

Chemometric Analysis of Bacterial Peptidoglycan Reveals Atypical Modifications which Empower the Cell Wall against Predatory Enzymes and Fly Innate Immunity

Akbar Espaillat[†], Oskar Forsmo[†], Khouzaima El Biari[§], Rafael Björk[‡], Bruno Lemaitre^{||}, Johan Trygg[‡], Francisco Javier Cañada[§], Miguel A. de Pedro[⊥], and Felipe Cava^{†*}

[†] Laboratory for Molecular Infection Medicine Sweden, Department of Molecular Biology, Umeå Centre for Microbial Research, Umeå University, 90187 Umeå, Sweden.

[§] Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Ramiro de Maeztu 9, 28040 Madrid, Spain.

[‡] Department of Chemistry. Umeå University, 90187 Umeå, Sweden.

^{||} Global Health Institute, Swiss Federal Institute of Technology, Station 19, CH-1015 Lausanne, Switzerland.

[⊥] Centro de Biología Molecular “Severo Ochoa”, Universidad Autónoma de Madrid-Consejo Superior de Investigaciones Científicas, Madrid 28049. Spain.

*To whom correspondence should be addressed.

E-mail: felipe.cava@molbiol.umu.se

Abstract

Peptidoglycan is a fundamental structure for most bacteria. It contributes to the cell morphology and provides cell wall integrity against environmental insults. While several studies have reported a significant degree of variability in the chemical composition and organization of peptidoglycan in the domain Bacteria, the real diversity of this polymer is far from fully explored. This work exploits rapid ultra-performance liquid chromatography and multivariate data analysis to uncover peptidoglycan chemical diversity in the Class Alphaproteobacteria, a group of Gram negative bacteria which are highly heterogeneous in terms of metabolism, morphology and life-styles. Indeed, chemometric analyses revealed novel peptidoglycan structures conserved in Acetobacteria: amidation at the α -(L)-carboxyl of meso-diaminopimelic acid and the presence of mucopeptides cross-linked by (1-3) L-Ala-D-(meso)-diaminopimelate cross-links. Both structures are growth-controlled modifications which influence sensitivity to Type VI secretion system peptidoglycan endopeptidases and recognition by the *Drosophila* innate immune system, suggesting relevant roles in the environmental adaptability of these bacteria. Collectively our findings demonstrate the discriminative power of chemometric tools on large cell wall-chromatographic datasets to discover novel peptidoglycan structural properties in bacteria.

Introduction

One of the defining characteristics of bacteria is the presence of a peptidoglycan layer as a critical component of the cell envelope¹. In Gram negative bacteria, a thin peptidoglycan layer (sacculus) lies in the periplasm between the cytoplasmic and outer membranes (Figure 1A). In contrast, Gram positive bacteria contain one thick, multi-layered peptidoglycan which envelopes the cytoplasmic membrane. The canonical monomeric subunit consists of the disaccharide pentapeptide *N*-acetylglucosamine $\beta(1\rightarrow4)$ *N*-acetylmuramic acid-L-Ala-D-Glu-(γ)-(di-amino acid)-D-Ala-D-Ala, where meso-diaminopimelic acid and L-lysine are the more frequent di-amino acids of Gram negative and Gram positive bacteria, respectively¹. Monomers are converted into linear polymers by means of $\beta(1\rightarrow4)$ glycosidic bonds, and then linear polymers are covalently linked by means of peptide bridges between the peptide moieties. The final result is a net-like macromolecule which encloses the cell body. Further metabolic activities result in a series of modifications in the chemical nature of peptidoglycan subunits, and on the relative proportions of the different subunits¹. The peptidoglycan sacculus plays crucial roles defining cell shape and preserving bacterial cell integrity. Its crucial role becomes apparent when peptidoglycan synthesis is inhibited by antibiotics (e.g. penicillins), which cause morphological alterations, cell lysis and death².

Since the cell wall is a barrier between the environment and the cell, it seems quite rational to expect both adaptive and evolutionary variability in this structure amongst the many existing bacterial species. In fact, even the low resolution techniques (e.g. thin layer chromatography³) available in the early days of cell wall research pointed out substantial peptidoglycan variability. Introduction of high resolution techniques such as High-Performance Liquid Chromatography (HPLC) and its recently improved “Ultra-Performance” version (UPLC)⁴, revealed new levels

of complexity demonstrating the dynamic nature of peptidoglycan structure and its adaptive growth-phase dependent peptidoglycan plasticity^{5,6}. Uncovering the breadth of peptidoglycan structural variability will lead to a better understanding of cell wall biology in nature, in particular its role in environmental adaptation and signaling. However, an in depth exploration of peptidoglycan variability requires the analysis of large numbers of samples and the application of chemometric tools to both, identify novel peptidoglycan traits and define clusters of organisms sharing these features. In other words, an “omic” approach conceptually similar to proteomics or metabolomics.

Cell wall structural variability might be an important element in the bacterial strategy to adapt to hostile environments, as are most natural habitats. For instance, many Gram negative bacteria have devised mechanisms to outcompete cohabitating species by releasing⁷ or injecting, via type VI secretion systems (T6SS)⁸, peptidoglycan-hydrolases. It has been suggested that some bacteria might escape by modifying the structure or composition of their peptidoglycan target^{8a}, i.e. chemical modifications⁹ and changes to the crosslink positions¹⁰. In higher organisms, peptidoglycan recognition proteins of the innate immune system provide an antibacterial defense mechanism targeting the cell wall. Because these proteins recognize highly conserved motifs in the peptidoglycan, some bacteria avoid recognition through specific modifications such as peptidoglycan deacetylation which mitigates host immune detection and thus favors pathogen persistence¹¹.

Here we report the development and application of chemometric tools to the analysis of peptidoglycan in representative species of the Gram negative bacterial class Alphaproteobacteria, one of the 6 classes which make up the phylum Proteobacteria¹². Alphaproteobacteria are appropriate sources for cell wall studies aiming at finding new structural variability because of

1
2
3 their high diversity in terms of metabolism, morphology, environmental niches and life-styles
4
5 spanning from free-living bacteria to symbionts and intracellular pathogens^{13,14}. The general
6
7 suitability of this method offers a rapid pipeline to uncover novel peptidoglycan traits in large
8
9 datasets. Multivariate data analysis of peptidoglycan UPLC spectra revealed three clusters, one
10
11 of them exclusively comprising members of the family Acetobacteraceae, a family in the
12
13 Alphaproteobacterial order Rhodospirillales characterized by the ability of its members to
14
15 produce acetic acid and grow at low pH¹⁵. The cell wall of these bacteria is characterized by the
16
17 presence of previously undescribed muropeptides amidated at the α -(L)-carboxyl group of meso-
18
19 diaminopimelic acid (DAP), and cross-linked through LD (1-3) peptide bridges. Functional
20
21 studies showed that these new muropeptides are poor substrates for Type VI secretion system
22
23 (T6SS) endopeptidases from the potentially co-inhabitant bacteria *Pseudomonas aeruginosa* and
24
25 *Acidovorax citrullii*. Furthermore, amidation at the α -carboxyl of DAP reduces the ability of
26
27 peptidoglycan to elicit the innate immune response in the vinegar fly *Drosophila melanogaster*, a
28
29 natural host for Acetobacteria.
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Results

Multivariate statistical analysis of peptidoglycan composition reveals three clusters in Alphaproteobacteria

As peptidoglycan chemical structure is inferred from mucopeptide chromatographic profiles, a clustering analysis was expected to reveal the subjacent compositional variability in large numbers of datasets. Therefore, we selected 20 different species representing five (Caulobacterales, Sphingomonadales, Rhodospirillales, Rhodobacterales and Rhizobiales) of the six different Orders of the Class Alphaproteobacteria (Table S1). We did not include species from Order Rickettsiales due to their growth requirements. Peptidoglycan mucopeptide composition from stationary phase cultures was analyzed by means of ultra-performance liquid chromatography (UPLC) (Figure S1) and its variability was assessed by hierarchical cluster analysis (HCA) ¹⁶ using principal component analysis (PCA). Both, clustering and interrelations were calculated according to the presence (or absence) and relative abundances of “peaks”, in the UV₂₀₄/retention time (t_R) records. For simplicity we used a t_R -window between 4 and 14.5 min, which includes all major subunit species, in particular the variants of the monomeric and the most abundant cross-linked mucopeptides (Figure 1, Figure S1).

An initial PCA distributed the studied species into three clusters (Cl.1, Cl.2 and Cl.3). Interestingly, members of Cl. 1 lay far away from those of Cl. 2 and Cl. 3 in the PCA diagram, suggesting substantial differences in their peptidoglycan nature (Figure 1B). To further explore the internal variability in each cluster we performed a deeper statistical analysis where all peaks and their relative abundances were considered. Application of these criteria generated heat-map

trees defining the common properties and established the internal relations amongst the species in each cluster (Figure 1C).

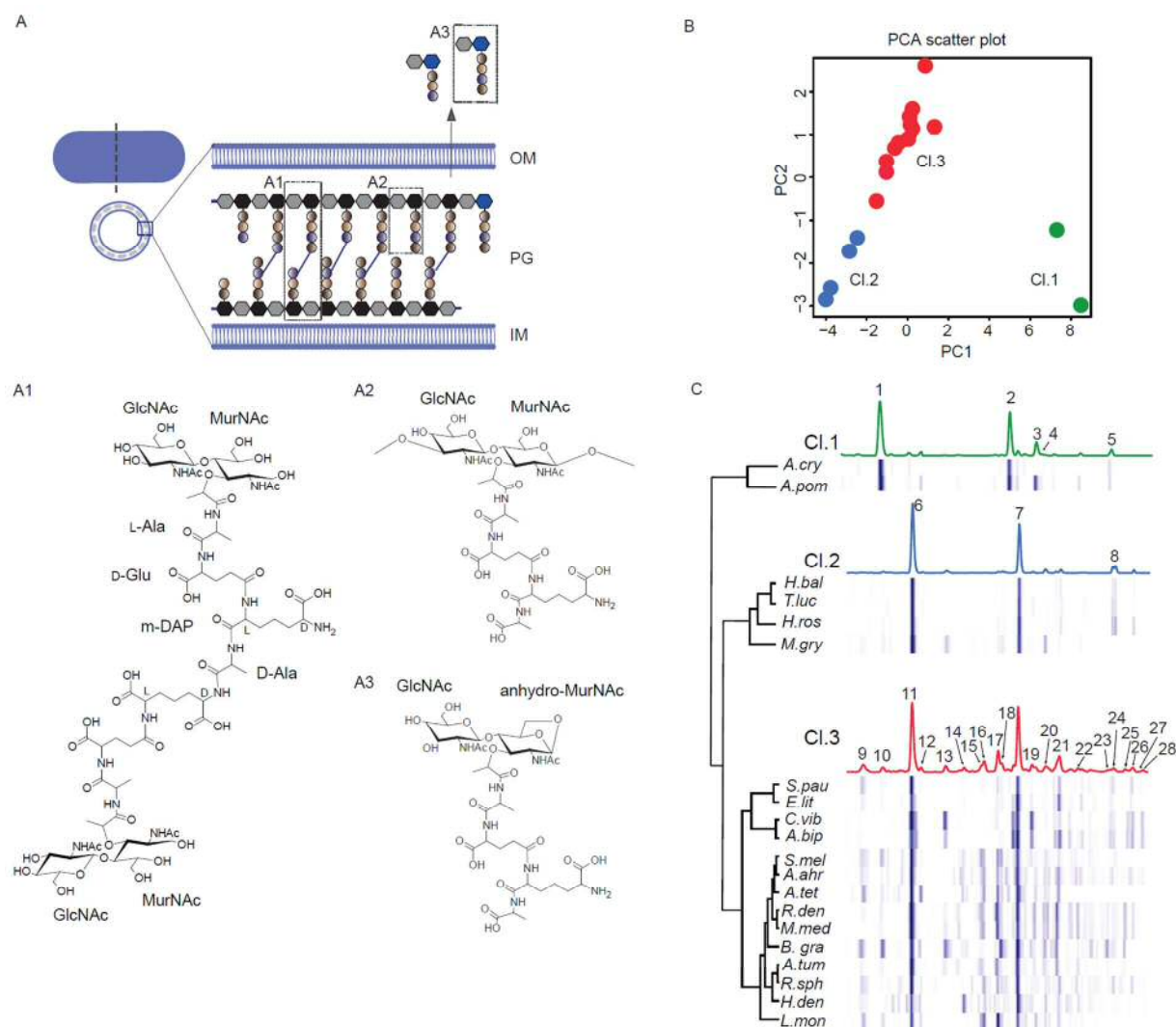


Figure 1. Chemometric clustering analyses on Alphaproteobacteria peptidoglycan samples.

(A) Schematic representation of Gram negative cell wall structure confined between inner (IM) and outer membrane (OM). Chemical structures for the dimeric (A1) and monomeric (A2) subunits in the polymer, and for the chain-terminal anhydro-muropeptides (A3). Hexagons represent *N*-acetylglucosamine (GlcNAc, gray), *N*-acetylmuramic acid (MurNAc, black) and

1
2
3 anhydro-MurNAc (blue). (B) Score scatter plot of the PCA. Among all possible combinations of
4 principal components, PC1 (most significant factor) *versus* PC2 (second most significant factor)
5 was found to explain more than 90% variance of the data. The samples were found to cluster in
6 three separate groups (green, blue and red). (C) Heat-map representing the relative abundance of
7 the muropeptides displayed by each bacteria (stationary phase cultures), from non-detected
8 (white) to highly abundant (deep blue). UPLC chromatogram on top shows the average pattern
9 as reference for each cluster and are color-coded as in the PCA.
10
11
12
13
14
15
16
17
18
19
20
21

22 To identify the peaks considered for clustering, each one was purified from the peptidoglycan of
23 an appropriate species and subjected to MALDI-MS. The mass values obtained were consistent
24 with *bona fide* muropeptides in all instances. Based on their masses, compounds were formulated
25 as follows: **1**= Tri (NH₂); **2**= Tetra-tri (NH₂); **3**= Mono-tri (NH₂,Ala); **4**= Mono-tetra (Ala); **5**=
26 Tetra-tri (NH₂,Anh); **6**= Tetra; **7**= Tetra-tetra; **8**= Tetra-tetra (Anh); **9**= Tri; **10**= Tetra (Gly⁴);
27 **11**= Tetra; **12**= Di; **13**= Penta; **14**= Nona (Gly⁴); **15**= Tri-tri (DAP); **16**= Tri-tetra (Gly⁴); **17**=
28 Tri-tetra (DAP); **18**= Tetra-tri; **19**= Tetra-penta; **20**= Tetra-tetra-tri; **21**= Tetra-tetra-tetra; **22**=
29 Tetra-tetra (Anh); **23**= Tetra-tetra (Anh,Anh); **24**= Tetra-tetra-tetra-tri; **25**= Tri-Tri-tetra
30 (Anh,DAP,DAP); **26**= Tetra-tetra-tri (Anh); **27**= Tetra-tetra-tetra(Anh); **28**= Tetra-tetra-tetra-
31 tetra(Anh) (Figure 1C, Table S2, Figure S2).
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

48 Analysis of peptidoglycan clustering in Alphaproteobacteria

49
50 Principal component analysis of UPLC data defined clusters of species sharing distinct
51 sets of peptidoglycan structural features, as well as a hierarchy of intra-cluster relations based on
52
53
54
55
56
57
58
59
60

more subtle, secondary variations. Figure 2 shows the hierarchical order derived from the PCA and its correlation with the iTOL-generated phylogenetic organization.

Species in Cl. 3 shared a peptidoglycan which was similar to that of the archetypical *E. coli*, characterized by the occurrence of DD (D-Ala-DAP) and LD (DAP-DAP) cross-links in both dimeric and trimeric mucopeptides, and with tetra- and tripeptides as stem peptides (Figure 2A, Table S2, Figure S2). The presence of LD-cross-linked mucopeptides was accompanied in all instances by at least a putative LD-transpeptidase encoded in the corresponding genomes (Table S3). One of the sub-groups within Cl. 3 (*Caulobacter vibrioides*, *Asticcacaulis biprosthecium*), was characterized by the accumulation of mucopeptides with D-Ala-D-Ala terminated stem pentapeptides (**13** and **19**) (Figure 2A, Table S2, Figure S2).

The peptidoglycan from Cl. 2 species was characteristically simpler, made up of tetrapeptides and cross-linked exclusively by DD-peptide bridges. Interestingly, *Magnetospirillum gryphiswaldense* displayed also all the additional Cl. 3 characteristic peptidoglycan features, although at much lower relative abundance than members of Cl. 3, and could be considered a transition species (Figure 2A). Components of Cl. 1 were unique. Compounds **1-5**, present in the members of Cl. 1, had t_R 's and masses which did not correspond with known mucopeptides (Table S2) making the affected species cluster far away of Cl. 2 and Cl. 3 in the PCA scatter plots (Figure 1B, Figure 2).

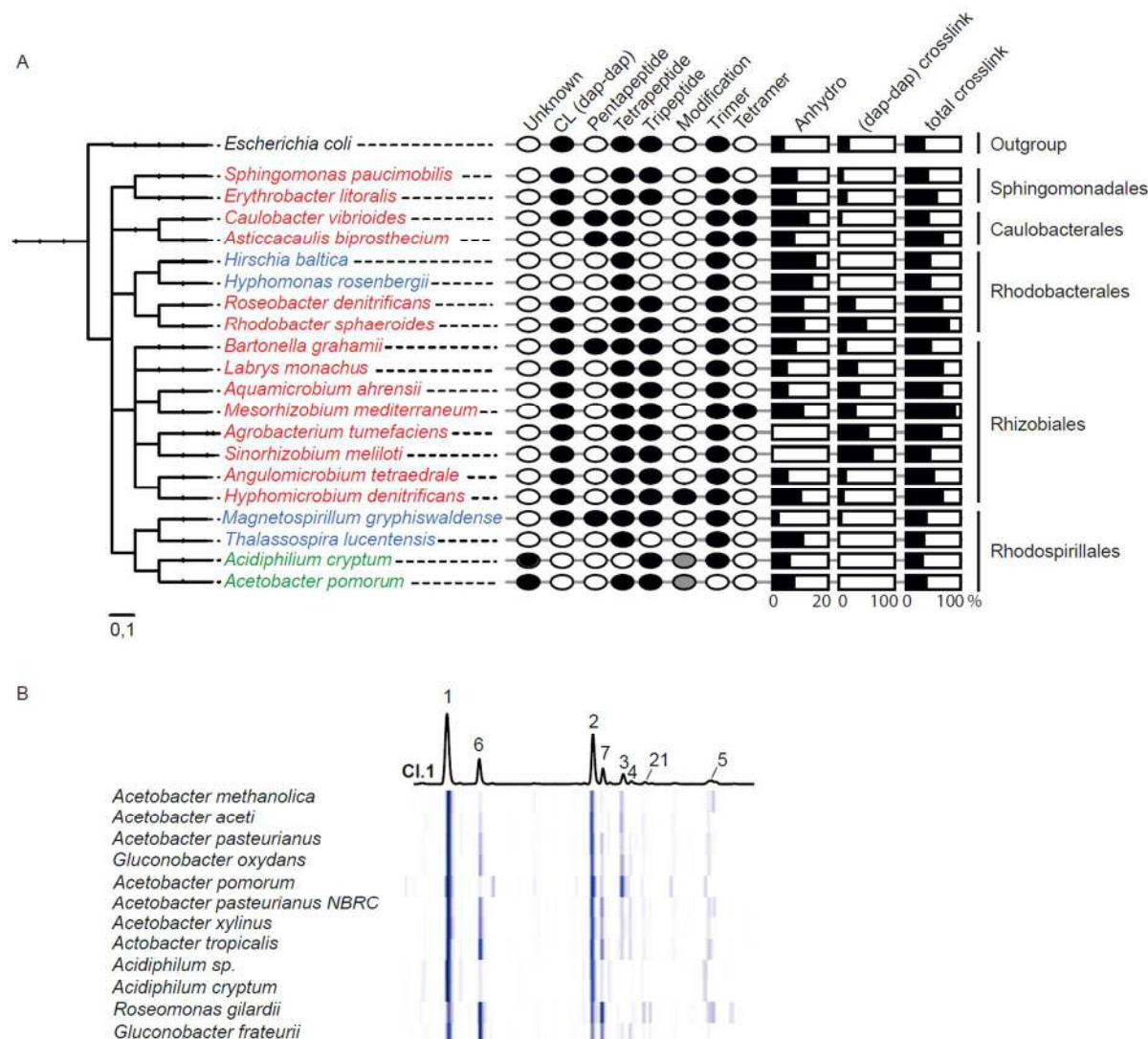


Figure 2. Comparative view of taxonomic organization and peptidoglycan clustering in Alphaproteobacteria. A) Phylogenetic organization of the different species was generated with iTOL software based on full genomic sequences. The names of species are color-coded to indicate their inclusion into the PCA derived Cl. 1, (green text); Cl. 2 (blue text); or Cl. 3 (red text). The array of ellipsoidal symbols illustrates the presence (black) or absence (white) of the differential structural traits indicated on top of each column. Two different “modifications” based on the MS data are illustrated by black (*H. denitrificans*) and grey (*A. pomorum* and *A. cryptum*) ellipsoidal symbols. The black portions in the rectangular bars indicate the percentual

1
2
3 abundance of anhydro-muropeptides (scale from 0 to 20%), and the values for total cross-
4 linkage, and LD (dap-dap) cross-linkage, expressed as cross-bridges per 100 muropeptides. White
5 bars mean not detected. (B) Muropeptide composition of the *Acetobacteria* species analyzed
6 (stationary phase cultures). The UPLC chromatogram on top shows the average pattern as
7 reference for the group. Numbers identify each muropeptide as described in table S2. The
8 relative abundance of muropeptides is presented as a heat-map from darker (more abundant) to
9 lighter (less abundant), as calculated from two independent experiments with triplicate samples.
10
11
12
13
14
15
16
17
18
19
20
21

22 Additionally, this bioinformatic approach has proved capable to identify peptidoglycan
23 variability even at the level of species, supporting PCA as an appropriate tool to both, group
24 bacteria according to relevant characteristics of peptidoglycan, and detect subtle discriminative
25 structural motifs.
26
27
28
29
30
31

32 **Acetobacteria have a distinctive peptidoglycan architecture**

33

34 Cluster 1 was represented by *Acidiphilum cryptum* and *Acetobacter pomorum*, which are
35 acetic acid bacteria. To determine whether the peptidoglycan structure characteristic for Cl.1 was
36 conserved throughout the Family Acetobacteraceae, ten additional species were investigated by
37 the same analytical methods. The results showed that all of them fitted within the peptidoglycan
38 pattern defined for Cl.1 (Figure 2B, Figure S3). Remarkably, introduction of new species to Cl.1
39 increased further the existing heterogeneity. Cluster 1 subdivision was associated with variations
40 on the relative abundance of the peaks, and the partial conservation of a second series of new
41 components, namely **3** and **4**, characteristic of *Acetobacter*, *Acidomonas* and *Gluconobacter*
42 members (Figure 2B). As shown in table S2, the more abundant muropeptides in Cl. 1 had
43 masses corresponding to **9**, **18** and its anhydro variant (Figure S2b), but with a deficit of one
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

mass unit (m.u.) in all instances, indicative of the substitution of an OH (17 m.u.) by an NH₂ (16 m.u.), often the result of an amidation reaction. In fact, ESI-MS/MS analysis of **1** traced the mass difference to the DAP residue suggesting amidation of one of the carboxylic groups in DAP (Figure S4A). To verify this, purified **1** and control mucopeptide **6** were subjected to NMR analysis (Figure 3A, Tables S4-S5 and Figures S5-S8). The amide region of the proton spectrum of **1** (Figure 3A upper panel) revealed two signals corresponding to each proton of the –CONH₂ amide group (labeled as G(NH₂)_{a/b} residue). These signals were absent in the control **6** spectrum. Instead, the characteristic signals for D-Ala (labeled as G) were the ones detected (Figure 3A lower panel).

The localization of the amide modification on the α -carboxyl of DAP (named L-amidation DAP from here onwards for referring to the carboxylate at the L-stereogenic center of DAP) (labeled as F_I) was deduced by observing the ROESY crosspeak connecting one of the NH₂ amide protons (labeled H_N-G_a) with the H $_{\alpha}$ of the L-end that in turn was identified by the TOCSY crosspeak with the amide proton of the peptidic bond connecting L-end DAP (H_N-F_L) with D-Glu residue (labeled as E) (Figure S9). This L-end DAP amide proton was identified by the corresponding ROESY crosspeak with the preceding CH₂ γ -protons of D-Glu residue (Figures S6, S9 and Tables S4, S5). Amidation of DAP residues has been reported in other instances¹⁷, but localized at the ϵ -carboxyl (named D-amidated DAP from here onwards for referring to the carboxylate at the D-stereogenic center of DAP). L-amidation DAP in **1** was further confirmed by additional NMR comparative analyses including the non-amidated form (**9**, Figures 3B, S10, S11 and Tables S4-S5) and the D-DAP amidated variant (**29**)^{17c}, which also presents amidation at glutamic α -carboxyl group (Figures 3B, S12, S13 and Tables S4, S5).

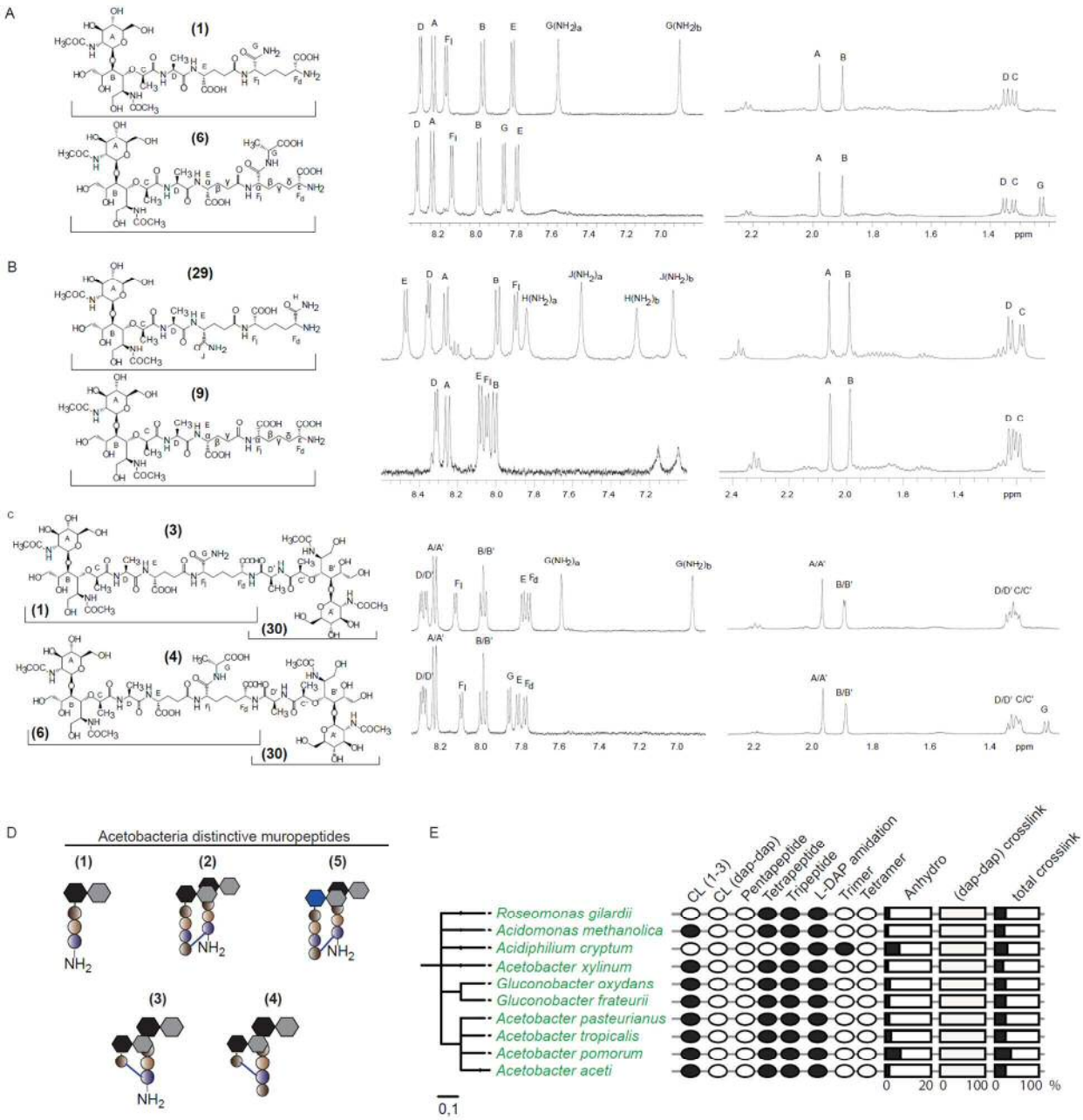


Figure 3. Characterization of peptidoglycan in Family Acetobacteraceae. (A-C) On the left molecular structures of **1**, **6**, **29**, **9**, **3** and **4**. On the right, amide and aliphatic regions of one dimensional ¹H NMR spectra of (A) **1** vs **6**, (Figures S6-S7), (B) **29** vs **9** (Figures S10 and S12) and, (C) **3** vs **4** (Figures S14-S15). See tables S4 and S5 for NMR assignments. **30**: GlcNAc-β-

(1→4)-MurNAc-L-Ala. (D) Schematics of the Fam. Acetobacteraceae distinctive muropeptides. Hexagons represent GlcNAc (gray), MurNAc (black) and anhydro-MurNAc (blue). (E) Comparative analysis of iTOL-based taxonomic organization and peptidoglycan clustering of Acetobacteraceae.

Analysis of the mass differences was consistent with **3** and **4** being alike except for a mass difference of 72 m.u. (Table S2). We hypothesized that such a difference in **3**, relative to **4**, was the result of the loss of the fourth D-alanine (-71 m.u.) in the peptide moiety as a consequence of an L-amidated DAP (-1 m.u.). Indeed, ESI-MS/MS analyses revealed a fragmentation sequence which corresponded to that of a **1** and **6**, respectively (Figure S3BC and Table S2).

A full NMR analysis of **1**, **6**, **29**, **9**, **3** and **4** (the amide and methyl regions of the corresponding spectra are presented in Figure 3A, 3B, 3C) by means a double set of homo- and heteronuclear spectra acquired either in water (with 10% deuterated water) or fully deuterated water (Figures S6-S7, S10, S12, S14-S15) allowed the assignment of the chemical shifts of all amide and non-exchangeable protons and almost all carbons (excluding carbonyls) of all six muropeptides (Tables S4 and S5). The full proton assignment allowed to deduce the connectivity of amino acid residues through the amide bonds by analyzing NOE (Nuclear Overhauser Effects) between the amide proton of a residue and the α proton (γ in the case of glutamic residue) of the previous residue (CH-CO-NH) by means of a combination of TOCSY and NOESY or ROESY spectra (Figures S6 to S17). NMR analysis demonstrated that **3** and **4** correspond to dimers LD-crosslinked via an atypical 1-3 bond (L-Ala-DAP), where **3** lacks the D-Ala in fourth position due to the amidation of the DAP L-center (Figure 3C, Figures S14-S17).

Proton NMR spectra of **3** and **4** (Figure 3C), in accordance with mass spectrometry data (Figure S4BC), clearly show the presence of duplicate signals for amide and methyl protons of sugars and L-Ala residues (labeled A/A', B/B', C/C' and D/D' for the NAc-glucosamine, reduced NAc-muramic acid, its D-Lactyl moiety and L-Ala residues, respectively) accounting for a repetition of the GlcNAc-MurNAc-L-Ala (**30** moiety). On the other side only signals for a unique D-Glu (labeled E) and DAP (labeled F₁ and F_d) amide protons were observed (Figure 3C, Figure S16 and S17). Additionally, two extra amide peaks in **3** (labeled G(NH₂)_a and G(NH₂)_b) corresponding to the amidation of the DAP carboxyl group at the L-end and, in case of **4** one extra amide proton and another extra methyl proton signals (labeled G) corresponding to the D-Ala residue at DAP carboxyl group at the L-end are observed (Figure 3C, Figure S16 and S17). The connectivity D-DAP/L-Ala was confirmed by the presence of the corresponding crosspeaks (HN ϵ -DAP: H α -Ala) in their NOESY spectra (Figure S16 (**3**), S17 (**4**)).

Therefore, we conclude that Cl. 1 of Alphaproteobacteria is made up of bacteria whose peptidoglycans are characterized by the presence of L-DAP amidated mucopeptides and, in some species, by the presence of LD (1-3) cross-linked mucopeptides, where 1-3 refers to the positions of the amino acids (L-Ala donor, (D-stereogenic center)-DAP acceptor) from each monomer involved in the crosslink (Figure 3DE).

Peptidoglycan structure is regulated in Acetobacteraceae

The variability in the relative abundances of L-amidation (from ca. 95% to 30%) and 1-3 crosslinking (from 0% to 15%) amongst Acetobacteraceae suggested that synthesis of these mucopeptides could be actively regulated at least in some species (Figure 4). To test this possibility, we investigated the effect of growth phase on the accumulation of the atypical

1
2
3 mucopeptides. Remarkably, L-amidated and 1-3 cross-linked mucopeptides follow very distinct
4
5 dynamics. The proportion of 1-3 dimers was highest in stationary phase (Figure 4A) in all the
6
7 strains. However, L-DAP amidation levels were apparently constitutive for all species except
8
9 *Gluconobacter frateurii* and *Roseomonas gilardii* (Figure 4B) where levels plummeted in
10
11 stationary phase. A more detailed analysis of the evolution of L-amidation in *G. frateurii* showed
12
13 that the reduction in amidation (from 80% down to 30%) was triggered early during the
14
15 transition from exponential into stationary phase (at an $OD_{600} \approx 0.8$ under our conditions (Figure
16
17 4CD, Figure S18A).

18
19
20
21
22 Dilution of cultures undergoing the reduction in L-amidation, immediately reverted the
23
24 trend, and amidation went up to the level characteristic for exponentially growing cultures
25
26 (Figure 4CD), clearly supporting the reversibility of the effect as well as its dependency on the
27
28 growth state of the cells. Interestingly, when cells were forced to stop growth at low OD because
29
30 of nutrient (mannitol) limitation, L-amidation remained high. This result supports a requirement
31
32 for the cells to grow at high densities for a certain time to accumulate non-amidated
33
34 mucopeptides, rather than an active de-amidation of existing L-amidated subunits (Figure S18A).
35
36 To assess nutrient exhaustion as the downregulation trigger, cultures of *G. frateurii* were
37
38 supplemented in parallel with individual media components, but no suppression of the effect was
39
40 found (Figure S18B). To ascertain whether extracellular signaling (e.g. quorum sensing) was
41
42 controlling downregulation of peptidoglycan amidation in *G. frateurii*, exponentially growing
43
44 cells were challenged with stationary phase preconditioned media, but again no effect on
45
46 peptidoglycan amidation was detected (Figure S18C). Thus we conclude that peptidoglycan
47
48 amidation is not downregulated by released factors nor starvation signals in *G. frateurii* and
49
50 further studies will be needed to shed light on its metabolic regulation in this bacterium.
51
52
53
54
55
56
57
58
59
60

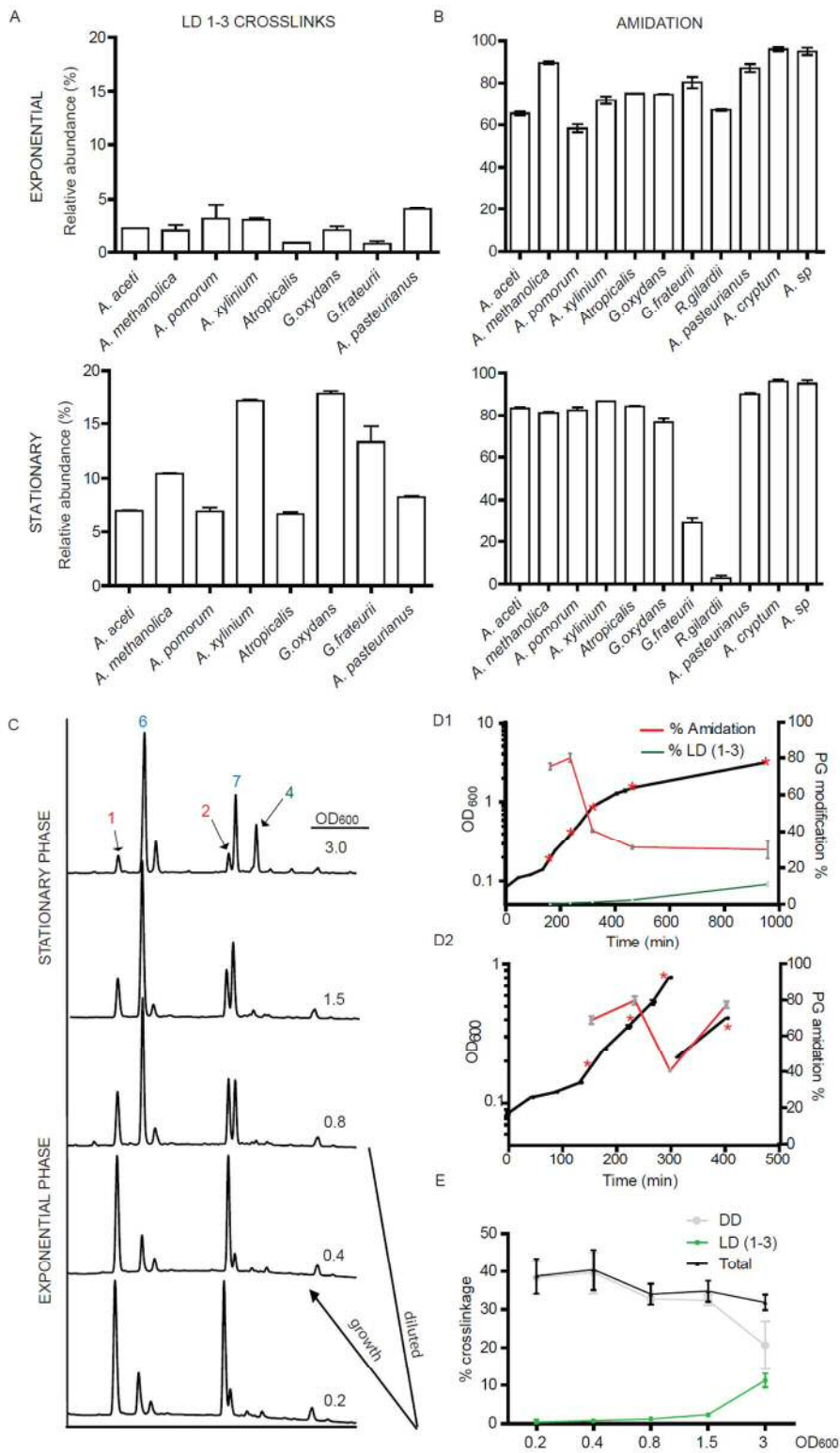


Figure 4. Regulation of peptidoglycan characteristic features in *G. frateurii*. Influence of the state of growth on the relative abundances of LD (1-3) crosslinked (A), and L-DAP amidated

muropeptides (B) in *Acetobacteria* spp. Upper panels display the values corresponding to mid-exponential ($OD_{600}=0.3$) cultures and bottom panels to 36 hours stationary phase cultures. Growth conditions were as described in the experimental section. Panel (C) displays the evolution of the UPLC peptidoglycan profile for *G. frateurii* throughout the transition from active growth into rest. The corresponding quantitative changes in amidation and LD (1-3) crosslinking are shown in panel (D1). Panel (D2) shows the evolution of amidation in a culture let to reach late exponential phase ($OD_{600}=0.8$), diluted in fresh, pre-warmed medium to $OD_{600}=0.2$ and allowed to grow up for a mass doubling ($OD_{600}=0.4$). Panel (E) displays the evolution of total (black), DD (4-3) (grey), and LD (1-3) (green) crosslinking associated to the transition from active growth into stationary phase. Experiments were done in triplicates.

Contrary to amidation, LD (1-3) dimer levels were very low in *G. frateurii* exponential phase cultures ($< 1\%$ of the total muropeptides) and increased sharply in stationary phase reaching values as high as 13% of total muropeptides, accounting for about 40% of the total cross-linkage (Figure 4CDE). Accumulation of LD (1-3) crosslinked muropeptides in stationary phase was compensated by a concomitant reduction in DD-cross-linked muropeptides, resulting in total cross-linkage remaining essentially constant throughout the exponential to stationary phase transition (Figure 4E). Interestingly, formation of LD (1-3) crosslinks is not associated to high optical densities but rather to growth arrest. In fact, cultures forced to stop growth at lower cell densities (from OD_{600} 0.5 to 3.0) because of nutrient limitation, triggered accumulations of LD (1-3) cross-linked muropeptides at cell densities where exponentially growing cells were devoid of them (Figure S18B). The fact that the accumulation of LD (1-3) cross-linked muropeptides was enhanced in stationary phase cells in all the species tested, suggests a common

regulatory mechanism, and a possible involvement of LD (1-3) mucopeptides in cell survival under nutrient scarcity and high density of population.

Acetobacteria-specific peptidoglycan modifications are protective to the action of type VI endopeptidases *in vitro*

Because peptidoglycan structural modifications can confer resistance against damaging enzymes (i.e. lysozyme)^{11a}, we decided to test whether or not the modifications found in the peptidoglycan of Acetobacteria could provide a protective effect against T6SS endopeptidases (Tse) from Family 1 and 2. Family 1 Tse have a characteristic CHAP domain and target the D-Glu- γ -meso-DAP bond, whilst Family 2 members display Nlp/P60 peptidase domains and cleave the meso-DAP-D-Ala bond in cross-linked mucopeptides^{8a} (Figure 5A, Figure S19A). Therefore, we purified Family 1-Tse from *Pseudomonas aeruginosa* (namely Tse1Pa^{8b}); and Family 2 Tse from *Acidovorax citrulli*, (namely Tse2Ac), and tested their activity against Acetobacteria dimeric mucopeptides. Both species, *P. aeruginosa* and *A. citrulli*, naturally coexist with Acetobacteria in the soil, and therefore may compete.

In vitro digestion assays of purified *G. frateurii* dimeric mucopeptides (**2**, **3**, **4** and **7**, Table S2) showed that neither Tse2Ac nor Tse1Pa were active on the LD-dimers (**3** and **4**) (Figure 5BC). Likewise, Tse2Ac endopeptidase showed no detectable activity on the L-amidated DD-crosslinked mucopeptide (**2**), but readily degraded the non-amidated counterpart (**7**) (Figure 5B). Finally, Tse1Pa showed a slight but still significant 20% lower activity on **2** compared to **7** (Figure 5C). The preference of Tse1Pa for canonical non-amidated DD-crosslinked dimers vs the L-DAP amidated version was further characterized in an *in vitro* endopeptidase time course assay using muramidase-digested *G. frateurii* peptidoglycan as substrate.

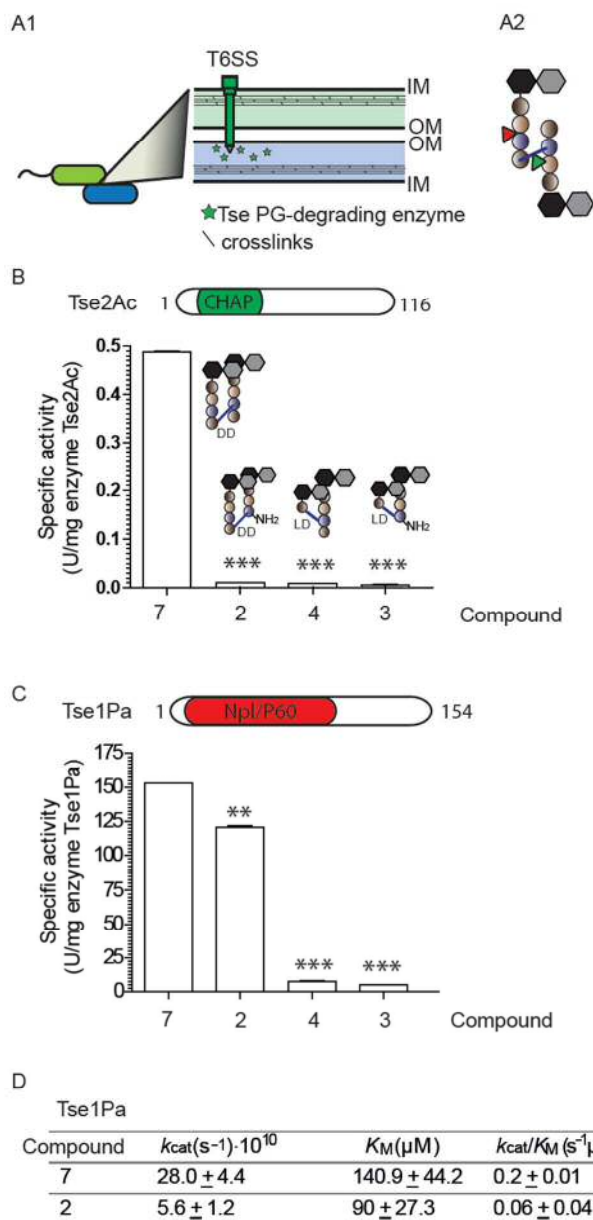


Figure 5. Activity of Tse1Pa and Tse2Ac endopeptidases on Acetobacteria peptidoglycan dimers. A1) Bacteria endowed with T6SS (green), an “inverted phage tail-like” structure that punctures through the OM, are able to inject effector molecules into the periplasmic space of a prey bacteria (blue). Effector endopeptidases (Tse) degrade the peptidoglycan of the prey cleaving the peptide bonds between D-Glu and DAP (as Tse1Pa, red arrowhead) or between D-Ala and DAP (as Tse2Ac, green arrowhead). Panels (B) and (C) illustrate the activity of Tse2Ac

and Tse1Pa activity on the indicated compounds, respectively (see Table S3). Schemes of the Tse protein (Tse2Ac 116 amino acids, Tse1Pa 154 amino acids) and domains, their cleaving selectivity and, **2**, **3**, **4** and **7** crosslinking stereochemistry (LD vs DD), are depicted. (D) Kinetic parameters of Tse1Pa-6His DD-endorpeptidase activities on **2** and **7**.

As expected, LD (1-3) dimers were completely resistant to Tse1Pa, and **7** was hydrolyzed faster than **2** (Figure S19BC). The reduction in the indicated dimers was associated to an accumulation of **12** and disaccharide hexapeptide (**31**) products expected for the γ -glutamyl-DAP endopeptidase activity of Tse1Pa (Figure S19BC). Finally, calculation of steady-state kinetics for the Tse1Pa enzyme on purified muopeptides revealed a k_{cat}/K_M 3.3-fold higher for **7** than for **2**, consistent with the reduced susceptibility of L-DAP amidated dimers to Tse1Pa (Figure 5D).

Together, these results indicate that the peptidoglycan modifications found in *Acetobacteria* may make them less susceptible to at least some T6SS effector peptidoglycan hydrolases.

L-DAP amidation reduces immunogenicity of *Acetobacteria* peptidoglycan in the fly *Drosophila melanogaster*.

Acetobacteraceae is one of the dominant bacterial Families of *Drosophila melanogaster* commensal microbiota¹⁸. *D. melanogaster* main antibacterial defense is provided by the innate immune response (IIR) system, which is triggered by peptidoglycan recognition and depends on the identity of the di-amino acid in third position (the DAP-specific IMD pathway)¹⁹. In consequence, we hypothesized that L-DAP amidation might be a mechanism devised by fly gut *Acetobacteria* to modulate the immune response of their host.

To test these ideas, we checked the effect of normal and L-DAP amidated anhydro-muropeptides since these are known to be the most immunogenic²⁰. Therefore, equimolar amounts of pure muropeptide control (**32**, Figure S2), identical to tracheal cytotoxin (TCT)²¹, and **33** (Tri (NH₂, Anh), Figure S2) from *G. oxydans* were injected into the fat body of *D. melanogaster* adult flies and the transcriptional induction of dipteridine, an antibacterial gene regulated by the IMD pathway, was followed and quantified (Figure 6).

Our results indicated that induction by **33** was about 70% of the value for **32**, a result supportive of the alleviating effect of amidation on the immunogenicity of peptidoglycan. For a more global estimation of the relative immunogenicity of Acetobacteria peptidoglycan, we injected identical amounts of muramidase digested peptidoglycan from either *E. coli* (non-amidated, 4.5 % anhydro-muropeptides) or *G. oxydans* (80% amidated, 4.1 % anhydro-muropeptides) in flies and measured the induced response as above (Figure 6).

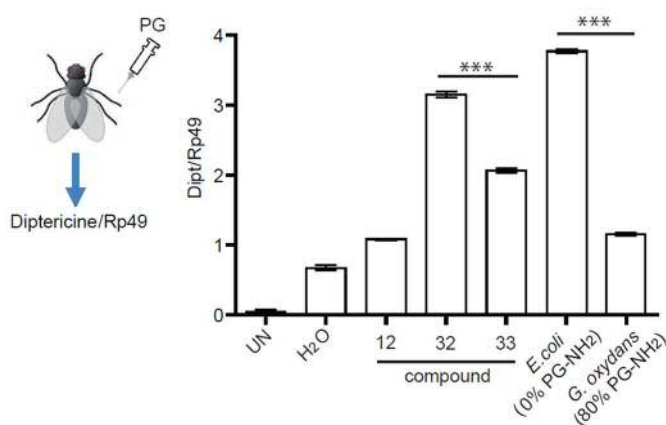


Figure 6. Induction of the IMD pathway of *D. melanogaster* by Acetobacteria amidated muropeptides and peptidoglycan. *D. melanogaster* flies were injected with the indicated muropeptides, or muramidase digested peptidoglycans, dissolved in water and production of Dipteridine was measured as the Dipteridine/Rp49 ratio. UN: untreated. H₂O: control flies

1
2
3 injected with water. Compound **12** was used as non-immunostimulatory muropeptide.
4
5 Experiments were done in triplicates (p-value: ***<0.0001).
6
7
8
9

10 Our results show that *E. coli* peptidoglycan induced 4-fold more dipteridine than *G.*
11
12 *oxydans* amidated peptidoglycan. Together, these results buttress that L-DAP amidation reduces
13
14 the immunogenicity of peptidoglycan on the IIR.
15
16
17
18
19

20 Discussion

21
22 The peptidoglycan cell wall is a fundamental structure in Bacteria whose real diversity is
23
24 still far from settled. In this work, we harness UPLC and multivariate data analysis to explore
25
26 peptidoglycan variability in Alphaproteobacteria, dominant organisms of aquatic environments
27
28 worldwide. Our analysis distributed the studied species in three clusters according to their
29
30 peptidoglycan properties. While Cl. 1 peptidoglycan properties made it substantially different to
31
32 that of the other clusters, it is important to keep in mind that both Cl. 2 and Cl. 3 also exhibited a
33
34 considerable internal degree of variability that can be associated with environmental adaptation.
35
36 One good example is provided by the distribution of pentapeptide and LD (DAP-DAP) cross-
37
38 linked muropeptides in these clusters. Since these properties have been reported in a number of
39
40 bacteria to be important in environmental adaptation and morphogenesis²², and in antibiotic
41
42 tolerance, respectively^{10,23}, their presence/absence in certain species might reflect specific
43
44 regulatory mechanisms important for the lifestyle of bacteria. In fact, some species as
45
46 *Hyphomonas* and *Roseomonas* encode LD-transpeptidase orthologues but have no detectable LD-
47
48 cross-linked muropeptides. Similarly, bacteria accumulating pentapeptides encode orthologues to
49
50 *E. coli* PBP5 in all cases (Figure 2, Table S3). This observation suggests either an unrelated
51
52
53
54
55
56
57
58
59
60

function for the putative enzymes, or control mechanisms strongly repressing their expression/activity under certain conditions.

Remarkably, Cl. 1 revealed two previously unreported peptidoglycan modifications characteristic of Acetobacteria: amidation at the L-center of DAP and LD (1-3) cross-linked muropeptides. Generation of these novel muropeptides may be presumed to require yet unknown enzymatic activities. Amidation of the ϵ -carboxyl group of meso-DAP (**29**, D-DAP amidation; Figure S2) has been described in some Gram positive bacteria^{17c, 24} mediated by the activity of cytoplasmic meso-DAP-amidotransferases, acting at the soluble precursor level²⁴. However, a similar mechanism seems unlikely to explain L-amidation (e.g. **1**). A reaction at the precursor level would imply the production of large amounts of truncated precursors leading to Lipid-II-disaccharide L-DAP amidated-tripeptides, which could only work as acceptor in transpeptidation reactions. Furthermore, in all the species studied, the α -carboxyl group was either amidated or bound to D-Ala, but free α -carboxyl groups were absent or below the detection limit (< 0.1%), suggesting a tight coupling between removal of the D-Ala residue and the amidation reaction. Generation of LD (1-3) cross-linked muropeptides might as well require the intervention of a new family of LD-transpeptidases, as none of the members of Cl. 1 had homologs to the LD-transpeptidases responsible for the synthesis of LD (DAP-DAP) cross-linked muropeptides or the attachment of Braun's lipoprotein in Enterobacteria²⁵ (Table S3). Future research should show if LD (1-3) crosslinking is catalyzed by a new type of LD-transpeptidase and whether its expression also confers tolerance to β -lactams as shown for (DAP-DAP) LD-transpeptidases¹⁰.

Both L-DAP amidation and 1-3 crosslinking synthesis, respond to specific growth cues which can even be species-specific. This plasticity likely implies regulatory mechanisms to continuously adjust the characteristics of the cell wall to the changing conditions and supports

that cell wall signatures are better reporters of environmental adaptation than taxonomic characters.

Functional studies of these peptidoglycan modifications pointed out potential defensive roles against cell wall hydrolytic enzymes and innate immune response. Our enzymatic studies of Tse factors on *Acetobacteria* mucopeptides clearly show that both L-DAP amidation and LD (1-3) crosslinking make peptidoglycan a less preferred substrate for T6SS Family 1 and 2 endopeptidases. These results suggest that peptidoglycan modifications in the peptide moieties might work as a protective mechanism against predatory peptidoglycan hydrolytic enzymes very much as modifications in the glycan backbone confer protection against lysozymes^{11a}. Microbial competition mediated by cell wall hydrolytic enzymes (e.g. mobilized by T6SS, outer-membrane vesicles, phages...) might be one driving force for peptidoglycan modifications where prey and predator play an arms race for niche colonization and survival.

Another interesting aspect of our findings is the potential role of L-DAP modifications in immunomodulation. Using a validated assay to assess modulatory action of peptidoglycan-derived fragments on *D. melanogaster* innate immune system, we showed that L-DAP amidation reduces the immunogenicity of peptidoglycan. Given that the IIR to DAP-type peptidoglycans relies on PGRP-LC, which preferentially recognizes anhydro-muropeptides, the lower immune-stimulation of *G. oxydans* peptidoglycan is likely the consequence of these being fully amidated. These results are in agreement with previous studies which showed that D-amidation of DAP was less stimulatory of the IIR of the fly²⁰. Therefore, as *Acetobacteria* is one of the dominant groups of the fly commensal microbiota together with *Lactobacilli*, which possess a D-amidated DAP peptidoglycan^{17c}, we hypothesize that DAP amidation might be an adaptive mean to prevent futile inductions of the antimicrobial defenses in the absence of a

1
2
3 pathogen. Microbial competition or host adaptation might be just two examples of evolutionary
4
5 pressures that have imprinted peptidoglycan variability in bacteria. Future investigations on
6
7
8 peptidoglycan variability in a wider range of bacteria will certainly increase the repertoire of
9
10 factors associated to cell wall biology and expand their biological meaning beyond microbial
11
12 ecology and innate immune response modulation.
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Conclusions

Collectively, our results support the potential of chemometric methods for kingdom-wide “mureinomic” studies to uncover novel peptidoglycan traits and their associated biosynthetic and regulatory activities. We describe unforeseen peptidoglycan traits in the Alpha subdivision of Proteobacteria: amidation at the α -carboxyl of meso-DAP and the presence of muropeptides L-Ala-D-(meso)-DAP cross-linked. These structures influence sensitivity to Type VI secretion system peptidoglycan endopeptidases and recognition by the *Drosophila* innate immune system, suggesting relevant roles in the environmental adaptability of these bacteria. Finally, we offer a powerful pipeline for a kingdom-wide peptidoglycan screening to boost knowledge of cell wall biology in Bacteria thereby opening new avenues towards the development of taxon-specific antimicrobial strategies.

Experimental section

Microbiology

E. coli strains were grown in Luria broth (LB) and on LB agar plates. Kanamycin (Kn) was used at 50 µg/ml. *L. plantarum* was grown in MRS broth at 30°C ^{17c}. Acetobacteria were grown on YP (yeast extract 1% and peptone 1%) supplemented with different sugars (all at 3%): glucose (G) (for the clustering samples), mannitol (M) or stationary phase preconditioned filtered media (PCM) and incubated for the indicated periods (Figure 3). Unless otherwise indicated, the different bacteria used in this work were grown (either exponentially, OD₆₀₀= 0.3; or to stationary phase, OD₆₀₀= 3.0, t= 36h) under optimal conditions (media and temperature) recommended by the DSMZ, ATCC and CECT bacterial collections.

Peptidoglycan isolation

Cells from 0.2 l cultures of overnight stationary phase, or 1 l exponential phase (OD₆₀₀= 4.0) were pelleted at 5,000 rpm and resuspended in 5 ml of PBS, added to an equal volume of 10% SDS in a boiling water bath and vigorously stirred for 4 h, then stirred overnight at *t_R*. The insoluble fraction (peptidoglycan) was pelleted at 100,000 rpm (400,000 x g), 15 min, 30°C (TLA-100.3 rotor; Optima™ Max ultracentrifuge, Beckman) and resuspended in MilliQ water. This step was repeated (4-5 times) until the SDS was washed out. Next, peptidoglycan was treated with Pronase E 0.1 mg/ml at 60°C for 1 h and further boiled in 1% SDS for 2 h to stop the reaction. After SDS was removed as described previously, peptidoglycan samples were resuspended in 200 µl of 50 mM sodium phosphate buffer pH 4.9 and digested overnight with 30 µg ml⁻¹ muramidase (Cellosyl, Hoechst) at 37 °C as described ²⁶. Muramidase digestion was stopped by incubation in a boiling water bath (5 min). Coagulated protein was removed by centrifugation. The supernatants were mixed with 150 µl 0.5 M sodium borate pH 9.5, and

1
2
3 subjected to reduction of muramic acid residues into muramitol by sodium borohydride treatment
4
5 (10 mg ml⁻¹ final concentration, 30 min at room temperature). Samples were adjusted to pH 3.5
6
7 with phosphoric acid. For purification of **29**, *L. plantarum* peptidoglycan was purified as
8
9 described by Bernard E, et al.^{17c}.

12 13 **Peptidoglycan analysis**

14
15 Chromatographic analyses of muropeptides were performed on an ACQUITY Ultra
16
17 Performance Liquid Chromatography (UPLC) BEH C18 column (130 Å, 1.7 µm, 2.1 mm by 150
18
19 mm; Waters, USA), and peptides were detected at Abs. 204 nm using an ACQUITY UPLC UV-
20
21 Visible Detector. Muropeptides were separated using a linear gradient from buffer A (phosphate
22
23 buffer 50 mM, pH 4.35) to buffer B (phosphate buffer 50 mM, pH 4.95 methanol 15% (v/v)) in
24
25 28 min, and flow 0.25 ml min⁻¹. Concentration of total murein was determined as described²⁷.
26
27 Individual muropeptides were quantified from their integrated areas using samples of known
28
29 concentration as standards. Experimental variability in muropeptide quantification by UPLC in
30
31 Figure 2 is <5% for minor muropeptides and <1% for more abundant muropeptides. See
32
33 supporting information for additional details.
34
35
36
37

38
39 For chemical analysis, muropeptides were purified by HPLC on an Aeris peptide column
40
41 (250 × 4.6 mm; 3.6 µm particle size; Phenomenex, USA) using adapted methods from the
42
43 described above used in the UPLC system. Pure muropeptides were concentrated and desalted
44
45 using a water-methanol gradient on the same column prior to MS and NMR analyses. The
46
47 identity of individual muropeptides was established by MALDI-TOF (Voyager DE-STR) and
48
49 electrospray-ion trap MS (Velos Pro Dual-Pressure Linear Ion Trap Mass Spectrometer system).
50
51

52 53 **NMR**

The structures of reduced mucopeptides **1**, **3**, **4**, **6**, **9** and **29** were defined by means of a collection of NMR experiments for each compound performed on samples dissolved in D₂O and H₂O:D₂O (90:10), acquired in 500, 600 or 700 MHz Bruker spectrometers and processed with TOPSPIN Bruker software.

Standard pulse sequences, included in TOPSIN 2 acquisition software, were used for one dimensional ¹H, two dimensional homonuclear NOESY (Nuclear Overhauser Effect Spectroscopy), ROESY (Rotating-frame NOE Spectroscopy), TOCSY (TOtal Correlation Spectroscopy) and DOSY (Diffusion Ordered Spectroscopy), and two dimensional heteronuclear ¹H-¹³C-HSQC (Heteronuclear Single Quantum Correlation). Solvent suppression modules and diffusion filtering (presaturation, waltzgate, excitation sculpting with gradients) were applied when needed.

For assignments of the chemical shifts of each compound, two sets of spectra were acquired at 35°C and 500MHz (22°C and 25 °C and 700MHz and 600MHz spectrometers equipped with cryo-probes respectively in the case of **4**): A set of spectra in deuterated water ¹H-1D, TOCSY with 20 and 60 ms spin-lock time, NOESY with 500 ms mixing time, HSQC and DOSY and a second set of spectra in water (with 10% deuterated water), ¹H-1D, TOCSY and NOESY in order to observe the amide protons.

For identification of the connectivity and position of substituents in the new mucopeptides **1**, **3** and **4**, in addition to previous spectra, some extra 1D-¹H-1D; TOCSY (20 and 60 ms spin-lock time); NOESY (500 or 600 ms mixing time) or ROESY with 300 or 400ms of spinlock time were acquired at different temperatures and magnetic fields. In some NOESY experiments the temperature of the sample was adjusted between (4°C to 35°C) to optimize the

correlation time of the mucopeptide in order to obtain observable NOESY crosspeaks. The chemical shifts were referenced to TSP (trimethyl silyl propanoic acid, $\delta=0$ ppm) or to water signal using the equation $\delta(\text{ppm}) = 5,051 - 0,011 \times T(^{\circ}\text{C})$ ²⁸.

For figures 3A-C and S6-S17, the different spins systems observed in the spectra are named: A, B, C, D, E, F (F_L and F_D), G, H, J, A', B', C' and D'. After assignment, this spins systems correspond with the different residues with the following correspondence:

A/A': N-acetyl glucosamine.

B/B': reduced N-acetyl muramic acid.

C/C': D-Lactyl moiety in N-acetyl muramic residue.

D/D': L-alanine residue.

E: D-glutamic acid residue.

F: meso-diaminopimelic acid with F_L spin system starting from the L-end and F_D spin system starting from the D-end.

G: spin system of residue connected to α -carboxyl at L-end of diaminopimelic.

H: spin system of residue connected to ε -carboxyl at D-end of diaminopimelic.

J: amide NH_2 spin system connected to α -carboxyl of D-glutamic acid.

Protein expression and purification.

A. citrulli tse, *P. aeruginosa tse*, and *slt70* from *E. coli* genes were cloned in pET28b (Novagen) for expression in *E. coli* BL21(DE3) cells ²⁹. For transformation, *E. coli* competence was induced following Inoue's method ³⁰. Expression was induced (at $\text{OD}_{600}=0.4-0.6$) with 1 mM IPTG for 2 h. Cell pellets were resuspended in 50 mM Tris HCl pH 7.5, 150 mM NaCl, and Complete Protease Inhibitor Cocktail Tablets (Roche), and lysed through a French press. Proteins were purified from cleared lysates (45 min, 23,000 rpm) on Ni-NTA agarose columns

(Qiagen), and eluted with a discontinuous imidazol gradient. Pure proteins were visualized by SDS-PAGE electrophoretic protein separation³¹.

***In vitro* peptidoglycan digestion by T6SS effectors**

The extent of T6SS-dependent peptidoglycan digestion was analysed over time (45, 90 and 180 min). 0.4 mg ml⁻¹ of stationary phase muramidase (30 µg ml⁻¹)-digested peptidoglycan from *G. frateurii* was incubated for 90 min on a Tris HCL 20 mM pH:8 with 0.1 mg ml⁻¹ of purified Tse enzyme from *P. aeruginosa* (Tse1Pa) or *A. citrulli* (Tse2Ac). Individual muropeptides were quantified from their integrated areas using samples whose concentration was previously determined by colorimetric quantification of DAP³².

Tse1Pa kinetic parameters were calculated by measuring production rates for different substrate concentrations (1–250 µM) of natural substrates (**2** and **7**). The activities for each substrate concentration were assessed in triplicate. All mean activities were plotted and adjusted to a Michaelis–Menten model using IGOR Pro (version 6.22A, WaveMetrics Inc.) to determine the apparent *K*_m, *V*_{max}, and *k*_{cat} by non-linear regression. *k*_{cat} was determined as *V*_{max}/[*E*₀], where [*E*₀] = nmol of protein ml⁻¹ (His-tagged Tse1Pa). Reactions were done with 0.67 µM Tse1Pa at 37°C in buffer 20 mM Tris HCl pH 8.0 buffer. Reactions were terminated by inactivation at 100°C for 10 min and centrifuged at 14,000 g for 30 min to discard the coagulated Tse1Pa. Next, Tse-treated muropeptides were analysed by UPLC. Determination of the extent of Tse-dependent degradation was performed by comparing the integration areas with respect to a non-treated sample. Statistical analyses were performed with unpaired Student's t-test. Differences were considered statistically significant at *p*<0.01.

Standard muropeptide isolation

Anhydromuropeptides **32** and **33** were produced by using stationary phase peptidoglycan from *G. frateurii* incubated with 100 $\mu\text{g ml}^{-1}$ of Slt70 O.N on Tris HCL 20 mM pH:8³³. Compounds **9**, **6** and **12** were obtained from muramidase-digested *E. coli*²⁶ and *G. frateurii* peptidoglycan, respectively. Compound **29** was obtained from muramidase-digested *L. plantarum* peptidoglycan^{17c}. The solubilized muropeptides were HPLC purified and their identity confirmed by MS.

Multivariate data analysis of peptidoglycan

Data consisted of UPLC Empower liquid chromatography spectra of peptidoglycan samples isolated from the 30 different species of Alphaproteobacteria used in this study and *E. coli* as outlayer. A minimum of 3 sample sets were used for each species. Data included two sets of frequency domains, the standard domain (Abs. 280 nm) and the muropeptide domain (Abs. 204 nm). Chromatograms were collected at a sampling rate of one point per second for 28 min. Internal standard (IS) peptides (Tyr-Ala, Val-Tyr, Val-Tyr-Val, Tyr-Ala-Ala; Tyr-Gly-Gly; Tyr-Gly-Ala; Sigma-Aldrich) were used to give the data fixed anchor points to aid in the alignment step which allows for reliable alignment of chromatograms without using peaks in the muropeptide domain. Baseline correction was performed using the Statistics-sensitive Non-linear Iterative Peakclipping algorithm (SNIP) provided by the MADLquant packet in R. The window size parameter (iterations=15) was set to 15 and the clipping window was set to decrease (decreasing=True) by each iteration. To remove the bleed-through of IS to the muropeptide domain, all, *k*, IS were detected by finding the local maximas within +/- 20 seconds (halfWindowSize=20) in the standard domain using the MALDIquant function detectPeaks. Detected peaks were organized by their signal to noise ratio (SNR) and the *k* peaks with the

highest SNR ratio were identified as the IS. Peak widths were found by looking at when the sum of a window of +/- 3 seconds ceases to decrease. The obtained peak indices were set to zero in the mucopeptide domain. Next, the chromatograms were cut so that retention times between 4 and 20 min were used. Chromatograms were normalized using total ion current normalization as implemented in MALDIquant. The data was then aligned using reference peaks in the samples and IS, then replicate spectra were averaged within species.

For initial clustering, only intensity values between 4 and 14.5 min were used since this time interval comprises all main peptidoglycan components. A PCA was performed on the processed chromatograms and the obtained score vectors were used for clustering. The number of principal components used corresponded to the number of chromatograms in the data. Clustering analysis with wards criterion was carried out on the score matrix of the principal components. Sample similarity was evaluated by means of Euclidean distance. Once initial clustering was generated, clustering on sub-groups was performed on the full time period between 4 and 20 min. Average distances from the mean (and standard deviation) in the 3 clusters were found to be 2.99 (0.84), 3.92 (1.77), 2.07 (0.76) indicating relatively homogeneous clusters. Average distance between the clusters were found to be 7.09 (sd 2.45) indicating substantial differences in the peptidoglycan structures of the samples.

All data analysis was performed using R version 3.1 and data visualization using the ggplot2 package.

Alignment was performed utilizing the retention time of v peaks common in all samples, p_{ij} , $i = 1, \dots, n$, $j = 0, \dots, v$, where $p_{i0} = 0$. Anchor points, a_0, \dots, a_v were then calculated as:

$a_j = \frac{1}{n} \sum_{i=1}^n p_{ij}$. For each sample, the shift, s_{ij} , was then calculated as: $s_{ij} = a_j - p_{ij}$.

Time was then transformed as:

$$f(t_I) = \begin{cases} t_I + \frac{1}{p_{ij} - p_{i,j-1}} [s_{i,j-1}(p_{ij} - t_I) + s_{ij}(t_I - p_{i,j-1})]; & \text{for } p_{i,j-1} \leq t_I < p_{ij} \\ t_I + s_{iv}; & \text{for } p_{iv} \leq t_I \end{cases}$$

Intensity was then interpolated to the new time domain by linear interpolation.

Peaks Widths were found utilizing the found peak index, I_p , and moving a sliding window of size, $winSize$, incrementally towards the left and the right. The windows were calculated for each sample p_{ij} , $i = 1, \dots, n$, as:

$$\begin{cases} I_L = I_P - winSize - 1 \\ I_R = I_P + winSize + 1 \end{cases}$$

$$\begin{aligned} \bar{S}_L &= \sum_{j=I_L}^{I_L+winSize} XX_{ij} \\ \bar{S}_R &= \sum_{j=I_R-winSize}^{I_R} XX_{ij} \end{aligned}$$

Where the peak edges were incrementally moved away from the peak index as:

$$\begin{cases} I_{Lk} = I_L - k \\ I_{Rk} = I_R + k \\ k = 0, \dots, k \end{cases}$$

The peak bounds, B , were found as:

$$\begin{cases} \text{if } \bar{S}_{Lk} > \bar{S}_{L(k-1)} \\ B_L = I_L + \frac{winSize}{2} \end{cases}$$

$$\begin{cases} \text{if } \bar{S}_{Rk} > \bar{S}_{R(k-1)} \\ B_R = I_R - \frac{winSize}{2} \end{cases}$$

Where s denotes the sum of the current window (Left, L , or Right, R). Peak bound is set to the center of the sliding window when the criteria for incremental sliding is not met anymore.

***Drosophila* innate immune system stimulation**

Quantitative analysis of Dipterincine expression was performed, as described previously²⁰. Briefly, total RNA was extracted from cells and cDNA was synthesized by using cDNA synthesis kit (Roche, Basel, Switzerland). Fluorescence real-time PCR was performed using dsDNA dye SYBR Green (PerkinElmer, Boston, MA). SYBR Green analysis was performed on an ABI PRISM 7700 system (PerkinElmer). Primer pairs for Dipterincine (readout) and Rac2 (control) were used to detect target gene transcripts²⁰. All samples were analyzed in triplicate, and the amount of mRNA detected was normalized relative to the control Rac2 values.

Phylogenetic organization correlating cell wall properties

Generation of trees correlating phylogeny and peptidoglycan structural properties of Alphaproteobacteria (Fig. 2 and Fig. 3) was performed using the iTOL (interactive Tree of Life) software version 2.1. (<http://itol.embl.de/>).

Notes

The authors declare no competing financial interests.

Acknowledgements

Thanks to Siv Andersson, Francisco Rodriguez Valera, Solange Oliveira, Encarnación Velázquez, Patrice Nordmann, André Lipski, Arnoux Pascal and Marie-Pierre Chapot for providing strains. We thank Mathew K. Waldor, Jose Berenguer, Patrik Rydén, Keshav Kumar, Waldemar Vollmer and Cava lab members for insightful discussions. We also thank the CBMSO proteomic facility and Thomas Kieselbach for helping with the MS and MS/MS analysis and Fanny Schüpfer for helping with the *Drosophila* immune stimulations. This work was supported by the Laboratory for Molecular Infection Medicine Sweden (MIMS), the Knut and Alice Wallenberg Foundation (KAW), and the Swedish Research Council to F. Cava, Kempe foundation scholarship to AE and UCMR Linneus professorship to MAP and grants CTQ2012-32025 and CTQ2015-64597-C2 of Spanish Ministry of Economy and Competitiveness to FJC.

References

1. Vollmer, W.; Blanot, D.; de Pedro, M. A., Peptidoglycan structure and architecture. *FEMS Microbiol Rev* **2008**, *32* (2), 149-67.
2. Schneider, T.; Sahl, H. G., An oldie but a goodie - cell wall biosynthesis as antibiotic target pathway. *Int J Med Microbiol* **2010**, *300* (2-3), 161-9.
3. Plapp, R.; Strominger, J. L., Biosynthesis of the peptidoglycan of bacterial cell walls. XVII. Biosynthesis of peptidoglycan and of interpeptide bridges in *Lactobacillus viridescens*. *The Journal of biological chemistry* **1970**, *245* (14), 3667-74.
4. (a) Desmarais, S. M.; Tropini, C.; Miguel, A.; Cava, F.; Monds, R. D.; de Pedro, M. A.; Huang, K. C., High-throughput, Highly Sensitive Analyses of Bacterial Morphogenesis Using Ultra Performance Liquid Chromatography. *The Journal of biological chemistry* **2015**, *290* (52), 31090-100; (b) Kuhner, 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.
5. Cava, F.; de Pedro, M. A., Peptidoglycan plasticity in bacteria: emerging variability of the murein sacculus and their associated biological functions. *Current opinion in microbiology* **2014**, *18*, 46-53.
6. Vollmer, W.; Bertsche, U., Murein (peptidoglycan) structure, architecture and biosynthesis in *Escherichia coli*. *Biochimica et biophysica acta* **2008**, *1778* (9), 1714-34.
7. Shah, I. M.; Dworkin, J., Induction and regulation of a secreted peptidoglycan hydrolase by a membrane Ser/Thr kinase that detects muropeptides. *Molecular microbiology* **2010**, *75* (5), 1232-43.
8. (a) Russell, A. B.; Singh, P.; Brittnacher, M.; Bui, N. K.; Hood, R. D.; Carl, M. A.; Agnello, D. M.; Schwarz, S.; Goodlett, D. R.; Vollmer, W.; Mougous, J. D., A widespread bacterial type VI secretion effector superfamily identified using a heuristic approach. *Cell Host Microbe* **2012**, *11* (5), 538-49; (b) Russell, A. B.; Hood, R. D.; Bui, N. K.; LeRoux, M.; Vollmer, W.; Mougous, J. D., Type VI secretion delivers bacteriolytic effectors to target cells. *Nature* **2011**, *475* (7356), 343-7.
9. (a) Cava, F.; de Pedro, M. A.; Lam, H.; Davis, B. M.; Waldor, M. K., Distinct pathways for modification of the bacterial cell wall by non-canonical D-amino acids. *The EMBO journal* **2011**, *30* (16), 3442-53; (b) Lam, H.; Oh, D. C.; Cava, F.; Takacs, C. N.; Clardy, J.; de Pedro, M. A.; Waldor, M. K., D-amino acids govern stationary phase cell wall remodeling in bacteria. *Science* **2009**, *325* (5947), 1552-5.
10. Gupta, R.; Lavollay, M.; Mainardi, J. L.; Arthur, M.; Bishai, W. R.; Lamichhane, G., The *Mycobacterium tuberculosis* protein LdtMt2 is a nonclassical transpeptidase required for virulence and resistance to amoxicillin. *Nature medicine* **2010**, *16* (4), 466-9.
11. (a) Boneca, I. G.; Dussurget, O.; Cabanes, D.; Nahori, M. A.; Sousa, S.; Lecuit, M.; Psyllinakis, E.; Bouriotis, V.; Hugot, J. P.; Giovannini, M.; Coyle, A.; Bertin, J.; Namane, A.; Rousselle, J. C.; Cayet, N.; Prevost, M. C.; Balloy, V.; Chignard, M.; Philpott, D. J.; Cossart, P.; Girardin, S. E., A critical role for peptidoglycan N-deacetylation in *Listeria* evasion from the host innate immune system. *Proc Natl Acad Sci U S A* **2007**, *104* (3), 997-1002; (b) Chaput, C.; Ecobichon, C.; Cayet, N.; Girardin, S. E.; Werts, C.; Guadagnini, S.; Prevost, M. C.; Mengin-Lecreulx, D.; Labigne, A.; Boneca, I. G., Role of AmiA in the morphological transition of *Helicobacter pylori* and in immune escape. *PLoS pathogens* **2006**, *2* (9), e97; (c) Viala, J.; Chaput, C.; Boneca, I. G.; Cardona, A.; Girardin, S. E.; Moran, A. P.; Athman, R.; Memet, S.; Huerre, M. R.; Coyle, A. J.; DiStefano, P. S.; Sansonetti, P. J.; Labigne, A.; Bertin, J.; Philpott,

- D. J.; Ferrero, R. L., Nod1 responds to peptidoglycan delivered by the *Helicobacter pylori* cag pathogenicity island. *Nat Immunol* **2004**, *5* (11), 1166-74.
12. Ruggiero, M. A.; Gordon, D. P.; Orrell, T. M.; Bailly, N.; Bourgoin, T.; Brusca, R. C.; Cavalier-Smith, T.; Guiry, M. D.; Kirk, P. M., A higher level classification of all living organisms. *PloS one* **2015**, *10* (4), e0119248.
13. (a) Sallstrom, B.; Andersson, S. G., Genome reduction in the alpha-Proteobacteria. *Current opinion in microbiology* **2005**, *8* (5), 579-85; (b) Le, P. T.; Pontarotti, P.; Raoult, D., Alphaproteobacteria species as a source and target of lateral sequence transfers. *Trends in microbiology* **2014**, *22* (3), 147-56.
14. (a) Venter, J. C.; Remington, K.; Heidelberg, J. F.; Halpern, A. L.; Rusch, D.; Eisen, J. A.; Wu, D.; Paulsen, I.; Nelson, K. E.; Nelson, W.; Fouts, D. E.; Levy, S.; Knap, A. H.; Lomas, M. W.; Nealson, K.; White, O.; Peterson, J.; Hoffman, J.; Parsons, R.; Baden-Tillson, H.; Pfannkoch, C.; Rogers, Y. H.; Smith, H. O., Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **2004**, *304* (5667), 66-74; (b) Morris, R. M.; Rappe, M. S.; Connon, S. A.; Vergin, K. L.; Siebold, W. A.; Carlson, C. A.; Giovannoni, S. J., SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **2002**, *420* (6917), 806-10.
15. Gillis, M., Intra- and intergeneric similarities of the rRNA cistrons of *Acetobacter* and *Gluconobacter* [proceedings]. *Antonie Van Leeuwenhoek* **1978**, *44* (1), 117-8.
16. Hassani, S.; Martens, H.; M., Q. E.; Kohler, A., degrees of freedom estimation in principal component analysis and consensus principal component analysis. *Chemom. Intell. Lab. Syst* **2012**, *118*, 246-259.
17. (a) Warth, A. D.; Strominger, J. L., Structure of the peptidoglycan from vegetative cell walls of *Bacillus subtilis*. *Biochemistry* **1971**, *10* (24), 4349-58; (b) Mahapatra, S.; Crick, D. C.; McNeil, M. R.; Brennan, P. J., Unique structural features of the peptidoglycan of *Mycobacterium leprae*. *Journal of bacteriology* **2008**, *190* (2), 655-61; (c) Bernard, E.; Rolain, T.; Courtin, P.; Hols, P.; Chapot-Chartier, M. P., Identification of the amidotransferase AsnB1 as being responsible for meso-diaminopimelic acid amidation in *Lactobacillus plantarum* peptidoglycan. *Journal of bacteriology* **2011**, *193* (22), 6323-30.
18. Chaston, J. M.; Newell, P. D.; Douglas, A. E., Metagenome-wide association of microbial determinants of host phenotype in *Drosophila melanogaster*. *mBio* **2014**, *5* (5), e01631-14.
19. Mengin-Lecreulx, D.; Lemaitre, B., Structure and metabolism of peptidoglycan and molecular requirements allowing its detection by the *Drosophila* innate immune system. *Journal of endotoxin research* **2005**, *11* (2), 105-11.
20. Stenbak, C. R.; Ryu, J. H.; Leulier, F.; Pili-Floury, S.; Parquet, C.; Herve, M.; Chaput, C.; Boneca, I. G.; Lee, W. J.; Lemaitre, B.; Mengin-Lecreulx, D., Peptidoglycan molecular requirements allowing detection by the *Drosophila* immune deficiency pathway. *Journal of immunology* **2004**, *173* (12), 7339-48.
21. Cookson, B. T.; Tyler, A. N.; Goldman, W. E., Primary structure of the peptidoglycan-derived tracheal cytotoxin of *Bordetella pertussis*. *Biochemistry* **1989**, *28* (4), 1744-9.
22. (a) Popham, D. L.; Young, K. D., Role of penicillin-binding proteins in bacterial cell morphogenesis. *Current opinion in microbiology* **2003**, *6* (6), 594-9; (b) Moll, A.; Dorr, T.; Alvarez, L.; Davis, B. M.; Cava, F.; Waldor, M. K., A D, D-carboxypeptidase is required for *Vibrio cholerae* halotolerance. *Environ Microbiol* **2015**, *17* (2), 527-40; (c) Ghosh, A. S.; Chowdhury, C.; Nelson, D. E., Physiological functions of D-alanine carboxypeptidases in *Escherichia coli*. *Trends Microbiol* **2008**, *16* (7), 309-17.

23. Mainardi, J. L.; Villet, R.; Bugg, T. D.; Mayer, C.; Arthur, M., Evolution of peptidoglycan biosynthesis under the selective pressure of antibiotics in Gram-positive bacteria. *FEMS Microbiol Rev* **2008**, 32 (2), 386-408.
24. Levefaudes, M.; Patin, D.; de Sousa-d'Auria, C.; Chami, M.; Blanot, D.; Herve, M.; Arthur, M.; Houssin, C.; Mengin-Lecreulx, D., Diaminopimelic Acid Amidation in Corynebacteriales: New insights into the role of lta in peptidoglycan modification. *The Journal of biological chemistry* **2015**, 290 (21), 13079-94.
25. (a) Magnet, S.; Bellais, S.; Dubost, L.; Fourgeaud, M.; Mainardi, J. L.; Petit-Frere, S.; Marie, A.; Mengin-Lecreulx, D.; Arthur, M.; Gutmann, L., Identification of the L,D-transpeptidases responsible for attachment of the Braun lipoprotein to Escherichia coli peptidoglycan. *Journal of bacteriology* **2007**, 189 (10), 3927-31; (b) Magnet, S.; Dubost, L.; Marie, A.; Arthur, M.; Gutmann, L., Identification of the L,D-transpeptidases for peptidoglycan cross-linking in Escherichia coli. *Journal of bacteriology* **2008**, 190 (13), 4782-5.
26. Glauner, B.; Holtje, J. V.; Schwarz, U., The composition of the murein of Escherichia coli. *The Journal of biological chemistry* **1988**, 263 (21), 10088-95.
27. Meadow, P.; Work, E., Bacterial transamination of the stereoisomers of diaminopimelic acid and lysine. *Biochimica et biophysica acta* **1958**, 28 (3), 596-9.
28. Gottlieb, H. E.; Kotlyar, V.; Nudelman, A., NMR Chemical Shifts of Common Laboratory Solvents as Trace Impurities. *The Journal of organic chemistry* **1997**, 62 (21), 7512-7515.
29. Rosenberg, A. H.; Lade, B. N.; Chui, D. S.; Lin, S. W.; Dunn, J. J.; Studier, F. W., Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene* **1987**, 56 (1), 125-35.
30. (a) Inoue, H.; Nojima, H.; Okayama, H., High efficiency transformation of Escherichia coli with plasmids. *Gene* **1990**, 96 (1), 23-8; (b) Hanahan, D., Techniques for transformation of E. coli. In *DNA cloning: A Practical Approach IRL Press, Oxford, United Kingdom* **1985**, vol. 1 (Ed. D.M. Glover), 109-135.
31. Laemmli, U. K.; Favre, M., Maturation of the head of bacteriophage T4. I. DNA packaging events. *Journal of molecular biology* **1973**, 80 (4), 575-99.
32. Work, E., Reaction of ninhydrin in acid solution with straight-chain amino acids containing two amino groups and its application to the estimation of alpha epsilon-diaminopimelic acid. *The Biochemical journal* **1957**, 67 (3), 416-23.
33. Beachey, E. H.; Keck, W.; de Pedro, M. A.; Schwarz, U., Exoenzymatic activity of transglycosylase isolated from Escherichia coli. *European journal of biochemistry / FEBS* **1981**, 116 (2), 355-8.

Table of contents

