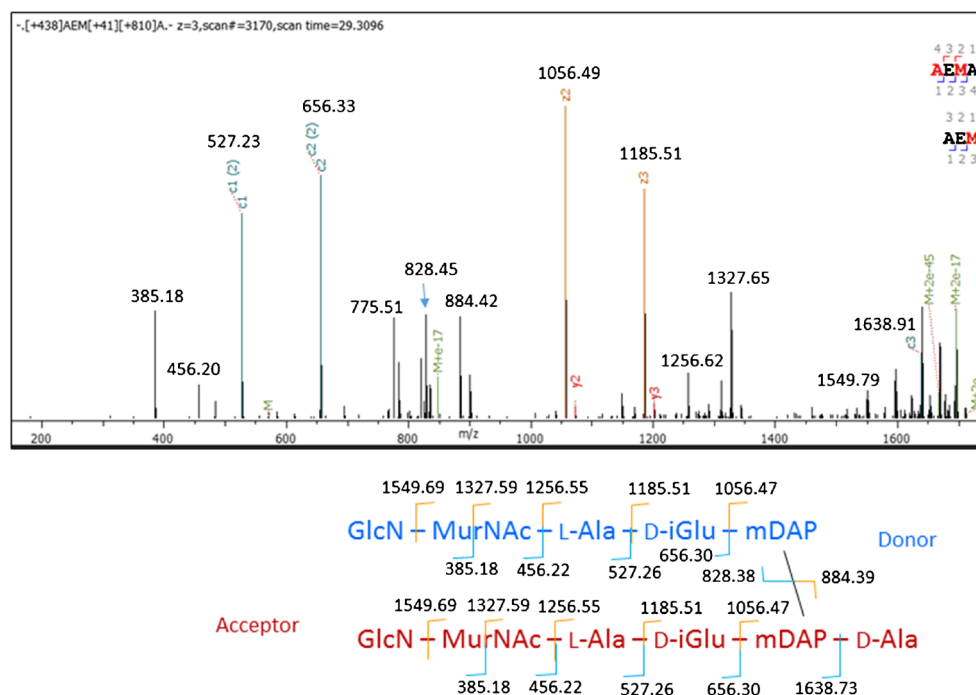
Conclusion

This study describes an effective data-dependent tandem mass spectrum acquisition strategy, along with an almost fully automated data analysis approach to identify muropeptide structures, employing *C. difficile* PG as a proof of concept. We showed that nano-LC separation combined with the low ppm *m/z* accuracy of the Orbitrap mass spectrometer allows the structural analysis of sub-microgram amounts of peptidoglycan extracted from the equivalent of as few as 10^4 cells in less than an hour. The objective automated analysis of MS/MS spectra by the Byonic software confirmed previous structural analyses and revealed several modifications found in low abundance that had not been reported so far in *C. difficile*, in particular a wide variety of C-terminal amino acids in acceptor peptides. Taken together, we propose that this novel analysis pipeline could replace the conventional chromatographic and mass spectrometric approaches used for the past 30 years.

Author contributions Marshall Bern and Stephane Mesnage designed the research. Richard Beniston and Stephane Mesnage prepared samples and optimized the nano-LC-MS conditions. Marshall Bern and Stephane Mesnage analysed the data and wrote the paper.

Compliance with ethical standards

Conflict of interest The authors declare no conflicts of interest.

References

1. Vollmer W, Blanot D, de Pedro MA. Peptidoglycan structure and architecture. FEMS Microbiol Rev. 2008;32(2):149–67.
2. Schleifer KH, Kandler O. Peptidoglycan types of bacterial cell walls and their taxonomic implications. Bacteriol Rev. 1972;36(4):407–77.
3. Vollmer W. Structural variation in the glycan strands of bacterial peptidoglycan. FEMS Microbiol Rev. 2008;32(2):287–306.
4. Glauner B. Separation and quantification of muropeptides with high-performance liquid chromatography. Anal Biochem. 1988;172(2):451–64.
5. Kuhner D, Stahl M, Demircioglu DD, Bertsche U. From cells to muropeptide structures in 24 h: peptidoglycan mapping by UPLC-MS. Sci Rep. 2014;4:7494.
6. Desmarais SM, Tropini C, Miguel A, Cava F, Monds RD, de Pedro MA, et al. High-throughput, highly sensitive analyses of bacterial morphogenesis using ultra performance liquid chromatography. J Biol Chem. 2015;290(52):31090–100.
7. Frirdich E, Biboy J, Adams C, Lee J, Ellermeier J, Gielda LD, et al. Peptidoglycan-modifying enzyme Pgp1 is required for helical cell shape and pathogenicity traits in Campylobacter jejuni. PLoS Pathog. 2012;8(3):e1002602.
8. Bui NK, Gray J, Schwarz H, Schumann P, Blanot D, Vollmer W. The peptidoglycan sacculus of Myxococcus xanthus has unusual structural features and is degraded during glycerol-induced myxospore development. J Bacteriol. 2009;191(2):494–505.
9. Patti GJ, Chen J, Schaefer J, Gross ML. Characterization of structural variations in the peptidoglycan of vancomycin-susceptible Enterococcus faecium: understanding glycopeptide-antibiotic binding sites using mass spectrometry. J Am Soc Mass Spectrom. 2008;19(10):1467–75.
10. Lambert C, Lerner TR, Bui NK, Somers H, Aizawa S, Liddell S, et al. Interrupting peptidoglycan deacetylation during Bdellovibrio predator-prey interaction prevents ultimate destruction of prey wall, liberating bacterial-ghosts. Sci Rep. 2016;6:26010.
11. Peltier J, Courtin P, El Meouche I, Lemee L, Chapot-Chartier MP, Pons JL. Clostridium difficile has an original peptidoglycan structure with a high level of N-acetylglucosamine deacetylation and mainly 3-3 cross-links. J Biol Chem. 2011;286(33):29053–62.
12. Mainardi JL, Fourgeaud M, Hugonnet JE, Dubost L, Brouard JP, Ouazzani J, et al. A novel peptidoglycan cross-linking enzyme for a beta-lactam-resistant transpeptidation pathway. J Biol Chem. 2005;280(46):38146–52.
13. Krokhin OV, Craig R, Spicer V, Ens W, Standing KG, Beavis RC, et al. An improved model for prediction of retention times of tryptic peptides in ion pair reversed-phase HPLC: its application to protein peptide mapping by off-line HPLC-MALDI MS. Molec Cell Proteom : MCP. 2004;3(9):908–19.
14. Catalina MI, Koeleman CA, Deelder AM, Wuhrer M. Electron transfer dissociation of N-glycopeptides: loss of the entire N-glycosylated asparagine side chain. Rapid Commun Mass Spectrom : RCM. 2007;21(6):1053–61.
15. Baker PR, Medzihradszky KF, Chalkley RJ. Improving software performance for peptide electron transfer dissociation data analysis by implementation of charge state- and sequence-dependent scoring. Molec Cell Proteom : MCP. 2010;9(9):1795–803.
16. Atrih A, Bacher G, Allmaier G, Williamson MP, Foster SJ. Analysis of peptidoglycan structure from vegetative cells of Bacillus subtilis 168 and role of PBP 5 in peptidoglycan maturation. J Bacteriol. 1999;181(13):3956–66.