Title: High-throughput Automated Muropeptide Analysis (HAMA) Reveals Peptidoglycan Composition of Gut Microbial Cell Walls

Authors: Ya-Chen Hsu[1], Pin-Rui Su[1,2], Lin-Jie Huang[1], Kum-Yi Cheng[1], Chun-hsien Chen[1], Cheng-Chih Hsu[1]*

Affiliations:

[1] Department of Chemistry, National Taiwan University, Taipei 10617, Taiwan

[2] Department of Molecular Genetics, Erasmus MC, Rotterdam, Netherlands

Corresponding Author:

* Cheng-Chih Hsu

Department of Chemistry, National Taiwan University, Taipei 10617, Taiwan

+886-2-3366-3844, ccrhsu@ntu.edu.tw

ORCID: Cheng-Chih Hsu: 0000-0002-2892-5326

ABSTRACT

Peptidoglycan (PGN), a net-like polymer constituted by muropeptides, provides protection for microorganisms and has been one of the major targets for antibiotics for decades. Researchers have explored host-microbiome interactions through PGN recognition systems and discovered key muropeptides modulating host responses. However, most common characterization techniques for muropeptides are labor-intensive and require manual analysis of mass spectra due to the complex cross-linked PGN structures. Each species has unique moiety modifications and inter-/intra-bridges, which further complicates the structural analysis of PGN. Here, we developed a high-throughput automated muropeptide analysis (HAMA) platform leveraging tandem mass spectrometry and *in silico* muropeptide MS/MS fragmentation matching to comprehensively identify muropeptide structures, quantify their abundance, and infer PGN cross-linking types. We demonstrated the effectiveness of the HAMA platform using well-characterized PGNs from *E. coli* and *S. aureus* and further applied it to common gut bacteria including species of *Bifidobacterium, Bacteroides, Lactobacillus, Enterococcus,* and *Akkermansia*. We thoroughly explored their PGN structures accurately identified muropeptide mono-/multi-mers, and even unambiguously discriminated the structural isomers via the HAMA platform. Furthermore, we found that the cell stiffness may be correlated to the compactness of the PGN structures through the length of interpeptide bridges or the site of transpeptidation within *Bifidobacterium* species. In summary, the HAMA framework exhibits an automated, intuitive, and accurate analysis of PGN compositions, which may serve as a potential tool to investigate the post-synthetic modifications of saccharides, the variation in interpeptide bridges, and the types of cross-linking within bacterial PGNs.

INTRODUCTION

Enveloping the cytoplasmic membrane, peptidoglycan (PGN) is an essential and distinct component of bacterial cell walls that maintains the defined bacterial shape, preserves cell integrity by counteracting the internal osmotic pressure,[1] and acts as a scaffold for other cell wall-anchored components such as proteins and teichoic acids.[2] PGN also provides a protective barrier for microorganisms against environmental threats, contributing to its structural rigidity. PGN is a net-like polymeric structure composed of various muropeptide molecules, with their glycans linearly conjugated and short peptide chains cross-linked through transpeptidation.[3] The muropeptide (i.e., PGN-repeating unit) comprises *N*-acetylglucosamine (GlcNAc), *N*-acetylmuramic acid (MurNAc) and a 2–5 amino acid peptide stem, where the two saccharides are linked by β-1,4 glycosidic bonds and the peptide chain covalently binds to the lactyl moiety of MurNAc. In Gram-negative bacteria, the peptide stem generally contains L-alanine, D-iso-glutamic acid, *meso*-diaminopimelic acid (*m*DAP), D-alanine, and D-alanine, while in Gram-positive bacteria, it contains L-alanine, D-iso-glutamine, L-lysine, D-alanine, and D-alanine. Although the structural composition of PGN is relatively conserved across species, the mature PGN structure remains heterogeneous and species-specific. This structural diversity is attributed to chemical modifications of the sugar backbone (e.g., *O*-acetylation, *N*-deacetylation, and *N*-glycolylation), alternative amino acids of the peptide stem, interpeptide bridge compositions (e.g., various number of glycine chain), and cross-linking types (e.g., 4-3 linkage, 3-3 linkage, and 2-4 linkages).[4, 5]

In recent years, characterizing muropeptides has become increasingly important, not only because PGN is a target for antibiotic drug design, but also because of the various roles muropeptides play as signaling molecules involved in microbial interaction, antimicrobial release,

and host innate immunity.[6-9] The gut microbiota, consisting of thousands of bacteria, is a key source of PGNs in the host organisms. The signaling functions of gut microbiota-derived PGN fragments in host-microbiota interactions have been extensively studied in inflammation, metabolism, autoimmune diseases, and brain development.[10-13] Therefore, analyzing the PGN structural composition and its chemical modifications is critical to understand the PGN recognition processes and subsequent activation mechanisms in the immune system.[14]

Nowadays, analytical methods for PGN architecture include HPLC, LC-MS, solid-state NMR, and AFM imaging, enabling a comprehensive study of PGN structure from chemical composition to 3-D architecture.[5] Online ultra-performance liquid chromatography coupled with electrospray ionization–tandem mass spectrometry (UPLC-ESI-MS/MS) is the primary strategy for muropeptide structural identification, requiring minimal sample amounts and time.[15-17] However, peak annotation and MS/MS spectra alignment still mostly relies on manual inspection of experienced biological chemists. Previous work published by Bern *et al.* demonstrated on the analysis of muropeptide monomers with MS/MS fragmentation spectra by commercial Byonic peptide identification software.[18] Later, PGFinder showed consistent and reproducible data analysis of Gram-negative bacterial PGN with MS1 library searching.[19] Although they both demonstrated great success in identifying muropeptide monomers, the accurate identification of muropeptide multimers and other various bacterial PGN structures still remains unresolved. This is because deciphering the compositions requires MS/MS fragmentation, but it is still challenging to automatically annotate MS/MS spectra from these complex muropeptide structures.

Here, we developed an in-house, high-throughput automated muropeptide analysis (HAMA) platform that simplifies muropeptide structures to sequence format, making it compatible with

the (glyco-)proteomics-like “bottom-up” approach. We then established a comprehensive *in silico* MS/MS fragmentation database for muropeptide identification. Using this platform coupled with high-resolution mass spectrometry, we revealed the PGN compositions of common gut bacteria (some of which are Gram-positive), including species from *Bifidobacterium*, *Bacteroides*, *Lactobacillus*, *Enterococcus*, and *Akkermansia*.

MATERIAL AND METHODS

Bacterial strains and cell culture. Bacteria and their respective media used in this work were listed in Supplementary Table 1. Anaerobic bacteria such as *Bifidobacterium*, *Bacteroides*, *Lactobacillus*, *Enterococcus*, and *Akkermansia muciiniphila* were cultured anaerobically by using an anaerobic workstation (Whitley DG 250, Don Whitley Scientific Limited, England).

Peptidoglycan extraction and mutanolysin digestion. The extraction of peptidoglycan was performed using a previously described method, with some modifications.[9, 16] Ten milliliters of the overnight culture that has reached the stationary phase were harvested and lysed by 2% sodium dodecyl sulfate (SDS) solution in 0.1 M Tris/HCl (pH 7.0). The SDS lysate was boiled for 30 minutes at 100 □ in a heating block and spun down at 10,000 rpm for 5 minutes. In order to wash out SDS thoroughly, the pellet was resuspended with 1 mL of H_2O, centrifuged at 10,000 rpm for 5 minutes, and removed supernatant. After washed with H_2O three times, the pellet was resuspended in 750 μL of H_2O and sonicated for 30 minutes. The suspension was spiked with 750 μL of 0.1 M Tris/HCl solution (pH 7.0) and treated with DNase and RNase at 37 □ in a shaker for 1 hour. After the removal of residual nucleic acids, the suspension was subsequently treated with pronase to digest cell wall-bound proteins (final concentration of 0.1 μg/mL for 16 hours at the same conditions). The pronase-treated cell walls were washed twice with H_2O before wall teichoic acids were released by 1 mL of 1 N HCl incubation for 4 hours. Insoluble peptidoglycans were washed with H_2O until pH 5-6 and resuspended in 200-500 μL of 12.5 mM sodium dihydrogen-phosphate buffer (pH 5.5) to an OD_{600} of 3.0. One hundred microliters of insoluble peptidoglycan suspensions were digested overnight with 1 μL of mutanolysin solution (1.000 U/mL in H_2O). Following mutanolysin inactivation (100 □ boiling for 3 minutes) and

centrifugation (5 minutes at 10,000 rpm), the supernatants were spiked with 50 μL of reduction solution (10 mg/mL sodium borohydride in 250 mM borate buffer at pH 9.0). After 20 minutes incubation at room temperature, the reduction reaction was stopped by adding 1 μL of phosphoric acid (98%), resulting in pH 2-3. Reduced muropeptide samples were subsequently analyzed by UPLC-MS/MS.

UPLC-MS/MS analysis of muropeptides. UPLC-MS/MS analysis was performed by a Dionex UltiMate 3000 UHPLC system coupled to a Q-Exactive Plus hybrid quadrupole-orbitrap mass spectrometer (Thermo Scientific). Soluble muropeptides were separated by an ACQUITY UPLC CSH C18 column (130 □, 1.7 μm, 2.1 mm × 100 mm), with a solvent A (0.1% formic acid in water) and solvent B (0.1% formic acid in acetonitrile) as the mobile phase. A flow rate of 0.25 mL/min and 1% mobile phase B were applied in column condition. The injection volume of each sample was 5 μL. An elution gradient was run for 25 minutes: starting with 1% solvent B for 0-2 min; 1-20% B (linear), 2-15 min, 20-95% B (linear), 15-17 min; 95% B, 17-19 min; 95-1% B (linear), 19-21 min; 1% B, 21-25 min for re-equilibration. The column temperature was controlled at 52 □ throughout the whole analysis program. A Heat electrospray ionization (HESI) source was operated in positive mode, with parameters automatically optimized under a flow rate of 250 μL/min (a capillary temperature of 300 □, a probe temperature of 300 □, and a spray voltage of 3.5 kV). Data-dependent acquisition (DDA) mode was used in the instrument. A mass range of MS1 acquisition was from 400 to 2,000 *m/z* at a resolution of 70,000 and the top 5 most abundant ions were subjected to higher-energy collision-induced dissociation (HCD) fragmentation with Δ*m/z* 3 isolation window, stepped normalized collision energy (NCE) at 20, 25, and 35 (A.U.), and dynamic exclusion time of 4 sec.

Data processing and data analysis. LC-MS/MS raw data was converted to an mzXML format using MSConvert (ProteoWizard) and then processed by HAMA for mass deconvolution and PGN identification.

Mutanolysin digestion assay. Purified peptidoglycan was adjusted to OD_{600} 1.0 in sodium dihydrogen-phosphate buffer and digested with 5 μL of mutanolysin (1.000 U/mL in H_2O). Absorbance was measured at OD_{600} of time points at 0, 3, 6, 9, and 16 hours in a Synergy H1 microplate reader (BioTek) with constant orbital shaking at 37 °C.

Immobilization of cultured bacteria for AFM imaging. The method of bacteria immobilization was performed as described previously, with some modifications.[20] Overnight cultured bacteria were washed twice and diluted to OD_{600} 0.3 with 1x PBS. Freshly cleaved Si wafers were coated by 500 μL of 0.1% (w/v) poly-L-lysine in H_2O and left overnight at room temperature. Afterward, the substrates were rinsed five times with ultrapure H_2O and dried with a nitrogen flow. One hundred microliters of the bacteria suspension were dropped on a silicon substrate, left for 10-20 minutes, and then rinsed three times in H_2O bath. The bacteria were dried onto the substrate with flowing nitrogen and rehydrated again in 1x PBS bath for liquid imaging.

Atomic force microscopy (AFM) imaging. Cultured bacteria images were acquired in 1x PBS bath at room temperature by AFM (Dimension Icon, Bruker) with ScanAsyst™. Data were collected by the PeakForce Quantitative NanoMechanics (QNM) mode with qp-CONT probes

(spring constant: 0.25 N/m, resonant frequency: 30 kHz, NANOSENSORS™). Images were acquired at a scan rate of 25 μm/s, an applied force of 500 pN, and with a resolution of 256 × 256 pixels per image frame. For mechanical analysis, the approaching part of the curves were fitted with the Derjaguin-Muller-Toporov (DMT) model. DMT modulus of each bacterium was recorded with 16 × 16 curves on top of the bacteria with the area of 250 × 250 nm^2.

RESULTS

HAMA platform: a High-throughput Automated Muropeptide Analysis for Identification of PGN Fragments

Identification and structural analysis of PGNs are time-consuming and challenging due to the complex compositions and modifications. Herein, we developed a high-throughput and automated platform for PGN structural analysis inspired by the (glyco-)proteomic "bottom-up" approach. However, unlike proteins are 1-D amino acid sequences, the PGN is composed by various muropeptides that are intertwined by their glycans and peptide stems. Hence, we first simplified the muropeptide structures as 1-line muropeptide sequences while preserving their complex linkage compositions (Figure 1a, b). Based on the prior knowledge of all possible compositions and modifications of the bacteria species, we used *DBuilder* to build the species-specific database. Subsequently, we analyzed the MS data with *Analyzer* along with the possible *b*- and *y*- ions of CID/HCD fragmentation generated *in silico*. The above-mentioned automated identification processes are compiled in the HAMA platform, a MATLAB-based software with a user-friendly graphic user interface (Supplementary Figure 1 and Additional Files) that includes three parts: *DBuilder*, *Analyzer*, and *Viewer*.

(i) ***DBuilder*** *for Species-specific muropeptide database construction.* In order to construct muropeptide databases *in silico*, we named each residue in the muropeptide with a distinct letter and simplified the muropeptide structure into a 1-line sequence format (all representation letters of monosaccharides and amino acids are listed in Supplementary Table 2). Muropeptide sequences were structured with dashes representing peptide/glycosidic bonds and parentheses to discriminate certain subunits. Based on their known (or putative) PGN structures, all possible combinations of GlcNAc, MurNAc and peptide were input into *DBuilder* to generate a

comprehensive database that contains monomeric, dimeric, and trimeric muropeptides (Figure 1b). Herein, we constructed PGN multimers using two types of polymerization events: transglycosylation connected via glycosidic bonds, and transpeptidation linked through peptide bonds. In *DBuilder*, the type of peptide linkage used was the common 4-3 cross-link, which could be achieved either directly through a covalent bond between the penultimate D-Ala and the third residue in the acceptor peptide stem, or via an interpeptide bridge. We also considered that the terminal D-Ala (at the fifth position) in the donor stem was not allowed during the transpeptidation reaction. In addition, choosing more possible modifications (anhydro MurNAc residues, deacetyl GlcNAc residues, acetyl MurNAc residues, and amidated iso-Glu) to construct a PGN database could lead to severe mass coincidence issues for MS1 searching in the *Analyzer*. Hence, we set a maximum modification number of six to avoid generating multi-modified muropeptide candidates. Finally, the species-specific muropeptide database was outputted to a *.csv format file, including muropeptide sequences and corresponding chemical formula and theoretical monoisotopic mass.

(ii) ***Analyzer*** *for analysis of muropeptide MS raw files. Analyzer* is an integrated tool for processing MS data and identifying muropeptides. The flowchart is shown in Figure 1c. First, the species-specific database, MS raw file (mzXML format),[21] and parameters were loaded into *Analyzer*. Parameters were originally set with MS1 range *m/z* 400-2000, retention time range 2-12 min, mass tolerance of 10 ppm (orbitrap mass spectrometry data), intensity threshold of 1e5, etc., which could be input by users for different experimental conditions. Then, the mzXML file was processed sequentially by peak picking, deisotoping, feature selection, and deconvolution. The observed masses were searched against the loaded database within 10 ppm tolerance. However, in cases where the observed mass matched more than one inferred sequence in MS1

searching, *Analyzer* compared the MS/MS spectra of those MS1 matched features to *in silico* MS/MS fragmentation spectra of corresponding inferred candidates through cosine similarity and matched peak score (MPS, the number of the matched peaks divided by the number of the predicted peaks) calculation. The observed mass was then identified as the sequence with the highest matching score. Lastly, all identified muropeptides were merged into an Excel spreadsheet reporting charge state, molecular weight, retention time, peak intensity, peak area, sequence, main scan number, cosine similarity, score, etc. *Analyze*r also outputted a base peak chromatogram with peak annotations and a result file (in MATLAB data) for *Viewer* input. The entire analysis was completed within a few minutes.

(iii) ***Viewer*** *for annotation and identification display. Viewer*, a visualization tool that allows users to browse extracted ion chromatograms (XICs) of the identified muropeptides and visualize the *in silico* MS/MS matching spectra annotated with *b*- and *y*- ions. In the muropeptide spectra match (PSM) page of each identified muropeptide, *Viewer* lists all inferred candidates and visualizes their individual MS/MS spectral matches to clearly demonstrate the process of scoring in *Analyzer*. Additionally, *Viewer* provides a data sorting function that allows users to classify the identification list by molecular weight, peak area, score (in ascending or descending order), and sequence (in alphabetical order).

Demonstration of the HAMA Platform Using Well-characterized PGNs of *E. coli* and *S. aureus*

As a proof-of-concept, we demonstrated the HAMA platform using well-characterized peptidoglycans from Gram-negative *Escherichia coli* DH5α and Gram-positive *Staphylococcus aureus* SA113. The typical PGN subunit of *E. coli* is GlcNAc – MurNAc – L-Ala – D-iGlu –

*m*DAP – D-Ala – D-Ala, represented as a sequence of B-M-l(-A-E-H-A-A) in the HAMA platform. For *S. aureus*, the classic PGN structure is a disaccharide-pentapeptide with the Gly_5 interpeptide bridge branching from the lysine, which is simplified as B-M-l(-A-Q-(G-G-G-G-G-)K-A-A. Based on the analyzed input dataset, approximately 70% of the peak area in the base peak chromatograms was assigned to muropeptide signals, which allowed for a comprehensive PGN mapping (Figure 2a, c). The HAMA platform successfully identified *E. coli* and *S. aureus* muropeptides, and their XICs and MS/MS matching spectra could be visualized in *Viewer* (Figure 2b, d). In total, 28 and 89 unique muropeptides were identified in PGNs of *E. coli* and *S. aureus*, respectively. Table 1 lists overviews of the main muropeptide peaks in the chromatograms. Upon examining the entire identification results, we found that some muropeptides were eluted in multiple retention times due to the existence of a few abundant stereoisomers. Additionally, a few monosaccharide muropeptides identified as the loss of GlcNAc could be recognized as in-source fragments by exact co-elution.

In the *E. coli* data, the most abundant muropeptides were disaccharide-tetrapeptide (Tetra monomer) and disaccharide-tetrapeptide-disaccharide-tetrapeptide (Tetra-Tetra dimer). We also identified two low-abundant muropeptides, B-M-l(-A-E-H-G) and B-M-l(-A-E-H-A-G), in which the fourth and fifth Ala are individually substituted with a Gly. This unique composition has been reported in strain *E. coli* Nissle 1917 from the previous analysis,[16] but not for DH5α strain used in our study. In the output of *S. aureus* peptidoglycan, the elution profile was the same as what has been previously known: the most abundant monomers were disaccharide-pentapeptide with a $(Gly)_5$ bridge and disaccharide-pentapeptide with a Gly bridge. The most abundant cross-linked dimer was disaccharide-tetrapeptide-disaccharide-pentapeptide, which contains a total of ten Gly residues.[16] We also characterized the known modifications and

structural variations within *S. aureus* peptidoglycan, such as *O*-acetylation of MurNAc, the presence of D-iGlu (non-amidated) in the stem peptides, and the length variation of the interpeptide bridge. The high-throughput analysis allowed for the identification of monomeric muropeptides consisting of one to nine Gly residues and dimeric muropeptides containing a total of five to fourteen Gly residues in a single analysis. However, due to the insufficient structural information in the peptide backbone provided by the HCD fragmentation spectra, the exact number of Gly residues harbored in each PGN unit could not be determined. Hence, further careful analysis and manual verification are required to confirm the identity of identified muropeptides, particularly for low-score dimers and trimers. Collectively, these results provide valuable insights into the PGN compositions and architectures.

Characterizing Gut Bacterial PGN Compositions and Resolving Isomeric Muropeptides

Over the past two decades, an increasing number of gut microbial species have been found to be associated with human health. In addition, emerging evidence has supported the notion that gut microbial muropeptides work as signaling molecules that mediate host−microbiome interactions in metabolism, gut homeostasis, and immunity.[22] However, little has been discussed about the structures of PGN fragments due to the diversity and complexity of gut bacterial PGNs and the lack of an efficient analytical tool.[14] To date, most of the structural information on gut bacterial peptidoglycan comes from reports published between 1970 to 2000,[23-25] whereas that of the health-promoting gut microbes discovered in the last two decades have been rarely reported. Therefore, we utilized the HAMA platform to investigate the specific PGN structures of several common gut bacterial species from *Bifidobacterium*, *Bacteroides*, *Lactobacillus*, *Enterococcus*, and *Akkermansia*. We collected LC-MS/MS data of purified peptidoglycans from ten gut

bacterial species and identified those with the self-built species-specific muropeptide databases. Overall, the base peak chromatograms of each species showed approximately 70% of peaks were annotated as muropeptides, and these elution profiles reflected the actual composition and organization within the peptidoglycan architecture. Additionally, the structural compositions of these purified peptidoglycans are consistent with previous reports, listed in Table 2. The muropeptide profiles of the species in this study are summarized in Supplementary Table 3-12.

In the identification results of Gram-positive gut bacterial peptidoglycans, the main muropeptide sequence is GlcNAc – MurNAc – L-Ala – D-iGln – L-Lys – D-Ala – D-Ala, in which the second amino acid of the stem peptide is usually amidated to D-iGln from D-iGlu. This chemical modification is done by the MurT or GatD biosynthetic enzyme and is supposed to control the PGN cross-linking levels, which has recently been demonstrated using labeled PGN stem mimics in certain species.[26] Another unique feature is L-ornithine (L-Orn) at the third position of the stem peptide. Unlike L-lysine, which is a usual feature of Gram-positive peptidoglycans, L-Orn is a non-proteinogenic amino acid and has been found in the peptidoglycan of certain *Bifidobacterium* species before.[23] As anticipated, we identified the L-Orn-harbored muropeptides in *B. bifidum* and *B. longum* peptidoglycans. The most species-specific variation among Bifidobacterial PGN was the architecture of the interpeptide bridge, such as Gly, Ser-Asp, and Ser-Ala-Thr-Ala, which corresponded to *B. breve*, *B. bifidum*, and *B. longum* species. This structural diversity arouses our interest in the relationship between the length of an interpeptide bridge and the physical property of the bacterial cell envelope, which will be discussed in a later section.

In the peptidoglycans of *E. faecium* and *L. acidophilus* (Supplementary Tables 4 and 5), we observed the monomeric structure of GlcNAc – MurNAc – L-Ala – D-iGln – L-Lys – D-Ala – D-

Ala, which harbored an interpeptide bridge of asparagine (D-Asn) or aspartate (D-Asp), shortened as B-M-l(-A-Q-(N-)K-A-A) or B-M-l(-A-Q-(D-)K-A-A). In this part, we found that both B-M-l(-A-E-(N-)K) and B-M-l(-A-Q-(D-)K) had identical molecular weights since the mass difference (+ 0.984 Da) between Asn (N) and Asp (D) is the same as between Gln (Q) and Glu (E). This kind of isomeric muropeptides makes identification more complicated, but it can still be addressed by MS/MS *in silico* fragmentation matching under an appropriate separation chromatography. For example, two disaccharide-tripeptides separately eluted at 6.18 min and 7.02 min were found as structural isomers that existed in *E. faecium* peptidoglycan (Figure 3 and Supplementary Table 4). Through *in silico* MS/MS fragmentation matching, these two isomers can be discriminated and identified as the sequences of B-M-l(-A-E-(N-)K) and B-M-l(-A-Q-(D-)K), respectively. To validate the correctness of the automated platform, we also extracted the experimental data of those isomers and manually inspected the MS/MS spectra (Supplementary Figure 2). However, this strategy did not work for the identification of *E. faecalis* peptidoglycan whose interpeptide bridge is composed of two alanine residues (Supplementary Table 3). In this case, two structurally isomeric muropeptides, a disaccharide-tripeptide with a bridge (B-M-l(-A-Q-(A-A-)K)) and a disaccharide-pentapeptide (B-M-l(-A-Q-K-A-A)), have similar *in silico* MS/MS fragmentation patterns, which often leads to misidentification. Nevertheless, the retention time still gave us information to identify them. We annotated the peak at a retention time of 3.66 min as disaccharide-pentapeptide (B-M-l(-A-Q-K-A-A)) by comparing it to an identical sequence that appeared in *L. acidophilus* peptidoglycan with a similar retention time (3.61 min). Therefore, the other peak at a retention time of 6.39 min could be determined as disaccharide-tripeptide with a bridge (B-M-l(-A-Q-(A-A-)K)), which had a longer retention time than the linear structure one.

Apart from Gram-positive bacteria, we also analyzed the peptidoglycan of several anaerobic Gram-negative bacteria, including *Bacteroides fragilis*, *Bacteroides ovatus*, *Bacteroides thetaiotaomicron*, and *Akkermansia muciniphila* (Table 2 and Supplementary Table 9-12). Structurally, the general stem peptide of Gram-negative peptidoglycan is L-Ala – D-iGlu – *m*DAP – D-Ala – D-Ala, with diaminopimelic acid in the third position being a representative feature. The peptidoglycan structures we identified were consistent with previously published ones, with *O*-acetyl-MurNAc found in *Bacteroides* species and *N*-deacetyl-GlcNAc (GlcN) in *A. muciniphila*.[27, 28] The output analysis showed that around 56-66% of the total muropeptides in *Bacteroides* species' peptidoglycans contained *O*-acetylated MurNAc, while approximately 87% of the total muropeptides in *A. muciniphila* peptidoglycan contained de-*N*-acetylated GlcNAc. The high occurrence of *N*-deacetylation in *A. muciniphila* peptidoglycan suggests that *A. muciniphila* might possess a homolog of oxidative stress-induced PGN deacetylase (PgdA) found in *Helicobacter pylori*.[29-31] Chemical modifications to the disaccharide backbone are known to provide resistance to lysozyme and protect bacteria against enzymatic attack from the host innate immune system.[32, 33]

Inferring PGN Cross-linking Types Based on Identified PGN Fragments

The substrates and catalyzed enzymes involved in muropeptide cross-linking have been targets for antibiotic development and antimicrobial resistance studies.[34] Consequently, we sought not only to identify the PGN structures but also to explore the peptide cross-linking types within these gut bacterial PGNs. In general, there are two types of PGN cross-linkage: 4-3 cross-links generated by D,D-transpeptidases (Ddts) and 3-3 cross-links created by L,D-transpeptidases (Ldts) (Figure 4a). Transpeptidation involves two stem peptides which function as acyl donor

and acceptor substrates, respectively. As the enzyme names imply, the donor substrates that Ddts and Ldts bind to are terminated as D,D-stereocenters and L,D-stereocenters, which structurally means pentapeptides and tetrapeptides. During D,D-transpeptidation, Ddts recognize D-Ala^4-D-Ala^5 of the donor stem (pentapeptide) and remove the terminal D-Ala^5 residue, forming an intermediate. The intermediate then cross-links the NH_2 group in the third position of the neighboring acceptor stem, forming a 4-3 cross-link. Following this distinctive rule of PGN biosynthesis, the possible combinations of 4-3 cross-linked dimers are disaccharide-tetrapeptide–disaccharide-pentapeptide (D45) and disaccharide-tetrapeptide–disaccharide-tetrapeptide (D44). During L,D-transpeptidation, Ldts cleave the terminal D-Ala^4 residue of the donor stem (tetrapeptide) and generate a peptide bond between the third residue and the NH_2 group in the third position of acceptor stem.[35] The possible structures of 3-3 cross-linked dimers are disaccharide-tripeptide–disaccharide-pentapeptide (D35), disaccharide-tripeptide–disaccharide-tetrapeptide (D34), and disaccharide-tripeptide–disaccharide-tripeptide (D33). Hence, we could infer the possible PGN cross-linkage types and the involved enzymes based on the main PGN fragments.

Take *B. breve* ATCC 15700, CSCC 1900, and ATCC 15698 as examples. The LC-MS profiles showed noticeable differences in muropeptide compositions among the three strains (Figure 4b). The relative muropeptide compositions of each strain are presented in a heatmap (Figure 4c). Obviously, the main muropeptides of *B. breve* ATCC 15700 are disaccharide-tripeptide with an interpeptide bridge (M3b), disaccharide-tetrapeptide with an interpeptide bridge (M4b), and disaccharide-tripeptide–disaccharide-tetrapeptide (D34), whereas the CSCC 1900 strain showed great abundances in disaccharide-tetrapeptide with an interpeptide bridge (M4b), disaccharide-pentapeptide with an interpeptide bridge (M5b), and disaccharide-

tetrapeptide–disaccharide-tetrapeptide (D44). These results suggested that 3-3 cross-links and 4-3 cross-links might be predominant in the peptidoglycans of ATCC 15700 and CSCC 1900 strains, respectively. In the case of ATCC 15698 strain, its peptidoglycan likely contains both types of cross-links since the abundances of M3b and M4b, as well as D34 and D44, are almost equivalent.

Broad-spectrum β-lactams are known to inhibit D,D-transpeptidases, such as penicillin-binding proteins (PBPs). However, L,D-transpeptidases are generally insensitive to β-lactams and offer alternative cross-links in the PGNs.[36] Numerous studies have investigated the role of L,D-transpeptidases in the maintenance and remodeling of mature peptidoglycan in organisms such as *E. faecium, C. difficile, E. coli,* and *M. tuberculosis*.[37-42] Based on the muropeptide compositional analysis mentioned above, we found high abundances of M3/M3b monomer and D34 dimer in the PGNs of *E. faecalis, E. faecium, L. acidophilus, B. breve, B. longum,* and *A. muciniphila*, which may be the PGN products formed by Ldts. While the homologs of Ldts in *L. acidophilus, B. breve, B. longum,* and *A. muciniphila* have been supported by genome sequence evidence,[43-46] biochemical evidence is needed to confirm the existence of L,D-transpeptidases in those species.

Exploring the Bridge Length-dependent Cell Envelope Stiffness in *B. longum* and *B. breve*

Peptidoglycan, a protective exoskeleton around cells, provides structural integrity to the cell. The porosity of the PGN scaffold, defined by the degree of cross-linking, influences the transport of larger molecules such as proteins. Therefore, modifications to PGN structure are anticipated to significantly affect bacterial cell mechanics and interactions.[47, 48] Previous evidence has indicated that a high PGN cross-linking level enhances the stiffness of the cell wall material in Gram-positive bacteria.[49] The PGN-lattice architecture based on the interpeptide bridge length has been

investigated using solid-state NMR in *S. aureus*.[50] However, the effect of the interpeptide bridge length variants on the porosity and stiffness of bacterial cell envelope has little been discussed. Interestingly, the identification table presented above showed that *Bifidobacterium* peptidoglycans have different architecture of the interpeptide bridges among species: a tetrapeptide bridge (Ser-Ala-Thr-Ala) found in *B. longum* and a monopeptide bridge (Gly) found in *B. breve* (Supplementary Figure 3). We wondered whether the length of interpeptide bridges may be related to the bacterial cell envelope's mechanical properties and hypothesized that the cross-linking with shorter bridges may form a tighter meshwork in peptidoglycan layers (Figure 5a).

In order to explore the above relevance, we selected six *Bifidobacterium* bacteria with similar cell morphology as research objects, including three *B. breve* strains and three *B. longum* strains. We summarized the muropeptide compositions of these species (Figure 5b). A slightly higher abundance of peptide-linked dimers in *B. breve* than in *B. longum* implies a higher degree of PGN cross-linking level in *B. breve*. Next, we directly measured the cell stiffness (defined by Young's modulus, E) of these six bacteria using the single-cell atomic force microscopy (AFM) technique. High-resolution elasticity maps of representative cells were recorded within the central regions to avoid edge effects (Figure 5c). Overall, all *B. breve* were stiffer than *B. longum* with five times higher Young's modulus values (Figure 5d). Furthermore, *B. breve* also showed a slower enzymatic hydrolysis rate in purified PGNs, implying that the cell wall structure of *B. breve* is characterized by a compact PGN architecture (Supplementary Figure 4). Since *B. breve* strains have shorter interpeptide bridges (one instead of four amino acids in *B. longum*), these data may suggest that the compactness of PGN and the cell stiffness are correlated. Interestingly, among three *B. breve* strains, the ATCC 15700 strain showed significantly stiffer than the other

strains, and it also contained more abundant 3-3 cross-linkages in its PGN (i.e., more M3b muropeptides). Computational modeling suggested a more rigid stem peptide in the conformation of L,D-cross-linking, implying the 3-3 cross-linkages may strengthen the PNG layer.[36, 51] Taken all together, we speculate that a tight peptidoglycan network woven by shorter interpeptide bridges or 3-3 cross-linkages could give bacteria stiffer cell walls. However, it is important to note that cell stiffness is a mechanical property that also depends on PGN thickness, overall architecture, and turgor pressure. These parameters may vary among different bacterial strains. Hence, carefully controlled, genetically engineered strains with similar characteristics will be needed to dissect the role of cross-bridge length in cell envelope stiffness.

DISCUSSION

The purpose of this study was to develop an automated platform for identifying and analyzing the muropeptide compositions of gut bacterial PGNs. Using high-resolution MS data, we characterized PGN structures of common gut microbes, investigated their PGN cross-linking types, and evaluated the effect of bridge-length variants on the stiffness of bacterial cell walls.

In early 1970s, the amino acid compositions of PGN stem peptides were characterized in many bacteria species, including certain gut microorganisms. At that time, PGNs were purified and hydrolyzed under harsh acidic conditions and then separated by two-dimensional paper chromatography to determine the amino acid sequence. Based on this characterization work, Schleifer and Kandler proposed a bacterial PGN classification system that is still used today.[23] In the *Bifidobacterium* and *Lactobacillu*s genera, only type A PGNs were identified, representing 4-3 cross-linking through interpeptide bridges.[25] However, our study of *B. breve*, *B. longum*, and *L. acidophilus* revealed a significant abundance of M3b and D34 in their muropeptide compositions, implying the incorporation of 3-3 cross-linkage type within PGN network. This cross-linking type is catalyzed by penicillin-insensitive L,D-transpeptidases. We speculate that this finding may be influenced by the comprehensive mass spectrometric approaches we employed or by variations in growth condition. Moreover, we utilized the well-established enzymatic method involving mutanolysin to cleave the β-*N*-acetylmuramyl-(1,4)-*N*-acetylglucosamine linkage, which preserves the original peptide linkage in intact PGN subunits. Importantly, our HAMA platform provides a powerful tool for mapping peptidoglycan architecture, giving structural information on the PGN biosynthesis system. This involves the ability to infer possible PGN cross-linkages based on the type of PGN fragments obtained from hydrolysis. For instance, the identification of 3-3 cross-linkage formed by L,D-transpeptidases (Ldts) is of particular

significance. Unlike 4-3 cross-linkages, the 3-3 cross-linkage is resistant to inhibition by β-lactam antibiotics, a class of antibiotics that commonly targets bacterial cell wall synthesis through interference with 4-3 cross-linkages. Therefore, by elucidating the specific cross-linkage types within the peptidoglycan architecture, our approach offers insights into antibiotic resistance mechanisms.

In most Gram-positive bacteria, the interpeptide bridges vary depending on the species and are likely associated with the PGN architectures. Although the chemical structure of PGN units has been extensively studied, the physical structure and 3-D architecture remain open questions.[52, 53] A previous review by Kim et al. suggested that the length of interpeptide bridges is a key factor in determining different PGN architectures and proposed three PGN-bridge length-dependent architectures based on the orientations of cross-linked peptide stems.[50] The parallel-stem model, which has the smallest pore size and the highest cross-linking level, is suited for bacteria with long-length bridges. The perpendicular-stem model, with intermediate pore size, is proposed for bacteria with intermediate-length bridges, while the antiparallel-stem model, which has the largest pore size, is for bacteria without bridges. Notably, our study suggested a potential correlation between the cell stiffness and the compactness of bacterial cell walls in *Bifidobacterium* species (Figure 5). *B. longum*, which predominantly harbors tetrapeptide bridges (Ser-Ala-Thr-Ala), exhibits a trend towards lower stiffness, whereas *B. breve*, characterized by PGN cross-linked with monopeptide bridges (Gly), demonstrates a trend towards higher stiffness. These findings suggested that it may be correlated between the increased rigidity and the more compact PGN architecture built by shorter cross-linked bridges. Future studies on the effect of quantitatively changing the PGN-bridge length on stiffness will require intensive research, considering the cross-linking densities, glycan strand lengths, PGN

architecture models, and other biomolecules on the cell wall, such as lipoprotein, wall teichoic acids, and lipoteichoic acids.[54-57] These studies may provide further insight into the physical structure and architecture of PGN, which would contribute to a better understanding of the factors determining bacterial cell wall properties.

In the HAMA platform, simplifying muropeptide structures into sequences has facilitated the generation of *in silico* MS/MS fragments for spectra matching, enabling the structural resolving of isomeric muropeptides. However, there are several limitations to our study design. Our species-specific PGN database is built based on known structures and modifications reported previously. While the identified muropeptides cover approximately 70% of peak area in the base peak chromatograms, certain low-abundant muropeptides composed of uncoded amino acids or saccharides required additional MS/MS identification with manual analysis. For peptide-linked multimers, the current *DBuilder* only builds up 4-3 cross-linked sequences because including the 3-3 cross-link type results in severe misidentifications, as it produces similar *in silico* MS/MS fragmentation patterns as the 4-3 cross-links (about 88-90% similarity). However, we can infer possible PGN cross-linkages based on the type of PGN fragments obtained from hydrolysis.[58] The primary limitation is that MS/MS spectra generated by HCD fragmentation carry mostly structural information from disaccharide moiety (glyco-oxonium ions) without providing sufficient peptide fragment ions to derive peptide sequence,[59] making it difficult to accurately identify multimers and determine the position of modifications. Nevertheless, accurate monoisotopic masses acquired from high-resolution mass spectrometry are still informative in peptidoglycan compositional components. Alternative approaches can be considered to elucidate branched structures of muropeptides, such as breaking down PGN with cell wall amidases (cleaving the amide bond between MurNAc and L-Ala residue) to target stem peptide analysis,[60,

[61] acquiring MS^3 spectra to obtain fragmentation maps, or using electron transfer dissociation (ETD) fragmentation mode to leave more peptide fragments.[18] Additionally, strategies commonly used in glycopeptide identification, such as multi-fragmentation modes (CID, HCD, and ETD), could also be applied to PGN identification to provide more information on peptide sequences and disaccharide structures.[62] To achieve a more comprehensive identification of muropeptides, future work is needed to generate an expanded database, *in silico-based* fragmentation patterns, and improved MS/MS spectra acquisition.

HAMA is an innovative automated platform that constructs species-specific muropeptide databases, validates structural identification using *in silico* MS/MS analysis, and provides visualized results for efficient and reliable identification of muropeptides, greatly reducing the time-consuming task of manual interpretation of LC-MS/MS data. The HAMA platform has the potential to be a valuable tool for various research fields, including microbiology, pathology, molecular biology, and immunology. It has potential applications in identifying activation ligands for antimicrobial resistance studies, characterizing key motifs recognized by pattern recognition receptors for host-microbiota immuno-interaction research, and mapping peptidoglycan in cell wall architecture studies. With the ease and efficiency offered by HAMA, we believe that muropeptide analysis will become more accessible and contribute to a deeper understanding of cell wall biology.

REFERENCES

1. Silhavy, T. J.; Kahne, D.; Walker, S., The bacterial cell envelope. *Cold Spring Harb. Perspect. Biol.* **2010,** *2* (5), a000414.
2. Vollmer, W.; Blanot, D.; de Pedro, M. A., Peptidoglycan structure and architecture. *FEMS Microbiol. Rev.* **2008,** *32* (2), 149–167.
3. Irazoki, O.; Hernandez, S. B.; Cava, F., Peptidoglycan muropeptides: release, perception, and functions as signaling molecules. *Front. Microbiol.* **2019,** *10*, 500.
4. Schumann, P., Peptidoglycan structure. In *Methods Microbiol.*, Elsevier: 2011; Vol. 38, pp 101-129.
5. Porfírio, S.; Carlson, R. W.; Azadi, P., Elucidating Peptidoglycan Structure: An Analytical Toolset. *Trends Microbiol.* **2019,** *27* (7), 653-654.
6. Wolf, A. J.; Underhill, D. M., Peptidoglycan recognition by the innate immune system. *Nat. Rev. Immunol.* **2018,** *18* (4), 243-254.
7. Nikolaidis, I.; Favini-Stabile, S.; Dessen, A., Resistance to antibiotics targeted to the bacterial cell wall. *Protein Sci* **2014,** *23* (3), 243-259.
8. Girardin, S. E.; Travassos, L. H.; Hervé, M.; Blanot, D.; Boneca, I. G.; Philpott, D. J.; Sansonetti, P. J.; Mengin-Lecreulx, D., Peptidoglycan molecular requirements allowing detection by Nod1 and Nod2. *J. Biol. Chem.* **2003,** *278* (43), 41702-41708.
9. Huang, Y. W.; Wang, Y.; Lin, Y.; Lin, C.; Lin, Y. T.; Hsu, C. C.; Yang, T. C., Impacts of penicillin binding protein 2 inactivation on β-lactamase expression and muropeptide profile in *Stenotrophomonas maltophilia*. *mSystems* **2017,** *2* (4), e00077-00017.
10. Cavallari, J. F.; Fullerton, M. D.; Duggan, B. M.; Foley, K. P.; Denou, E.; Smith, B. K.; Desjardins, E. M.; Henriksbo, B. D.; Kim, K. J.; Tuinema, B. R.; Stearns, J. C.;

Prescott, D.; Rosenstiel, P.; Coombes, B. K.; Steinberg, G. R.; Schertzer, J. D., Muramyl dipeptide-based postbiotics mitigate obesity-induced insulin resistance via IRF4. *Cell Metab.* **2017,** *25* (5), 1063-1074.

11. Inohara, N.; Ogura, Y.; Fontalba, A.; Gutierrez, O.; Pons, F.; Crespo, J.; Fukase, K.; Inamura, S.; Kusumoto, S.; Hashimoto, M.; Foster, S. J.; Moran, A. P.; Fernandez-Luna, J. L.; Nuñez, G., Host recognition of bacterial muramyl dipeptide mediated through NOD2. Implications for Crohn's disease. *J. Biol. Chem.* **2003,** *278* (8), 5509-5512.

12. Huang, Z.; Wang, J.; Xu, X.; Wang, H.; Qiao, Y.; Chu, W. C.; Xu, S.; Chai, L.; Cottier, F.; Pavelka, N.; Oosting, M.; Joosten, L. A. B.; Netea, M.; Ng, C. Y. L.; Leong, K. P.; Kundu, P.; Lam, K. P.; Pettersson, S.; Wang, Y., Antibody neutralization of microbiota-derived circulating peptidoglycan dampens inflammation and ameliorates autoimmunity. *Nat. Microbiol.* **2019,** *4* (5), 766-773.

13. Arentsen, T.; Qian, Y.; Gkotzis, S.; Femenia, T.; Wang, T.; Udekwu, K.; Forssberg, H.; Diaz Heijtz, R., The bacterial peptidoglycan-sensing molecule Pglyrp2 modulates brain development and behavior. *Mol. Psychiatry* **2017,** *22* (2), 257-266.

14. Bersch, K. L.; DeMeester, K. E.; Zagani, R.; Chen, S.; Wodzanowski, K. A.; Liu, S.; Mashayekh, S.; Reinecker, H. C.; Grimes, C. L., Bacterial peptidoglycan fragments differentially regulate innate immune signaling. *ACS Cent. Sci.* **2021,** *7* (4), 688-696.

15. Alvarez, L.; Hernandez, S. B.; de Pedro, M. A.; Cava, F., Ultra-sensitive, high-resolution liquid chromatography methods for the High-throughput quantitative analysis of bacterial cell wall chemistry and structure. *Methods Mol. Biol.* **2016,** *1440*, 11-27.

16. Kühner, D.; Stahl, M.; Demircioglu, D. D.; Bertsche, U., From cells to muropeptide structures in 24 h: peptidoglycan mapping by UPLC-MS. *Sci. Rep.* **2014,** *4*, 7494.

17. Glauner, B., Separation and quantification of muropeptides with high-performance liquid chromatography. *Anal. Biochem.* **1988,** *172* (2), 451-464.

18. Bern, M.; Beniston, R.; Mesnage, S., Towards an automated analysis of bacterial peptidoglycan structure. *Anal. Bioanal. Chem.* **2017,** *409* (2), 551-560.

19. Patel, A. V.; Turner, R. D.; Rifflet, A.; Acosta-Martin, A. E.; Nichols, A.; Awad, M. M.; Lyras, D.; Gomperts Boneca, I.; Bern, M.; Collins, M. O.; Mesnage, S., PGFinder, a novel analysis pipeline for the consistent, reproducible, and high-resolution structural analysis of bacterial peptidoglycans. *eLife* **2021,** *10*, e70597.

20. Pasquina-Lemonche, L.; Burns, J.; Turner, R. D.; Kumar, S.; Tank, R.; Mullin, N.; Wilson, J. S.; Chakrabarti, B.; Bullough, P. A.; Foster, S. J.; Hobbs, J. K., The architecture of the Gram-positive bacterial cell wall. *Nature* **2020,** *582* (7811), 294-297.

21. Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M.-Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P., A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **2012,** *30* (10), 918-920.

22. Bastos, P. A. D.; Wheeler, R.; Boneca, I. G., Uptake, recognition and responses to peptidoglycan in the mammalian host. *FEMS Microbiol. Rev.* **2020,** *45* (1), fuaa044.

23. Schleifer, K. H.; Kandler, O., Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol. Rev.* **1972,** *36* (4), 407-477.

24. Veerkamp, J. H., The structure of the cell wall peptidoglycan of *Bifidobacterium bifidum* var. *pennsylvanicus*. *Arch. Biochem. Biophys.* **1971,** *143* (1), 204-211.

25. Pyclik, M.; Srutkova, D.; Schwarzer, M.; Górska, S., Bifidobacteria cell wall-derived exo-polysaccharides, lipoteichoic acids, peptidoglycans, polar lipids and proteins - their chemical structure and biological attributes. *Int. J. Biol. Macromol.* **2020,** *147*, 333-349.

26. Pidgeon, S. E.; Apostolos, A. J.; Nelson, J. M.; Shaku, M.; Rimal, B.; Islam, M. N.; Crick, D. C.; Kim, S. J.; Pavelka, M. S.; Kana, B. D.; Pires, M. M., L,D-transpeptidase specific probe reveals spatial activity of peptidoglycan cross-linking. *ACS Chem. Biol.* **2019,** *14* (10), 2185-2196.

27. Weadge, J. T.; Clarke, A. J., Identification and characterization of *O*-acetylpeptidoglycan esterase: a novel enzyme discovered in *Neisseria gonorrhoeae*. *Biochemistry* **2006,** *45* (3), 839-851.

28. Garcia-Vello, P.; Tytgat, H. L. P.; Gray, J.; Elzinga, J.; Di Lorenzo, F.; Biboy, J.; Vollmer, D.; De Castro, C.; Vollmer, W.; de Vos, W. M.; Molinaro, A., Peptidoglycan from *Akkermansia muciniphila* MucT: chemical structure and immunostimulatory properties of muropeptides. *Glycobiology* **2022,** *32* (8), 712-719.

29. Moynihan, P. J.; Sychantha, D.; Clarke, A. J., Chemical biology of peptidoglycan acetylation and deacetylation. *Bioorg. Chem.* **2014,** *54*, 44-50.

30. Wang, G.; Olczak, A.; Forsberg, L. S.; Maier, R. J., Oxidative stress-induced peptidoglycan deacetylase in *Helicobacter pylori*. *J. Biol. Chem.* **2009,** *284* (11), 6790-6800.

31. Wang, G.; Maier, S. E.; Lo, L. F.; Maier, G.; Dosi, S.; Maier, R. J., Peptidoglycan deacetylation in *Helicobacter pylori* contributes to bacterial survival by mitigating host

immune responses. *Infect. Immun.* **2010,** *78* (11), 4660-4666.

32. Davis, K. M.; Weiser, J. N., Modifications to the peptidoglycan backbone help bacteria to establish infection. *Infect. Immun.* **2011,** *79* (2), 562-570.

33. Sychantha, D.; Brott, A. S.; Jones, C. S.; Clarke, A. J., Mechanistic pathways for peptidoglycan *O*-acetylation and de-*O*-acetylation. *Front. Microbiol.* **2018,** *9*, 2332.

34. Srisuknimit, V.; Qiao, Y.; Schaefer, K.; Kahne, D.; Walker, S., Peptidoglycan cross-linking preferences of *Staphylococcus aureus* penicillin-binding proteins have implications for treating MRSA Infections. *J. Am. Chem. Soc.* **2017,** *139* (29), 9791-9794.

35. Peltier, J.; Courtin, P.; El Meouche, I.; Lemée, L.; Chapot-Chartier, M. P.; Pons, J. L., *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), 29053-29062.

36. Aliashkevich, A.; Cava, F., LD-transpeptidases: the great unknown among the peptidoglycan cross-linkers. *FEBS J.* **2022,** *289* (16), 4718-4730.

37. Mainardi, J. L.; Morel, V.; Fourgeaud, M.; Cremniter, J.; Blanot, D.; Legrand, R.; Frehel, C.; Arthur, M.; Van Heijenoort, J.; Gutmann, L., Balance between two transpeptidation mechanisms determines the expression of β-lactam resistance in *Enterococcus faecium*. *J. Biol. Chem.* **2002,** *277* (39), 35801-35807.

38. Lecoq, L.; Dubée, V.; Triboulet, S.; Bougault, C.; Hugonnet, J. E.; Arthur, M.; Simorre, J. P., Structure of Enterococcus faecium l,d-transpeptidases acylated by ertapenem provides insight into the inactivation mechanism. *ACS Chem. Biol.* **2013,** *8* (6), 1140-1146.

39. Magnet, S.; Arbeloa, A.; Mainardi, J. L.; Hugonnet, J. E.; Fourgeaud, M.; Dubost, L.;

Marie, A.; Delfosse, V.; Mayer, C.; Rice, L. B.; Arthur, M., Specificity of l,d-transpeptidases from Gram-positive bacteria producing different peptidoglycan chemotypes. *J. Biol. Chem.* **2007,** *282* (18), 13151-13159.

40. Sütterlin, L.; Edoo, Z.; Hugonnet, J. E.; Mainardi, J. L.; Arthur, M., Peptidoglycan cross-linking activity of l,d-transpeptidases from *Clostridium difficile* and inactivation of these enzymes by β-lactams. *Antimicrob. Agents Chemother.* **2018,** *62* (1), e01607-01617.

41. Magnet, S.; Dubost, L.; Marie, A.; Arthur, M.; Gutmann, L., Identification of the l,d-transpeptidases for peptidoglycan cross-linking in *Escherichia coli*. *J. Bacteriol.* **2008,** *190* (13), 4782-4785.

42. Lavollay, M.; Arthur, M.; Fourgeaud, M.; Dubost, L.; Marie, A.; Veziris, N.; Blanot, D.; Gutmann, L.; Mainardi, J. L., The peptidoglycan of stationary-phase *Mycobacterium tuberculosis* predominantly contains cross-links generated by l,d-transpeptidation. *J. Bacteriol.* **2008,** *190* (12), 4360-4366.

43. Jeon, S.; Kim, H.; Choi, Y.; Cho, S.; Seo, M.; Kim, H., Complete genome sequence of the newly developed *Lactobacillus acidophilus* strain with improved thermal adaptability. *Front. Microbiol.* **2021,** *12*, 697351.

44. Morita, H.; Toh, H.; Oshima, K.; Nakano, A.; Omori, E.; Hattori, Y.; Arakawa, K.; Suda, W.; Honda, K.; Hattori, M., Complete genome sequence of *Bifidobacterium breve* JCM 1192(T) isolated from infant feces. *J. Biotechnol.* **2015,** *210*, 81-82.

45. Poyet, M.; Groussin, M.; Gibbons, S. M.; Avila-Pacheco, J.; Jiang, X.; Kearney, S. M.; Perrotta, A. R.; Berdy, B.; Zhao, S.; Lieberman, T. D.; Swanson, P. K.; Smith, M.; Roesemann, S.; Alexander, J. E.; Rich, S. A.; Livny, J.; Vlamakis, H.; Clish, C.; Bullock, K.; Deik, A.; Scott, J.; Pierce, K. A.; Xavier, R. J.; Alm, E. J., A library of

human gut bacterial isolates paired with longitudinal multiomics data enables mechanistic microbiome research. *Nat. Med.* **2019,** *25* (9), 1442-1452.

46. Ottman, N. A. Host immunostimulation and substrate utilization of the gut symbiont *Akkermansia muciniphila*. PhD thesis, Wageningen University, Wageningen, 2015.

47. Auer, G. K.; Weibel, D. B., Bacterial cell mechanics. *Biochemistry* **2017,** *56* (29), 3710-3724.

48. Chang, J. D.; Wallace, A. G.; Foster, E. E.; Kim, S. J., Peptidoglycan compositional analysis of *Enterococcus faecalis* biofilm by stable isotope labeling by amino acids in a bacterial culture. *Biochemistry* **2018,** *57* (7), 1274-1283.

49. Loskill, P.; Pereira, P. M.; Jung, P.; Bischoff, M.; Herrmann, M.; Pinho, M. G.; Jacobs, K., Reduction of the peptidoglycan crosslinking causes a decrease in stiffness of the *Staphylococcus aureus* cell envelope. *Biophys. J.* **2014,** *107* (5), 1082-1089.

50. Kim, S. J.; Chang, J.; Singh, M., Peptidoglycan architecture of Gram-positive bacteria by solid-state NMR. *Biochim Biophys Acta Biomembr* **2015,** *1848* (1 Pt B), 350-362.

51. de Pedro, M. A.; Cava, F., Structural constraints and dynamics of bacterial cell wall architecture. *Front. Microbiol.* **2015,** *6*, 449.

52. Vollmer, W.; Seligman, S. J., Architecture of peptidoglycan: more data and more models. *Trends Microbiol.* **2010,** *18* (2), 59-66.

53. Turner, R. D.; Vollmer, W.; Foster, S. J., Different walls for rods and balls: the diversity of peptidoglycan. *Mol. Microbiol.* **2014,** *91* (5), 862-874.

54. Mathelié-Guinlet, M.; Asmar, A. T.; Collet, J.-F.; Dufrêne, Y. F., Lipoprotein Lpp regulates the mechanical properties of the *E. coli* cell envelope. *Nat. Commun.* **2020,** *11* (1), 1789.

55. Auer, G. K.; Lee, T. K.; Rajendram, M.; Cesar, S.; Miguel, A.; Huang, K. C.; Weibel, D. B., Mechanical genomics identifies diverse modulators of bacterial cell stiffness. *Cell Syst.* **2016,** *2* (6), 402-411.

56. Santa Maria Jr., J. P.; Sadaka, A.; Moussa, S. H.; Brown, S.; Zhang, Y. J.; Rubin, E. J.; Gilmore, M. S.; Walker, S., Compound-gene interaction mapping reveals distinct roles for *Staphylococcus aureus* teichoic acids. *Proc. Natl. Acad. Sci. U.S.A.* **2014,** *111* (34), 12510-12515.

57. Saar-Dover, R.; Bitler, A.; Nezer, R.; Shmuel-Galia, L.; Firon, A.; Shimoni, E.; Trieu-Cuot, P.; Shai, Y., D-alanylation of lipoteichoic acids confers resistance to cationic peptides in group B *streptococcus* by increasing the cell wall density. *PLoS Pathog.* **2012,** *8* (9), e1002891.

58. Rimal, B.; Senzani, S.; Ealand, C.; Lamichhane, G.; Kana, B.; Kim, S. J., Peptidoglycan compositional analysis of *Mycobacterium smegmatis* using high-resolution LC-MS. *Sci. Rep.* **2022,** *12* (1), 11061.

59. Liu, G.; Cheng, K.; Lo, C. Y.; Li, J.; Qu, J.; Neelamegham, S., A comprehensive, open-source platform for mass spectrometry-based glycoproteomics data analysis. *Mol. Cell Proteomics.* **2017,** *16* (11), 2032-2047.

60. Greene, N. G.; Narciso, A. R.; Filipe, S. R.; Camilli, A., Peptidoglycan branched stem peptides contribute to *Streptococcus pneumoniae* virulence by inhibiting pneumolysin release. *PLoS Pathog.* **2015,** *11* (6), e1004996.

61. Biswas, R.; Voggu, L.; Simon, U. K.; Hentschel, P.; Thumm, G.; Götz, F., Activity of the major staphylococcal autolysin Atl. *FEMS Microbiol. Lett.* **2006,** *259* (2), 260-268.

62. Reiding, K. R.; Bondt, A.; Franc, V.; Heck, A. J. R., The benefits of hybrid

fragmentation methods for glycoproteomics. *Trends Analyt Chem* **2018,** *108*, 260-268.

ACKNOWLEDGMENTS

This research was supported by Ministry of Science and Technology (MOST), R.O.C. (Grants MOST 108-2636-M-002-008-, 109-2636-M-002-005-, and 110-2636-M-002-014-). Y.-C. H. acknowledged the financial support of "The Program of Research Performance Enhancement via Students Entering Ph.D. Programs Straight from an Undergraduate/Master's Program" from National Taiwan University. The instrument support from the NTU Mass Spectrometry Platform was acknowledged.

AUTHOR CONTRIBUTIONS

Y.-C.H., P.-R.S., and C.-C.H. designed the experiments. P.-R.S. and Y.-C.H. built the HAMA software. Y.-C.H. carried out all the MS experiments and analyzed the data. L.-J.H., K.-Y.C., and C.-h.C. designed and carried out the AFM experiments. C.-C.H. supervised the study. Y.-C.H., P.-R.S. and C.-C.H. wrote the paper.

AUTHOR INFORMATION

Corresponding Author

* Cheng-Chih Hsu: ccrhsu@ntu.edu.tw

ORCID

Ya-Chen Hsu: 0000-0003-1202-0132

Chun-hsien Chen: 0000-0001-5507-3248

Cheng-Chih Hsu: 0000-0002-2892-5326

COMPETING INTERESTS

The authors declare no competing financial interest.

FIGURES

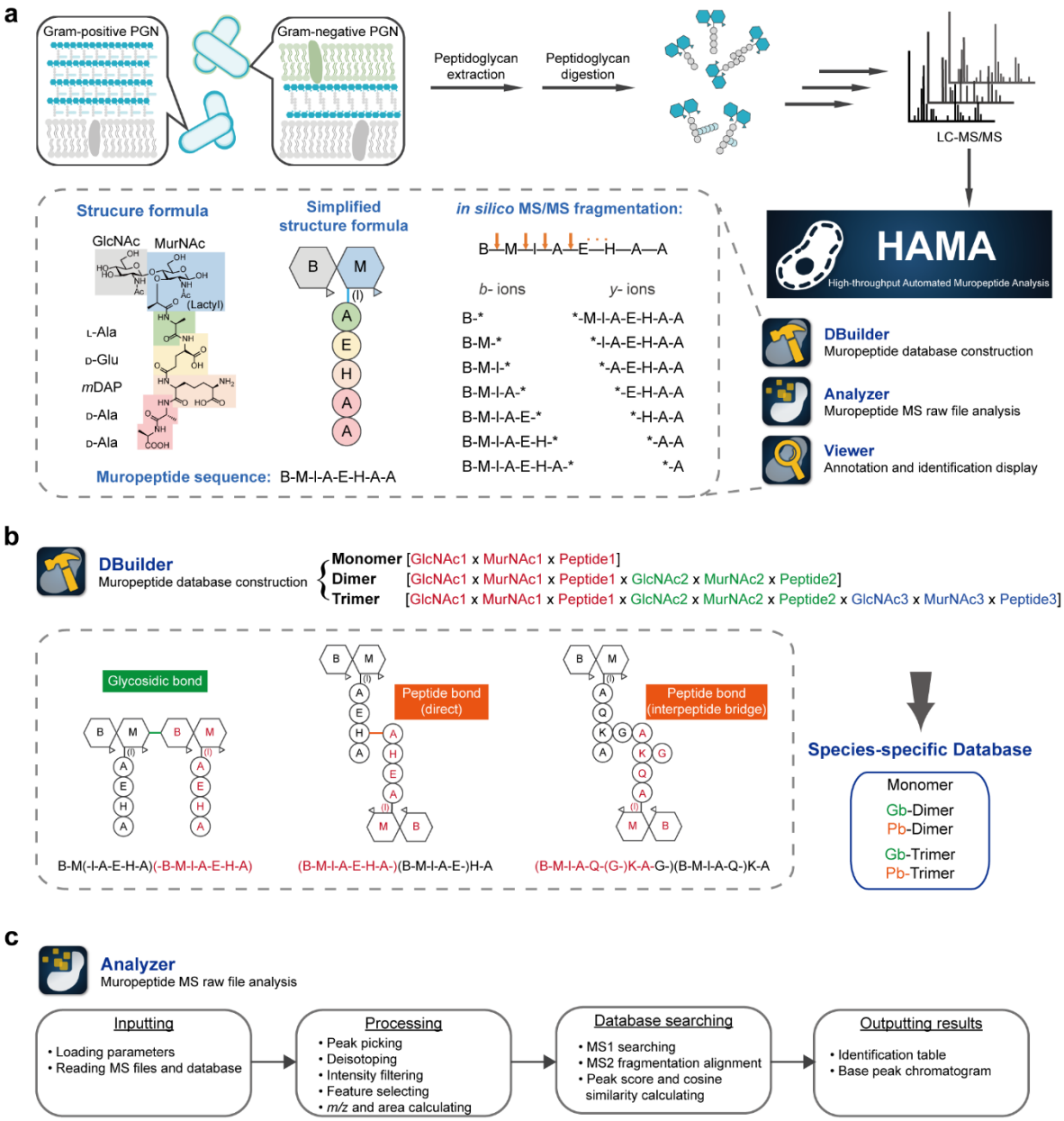

Figure 1. Schematic representation of the high-throughput automated muropeptide analysis (HAMA) framework. (a) The peptidoglycans of bacteria were extracted and purified, followed by mutanolysin digestion. The resulting muropeptide products are analyzed by UHPLC-MS/MS and identified using the HAMA platform. The HAMA strategy involves simplifying muropeptide structures to sequence format, which facilitates the database

construction and *in silico* generation of *b*- and *y*- ion fragmentation spectra for matching. Muropeptide symbols: B, *N*-acetylglucosamine; M, *N*-acetylmuraminitol (without lactyl group); l, lactic acid; A, alanine; E, glutamic acid; H, diaminopimelic acid. (b) *DBuilder* constructs a muropeptide database containing monomers, dimers, and trimers with two types of linkage: glycosidic bonds (Gb) and peptide bonds (Pb). For peptide linkages, the direct way is through a direct covalent bond between the penultimate D-Ala of a donor stem and *m*DAP residue in the neighboring acceptor stem, and the indirect way is via an interpeptide bridge branching from the lysine. Donor peptides are labeled in red, and acceptor peptides are labeled in black. (c) The flowchart outlines the LC-MS data processing in *Analyzer*.

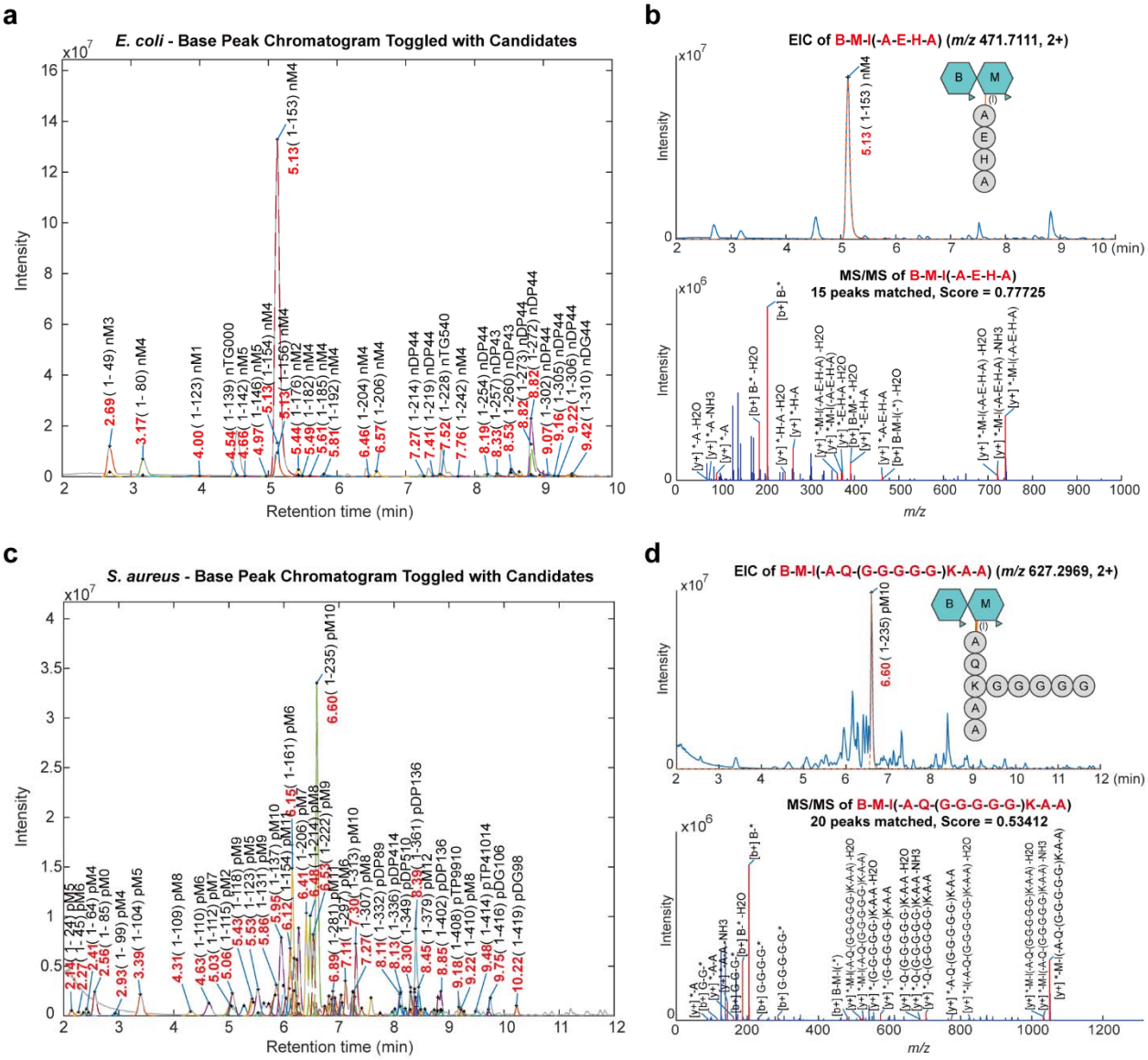

Figure 2. Automated identification of well-characterized peptidoglycans from *E. coli* and *S. aureus* using the HAMA platform. (a, c) Base peak chromatograms showing the muropeptide analysis of *E. coli* and *S. aureus*. The label content includes retention time (in red), feature index, and muropeptide class (in black). (b, d) Extracted ion chromatograms of the most abundant muropeptide and their MS/MS spectra annotated with *b*- and *y*- fragments were visualized in *Viewer*. Muropeptide symbols: B, *N*-acetylglucosamine; M, *N*-acetylmuraminitol (without lactyl group); l, lactic acid; A, alanine; E, glutamic acid; Q, glutamine; K, lysine; G, glycine.

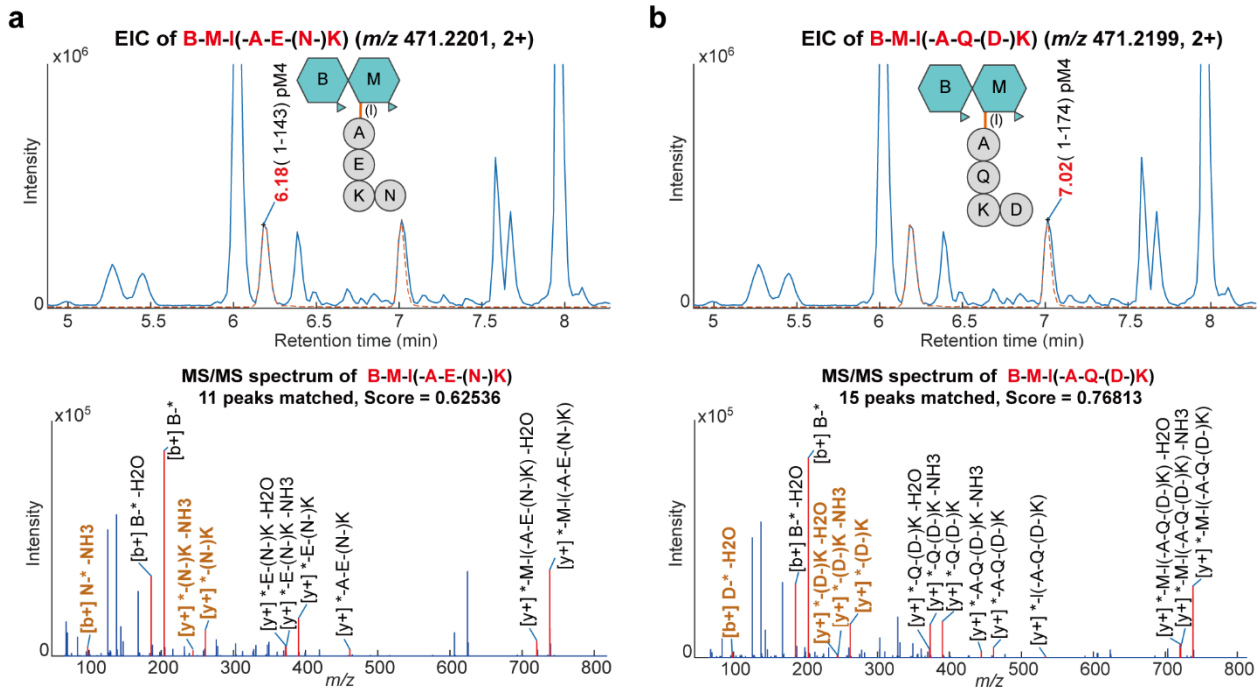

Figure 3. Resolving isomeric muropeptides by *in silico* MS/MS fragmentation matching. Two isomeric muropeptides with the same parent ion, *m/z* 471.22, were identified as two disaccharide-tripeptides: (a) B-M-l(-A-E-(N-)K eluted at 6.18 min, and (b) B-M-l(-A-Q-(D-)K eluted at 7.02 min. The sequence of each isomer was determined using *in silico* MS/MS fragmentation matching, with the identified sequence having the highest matching score. The key MS/MS fragments that discriminate between two isomers are labeled in bold brown. Muropeptide symbols: B, *N*-acetylglucosamine; M, *N*-acetylmuraminitol (without lactyl group); l, lactic acid; A, alanine; E, glutamic acid; Q, glutamine; K, lysine; N, Asparagine; D, Aspartic acid.

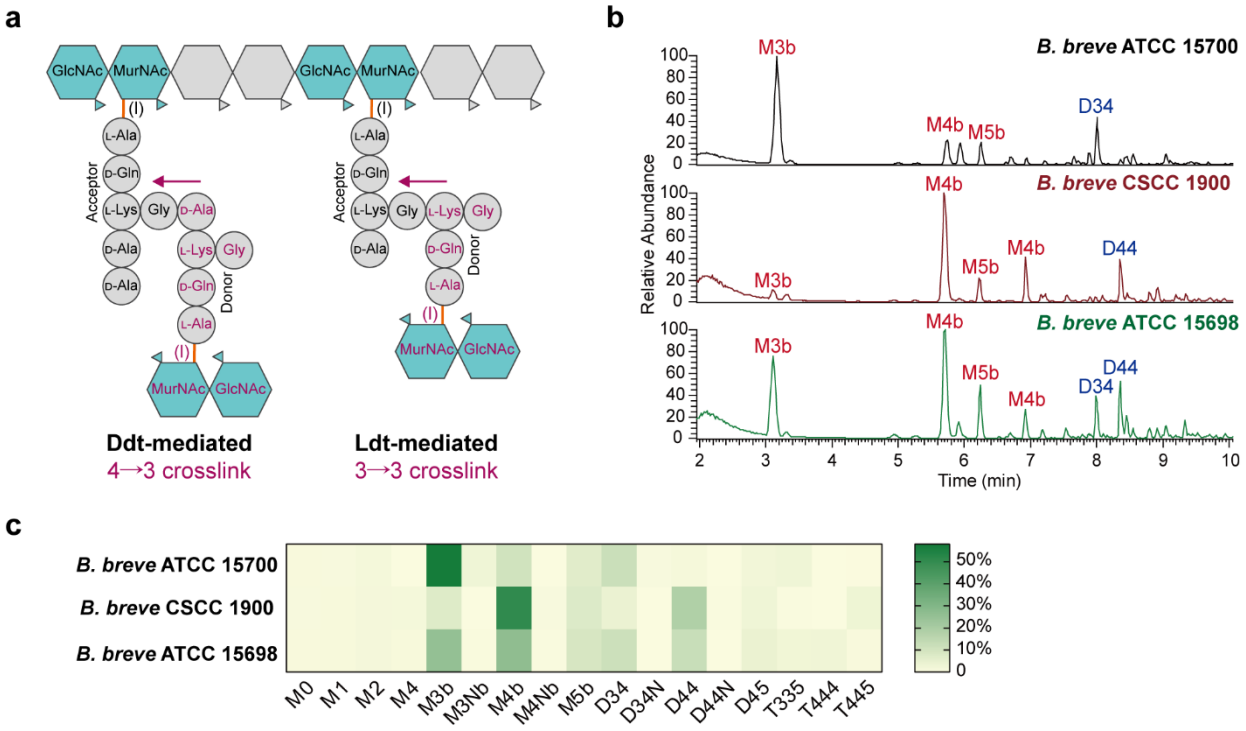

Figure 4. Muropeptide composition analysis of *Bifidobacterium breve* strains. (a) Schematic representation of the two possible cross-linking types in the PGN of *B. breve*: Ddt-mediated 4-3 cross-link and Ldt-mediated 3-3 cross-link. Donor peptide stems are labeled in red. The arrow indicates the direction of cross-links catalyzed by transpeptidases. (b) Base peak chromatograms of muropeptide analysis of *B. breve* ATCC 15700, CSCC 1900, and ATCC 15698 strains. The main peaks were annotated with muropeptide symbols. (c) Heatmap showing the muropeptide compositions (% of total) of the PGN of three *B. breve* strains. Symbols: M, monomer; D, dimer; T, trimer (numbers following the letters indicate the number of amino acids in stem peptides). M0, disaccharide; M1, disaccharide-monopeptide; M2, disaccharide-dipeptide; M4, disaccharide-tetrapeptide; M3b, disaccharide-tripeptide with an interpeptide bridge; M3Nb, disaccharide-tripeptide with an anhydro-MurNAc and an interpeptide bridge; M4b, disaccharide-tetrapeptide with an interpeptide bridge; M4Nb, disaccharide-tetrapeptide with an anhydro-MurNAc and an interpeptide bridge; M5b, disaccharide-pentapeptide with an interpeptide bridge;

D34, disaccharide-tripeptide–disaccharide-tetrapeptide with a peptide cross-link; D34N, disaccharide-tripeptide–disaccharide-tetrapeptide with a peptide cross-link and an anhydro-MurNAc; D44, disaccharide-tetrapeptide–disaccharide-tetrapeptide with a peptide cross-link; D44N, disaccharide-tetrapeptide–disaccharide-tetrapeptide with a peptide cross-link and an anhydro-MurNAc; D45, disaccharide-tetrapeptide–disaccharide-pentapeptide with a peptide cross-link; T335, disaccharide-tripeptide–disaccharide-tripeptide-disaccharide-pentapeptide with two peptide cross-links; T444, disaccharide-tetrapeptide–disaccharide-tetrapeptide-disaccharide-tetratapeptide with two peptide cross-links; T445, disaccharide-tetrapeptide–disaccharide-tetrapeptide-disaccharide-pentapeptide with two peptide cross-links.

Figure 4 - Source data 1

Heatmap data of Figure 4c.

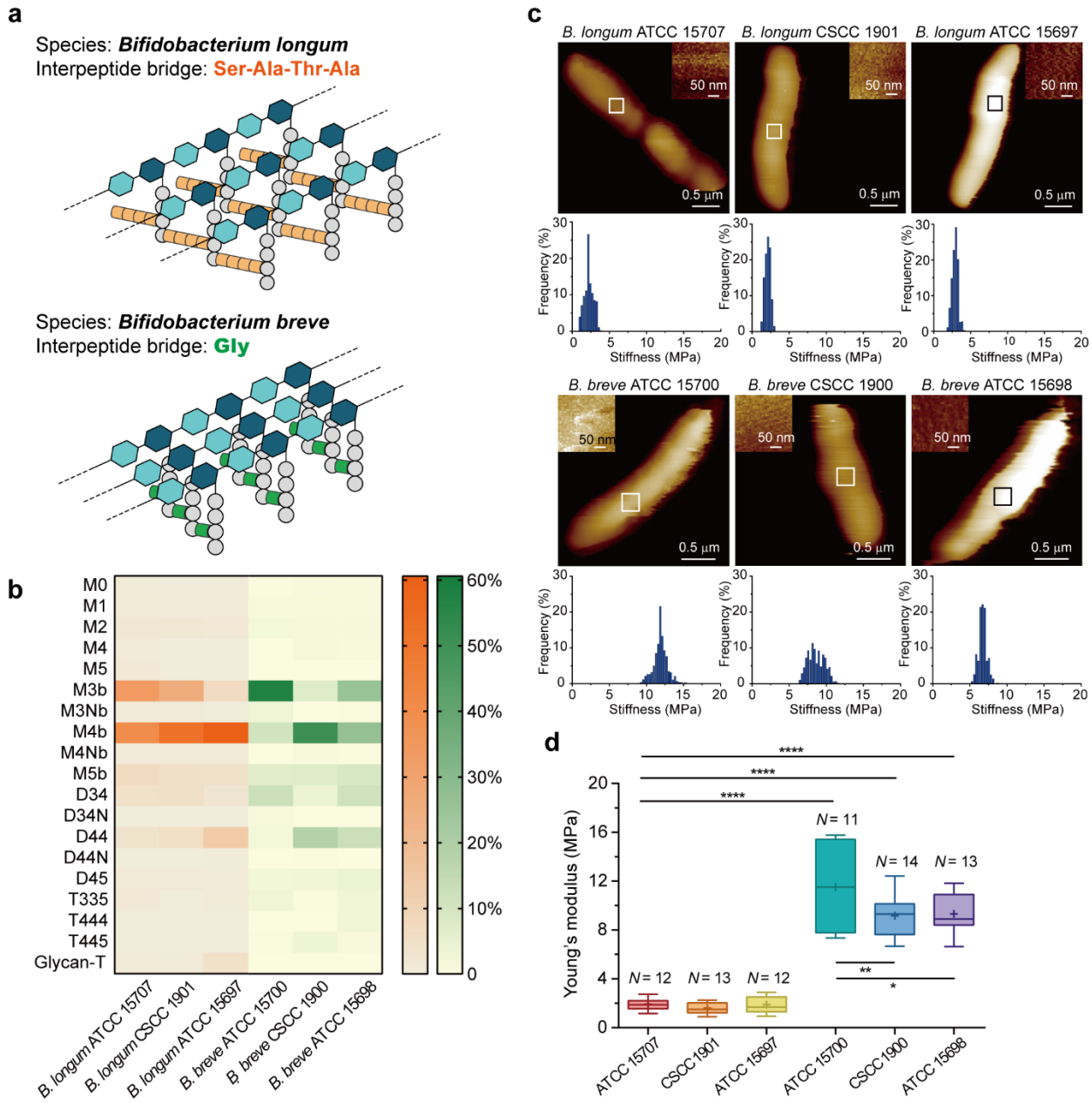

Figure 5. Bridge length-dependent cell envelope stiffness in *B. longum* and *B. breve*. (a) Schematic illustration of the Gram-positive PGN architecture of *B. longum* (orange tetrapeptide bridges) and *B. breve* (green monopeptide bridges). Glycan strands composed of repeating units of the β-1,4-linked disaccharides are cross-linked by interpeptide bridges, forming 3-D peptidoglycan layers. (b) Heatmap displaying the muropeptide compositions (% of total) of the PGNs in *B. longum* and *B. breve* (three strains each). Symbols: M, monomer; D, dimer; T, trimer

(numbers indicate amino acids in stem peptides). Description of symbol abbreviations as in Figure legend 4, with the addition of Glycan-T representing trimers linked by glycosidic bonds. (c) AFM imaging of living *Bifidobacterium*. Topographical images in PBS buffer and the inset shows elasticity images from the top of the cell. Distribution of Young's modulus values corresponding to the elasticity images in the inset. (d) Statistical analysis was performed for each strain, showing the distribution of two groups with the shorter interpeptide bridge corresponded to higher stiffness of cell envelope. Shown here are the mean values (cross), the median, and the 25 and 75% quartiles (boxes) obtained from *N* independent cells over at least three independent experiments. P values were calculated using a one-way ANOVA analysis. *$P \leq 0.05$, **$P \leq 0.01$, and ****$P \leq 0.0001$.

Figure 5 - Source data 1

Heatmap data of Figure 5b and AFM data of Figure 5c and 5d.

TABLES

Table 1. Muropeptides of *Escherichia coli DH5α* / *Staphylococcus aureus* SA113 analyzed by UPLC-MS/MS.

	Muropeptide sequence	Retention time (min)	Class[a]	Observed MW (Da)	Theoretical MW (Da)	Δppm	Stem peptides
Escherichia coli	B-M-l(-A-E-H)	2.69	nM3	870.3711	870.3706	-0.60	Tri
	B-M-l(-A-E-H-G)	3.17	nM4	927.3918	927.3921	0.25	Tetra-Gly(4)
	B-M-l(-A-E-H-A)*	**5.13**	**nM4**	**941.4076**	**941.4077**	**0.14**	**Tetra**
	M-l(-A-E-H-A)[b]	5.13	nM4	738.3277	738.3283	0.91	Tetra
	(B-M-l-A-E-H-A-)(B-M-l-A-E-)H	8.53	nDP43	1793.7739	1793.7677	-3.43	Tri-Tetra
	(B-M-l-A-E-H-A-)(B-M-l-A-E-)H-A*	**8.82**	**nDP44**	**1864.8074**	**1864.8048**	**-1.39**	**Tetra-Tetra**
Staphylococcus aureus	B-M-l(-A-Q-K-A-A)	3.39	pM5	967.4721	967.4710	-1.20	Penta
	B-M-l(-A-Q-(G-G-G-G-G-G-)K-A)	5.95	pM10	1238.5641	1238.5626	-1.14	Tetra
	B-M-l(-A-Q-(G-)K-A-A)*	**6.15**	**pM6**	**1024.4925**	**1024.4924**	**-0.02**	**Penta**
	B-M-l(-A-Q-(G-G-G-G-G-G-G-G-)K-A)	6.18	pM12	1352.6041	1352.6056	1.12	Tetra
	B-M-l(-A-Q-(G-G-)K-A-A)	6.41	pM7	1081.5163	1081.5139	-2.26	Penta
	B-M-l(-A-Q-(G-G-G-)K-A-A)	6.48	pM8	1138.5362	1138.5354	-0.75	Penta
	B-M-l(-A-Q-(G-G-G-G-)K-A-A)	6.53	pM9	1195.5577	1195.5568	-0.71	Penta
	B-M-l(-A-Q-(G-G-G-G-G-)K-A-A)*	**6.60**	**pM10**	**1252.5806**	**1252.5783**	**-1.86**	**Penta**
	B-M-l(-A-E-(G-G-G-G-G-)K-A-A)	7.30	pM10	1253.5648	1253.5623	-2.02	Penta
	(B-M-l-A-Q-(G-G-G-G-G-)K-A-G-G-G-G-G-)(B-M-l-A-Q-)K-A-A*	**8.39**	**pDP136**	**2416.1030**	**2416.1089**	**2.44**	**Tetra-Penta**

[a] Muropeptides are cataloged by classes of Gram-negative monomer, nM, Gram-positive monomer, pM, Gram-negative dimer (peptide bond), nDP, Gram-positive dimer (peptide bond), pDP, Gram-negative dimer (glycosidic bond), nDG, Gram-positive dimer (glycosidic bond), pDG, and so on. Numbers following the letters are the total number of amino acids in each peptide strand.

[b] Monomer formed in the electrospray source by loss of monosaccharides and recognized as an in-source fragment by exact co-elution.

[*] The main products from PGN hydrolysis were labeled with bold font.

Table 1 - Source data 1

Raw output table of the whole muropeptide identification of *E. coli*.

Table 1 - Source data 2

Raw output table of the whole muropeptide identification of *S. aureus*.

Table 2. The characterized peptidoglycan types of gut bacteria used in this study.

		Glycan backbone		Stem peptide amino acids					
Species	**Strains**	**GlcNAc**	**MurNAc**	**1**	**2**	**3**	**4**	**5**	**Interpeptide bridge** (Acceptor ← Donor)
Bifidobacterium bifidum	DSM 20082	+	+	L-Ala	D-iGln/Glu	L-Orn	D-Ala	D-Ala	D-Ser-D-Asp
Bifidobacterium breve	ATCC 15700 infant intestine	+	+	L-Ala	D-iGln/Glu	L-Lys	D-Ala	D-Ala	Gly
	CSCC 1900 nursling stools	+	+	L-Ala	D-iGln/Glu	L-Lys	D-Ala	D-Ala	Gly
	ATCC 15698 infant intestine	+	+	L-Ala	D-iGln/Glu	L-Lys	D-Ala	D-Ala	Gly
Bifidobacterium longum	ATCC 15707 adult intestine	+	+	L-Ala	D-iGln/Glu	L-Orn	D-Ala	D-Ala	L-Ser-L-Ala-L-Thr-L-Ala
	CSCC 1901 nursling stools	+	+	L-Ala	D-iGln/Glu	L-Orn/	D-Ala	D-Ala	L-Ser-L-Ala-L-Thr-L-Ala
	ATCC 15697 infant intestine	+	+	L-Ala	D-iGln/Glu	L-Orn	D-Ala	D-Ala	L-Ser-L-Ala-L-Thr-L-Ala
Lactobacillus acidophilus	ATCC 4356	+	+	L-Ala	D-iGln/Glu	L-Lys	D-Ala	D-Ala	D-iAsx (Asn+Asp)
Enterococcus faecalis	ATCC 19433	+	+	L-Ala	D-iGln/Glu	L-Lys	D-Ala	D-Ala	L-Ala- L-Ala
Enterococcus faecium	ATCC 19434	+	+	L-Ala	D-iGln/Glu	L-Lys	D-Ala	D-Ala	D-iAsx (Asn+Asp)
Bacteroides fragilis	ATCC 25285	+	*O*-acetylation	L-Ala	D-iGlu	*m*DAP	D-Ala	D-Ala/ Gly	-
Bacteroides ovatus	ATCC 8483	+	*O*-acetylation	L-Ala	D-iGlu	*m*DAP	D-Ala	D-Ala/ Gly	-
Bacteroides thetaiotaomicron	ATCC 29741	+	*O*-acetylation	L-Ala	D-iGlu	*m*DAP	D-Ala	D-Ala/ Gly	-
Akkermansia muciniphila	ATCC BAA-835	*N*-deacetylation	+	L-Ala	D-iGlu	*m*DAP	D-Ala	D-Ala/ Gly	-

ADDITIONAL FILES

Supplementary File

Supplementary file includes a list of bacteria strains, supplementary figures (MS/MS spectra), and supplementary tables (lists of automated identification of gut bacterial muropeptides).

The HAMA software package and *Analyzer* output files are available at: https://drive.google.com/drive/folders/17LVGOm-LEHzNmw7nPULXMEZvg_cnO3k1?usp=share_link