

Analysis of peptidoglycan structural diversity using LC-MS/MS and peptidoglycomics software

Marcel Gustavo Alamán Zárate

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield Faculty of Science Department/School of Biosciences.

March 2025

II

Acknowledgement of collaborative work within the thesis

This doctoral thesis is formatted and submitted as a compilation of publications. The candidate hereby declares that all work presented herein is their original contribution, except where explicitly indicated in jointly authored publications. In those instances, the specific contributions of the candidate and co-authors are detailed at the beginning of each respective chapter. The candidate confirms that appropriate credit has been provided throughout this thesis for all the work made by others.

This thesis is structured into 7 chapters. Chapter 3 constitutes an original manuscript prepared specifically for this thesis. Chapters 4 and 5 consist of previously published works. Chapter 6 represents upcoming manuscript with the candidate's contribution.

Personal Acknowledgements

The author gratefully acknowledges the financial support provided by the **Secretaría de Ciencia, Humanidades, Tecnología e Innovación (Secihti),** formerly the Consejo Nacional de Ciencia y Tecnología (Conacyt), through scholarship **2021-000007-01EXTF-00221**.

The author extends sincere gratitude to **Dr Stéphane Mesnage** for his very capable direction during my PhD and his unfailing expressions of support, confidence and interest which were instrument in the successful completion of this work.

The author is grateful to **Dr Caroline Evans** for being by second supervisor and for her involvement in all published research undertaken by the author.

The author is grateful with **Prof Simon Foster** for his invaluable advice and guidance throughout my doctoral studies, and for his willingness to be my examiner.

The author takes this opportunity to thank **Dr Graham Christie** for accepting being my examiner.

The author expresses profound appreciation to all Principal Investigators, colleagues and friends encountered during the past three years of my PhD. Special thanks are due to Abdullah, Alex, Ana, Anirudh, Ankur, Anne, Becca, Becky, Berardo, Brooks (the best and only first coauthor I know), Caro, Charly, Conrad, Danae, Darren, David, Dilah, Ding, Dr Raj, Finn, Gloria, Hannah, Hannah F, Hannah M, Dr I.D.E.A.L, Itzel, Jess B, Jess H, Jessie D, Joe K, Josh, Katie, Laila, Laura, Lea, Lucy, Lucy L, Lucy T, Lucy U, Luz, Mandie, Mariana, Mark, Mawgan, Mia, Miss Yue, Molly, Mr Krishna, Nicola, Pablo, Rosa, Rob, Sol, Sophie, Soroush, Tati, Tayla, Tia, Wilko, Zhihao. Your patience, tolerance, and, most importantly, the time and presence you shared throughout my doctoral journey are deeply cherished. The author apologizes for any unintentional omissions.

Finally, the author has immeasurable gratitude to his family, whose unwavering support was essential to this accomplishment. Though briefly stated, may this acknowledgment convey the depth of appreciation and anticipation for their embrace upon the author's return.

Declaration

I, the author, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (<u>www.sheffield.ac.uk/ssid/unfair-means</u>). This work has not previously been presented for an award at this, or any other, university.

Publications arising from this work

Galley, N. F., Greetham, D., **Alamán-Zárate, M. G.**, Williamson, M. P., Evans, C. A., Spittal, W. D., Buddle, J. E., Freeman, J., Davis, G. L., Dickman, M. J., Wilcox, M. H., Lovering, A. L., Fagan, R. P., & Mesnage, S. (2024). *Clostridioides difficile* canonical L,Dtranspeptidases catalyze a novel type of peptidoglycan cross-links and are not required for β-lactam resistance. The Journal of Biological Chemistry, 300(1), 105529. https://doi.org/10.1016/j.jbc.2023.105529

Alamán-Zárate, M. G^{*}., Rady, B. J.*, Evans, C. A., Pian, B., Greetham, D., Marecos-Ortiz, S., Dickman, M. J., Lidbury, I. D. E. A., Lovering, A. L., Barstow, B. M., & Mesnage, S. (2024). Unusual 1-3 peptidoglycan cross-links in *Acetobacteraceae* are made by L,D-transpeptidases with a catalytic domain distantly related to YkuD domains. The Journal of Biological Chemistry, 300(1), 105494. <u>https://doi.org/10.1016/j.jbc.2023.105494</u> *First co-authors

Alamán-Zárate, M. G^{*}., Rady, B. J.*, Ledermann, R., Shephard, N., Evans, C.A., Dickman, M.J., Turner, R.D., Rifflet, A., Patel, A.V., Gomperts Boneca, I., Poole, P.S., Bern, M., & Mesnage, S. (2025). A software tool and strategy for peptidoglycomics, the high-resolution analysis of bacterial peptidoglycan via LC-MS/MS. Communications Chemistry, 8 (1), 91. <u>https://doi.org/10.1038/s42004-025-01490-6</u>

*First co-authors

Abstract

Peptidoglycan (PG) is a complex and essential macromolecule found in the bacterial cell envelope. The structural analysis of PG requires the purification of this molecule and its enzymatic digestion to produce disaccharide-peptides. PG fragments (called muropeptides) are analysed by liquid chromatography coupled to mass spectrometry (LC-MS/MS) and the structure of the intact molecule is inferred from the composition of muropeptides and their crosslinking patterns. Despite tremendous progress in the LC-MS/MS instrumentation, existing 'Omics software are not suitable, and PG analysis therefore, remains a mostly manual, tedious, biased, and error-prone process. This work builds on the recent development of PGFinder, an open-access software tool dedicated to PG analysis. Using bacteria displaying distinct PG compositions as model systems, this work contributed to optimise LC-MS/MS conditions, the capabilities of PGFinder and provided a detailed and consistent strategy for PG structural analysis. This represents a major step towards making peptidoglycomics a full-fledged discipline.

Table of Contents

LIST OF FIGURES	XI
LIST OF TABLES	XIII
ABBREVIATIONS	XV
1. BACTERIAL CELL ENVELOPES: ARCHITECTURE, FUNCTION, AND DYNAMICS	1
1.1 Cell Envelope Organization: Monoderms and Diderms	1
1.1.1 Monoderms (Gram-positive bacteria)	2
1.1.2 Diderms (Gram-negative bacteria):	3
1.1.3 Other cell envelope types	4
1.2 Peptidoglycan Architecture and Function	5
1.2.1 PG's Role: Exoskeleton, scaffold, environmental interaction	6
1.2.2 PG Architecture: Overall structure and variations	6
1.3 PG BUILDING BLOCKS (DISACCHARIDE-PEPTIDES)	8
1.3.1 Glycan Moiety	8
1.3.2 Peptide Moiety	9
1.4 PG BIOSYNTHESIS	9
1.4.1 Cytoplasmic Steps : UDP-N-Acetylmuramyl Pentapeptide precursor synthesis	10
1.4.2 Membrane Steps: Lipid II production and translocation	10
1.4.3 Extracellular Polymerization:	12
1.4.4 Complexes involved in synthesis and hydrolysis	14
$1.5\mathrm{PG}\mathrm{Remodelling}$	16
1.5.1 Role of Remodelling in Bacterial Physiology and Symbiosis	16
1.5.2 Glycan Chain Modifications: O-acetylation, N-deacetylation, Glycolylation, Muram	ic δ-
lactam rings, and AnhydroMurNAc groups	17
1.5.3 Cleavage by PG Hydrolases: Peptidases, Glycosylhydrolases, Lytic transglycosylases	20
1.5.4 Peptide Stem Modifications	23
1.5.5 Outer and Inner Membrane Tethering (Diderms)	25
1.5.6 Glycan chain length and its regulation	26
1.6 PG STRUCTURAL ANALYSIS METHODS	27
1.6.1 Nuclear Magnetic Resonance (NMR): Solid-state for PG structure and Solution NMI	R for
purified fragments	27
1.6.2 PG Purification and Digestion	28
1.6.3 Chemometric Analysis (rp-HPLC, UV detection)	28
1.6.4 LC-MS/MS Analysis	29
1.6.5 LC-MS/MS Data Analysis Strategies:	29
1.6.6 Muropeptide Analysis limitations	33
1.7 MODEL SYSTEM: RHIZOBIUM LEGUMINOSARUM	34
1.7.1. Importance of R. leguminosarum in agriculture	34
1.7.2 R. leguminosarum Life Cycle	35
1.7.3 Evidence for PG remodelling during R. leguminosarum life cycle	37
1.7.4 R. leguminosarum multigene families encoding L,D and D,D-transpeptidases	37
1.7.5 Potential Role of β -Barrel Proteins	39
1.8 MODEL SYSTEM: CLOSTRIDIOIDES DIFFICILE	39
1.8.1 C. atfficile from microbiota to pathogen	39
1.8.2 Multigene Families: L,D -Transpeptidases and its role	40
1.9 MODEL SYSTEM: GLUCONOBACTER OXYDANS	41
1.9.1 G. oxydans, an Acetobactericeae with industrial relevance	41
1.9.2 G. oxydans' PG shows the presence of 1-3 crosslinks (L,D -Transpeptidase)	42
1.9.3 Possible role of uncommon L,D-transpeptiaation	42
AIMS AND OBJECTIVES	43

2.	MATERIAL AND METHODS	44
	2.1 BACTERIAL STRAINS, PLASMIDS AND OLIGONUCLEOTIDES	44
	2.2 CHEMICALS AND ENZYMES	45
	2.3 GROWTH MEDIA AND BUFFERS	45
	2.3.1 Growth media and conditions	45
	2.3.2 Sodium Borate Buffer for muropentide reduction	
	2.3.3 HPLC buffers for LC and LC-MS/MS	
	2.3.4 Mutanolusin buffer	46
	2.4 METHODS	46
	2.4.1 Extraction and mirification of pentidoalucan	
	2.4.2 Preparation of soluble muronentides	47
	2.4.3 Separation of soluble muronentides by HPLC	47
	2.4.4 LC-MS/MS	
	2.4.5 Plasmid constructions	
	2.4.6 Preparation of G. oxydans competent cells and transformation	49
3	OPTIMIZATION OF LC-MS/MS DATA ACQUISITION	50
	3.1 INTRODUCTION	50
	3.21 MS sourch I dontification of DC monomore using DCE in dom with the Tone framewat	51
	setun database	1000 51
	3.2.2 MS/MS search: Identification of PG monomers using Buos® and a Top5 framentation	setun
	3.2.3 Meraina PGFinder and Buos® analyses using a Top5 fragmentation setup	57
	3.2.4 MS search, identification of PG monomers using PGFinder and a Top25 fraament	ation
	y = 1 $y = 1$	62
	3.2.5 MS/MS search using Buos® and a Top25 fragmentation setup	64
	3.2.6 Merging PGFinder and Byos® analyses for Top25 and Top5 fragmentation setups	65
	3.2.7 Top 5 and Top 25 QualBrowser analysis	66
	3.2.8 Amidation products	69
	3.3 DISCUSSION	73
	3.4 Concluding Remarks	74
4.	CHARACTERIZATION OF CLOSTRIDIOIDES DIFFICILE'S PG COMPOSITION	AND
R	LDT _{CD} S ACTIVITY IN VITRO	75
		77
	4.1 ABSTRACT	/ /
	4.2 INTRODUCTION	8/ 70
	4.3 RESULTS	/9
	4.3.1 In vitro assays win recombinant Laical and Laicas reveat distinct activities and a stamp of L D transmential stime	novei
	42.2 MS/MS and NMP angleson of L dtais and L dtais transportidation products reveal a	/9 monal
	4.5.2 MIS/MIS and MMIL analyses of Lawar and Lawar in anspeptication produces reveal as tame of nontridoglacan cross link	1000ei 01
	A 3.3 TraDIS and usis and delation reweal that the three equivies of C difficile L dts are	01
	4.5.5 Tradis analysis and gene deletion reveal that the three canonical C. alfred dats are	-11011 85
	A 3 A High-resolution structure of the WT and triple mutant pentidoglucans	دہ ۶۶
	4.3.4 Ilight resolution structure of the narental R20291 strain and its isogenic 3Aldt derivati	00 me 90
	4.4 Discussion	νο.90 ΔΛ
	A 5 EXDERIMENTAL PROCEDURES	ب ر مو
	4.5.1 Bacterial strains nlasmids and growth conditions	0و عم
	4.5.2 Construction of C. difficile deletion mutants	90 97
	453 Determination of minimum inhibitory concentrations	رد مە
	4.5.4 Chromosomal DNA extraction sequencing and genome analysis	פר פפ
	455 TraDIS analysis	טיפ ספ
	456 Pentidoalucan extraction	טכ מפ
	4.5.7 Ldt a production and murification	
	4.5.8 Purification of substrates for in vitro assaus	
	J	

4.5.9 In vitro Ldt assaus	99
4.5.10 Preparation of soluble muropeptides for PG structural analysis	
4.5.11 Ultrahigh-Performance chromatography coupled to tandem mass spectrometry	100
4.5.12 Nuclear magnetic resonance	100
4.5.13 Analysis of PG structure	100
4.5.14 Flow cytometry	101
4.5.15 Sporulation	101
4.5.16 Toxin release assays	101
4.6 CONCLUDING REMARKS	102
5. IDENTIFICATION AND CHARACTERIZATION OF AN ENZYME CATALYZING FORMATION OF 1-3 CROSSLINKS IN <i>GLUCONOBACTER OXYDANS</i> PG	THE 104
5.1 Abstract	106
5.2 INTRODUCTION	107
5.3 Results	108
5.3.1 Building a software tool for the structural analysis of 1-3 crosslinked peptidoglycans 5.3.2 High resolution analysis of G. oxydans B58 peptidoglycan	3 108 109
5.3.3 Identification of the L,D-transpeptidase catalyzing the formation of 1-3 crosslinks oxydans	in G. 116
5.3.4 Complementation and heterologous expression experiments show that the LdtGo2 er	ızyme
is sufficient to catalyze peptidoglycan 1-3 crosslinks	118
5.3.5 Ldt_{Go2} is characterised by an atypical YkuD-like catalytic domain that can be fou	nd in
distant families of bacteria with 1-3 $peptidogly can crosslinks$	122
5.4 DISCUSSION	125
5.5 Experimental Procedures	126
5.5.1 Bacterial strains, plasmids, oligonucleotides, and growth conditions	126
5.5.2 Plasmid constructions	127
5.5.3 Preparation of G. oxydans competent cells and transformation	128
5.5.5 Pronanation of coluble magnementides	128 120
5.5.6 LC-MS/MS	120 178
5 5 7 Analysis of nentidoalycan structure	120
5.5.8 Comparative genomics and bioinformatic analysis	129
5.6 CONCLUDING REMARKS	130
6 NOVEL SOFTWARE TOOLS AND STRATEGY FOR DC ANALYSIS AND ITS ADDITCA	τιον
TO EXPLORE PG REMODELLING IN <i>RHIZOBIUM LEGUMINOSARUM</i> PG	110N
	199
6.1 ABSTRACT	135
6.2 INTRODUCTION	136
6.3 RESULTS	13/ J
0.3.1 Ennancing PGFinder's existing functionality with an improved MB output and	1 We0
6 3 2 Condensing complex muropentide structures into PGLang, a concise formal language	157 10 128
6.3.3 Ernanding PAFinder to automate mass calculation and MS/MS fragment prediction	1/0
6.3.4 Rhizobium leguminosarum as a model system for describing a five-step PG an	alysis
6.3.5 Sten1: Identifying PG monomers using MS and MS/MS	145
6.3.4 Step 2. Identifying PG modifications	145
6.3.5 Step 3 Identifying outer membrane proteins covalently anchored to the PG	158
6.3.6 Step 4 Identifying PG multimers and confirming their structure	164
6.3.7 Step 5: Final quantification of muropeptides and comparison of growth conditions.	174
6.4 DISCUSSION	176
6.5 EXPERIMENTAL PROCEDURES	178
6.5.1 Bacterial strains and growth conditions	178
6.5.2 Peptidoglycan extraction and muropeptide preparation 6.5.3 LC-MS/MS analysis	179 179
6.5.4 Determination of glycan chain length and crosslinking index	179

6.5.5 Byos® searches	
6.6 CONCLUDING REMARKS	
7. GENERAL DISCUSSION	
7.1 Contributions to Peptidoglycomics Research: Enhancing Mass Spectrometry Dat	'A ANALYSIS
7.2 PGFINDER IMPROVEMENTS OVER THE COURSE OF THIS PROJECT	184
7.2.1 Moving to more user-friendly platform	184
7.2.2 Making database construction, standard and reliable	184
7.2.3 Using accuracy of mass identification (${\it \Delta}ppm$)	185
7.2.4 Fragment predictor module for MS/MS analyses	185
7.2.5 Analysis of 1-3 crosslinks	185
7.2.6 Batch processing of datasets	186
7.2.7 Drawing PG fragments and predicting their structure using SMILES	
7.2.8 PGFinder analysis caveats	186
7.3 How can we further improve PGFinder?	188
7.3.1 Database building	188
7.3.2 Database building	188
7.3.3 Integrating MS/MS analysis to the search output	188
7.3.4 Novel strategies to Identify PG "dark matter"	
$7.4\mathrm{How}\mathrm{did}\mathrm{PGF}$ inder and this work contribute to our understanding of PG biology	?190
7.4.1 The high sensitivity of LC-MS/MS can identify low abundance features	190
7.4.2 Using high-resolution LC-MS to revisit the specific activity of enzymes invo	lved in PG
synthesis	190
7.4.3 C. difficile, 3-3 crosslinks and antimicrobial resistance	191
7.4.4 Functional diversity of Ldts	191
7.4.5 Exchange reaction	
8. REFERENCES	194
APPENDIX 1	224
APPENDIX 2	237
APPENDIX 3	250

List of Figures

Chapter 1

Figure 1-1. Schematic organization of bacteria cell envelopes.	2
Figure 1-2 PG Glycan moiety orientation and composition	8
Figure 1- 3. Compostion of PG builling blocks.	10
Figure 1- 4. Cytoplasmic steps for PG precursor	11
Figure 1-5. Cytoplasmic transglycosylation during PG biosynthesis	12
Figure 1-6. Two major types of PG crosslinking found in the model bacterium E. coli	13
Figure 1-7 Examples of crosslinks and peptide bridges in PG	15
Figure 1-8 Glycan chain modifications	20
Figure 1-9 PG hydrolases cleavage points	23
Figure 1-10. MurT-GatD amidation in S. aureus	24
Figure 1-11 Timeline showing the major improvements to PGFinder	33
Figure 1-12. Bacterial morphology of <i>S. melioti</i> as free-living cells or as bacteroids in <i>Medicago</i>	
sativa nodules	35
Figure 1-13 Rhizobium leguminosarum lifecycle	36
Figure 1-14 Clostridioides difficile SEM morphology	40

Chapter 2

_		
Figure 2	2- 1. HPLC gradient to separate muropeptides	8

Chapter 3

Figure 3-1. Description of databases used for PGFinder (A) and Byos® (B) searches	52
Figure 3-2. PGFinder v. 0.1.1a Jupyter notebook visual interface (no longer available)	54
Figure 3-3. Top5 Byos® search analysis results	58
Figure 3-4. Summary of Top25 PGFinder search and comparison with Top5 results	62
Figure 3-5. Summary of Top25 Byos® search and comparison to Top5	64
Figure 3- 6. Top25 Byos® search analysis results.	66
Figure 3-7. Top5 and Top25 fragmentation setup compared via Total Ion Chromatograms (TIC).	68
Figure 3-8. PGFinder amidation search	71
Figure 3-9. Strategy for amidation identification	72

Figure 4-1. Ldt <i>in vitro</i> assays	80
Figure 4-2., MS/MS and NMR analysis of unusual muropeptides a and b.	83
Figure 4-3. Unusual peptidoglycan cross-links catalysed by Ldt _{Cd2} and Ldt _{Cd3}	
Figure 4-4. TraDIS analysis of the ldt_{cd} loci.	
Figure 4-5. HPLC-MS chromatogram of C. difficile reduced disaccharide-peptides	88

Figure S4- 1. One-dimensional 1H NMR spectra	
Figure S4- 2. Genome analysis of <i>C. difficile</i> WT and its Δ3ldt derivative	
Figure S4-3. Strategy for C. difficile peptidoglycan structural analysis	
Figure S4- 4. Comparative analysis of WT and $\Delta 3ldt$ cell size by flow cytometry.	92
Figure S4- 5. Comparative analysis of WT and Δ3ldt sporulation efficiency	93
Figure S4- 6. Analysis of toxin releases by C. difficile R20291 and Δ3ldt the mutant	93

Chapter 5

Figure 5-1. Analysis of 3-1 crosslinked peptidoglycans using PGFinder v1.1.01	108
Figure 5-2. HPLC-MS chromatogram of <i>Gluconobacter oxydans</i> reduced disaccharide-peptides1	109
Figure 5-3. Schematic representation of the L,D-transpeptidation reactions leading to the	
formation of 3-3 and 1-3 crosslinks1	111
Figure 5-4. LC-MS detection of 1-3 crosslinks in <i>G. oxydans</i> peptidoglycan	117
Figure 5-5. Heterologous protein synthesis of Ldt _{Go2} in <i>E. coli</i> BL21(DE3) increases the proportion	n
of 1-3 <i>L,D</i> - transpeptidation products1	122
Figure 5-6. LdtGo2 represents a distinct <i>L,D</i> -transpeptidase subclade with a divergent catalytic	
domain that is found primarily in the Acetobacteraceae and Burkholderiaceae1	123
Figure 5-7. Structural analysis of <i>G. oxydans</i> Ldt_{Go2} 1	126
Figure S5-1 . Strategy for <i>G. oxydans</i> peptidoglycan structural analysis1	110
Figure S5- 2. Tandem mass spectrometry analysis of <i>G. oxydans</i> monomers using the Byonic [™]	
module1	114
Figure S5- 3. Complementation of the ldt_{Go2} transposon insertion restores 1-3 crosslinks1	L18
Figure S5- 4. Strategy for peptidoglycan structural analysis of <i>E. coli</i> producing $\mathrm{Ldt}_{^{Go2}}$ 1	120

Figure 6-1. PGFinder's web interface makes MS analysis easier than ever and enables new	
functionality	138
Figure 6-2. Muropeptides are represented as a set of chemically linked, optionally modified	
residues.	139
Figure 6-3. PGLang is a simple language for describing potentially modified, branched, or cross-	
linked muropeptides	140
Figure 6-4. R. leguminosarum's PG compositon is comples and varies between growth condition	ns.
	143
Figure 6- 5. An end-to-end strategy for PG structural analysis via LC-MS/MS	144
Figure 6-6. The existence of AnhydroMurNAc-modified monomers can be confirmed via MS/MS.	158
Figure 6- 7. Validation of gm-AEJ=gm-AEJA and gm-AEJA=gm-AEJ mass coincidence	165
Figure 6-8. <i>R. leguminosarum</i> grown in minimal media (MM) have different PG composition	
compared to those grown in rich media (TY).	176
Figure S6- 1. Railroad syntax diagram for PGLang	141
Figure S6- 2. Validated and non-validated MS/MS spectra	154
Figure S6- 3. Summary of modifications analyses.	159
Figure S6- 4. Putative β-barrel proteins encoded by <i>Rhizobium leguminosarum</i> bv. <i>viciæ</i> strain	
3841	161
Figure S6- 5. The existence of β -barrel anchoring scars is confirmed via MS/MS	163

List of Tables

Chapter 1

<u>+</u>	
Table 1-1 Offline LC-MS/MS manual data analysis (Glauner et al., 1988)	
Table 1-2. List of putative PBPs encoded by <i>Rlv</i> 3841	
Table 1-3. List of putative Ldts encoded by Rlv3841	
Table 1-4. C. difficile Ldts enzymatic activities reported by Sütterlin, et al., 2018	41

Chapter 2

-			
Table 2-1. Bacterial strains,	plasmids, and	oligonucleotides	 4

Chapter 3

▲	
Table 3-1. Db_Cd1 composition	53
Table 3- 2 Top5 Db_Cd1 PGFinder Search summary	55
Table 3-3. Top5 PGFinder search output shows 14 mass coincidences (highlighted in blue)	
Table 3- 4. Top5 Db_Cd10 PGFinder Search summary	
Table 3-5. Consolidated PGFinder and Byos® results for Top5 dataset	60
Table 3-6. All Top25 PGFinder search are present in Top5. 9 mass coincidences highlighted	in blue
	63
Table 3-7. Consolidated PGFinder and Byos® results for Top25 dataset and Top5 Byos® va	lidated
monomers	67
Table 3-8 Qual Browser MS1 statistics to highest and low abundant peaks	68
Table 3-9. DB_Cd2: 95% of the total monomers' intensity	69

Chapter 4

Table 4-1. Muropeptide analysis of C. difficile R20291 and triple ldt_{cd2} , ldt_{cd2} , ldt_{cd3} mutant ($\Delta 3l$	dt)
Table 4-2. Summary of WT and $3\Delta ldt$ PG properties	91 92
Table S4-1. Chemical shift assignments for the dimer peptidoglycan fragment a Table S4- 2. Minimum inhibitory concentration (MIC) of β-lactams against <i>C. difficile</i> R20291 a:	85 nd
ldt_{Cd} isogenic mutants.	94
Table S4-3. Quantification of Ldt _{Cd2} endopeptidase activity against a 3-4 crosslinked dimer ^a Table S4-4 Bacterial strains, plasmids, and oligonucleotides	95 97

Table 5- 1. G. oxydans peptidoglycan composition	115
Table S5-1. List of peptide sequences used for <i>G. oxydans</i> MS/MS searches	112
Table S5-2. Databases for G. oxydans PGFinder searches	116
Table S5-3. List of peptide sequences used for <i>E. coli</i> MS/MS searches.	
Table S5- 4. DB0_Ec.	
Table S5-5. Species, TaxIds, and genome accessions used for comparative genomics	
Table S5- 6. Strains, plasmids and oligonucleotides	127

Table 6-1. A list of the PG modifications searched for, and the strategy used to validate them	157
Table 6-2. PGFinder identification of <i>N</i> -terminal peptides from RopA1,2,3,RopB and pRL90069	
anchored to gm-AEJ	164
Table 6-3. Comparative muropeptide analysis of PG extracted from cells grown in TY or MM	175
Table S6- I. Railroad syntax diagram for PGLang	141
Table S6- 2. Residues built into PGFinder	142
Table S6- 3. Modifications built into PGFinder	142
Table S6- 4. Database 1 (DB_1) Description	145
Table S6- 5. Rhizobium leguminosarum unbiased search with DB_1.	147
Table S6- 6. Monomers identified by MS/MS in TY datasets	155
Table S6-7. Cumulated abundance of monomers validated by MS/MS	156
Table S6- 8. Database 3 (DB_3)	156
Table S6- 9. Database 4 (DB_4)	162
Table S6- 10. PGFinder identification of dimers and trimers	166
Table S6- 11. <i>In silico</i> fragmentation of the two dimers (gm-AEJ=gm-AEJA (4-3) ang gm-AEJA=gm-	-
AEJ (3-3)) and consolidated list of expected ions	168
Table S6- 12. List of validated dimers	170
Table S6-13. List of unmodified muropeptides in TY1, TY2 and TY3 samples	171

Abbreviations

C°	Degree Celsius
aa	amino acids
AFM	Atomic Force Microscopy
Amp	Ampicillin
AnhydroMurNAc	anhydromuramic acid
B. subtilis	Bacillus subtilis
C-	Carboxyl
C. difficile	Clostridioides difficile
Cm	Chloramphenicol
COSY	COrrelation SpectroscopY
C-terminal	Caboxyl-terminus
Da	Dalton
D-Ala	D-Alanine
D-Ala-D-Ala	D-Alanyl-D-Alanine
D-Gln	D-Glutamine
D-Glu	D-Glutamate
DNA	Deoxyribonucleic acid
E. coli	Escherichia coli
E. faecalis	Enterococcus faecalis
E. faecium	Enterococcus faecium
FDAA	Fluorescent D-amino acid
G. oxydans	Gluconobacter oxydans
GlcNAc	N-acetylglucosamine
HPLC	High Performance Liquid Chromatography
IPTG	lsopropyl β -D-1-thiogalactopyranoside
iso-Gln	iso-Glutamine
Kan	Kanamycin

kDa	kiloDalton
L	liter
L-Ala	L-Alanine
LB	Luria Bertani
LC	Liquid Chromatography
LC-MS	Liquid chromatography coupled to mass spectrometry
Ldt	L,D-transpeptidase
L-Lys	L-Lysine
LPS	Lipopolysaccharides
L-Ser	L-Serine
LTA	Lipoteichoic acids
Μ	Molar, (mol/liter)
m/z	Mass to charge ratio
meso-DAP	meso-diaminopimelic acid
min	minutes
MM	Minimum Media
MS	Mass Spectrometry
MS/MS	Tandem mass spectrometry
MurNAc	N-acetylmuramic acid
N-	Amino
NaBH ₄	Sodium borohydride
ND	Not Detected
NMR	Nuclear Magnetic Resonance
N-terminal	Amino-terminus
OD ₆₀₀	Optical Density at 600 nm
PBP	Penicillin-Binding Protein
PCR	Polymerase Chain Reaction
PG	peptidoglycan

рН	log ₁₀ [H ⁺]		
R. leguminosarum	Rhizobium leguminosarum		
RIv3841	Rhizobium leguminosarum biovar viciae strain:3841		
ROESY	Rotating frame Overhauser Effect SpectroscopY		
rpm	revolutions per minute		
RT	Retention time		
S. aureus	Staphylococcus aureus		
SMILES	Simplified Molecular Input Line Entry System		
ТА	Teichoic Acids		
TEM	Transmission Electron Microscopy		
TG	Transglycosylation		
TheoMw	Theoretical Molcular Weight		
TLC	Thin Layer Chromatography		
TOCSY	TOtal Correlation SpectroscopY		
ТР	Transpeptidation		
TY	Tryptone Yeast extract media		
UDP	Uridine 5'-diphosphate		
UV	ultraviolet		
V.	version		
v/v	Volume per volume		
w/v	Weight per volume		
WT	Wild Type		
WTA	Wall teichoic acids		
xg	centrifugal force		
Δ ppm	difference between the measured mass and the theoretical mass, expressed in ppm		
μ	micro		
-OH	hydroxyl group		

Other abbreviations are explained in the text where appropriate

Amino acid	Three letter code	One letter code
Alanine	Ala	А
Arginine	Arg	R
Asparagine	Asn	Ν
Aspartic acid	Asp	D
Asparagine or aspartic acid	Asx	В
Cysteine	Cys	С
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glutamine or glutamic acid	Glx	Z
Glycine	Gly	G
Histidine	His	Н
Isoleucine	lle	I
Leucine	Leu	L
Lysine	Lys	K
meso- diaminopimelic acid	mDAP	յ
Methionine	Met	М
Phenylalanine	Phe	F
Proline	Pro	Р
Serine	Ser	S
Threonine	Thr	Т
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V
amidated <i>meso-</i> diaminopimelic acid	amidated mDAP	J _{NH2}

Chapter 1 Introduction

1. Bacterial Cell Envelopes: Architecture, Function, and Dynamics

1.1 Cell Envelope Organization: Monoderms and Diderms

The bacterial cell envelope is a complex, multilayered structure essential for bacterial survival, shape, and environmental interactions (Cochrane and Lohans, 2020; Dörr et al., 2019). Most bacteria have a peptidoglycan (PG) cell wall, which is crucial for osmotic stability. PG sacculi is a mesh-like macromolecule of cross-linked glycan chains to provide mechanical strength and anchor to cell wall components. It is dynamic, undergoing remodelling during growth and division (Beveridge, 1999; Egan et al., 2015),. The PG layer, forms the basis of the traditional Gram classification, differentiating bacteria based on differential staining properties (Peabody et al., 2016; Seltmann and Holst, 2002).

While both Gram-positive and Gram-negative bacteria have a PG layer, key distinctions exist. These differences rely in the peptide composition, crosslinking extent, and PG chain length, rather the fundamental chemical structure (Harper and Hernandez, 2020; Tocheva et al., 2016). Critically, not all bacteria adhere strictly to the Gram classification. Mycoplasma species, for example, lack a cell wall and do not retain Gram stain (Desvaux et al., 2018b; Kieser and Rubin, 2014). Therefore, a refined understanding of cell envelope architecture, encompassing PG characteristics and membrane organization, is essential for accurate bacterial classification.

The number of surrounding membranes is a way to define cell envelope architecture. The terms "monoderm" and "diderm" describe bacteria with one or two biological membranes, respectively (Desvaux et al., 2018b; Gupta, 1998). Gram positive (monoderm) usually have a thick PG layer and Gram negative (diderm) bacteria usually have thin PG layer but this is not a strict rule (Léonard et al., 2022). Both groups can synthesize varying thickness of PG and even remodel one form into the other (Tocheva et al., 2016).



Figure 1-1. Schematic organization of bacteria cell envelopes.

A. Gram-positives have one cell membrane with a thick PG layer. B. Gram-negatives are enveloped by two membranes with a thin PG layer between them. C. Other cell envelopes such as Mycobacteria incorporate arabinogalactan and mycolic acids into their PG staining as Gram-positive organisms. (Modified from Beaud Benyahia et al., 2025)

1.1.1 Monoderms (Gram-positive bacteria)

Monoderms, commonly referred to as Gram positive bacteria, are characterized by a relatively simple cell envelope architecture. They display a thick, multilayered PG layer, typically 20–80 nm thick and composed of 10–20 layers (Beaud Benyahia et al., 2025; Silhavy et al., 2010b). The PG layer can constitute 30–90% of the dry weight of the bacterial cell wall (Orsini Delgado et al., 2024). It is often directly exposed to the external environment (Pazos and Peters, 2019) and is responsible for Gram positive staining (Rogers and Perkins, 1968) (Figure 1-1A).

The thick, hydrophilic, and porous PG layer allows for high water capacity (Chancock, 2002). Several components are associated with it, including proteins, polysaccharides, and cations. Polymers can be covalently associated or non-associated with PG, including teichoic acids (TAs), capsules, secondary cell wall polymers, and S-layers (Hashimi and Tocheva, 2024; Siegel et al., 2016). Wall teichoic acids (WTAs) are phosphate-containing polymers (often polyglycerol or polyribitol phosphate) covalently linked to PG, while lipoteichoic acids (LTAs) are phosphate-containing polymers anchored to the cytoplasmic membrane and embedded within the cell wall (Desvaux et al., 2018b; Silhavy et al., 2010a). WTAs are anionic and can comprise 30-60% of the monoderm cell wall's dry weight (Beveridge, 1978).

The linkage between WTAs and PG is typically a phosphodiester bond involving the C6 hydroxyl group of the N-acetylmuramic acid (MurNAc) residue (Araki and Ito, 1989; Brown et al., 2013). While WTAs are commonly linked to MurNAc, less commonly, certain capsular polysaccharides (CPS), such as the capsule of *Streptococcus pneumoniae*, are linked to N-acetylglucosamine (GlcNAc) residues of PG (Larson and

Yother, 2017). The linkage for the S. pneumoniae capsule to GlcNAc in the PG is a direct glycosidic bond to the C6 of GlcNAc (Larson and Yother, 2017).

PG-anchored polymers maintain cell shape, participate in cell division, and impart a negative charge, facilitating extracellular metal cation binding (Brown et al., 2013; Rohde et al., 2019; Sonnenfeld et al., 1985). In pneumococci , virulence-related proteins attached to choline residues of TAs, aiding host colonization (Vollmer et al., 2019). WTAs are essential for maintaining cell shape, regulating cell division, coordinating growth and imparting a negative charge that facilitates extracellular cation binding (Brown et al., 2013; Rohde et al., 2019; Sonnenfeld et al., 1985). WTAs are vital for the proper localization, assembly, and activation of cell wall machinery, including penicillinbinding proteins (PBPs) (Brown et al., 2013). Furthermore, WTAs are key in regulating autolytic activity (autolysis), often by controlling the localization or activity of peptidoglycan hydrolases (Brogan and Rudner, 2023; Rohde et al., 2019; Wu et al., 2016). For instance, in *S. pneumoniae*, teichoic acid modifications influence where PG hydrolases bind, affecting autolysis (Brogan and Rudner, 2023).

Bacteria lacking WTAs exhibit significant defects in cell morphology and division (Brown et al., 2013; Weidenmaier and Peschel, 2008). These defects can arise from mutations in essential WTA biosynthesis genes, such as *tagO*, which catalyzes an initial, conserved step in the pathway. In *Bacillus subtilis and Listeria monocytogenes*, for example, cells with such mutations may lose their characteristic shape and become spherical (Brown et al., 2013). Teichuronic acids represent another class of anionic polymers covalently linked to the PG layer via phosphodiester bonds to MurNAc residues (Deng et al., 2010; Ward and Curtis, 1982).Teichuronic acids can be incorporated into the cell wall under phosphate-limiting conditions (Navarre and Schneewind, 1999; Wright and Heckels, 1975). In *Micrococcus luteus*, teichuronic acids are composed of glucuronic and mannuronic acid repeats (Deng et al., 2010).

LTAs regulate division and osmoprotection; in *B. subtilis*, mutants lacking LTAs are sensitive to low manganese concentrations (Schirner et al., 2009), and *S. aureus* LTA mutants require high sucrose or salt for viability (Corrigan et al., 2011). However, TAs synthesis varies among monoderms; instead some produce PG-anchored polysaccharides that decorate the external surface of the cell (Kampff et al., 2023). Some monoderms feature protein crystalline S-layers, these are non-covalently interacting with PG or TAs (Chancock, 2002; Desvaux et al., 2018b). Compared to diderms which possess an outer membrane, the direct accessibility of the monoderm cell wall and cytoplasmic membrane potentially increases susceptibility to β -lactam antibiotics targeting PG assembly, (Kawai and Errington, 2023). β -lactam resistance in monoderms is often mediated by altered penicillin-binding proteins (PBPs) acquired via horizontal gene transfer (Armstrong et al., 2021).

1.1.2 Diderms (Gram-negative bacteria):

Diderms are commonly referred to as Gram-negative bacteria. They possess an inner cytoplasmic membrane, a thin PG layer and an outer membrane that provides additional protection to the bacteria (Figure 1-1B) (Lithgow et al., 2023). This outer membrane prevents the retention of the crystal violet during Gram staining, leading to

their classification as Gram-negative (Rogers and Perkins, 1968). The cytoplasmic membrane is a phospholipid bilayer that acts as a barrier surrounding the entire cell (Garde et al., 2021, 2021). The thin PG layer occupies the periplasmic spaces, sandwiched between the cytoplasmic and outer membranes. The PG layer in diderms is typically a few nanometres thick (Beaud Benyahia et al., 2025; Mathelié-Guinlet et al., 2020; Rohde et al., 2019) and often consists of a single layer of PG (Turner et al., 2018).

The periplasmic space, located between the cytoplasmic membrane and the outer membrane, not only contains PG it also has a high concentration of proteins (Stock et al., 1977). The outer membrane is an asymmetric lipid bilayer, composed of phospholipids on the inner leaflet and lipopolysaccharides (LPS) in the outer leaflet (Zhang et al., 2013). It provides an additional permeability barrier, reducing susceptibility to antibiotics (Grasekamp et al., 2023; Hwang et al., 2018; Silhavy et al., 2010b; Webby et al., 2022). The outer membrane contains β -barrel proteins that play an important role in transport (Silhavy et al., 2010b; Sun et al., 2022). In *E. coli*, a model diderm bacterium, the PG within the periplasm is anchored to the outer membrane by a C- terminal lysine in Braun's lipoprotein (Lpp) (Braun, 1975; Pazos and Peters, 2019; Silhavy et al., 2010b). Lipopolysaccharides function as endotoxins and are essential for antibiotic survival (Silhavy et al., 2010b; Tamaki et al., 1971).

Some diderms produce extracellular polysaccharide capsules, which are variable in structure for phagocytosis resistance (Desvaux et al., 2018b; Waz et al., 2022). These capsules composition range from linear to branched polymer, are major virulence factors and aid to serotype characterization (Kampff et al., 2023; Mistou et al., 2016).

The diderm architecture is considered the most ancestral (Grasekamp et al., 2023). While both monoderm and diderm PG share a similar glycan structure, their peptide compositions differ (Yadav et al., 2018). Diderm PG typically contains *meso*-diaminopimelic acid (mDAP) in peptide stems, whereas monoderms frequently contains L-Lysine (Ghuysen, 1968). However, some diderms, such as *Deinococcus* and *Thermus*, have thick PG layers despite having an outer membrane and lacking LPS (Hashimi and Tocheva, 2024; Léonard et al., 2022)

1.1.3 Other cell envelope types

The Gram classification is a useful tool that provides information about the organisation of bacterial cell envelopes. However, many bacteria cannot be easily classified using this method. *Cyanobacteria* (a diderm) often stain Gram-positive despite having an outer membrane. Mollicutes (*Spiroplasma, Mycoplasma,* and *Acholeplasma*) appear Gram-positives but contain only lipids and cholesterol in their envelope, and no peptidoglycan (Otten et al., 2018; Trachtenberg, 2005). *Actinomycetia* (including *Mycobacteria*) incorporate arabinogalactan and mycolic acids into their PG, resulting in a unique cell wall architecture (Figure 1-1C) (Pazos and Peters, 2019; Walter and Mayer, 2019).

Mycobacteria represent an interesting case. They are grouped as Gram positive, due to genetic similarities and having a thick PG layer, but share some Diderm characteristics, such as the presence of porins in their outer lipid layer (Hett and

Rubin, 2008; Walter and Mayer, 2019). *Mycobacteria* have a thick mycolyl arabinogalactan-PG layer and an outer lipid layer containing porins (Radkov, 2018; Hett, 2008; Porfirio, 2018). This outer layer (mycomembrane), rich in mycolic acids, act as a pseudo-outer membrane, creating a permeability barrier similar to Gram negatives (Alderwick et al., 2015; Brown et al., 2020). *Corynebacteriales*, which are closely related to *Mycobacteria*, also possess a mycolic acid outer layer but lack other outer membrane markers, classifying them as "neoderms" (Beaud Benyahia et al., 2025; Zuber et al., 2008).

Thermotogae shows an outermost layer but lack LPS, suggesting an evolutionary intermediate between monoderm and diderm structures (Hashimi and Tocheva, 2024). It is hypothesized that the last bacterial common ancestor possessed a diderm envelope, with monoderms arising from multiple outer membrane losses (Beaud Benyahia et al., 2025; Tocheva et al., 2016). *Cyanobacteria* and *Deinococcota* are diderms with thick PG layers. *Chloroflexota* further complicates typical diderm definitions by lacking outer membrane but probably having an outer lipidic layer (Beaud Benyahia et al., 2025). *Planctomycetes*, while previously thought to lack PG, are now known to possess discontinuous PG sacculi, generating internal compartmentalization (Jeske et al., 2015; Otten et al., 2018). Obligate intracellular bacteria (*e.g., Chlamydiales, Anaplasmataceae*) often have a limited PG content (Otten et al., 2018).

Archaea are distinct from bacteria, typically lacking PG *sensu stricto* (Kandler and König, 1978; Mukhopadhyay, 2024; Walter and Mayer, 2019). Their cell walls are composed of diverse materials, including S-layers (often the sole cell wall component) (Rodrigues-Oliveira et al., 2017; Sleytr et al., 2014), pseudomurein (in methanogens) (van Wolferen et al., 2022), linked lipids, and other polymers (Albers and Meyer, 2011). Some archaea, such as *Ignicoccus hospitalis*, have a double membrane (van Wolferen et al., 2022), while others, like *Thermoplasma*, lack a cell wall entirely. Archaea also exhibit unique appendages like cannulae (Albers and Meyer, 2011).

1.2 Peptidoglycan Architecture and Function

PG, also known as murein, is a unique and essential component of most bacterial cell walls. This elastic mesh-like polymer plays a critical role in maintaining cell shape, providing protection against osmotic pressure, and serving as a scaffold for anchoring proteins and other cell surface molecules in bacteria (Pazos and Peters, 2019; Vollmer et al., 2008a). While the core structure of PG is conserved, variations in its composition contribute to the diverse characteristics of different bacterial species.

PG consists of long glycan chains composed of alternating *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) residues linked by β -1,4-glycosidic bonds. These glycan chains are crosslinked by short peptide side chains attached to the MurNAc residue via a lactyl group. A variation is the presence of *N*-glycolylmuramic acid in mycobacteria, where MurNAc is modified with a glycol group. The peptide side chains can also vary in composition across species. For example, *E. coli* typically has L-

Ala-D-*i*Glu-mDAP-D-Ala-D-Ala, while *S. aureus* uses L-Ala-D-iGln-L-Lys-D-Ala-D-Ala and *Mycobacteria* have L-Ala-D-iGlu-mDAP-D-Ala-D-Ala (Garde et al., 2021; Johannsen, 1993; Lederer et al., 1975).

PG synthesis initiates in the cytoplasm with the formation and transport of a precursor (lipid II). This precursor is then flipped across the cell membrane and integrated into existing PG through transglycosylation (glycan chain polymerization) and transpeptidation (peptide crosslinking) (Garde et al., 2021; Johannsen, 1993). This process occurs outside the cell membrane and is mediated by various enzymes, including D,D-transpeptidases (Penicillin binding proteins, PBPs) and L,D-transpeptidases (Ldts).

The structure of PG is dynamic, undergoing constant remodelling during growth and in response to environmental changes. This dynamic nature is essential for cell division and adaptation. Enzymes like PG hydrolases (autolysins) are responsible for breaking, remodelling, and recycling PG (Rohde et al., 2019). Lytic transglycosylases control glycan strand length (Yadav et al., 2018). These processes are tightly regulated and represent targets for antibiotics, particularly β -lactams, which inhibit transpeptidation (Tipper and Strominger, 1965).

While PG is a major component of eubacterial cell walls, other molecules contribute to the overall structure and function of the cell envelope. These include teichoic acids in monoderm bacteria, lipopolysaccharides (LPS) in diderm bacteria, and mycolic acids and arabinogalactan in mycobacteria. For example, the presence of mycolic acid in *Corynebacterium glutamicum* is associated with glutamate excretion (Lanéelle et al., 2013).

1.2.1 PG's Role: Exoskeleton, scaffold, environmental interaction

PG has a role as an exoskeleton which confers cell shape and resistance to the intracellular osmotic pressure, also is used as a scaffold for the display of proteins, polysaccharides or teichoic acids at the cell surface (Bugg, 1999; Godessart et al., 2021). Among the variety of bacterial products, PG plays an important role in modulation of immune response, because PG is present in nearly all bacterial species (Ealand et al., 2018; Kühner et al., 2014).

1.2.2 PG Architecture: Overall structure and variations

The PG sacculus, a closed structure of crosslinked glycan strands, was initially proposed to have a regular and quasi-crystalline organization (Tipper, 1970). However, structural analyses (X-ray diffraction, EM, AFM) have demonstrated that the PG sacculus is rather an elastic network. AFM studies have further revealed heterogeneous regions within the PG sacculus, including ring, knobbles, and pores of variable size (Vollmer et al., 2008a).

Two main models have been proposed to describe the PG architecture: the layered model and the scaffold model. In the layered model the glycan chains run parallel to the cytoplasmic membrane (Figure 1-2A) (Höltje, 1998; Koch, 1998; Yao et al., 1999).

Conversely, the scaffold model proposes that glycan chains run perpendicular to the cytoplasmic membrane (Figure 1-2B) (Dmitriev et al., 2003; Dmitriev et al., 2005). Advancements in Cryo-TEM and AFM have provided insights in both monoderms and diderms (Gan et al., 2008; Hayhurst et al., 2008; Turner et al., 2010; Wheeler et al., 2011).

Based on Cryo-TEM (Cryo-transmission electron microscopy) and AFM (Atomic force microscopy), evidence suggest the classical layered model for diderms, where both glycan strands and crosslinks run parallel to the cytoplasmic membrane (Gan et al., 2008). The scaffold model is not compatible to diderms because vertical alignment of the long glycan chains would exceed the maximum sacculus thickness (Dmitriev et al., 2003; Szwedziak et al., 2014; Vollmer and Höltje, 2004). Accordingly, cryo-electron tomography (CET) observations in *Caulobacter crescentus* revealed loosely packed and disordered PG sheets running in parallel to the bacteria membrane, favouring the layered model (Gan et al., 2008). AFM studies have found large pore structures distributed randomly in the sacculus, consistent with loose configuration (Turner et al., 2013).

Monoderms have more elaborate and species-specific PG architectures (Hayhurst et al., 2008; Turner et al., 2010; Wheeler et al., 2011). In *Bacillus subtilis*, a model proposing coiled PG cables encircling the cell cylinder has been suggested (Hayhurst et al., 2008). However, CET studies have proposed a uniformly dense cell wall model with circumferentially oriented glycan strands (Beeby et al., 2013). Using a combination of AFM data from live-cells and isolated sacculi alongside Cryo-TEM data obtained from purified sacculi, it was shown that *B. subtilis* cell wall is not a uniform layer. Instead, it has a structure with two regions of distinct densities: the outer surface having a dynamic gel-like structure and the inner surface having a denser and more organised arrangement (Pasquina-Lemonche et al., 2020).

Staphylococcus aureus, a monoderm coccus, displays a distinctive architecture. The first model proposed septal PG synthesis in a spiral pattern, stretching peptides during expansion (Seligman, 1987; Seligman and Pincus, 1987). Electron microscopy and AFM revealed concentric rings at division sites and a fibrous network at old poles (Giesbrecht et al., 1998; Touhami et al., 2004; Turner et al., 2010). Further AFM studies showed that S. aureus undergoes dynamic PG remodelling at the division site, resulting in a knobbly architecture and features that delimit previous divisions (Turner et al., 2010). These features are suggested to function as epigenetic factors for division site localization, allowing division in three planes (Turner et al., 2010). Given the short glycan chains in S. aureus (6 disaccharide units), a vertical alignment of highly crosslinked PG would offer greater resistance to mechanical stress, favouring the scaffold model (Dmitriev et al., 2003). However, solid-state NMR revealed an ordered and densely packed structure with a parallel orientation to the membrane, consistent with the layered model (Kim et al., 2013). Like B. subtilis, the mature surface of live S. aureus cells is characterised by a disordered porous gel of peptidoglycan, in contrast to the much denser inner PG surface. The nascent septal material forms a dense ring architecture with circumferentially oriented strands. These rings transition to a more porous state during cell wall maturation and are influenced by peptidoglycan synthesis and hydrolysis enzymes like PBP4 and SagB (Pasquina-Lemonche et al., 2020).

The three-dimensional architecture of glycan strands is a topic of ongoing investigation. Despite significant progress, other cell wall components like teichoic acids are not fully integrated in these PG architecture models.

1.3 PG Building Blocks (Disaccharide-Peptides)

1.3.1 Glycan Moiety

The glycan moiety of PG is composed of a disaccharide made of *N*-acetylglucosamine (2-Acetamido-2-Deoxy-D-Galactose, GlcNAc) and *N*-acetylmuramic acid (2-acetamido-3-O-[(1R)-1-carboxyethyl]-2-deoxy- β -D-glucopyranose, MurNAc). These two sugars are linked together via β -1-4 bonds (Figure 1-2C) (Vollmer et al., 2008a), forming a right-handed helix with 3 disaccharide molecules per turn, this β -1-4 linkage distinguish PG from other glycan polymers like cellulose and chitin (Meroueh et al., 2006).





The length of these glycan strand can vary between species, monoderms like *S. aureus* has mostly short glycan polymers (around 6 disaccahride units) (Harz et al., 1990), whereas *E. coli* (diderm) is 5-30 and *B. subtilis* (monoderm) have chains ranging from 96-500 disaccharide units (Hayhurst et al., 2008; Ward, 1973). It is important to note that both *E. coli* and *S. aureus*, also produce very long glycan chains (Turner et al., 2018; Wheeler et al., 2011; Wheeler et al., 2015), providing strong evidence against the scaffold model.

1.3.2 Peptide Moiety

The peptide moiety is attached to MurNAc subunit via a lactyl group. The peptide stem precursor is typically a chain of 5 amino acids attached to MurNAc, these pentapeptide stems contain both L- and D- amino acids, the latter being absent in proteins (Kühner et al., 2014; Vollmer et al., 2008a). The stem composition is conserved in a given bacterium but varies across species.

A common pentapeptide stem composition is L-Ala- γ -D-Glu-mDAP (or L-Lys)-D-Ala-D-Ala (mDAP, *meso*-diaminopimelic acid) (Figure 1-3A). The common configuration starts in position 1 with L-Alanine and it is highly conserved; the Glutamic acid usually follows in position 2 linking via its gamma carbon and leaving the alpha carbon to be amidated after the incorporation of this residue, like in *Streptococcus, Enterococcus and Lactobacillus* (Bugg, 1999; Schleifer and Kandler, 1972).

A major difference is found in position 3, which corresponds to a di-amino acid. In most diderm bacteria, such as *E. coli*, this is *meso*-DAP. However, an exception to this is *Bacillus subtilis*, a monoderm bacterium (Scheffers and Pinho, 2005). In other diderm organisms, position 3 can be L-Lys (*e.g. Enterococcus and Lactobacillus*) or L-ornithine/homoserine (Figure 1-3B). Lateral chains (Figure 1-3C) are found mostly in Gram-positive pathogenic cocci, with a variable composition and length (1 to 5 aa).

The two D-Ala residues at the C-terminus of the peptide stems are usually attached as a dipeptide (Figure 1-3B). The terminal D-Ala residue is removed in the mature molecule (Barreteau et al., 2008; Vollmer et al., 2008a). Some vancomycin-resistant strains (*Enterococcus faecium, Lactobacillus casei*) have D-lactate in position 5. In this case, vancomycin binds less efficiently to the peptide stem (Allen and Hobbs, 1995; Handwerger et al., 1994).

1.4 PG Biosynthesis

PG biosynthesis is a highly conserved process among bacteria. It involves nearly 20 enzymatic reactions occurring in three cellular locations: the cytoplasm (nucleotide precursor synthesis), the inner and outer face of the cytoplasmic membrane (lipid-linked intermediate synthesis and polymerization, respectively) (Barreteau et al., 2008; Garde et al., 2021; Lovering et al., 2012; Pazos and Peters, 2019; Shaku et al., 2020; Vollmer et al., 2008a).

1.4.1 Cytoplasmic Steps : UDP-N-Acetylmuramyl Pentapeptide precursor synthesis The initial stage of PG synthesis involves the cytoplasmic assembly of monomers. Fructose-6-phosphate is converted to UDP-GlcNAc by GlmSMU. Subsequently UDP-MurNAc is synthesized from UDP-GlcNAc. Amino acids of the pentapeptide stems are assembled by successive ligation reactions and added to UDP-MurNAc by MurC, MurD, MurE and MurF to produce UDP-MurNAc-pentapeptide. In most bacterial species, the first amino acid is L-alanine, as well as several compounds that are structurally related to L-Ala, for example in *Mycobacterium leprae*, the first amino acid of the stem peptide is glycine instead of L-Ala (Bouhss et al., 2008; Lovering et al., 2012) (Figure 1-4).



Figure 1-3. Composition of PG building blocks.

A. The disaccharide-pentapeptide PG motif found in the model bacterium *E. coli* is shown. Sugars are shown in black with the lactyl group of the MurNAc residue in red. Amino acids in the pentapeptide stem are shown in blue (1st position, R1), pink (2nd position, R2), green (3rd position, R3), orange (4th position, R4) and brown (5th position, R5). Some bacteria contain a lateral chain crossbridge (Rx), attached to the R3 residue, that can vary in composition and length. **B.** Variations in the peptide stem composition varies among species; some examples are shown in the table. **C.** Examples of lateral chain crossbridges, present in different organisms. meso-A2pm, meso-dioaminopimelic acid. D-Asx, partially amidated D-Asp.

1.4.2 Membrane Steps: Lipid II production and translocation

The next step occurs on the inner leaflet of the cytoplasmic membrane (Figure 1-4). It begins when UDP-MurNAc-pentapeptide is linked to an undecaprenyl lipid carrier (C_{55} -P), a reaction catalysed by the MraY transferase. The molecule formed is known as Lipid I. The MurG transferase then attaches UDP-GlcNAc to Lipid I, thereby generating Lipid II (Pazos & Peters, 2019; Shaku et al., 2020; Bouhss et al., 2008).

Next, Lipid II is translocated across the cytoplasmic membrane by MurJ (Kruijff et al., 2008; Meeske et al., 2015; Pazos and Peters, 2019; Sham et al., 2014). The disaccharide monomers in the outer cytoplasmic space can then be linked to pre-existing glycan strands via glycosyltransferases; transpeptidases (PBPs and Ldts) form peptide crosslinks (Böth et al., 2013). The lipid carrier with pyrophosphate is shuttled back to the cytosolic side of the membrane, where it can be reused for another round of synthesis (Bouhss et al., 2008; Kruijff et al., 2008).





The figure shows the enzymes required for Lipid II biosynthesis and the substrates. Synthesis starts with formation of UDP-GlcNAc and UDP-MurNAc in the cytoplasm. UDP -GlcNAc is synthesized from fructose 6-phosphate (F-6-P) by GlmSMU, UDP-MurNAc is synthesized by the addition of enolpyruvate to UDP-GlcNAc. The pentapeptide moiety is added to UDP-MurNAc in sequential ATP dependent reactions. Phospho MurNAc-pentapeptide is transferred to a lipid carrier resulting in the formation of lipid I and releasing UMP. The subsequent transfer of GlcNAc to lipid I results in the formation of lipid II and release of UDP. In the final step lipid II is translocated to the outer part of the membrane, where it is used as substrate for new PG (Modified from Shaku et al., 2020).

1.4.3 Extracellular Polymerization:

Once Lipid II is translocated into the periplasm, lipid precursors are incorporated into the existing PG. The incorporation of glycan chains and crosslinking of peptide stems is done by the activity of 2 classes of Penicillin Binding Proteins (PBP) (Figure 1-5 and 1-6)

1.4.3.1 Transglycosylation

Class A PBPs, such as PBP1a and PBP1b in *S. pneumoniae* or PonA1 and PonA2 in *M. tuberculosis*, play a key role in PG strand polymerization (Kieser and Rubin, 2014; Machowski et al., 2014). These PBPs have *N*- terminal domains that catalyze glycosyltransferase reactions, using monomeric-, tri-, tetra- and pentapeptide glycan chains as acceptors. The reaction catalyzed involves the displacement of the α -diphospho-undecaprenyl group by the C4 hydroxyl of GlcNAc, resulting in a β -1,4-linkage (Figure 1-5). Glycan chain elongation proceed through successive attacks of the growing glycan chain (donor) at the reducing end of lipid II (acceptor) (Bugg, 1999; Sauvage et al., 2008). Recent findings have shown that the enzymes FtsW and RodA are capable of catalyze glycan chain polymerization using Lipid II as substrate (Miyachiro et al., 2019; Straume et al., 2021).



Figure 1-5. Cytoplasmic transglycosylation during PG biosynthesis

The transglycosylation reaction involves the displacement of the α -diphospho-undecaprenyl group by the C4 hydroxyl group of GlcNAc, forming a β 1-4 linkage (Adapted from Bugg, 1999).

1.4.3.2 D,D-Transpeptidation (PBPs)

Penicillin Binding Proteins (PBPs) from class A and class B, have *D*,D-transpeptidase domains. *D*,D-transpeptidases recognize the D-Ala-D-Ala residues of a donor stem peptide. They catalyze the formation of a covalent intermediate with the D-Ala in position 4, subsequently linking the carboxyl (-COOH) group of this residue to the amino (NH₂) group of the mDAP residue at position 3 of an acceptor peptide stem, resulting in the formation of 4–3 cross-links (Figure 1-6A). As PBPs, D,D-transpeptidases are inhibited by β -lactam antibiotics, which are structural analogues of D-Ala-D-Ala (Figure 1-6C) (Ealand et al., 2018; Zapun et al., 2008a).

The transpeptidation reactions catalyzed by PBPs follow a three-step mechanism: (i) the rapid, reversible formation of a noncovalent complex between the enzyme and a PG stem pentapeptide [L-Ala- γ -D-Glu-mDAP (or L-Lys)-D-Ala-D-Ala], termed the donor strand; (ii) the formation of the covalent acyl-enzyme intermediate; (iii) diacylation, which involves crosslink formation with an acceptor PG stem peptide (transpeptidation). In bifunctional PBPs, transglycosylation can proceed while the transpeptidase domain is inhibited by penicillin, mutated, or deleted. Inactivation of the glycosyltransferase domain, completely blocks the PG polymerization (Ealand et al., 2018; Sauvage et al., 2008).

1.4.3.3 L,D-Transpeptidation (Ldts)

PG crosslinks can be formed by D,D-transpeptidases (PBPs) as well as L,Dtranspeptidases (Ldts) (Fig 1-6B). The latter enzymes recognize the last two residues of a tetrapeptide donor stem substrate, which is generated by a D,D-carboxypeptidase. Ldts cleave the mDAP-D-Ala bond of the donor tetrapeptide, forming a covalent intermediate between the mDAP at position 3 and the enzyme. Subsequently, the Ldt links the carboxyl group of the donor mDAP to the amino group of the mDAP of an acceptor peptide stem, creating 3-3 crosslinks. Ldts are inhibited by carbapenems, a specific β -lactam antibiotic class, which form a thioester bond with the cysteine residue in the active site (Fig 1-6D) (Aliashkevich and Cava, 2021; Dubée et al., 2012; Lecoq et al., 2013; Mainardi et al., 2005a).



Figure 1-6. Two major types of PG crosslinking found in the model bacterium *E. coli*

A. D-D transpeptidases catalyse crosslinking between the 4th amino acid (D-Ala) in the donor peptide stem and the 3rd residue (mDAP) of the acceptor stem. **B.** L-D transpeptidases catalyze crosslinking between the 3rd residue (mDAP) of the donor peptide stem and the 3rd residue (mDAP) of the donor peptide stem and the 3rd residue (mDAP) of the acceptor stem **C.** Penicillin resembles D-Ala-D-Ala from pentapeptide.

While Ldts are not always essential (e.g., in *E. coli*), they perform critical roles in maintaining cell wall integrity and morphology. Their diverse functions include incorporating non-canonical D-amino acids into PG, contributing to β -lactam resistance, maintaining cell envelope integrity through L,D-transpeptidation, and tethering outer membrane proteins to PG (Aliashkevich and Cava, 2021; Espaillat et al., 2024). However, in some bacteria, L,D-transpeptidation is essential. For example, *C. difficile* and *A. tumefaciens*, which possess 5 and 14 *ldt* genes, respectively, cannot survive without at least one of these enzymes (Aliashkevich et al., 2024; Bollinger et al., 2024).

1.4.3.4 Crosslinking: 3-4, 3-3, bridges and unusual crosslinks (1-3, 2-4)

The most common PG crosslink is the product of D,D-transpeptidation, 4-3 crosslinks (Sauvage et al., 2008) (Figure 1-7A), bacteria exhibit a diverse array of PG crosslinks, often modulated by environmental conditions and growth stage. The second most prevalent type is the 3-3 crosslink, product of L,D-transpeptidase, implicated in antibiotic resistance, survival, and cell wall remodelling (Aliashkevich et al., 2024) (Figure 1-7B). A third, unusual class, the 1-3 crosslink, is found in *Acetobacteraceae*, though the enzyme remains unidentified (Espaillat et al., 2016) (Figure 1-7C).

Monoderm bacteria frequently produce branched PG precursors due to lateral chains added to the diamino acid at the third position. These lateral chains vary in length and sequence. Common crossbridges include L-Ala-L-Ser in *S. pneumoniae* (Filipe and Tomasz, 2000) (Figure 1-7D), D-Asx in *E. faecium* (Billot-Klein et al., 1996) (Figure 1-7E), and Gly-L-Ser pentapeptides in *S. aureus* (Jonge et al., 1993) (Figure 1-7F). These variations contribute to diverse PG properties and likely influence cell physiology and environmental interactions (van Heijenoort and Gutmann, 2000). Branching enzymes, responsible for bridge synthesis, utilize substrates like lipid II (e.g., *S. aureus*), UDP-MurNAc-pentapeptide, or both (Vollmer et al., 2008a).

Finally, 2-4 crosslinks, unique to *Corynebacteria*, connect D-Glu at position 2 of the acceptor to D-Ala at position 4 of the donor, involving a diamino acid (D-Orn) interpeptide bridge through D,D-transpeptidation (Vollmer et al., 2008a) (Figure 1-7G).

1.4.4 Complexes involved in synthesis and hydrolysis

Bacterial cell wall biosynthesis is a dynamic process essential for coordinating cell growth, division, and shape maintenance (Daitch and Goley, 2020; Rohs and Bernhardt, 2021). Peptidoglycan (PG) undergoes continuous remodelling, requiring tightly regulated enzymatic synthesis, modification, and hydrolysis (Hayhurst et al., 2008; Rohs and Bernhardt, 2021). This coordination prevents lethal lysis and ensures structural integrity (Miyachiro et al., 2019; Wang et al., 2021).



Figure 1-7 Examples of crosslinks and peptide bridges in PG.

A. 4-3 crosslinks, B. 3-3 crosslinks, C. 1-3 crosslinks, D. Alanine-Serine crossbridge, E. Aspx crossbridge,
F. Glycine-Serine crossbridge, G. 2-4 crossbridge. Crosslinks shown in green.

PG synthesis relies on two major enzymatic groups: bifunctional class A penicillinbinding proteins (aPBPs) and SEDS (shape, elongation, division, sporulation) glycosyltransferases, which partner with class B PBPs (bPBPs) for transpeptidation (Egan et al., 2020; Sjodt et al., 2020). These SEDS-bPBP complexes form the core machinery for PG synthesis within two primary structures: the elongasome (for lateral cell wall expansion) and the divisome (for septal peptidoglycan formation) (Egan et al., 2020; Straume et al., 2020; Wang et al., 2021). Additionally, L,D-transpeptidases (LDTs) and aPBPs contribute to dispersed PG synthesis, facilitating expansion, reinforcement, and repair (Daitch and Goley, 2020; Garner, 2021; Straume et al., 2021) (Fig. 1-7cont.).

Multiprotein complexes regulate PG remodeling by linking PG synthases and hydrolases, ensuring hydrolytic activity occurs exclusively at sites of new synthesis (Miyachiro et al., 2019; Rohs and Bernhardt, 2021). Protein-protein interactions finely control enzyme activity, with scaffolding proteins like MreB (elongasome) and FtsZ (divisome) precisely positioning the machinery (Egan et al., 2020; Tinajero-Trejo et al., 2025). Specific interactions include *Escherichia coli* PBP1B associating with the lytic transglycosylase MItA and scaffolding protein MipA, while outer membrane lipoproteins (LpoA and LpoB) stimulate PG synthase activity (Egan et al., 2014; Egan et al., 2015; Garner, 2021)

PG hydrolase regulation occurs via activation, auto-inhibition, and proteolytic degradation (Egan et al., 2020; Rohs and Bernhardt, 2021; Yang et al., 2012). In *E. coli*, the D,D-endopeptidase MepS undergoes rapid degradation by Prc, facilitated by the

lipoprotein NIpI (Jeon and Cho, 2022). Additionally, PG hydrolases exhibit distinct subcellular localizations; in *Bacillus subtilis*, LytE is positioned helically along the cylindrical wall by MreBH, ensuring proper cell wall maturation and turnover (Carballido-López et al., 2006). This highly coordinated network of enzymatic interactions ensures bacterial cell wall integrity while adapting to growth and environmental stressors.



Figure 1-7 (cont.) Peptidoglycan biogenesis

In rod-shaped bacteria, peptidoglycan (PG) synthesis occurs via three major modes. Elongation relies on the actin-like protein MreB to guide patchy PG insertion along the lateral walls, away from the poles. Septal synthesis during division is mediated by the tubulin-like homologue FtsZ, which forms a ring at the cell center, promoting zonal PG synthesis for septum formation. Dispersive synthesis can function independently from elongation and septal synthesis, and contributes to PG enlargement, maintenance, and modification (Adapted from Daitch and Goley, 2020).

1.5 PG Remodelling

Although the composition of PG building blocks is always the same, both glycan chains and peptide stems undergo modifications either on the lipid II or after they are incorporated in the existing PG network. This process is called remodelling. It is a dynamic process involving continuous modifications and turnover of the existing PG structure, and is important for the cell to grow, divide, maintain their shape and adapt to environmental changes, such as sporulation or motility (Vermassen et al., 2019).

1.5.1 Role of Remodelling in Bacterial Physiology and Symbiosis

Diderms have an envelope stress response system to monitor and translate environmental stress into physiological adjustments for survival (Delhaye et al., 2019). In *E. coli*, antibiotics can trigger responses, such as activating lytic transglycosylases (Slt) and inducing 3-3 crosslinking, which is regulated by YgaU (Bernal-Cabas et al., 2015). The specific YgaU function potentially could degrade the cell wall (Delhaye et al., 2019).

Monoderms like *Baciilus* and *Clostridium* can form spores in response to nutrient starvation. These spores can germinate when the conditions become optimal (Tobin

et al., 2023). This capability is essential for transmission, infection, and antibiotic evasion by pathogens like *Clostridioides difficile* (Vonberg et al., 2008). Mature spores have 2 PG layers: germ cell wall and a thick outer layer named cortex which is deacetylated, loosely cross-linked, and evenly spaced with muramic- δ -lactam residues (up to 50% in *B. subtilis*) (Driks and Eichenberger, 2016; Popham et al., 1996a; Tobin et al., 2023).

Peptidoglycan is a crucial microorganism-associated molecular pattern (MAMP) that mediates the association between the non-pathogenic symbiont *Vibrio fischeri* and the Hawaiian bobtail squid (*Euprymna scolopes*). A peptidoglycan fragment released by lytic transglycosylase activity (a disaccharide-tetrapeptide referred to as the tracheal cytotoxin) triggers developmental changes in the squid. Sensing of this fragment by the squid induces the morphogenesis of the juvenile squid's light organ. Early changes include the removal of superficial ciliated epithelium to promote the bacteria recruitment and the construction of internal structures (Nyholm and McFall-Ngai, 2021). The influence of the symbiont bacteria extends beyond the light organ, modulating host gene expression in distant tissues such as eyes and gills, facilitating the maturation of the host's innate immune system by altering haemocyte responses (Nyholm and McFall-Ngai, 2021).

1.5.2 Glycan Chain Modifications: O-acetylation, N-deacetylation, Glycolylation, Muramic $\,^\delta$ -lactam rings, and AnhydroMurNAc groups

PG glycan chains undergo modifications on both GlcNAc and MurNAc sugars (Figure 1-8). These modifications occur at the C2 (amino group, -NH2) and C6 (hydroxyl, -OH) positions, where original groups are substituted. At C6 position occurs *O*-acetylation, and 1,6-anhydro-MurNAc formation, while *N*-deacetylation, *N*-glycolylation, and δ lactam ring formation occur at the C2 position. These glycan chain modifications are involved in bacterial survival, sporoulation, host-pathogen interactions, and the regulation of PG remodelling during growth (Vollmer, 2008).

O-acetylation

The addition of an *O*-acetyl group to the C6 carbon of MurNAc and less commonly to GlcNAc (Bernard et al., 2011) occur in both diderms and monoderms (Figure 1-8B and 1-8D). First described in *S. aureus* (Bera et al., 2005), *O*-acetylation plays an important role for pathogen resistance to host immune responses (Herbert et al., 2007). In monoderms, this modification is catalyzed by a single enzyme, OatA (Bera et al., 2005), whereas in diderms, it involves two enzymes, PatA and PatB (Sychantha et al., 2018; Weadge et al., 2005). *O*-acetylation inhibits the activity of antimicrobial cationic peptides and the PG hydrolytic activity of lysozyme, which is a key effector of the host innate immune system . This inhibition occurs by sterically hindering access to the glycan β -1,4 linkages (Clarke, 1993; Sychantha et al., 2018). Furthermore, because the C6 hydroxyl group of MurNAc is also modified in 1,6-anhydro-MurNAc, in *Neissseria gonorrehae O*-acetylation can regulate the activity of the lytic transglycosylase, which are responsible for 1,6-anhydro-MurNAc formation (Weadge and Clarke, 2006).
N-deacetylation

N-deacetylation, the removal of the C2 amino group from GlcNAc or MurNAc, mostly occurs in monoderms and was first described in *S. pneumoniae* (Vollmer and Tomasz, 2000). This process, catalyzed by deacetylases, involves the removal of the acetyl group at C2 position (Figure 1-8E and 1-8G) and occurs after lipid II translocation. *N*-deacetylation confers resistance against lysozyme and is important for virulence (Boneca et al., 2007; Sychantha et al., 2018).

For example, in monoderms like *C. difficile*, PG is highly deacetylated, with up to 93% glucosamine (GlcN) (Peltier et al., 2011). Similarly, in diderms such as *H. pylori*, *N*-deacetylation by PgdA is upregulated under oxidative stress during virulence, reducing recognition by the host receptors (Boneca et al., 1997; Coullon et al., 2020; Wang et al., 2009).

Glycolylation

N-glycolylation is a MurNAc modification exclusively found in *Actinomycetales* species such as mycobacteria (Azuma et al., 1970; Yadav et al., 2018). This modification involves the substitution of the acetyl group by a hydroxyl group at the C2 position (Figure 1-8F) and occurs during the cytoplasmic stage of PG synthesis. *N*-glycolylation confers resistance to β -lactam antibiotics and lysozyme. In *Mycobacterium smegmatis*, this modification is carried out by a single enzyme called NamH (Raymond et al., 2005). While the functional relevance of this modification remains unknown, it is hypothesized that it provides additional strength to the PG mesh through hydrogen bond formation (Brennan and Nikaido, 1995).

AnhydroMurNAc groups

In diderms, glycan strands are terminated with 1,6-anhydro muramic acid (AnhydroMurNAc) as a component of PG maturation. This modification results from lytic transglycosylase (LT) activity (Höltje et al., 1975). It involves the formation of an intramolecular ring between the C6 and C1 carbons of the MurNAc sugar ring. The presence of 1,6-anhydro muramic acid marks the end of glycan strands, preventing further polymerisation (Höltje et al., 1975). It also serves as a signalling molecule, inducing PG recycling and β -lactamase synthesis in *E. coli* (Jacobs et al., 1997; Johnson et al., 2013). Furthermore, this modification inhibits the activity of autolysins (Figure 1-8C) (van Heijenoort, 2011). Most diderms encode multiple lytic transglycosylase genes; for instance, *E. coli* encodes 12 of such enzymes, two of which are essential (Lee et al., 2024).

In diderms AnhydroMurNAc-muropeptides are the major released muropeptide species (Gilmore and Cava, 2025). Although *O*-acetylation of AnhydroMurNAc is not possible, a deacetylase (MdaA) capable of cleaving the acetyl group from AnhydroMurNac was recently described in *Agrobacterium tumefaciens* (Gilmore et al., 2024). While the biological role of this enzyme remains under investigation, it

suggests the potential for additional, previously undescribed muropeptide modifications.

Muramic δ -lactam rings

Muramic δ -lactam rings are a unique structural modification found in the PG of bacterial spores, specifically within the thick outer layer known as cortex (Tobin et al., 2023). This modification involves the formation of an intramolecular amide bond within a muramic acid residue (Gilmore et al., 2004) (Fig 1-8H). Biosynthesis occurs through a two-step enzymatic process involving two enzymes. First, CwID, an amidase, removes the peptide stem from MurNAc (Kim et al., 2023). Subsequently, PdaA catalyzes the deacetylation of MurNAc to MurN and the subsequent cyclization of MurN to form the muramic δ -lactam ring, by creating a bond between the lactyl group at C3 of MurN and the amino group at C2. (Gilmore et al., 2004; Tobin et al., 2023).

This modification plays a role in spore germination. In *B. subtilis*, mature spores have up to 50% of the MurNAc in the cortex converted to muramic- δ -lactam (Atrih et al., 1996; Popham et al., 1996a). During germination, these muramic- δ -lactam residues are recognized by spore cortex lytic enzymes, facilitating cortex degradation while preventing degradation of the cell wall PG (Christie and Setlow, 2020; Francis and Sorg, 2016). In *B. subtilis*, CwID mutants produce spores lacking muramic δ -lactam rings and are unable to properly degrade the cortex (Gilmore et al., 2004).



Figure 1-8 Glycan chain modifications

The structure of the unmodified GlcNAc-MurNAc disaccharide (middle) and of selected modifications in the GlcNAc (red) and MurNAc (blue) are shown. Modifications are highlighted in green. The O-acetylation of GlcNAc or MurNAc is reversible. Pep, stands for a pentapeptide linked to MurNAc. Based on Yadav et al.,2018

1.5.3 Cleavage by PG Hydrolases: Peptidases, Glycosylhydrolases, Lytic transglycosylases

PG remodelling is primarily mediated by a diverse group of enzymes known as PG hydrolases or autolysins. These enzymes can cleave glycan chains or peptide stems; in fact, a hydrolase has been identified for virtually every bond linking the PG amino acid and sugar components. Occasionally, multiple enzymatic activities are present within a single protein (Vollmer et al., 2008b). These enzymes are essential for various bacterial processes that require modification of the existing PG structure, including cell growth, cell division, cell separation, cell wall turnover and maintenance of cell shape (Blackman et al., 1998; Foster, 1992). PG hydrolases often belong to multigene families and exhibit functional redundancy. They are classified based on the type of bond they hydrolyse (van Heijenoort, 2011).

Peptidases

<u>Amidases</u> hydrolyse the amide bond between MurNAc and the first amino acid (L-Ala) from the peptide stem, separating the glycan strand from the peptide stem. (Figure 1-9A) (Turner et al., 2014). *E. coli* possesses five amidases: AmiA, AmiB, AmiC, AmiD and AmpD (Bernhardt and Boer, 2003). The periplasmic amidases AmiA, AmiB and AmiC are important for cell separation at the division site (Heidrich et al., 2001). AmiD is a lipoprotein anchored to the outer membrane and does not participate in cell separation (Uehara and Park, 2007). AmpD, a cytoplasmic enzyme, cleaves 1,6-anhydro muramic acid from the peptide stem, facilitating PG recycling (Uehara and Park, 2007).

<u>Endopeptidases</u> are enzymes that can be either soluble or membrane associated, and they cleave amide bonds between amino acids of different stem peptides (Figure 1-9B). They are classified based on the bond they hydrolyze: D,D-endopeptidases (DD-EPs) cleave between two D-amino acids, L,D-endopeptidases (LD-EPs) cleave between an L- and a D-amino acid, and D,L-endopeptidases (DL-EPs) when they cleave between a D and an L-amino acid.

D,D-EPs cleave the 4-3 crosslink between D-Ala at position four and the residue at position three (mDAP in *E. coli*) of a different peptide stem (Sauvage et al., 2008). *E. coli* has two types of DD-EPs: low molecular weight PBPs which are inhibited by β - lactam antibiotics (PBP4 and PBP7) and DD-EP unrelated to PBPs (MepA) (Keck et al., 1990).

L,D-EPs cleave the 3-3 cross-link bond between residues at position three of different stem peptides (Turner et al., 2014). In diderms, LD-EPs degrade the L-alanine-iso-D-glutamate bond and are important for specific growth morphologies (Fukushima et al., 2007). In *E. coli*, MepK possess LD-EP activity (Chodisetti and Reddy, 2019), although monofunctional L,D-EPs are reported to be less common in diderms (Gondré et al., 1973).

DL-EPs cleave the iso-D-glutamate-mDAP bond and are important for cell division and virulence in monoderms (Rico-Pérez et al., 2016). RipA in *M. tuberculosis* is a DL-EP involved in cell separation (Healy et al., 2020). Tse1 in *P. aeruginosa* is a DL-EP used in bacterial warfare, degrading competing bacteria's PG and causing lysis (Shang et al., 2012).

<u>Carboxypeptidases</u> (CPs) remove carboxyl-terminal residues from stem peptides (Figure 1-9C). Like endopeptidases, carboxypeptidases are classified based on the type of bond they hydrolyse (Turner et al., 2014).

D,D-carboxypeptidases (DD-CPs) cleave the bond between two D-amino acids, specifically the terminal D-Ala residues in a pentapeptide stem, resulting in a tetrapeptide stem (Ghosh et al., 2008). DD-CPs limit the amount of pentapeptides available for DD-transpeptidation; in their absence, excessive transpeptidation occurs, leading to unbalanced PG insertion (Glauner, 1988). Many low molecular weight PBPs in *E. coli* exhibit D,D-carboxypeptidase activity, including PBP4, PBP4b, PBP5, PBP6, PBP6b and AmpH (Ghosh et al., 2008).

L,D-carboxypeptidases (LD-CPs) cleave the terminal D-Ala from tetrapeptide stems, forming tripeptides (van Heijenoort, 2011). The tripeptide product is important for PG recycling system in diderms, providing a substrate for MurF to synthesize UDP-MurNAc-peptapeptide (Das et al., 2013; Templin et al., 1999).

Glycosylhydrolases

Glycosylhydrolases, muramidases or glycosidases, are enzymes that cleave glycosidic bonds within the glycan strands of PG (Vollmer et al., 2008b). They include:

<u>*N*-acetylglucosaminidases</u> hydrolyse the glycosidic bond between GlcNAc and MurNAc, leaving GlcNAc with a reducing end (Figure 1-9D). A well-characterised PG *N*-acetylglucosaminidase from *S. aureus*, AtlA, is a bifunctional enzyme containing an amidase domain (Vollmer et al., 2008b). This activity is critical for cell growth without new synthesis by hydrolysing septal PG (Turner et al., 2010; Wheeler et al., 2015).

<u>*N*-acetylmuramidases</u> (lysozymes) cleave the β -1,4- glycosidic bond between MurNAc and GlcNAc, generating MurNAc with a reducing end. Lysozymes hydrolyze the glycosidic bond, resulting in a product with a terminal reducing MurNAc residue (Figure 1-9E). These enzymes are produced by phages, bacteria, fungi, vertebrates, and invertebrates (Herlihey et al., 2014). For instance, pesticin, a bacteriocin from diderms, is described as muramidase (Sibinelli-Sousa et al., 2021).

Lytic transglycosylases

Lytic transglycosylases (LTs) are not hydrolases since they do not use water during catalysis. These periplasmic enzymes cleave the β -1,4 glycosidic bond between MurNAc and GlcNAc, simultaneously performing an intramolecular transglycosylation reaction. This results in the formation of a 1,6-anhydro ring at the MurNAc residue of the product (Figure 1-9F) (van Heijenoort, 2011). Notably, because LTs require a free C6 hydroxyl group (C6-OH) to form the 1,6-anhydro ring (Vollmer, 2008), they cannot cleave the glycan group of *O*-acetylated MurNAc. Therefore, it is hypothesized that the equilibrium between *O*-acetylation and de-*O*-acetylation serves as a regulatory mechanism for lytic transglycosylase activity in diderms (Frirdich and Gaynor, 2013).



Figure 1-9 PG hydrolases cleavage points

A. Amidases cleave the link between MurNAc and the first amino acid in the pentapeptide. **B.** Endopeptidases cleave the links in the middle of the pentapeptide. **C.** Carboxypeptidases cleave the link in the carboxyl-terminal amino acid of the pentapeptide. **D.** N-acetylglucoaminidases cleavage leaves GlcNAc with free reducing end in C1 and MurNAc with a free reducing end in C4. **E.** N-acetylmuramidases cleavage leavage leaves MurNAc with a free reducing end in C1 and GlcNAc with a free reducing end in C4. **F.** Lytic transglycolases cleave GlcNAc and MurNAc with the formation of the 1,6-anhydro ring at the MurNAc residue of the product.

1.5.4 Peptide Stem Modifications Amidation

Among PG-forming amino acids, D-Glu and *m*DAP in the stem peptide and D-Asp on the lateral chain are amino acids with free carboxyl groups than can be amidated (Figure 1-3). These modifications take place intercellularly, before the translocation through the membrane through the amidation of UDP-MurNAC-pentapeptide or lipid intermediates (Figure 1-10). Both *Lactobacillus lactis* and *E. faecium* have been found to amidate D-Asp, which make them less sensitive to lysozyme and cationic antimicrobials. *Lactobacillus plantarum* has shown amidation of *m*DAP to regulate the septation process and carboxypeptidase activity (Chapot-Chartier and Kulakauskas, 2014) In *S. aureus* the complex MurT-GatD converts D-iGln to D-iGlu in the internal cell membrane (Figure 1-10). GatD metabolizes free glutamine and provides an ammonia group to MurT, a ligase that converts D- Glu to D- Gln (Leisico et al., 2018; Münch et al., 2012). The inhibition of amidation reduces bacterial growth rate, resistance to β -lactams and increases sensitivity to lysozyme (Chapot-Chartier and Kulakauskas, 2014).



Figure 1-10. MurT-GatD amidation in *S. aureus*

The MurT amidation mechanism is similar to a ligase that requires the ammonia from the glutaminase GatD (Modified from Leisico et al., 2018).

AA exchange in position 4 (exchange by Ldts)

In addition to their crosslinking activity, Ldts can catalyze exchange reactions and replace the C-terminal D-Ala residue in tetrapeptide stems by a non-canonical D-amino acid (Caparrós et al., 1992; Cava et al., 2011). The high-resolution PG structure in *C. diffcile* revealed that many amino acids can be incorporated in the peptide stem (Bern et al., 2017), suggesting that Ldts are not selective when it comes to use of D-amino acids for exchange reaction. This agrees with the *in vitro* experiments carried out to characterize *E. faecium* Ldt_{fm}, since this enzyme can use the dipeptide L-Lys-D-Ala for exchange with D-amino acid and D-2-hydroxy acid. Ldt_{fm} displayed no exchange activity when L- amino acids were used (Mainardi et al., 2005b).

AA exchange in position 5 (exchange by LMW PBPs in enterococci)

PBPs can catalyze exchange reactions in PG. *S. aureus* PBP4 can exchange D-amino acids into peptide chains *in vitro* using lipid I and II precursors as a substrate (Qiao et al., 2014). This reaction has been monitored using fluorophores. Two LMW PBPs from *E. faecalis* and *S. gordonii* with a similar exchange activity have also been identified. They can exchange the fifth residue of PG pentapeptide precursors with D-Lys *in vitro* (Welsh et al., 2017). Evidence of this activity in *E. coli* was first described as D-amino acid incorporation in the fifth position of purified peptide stems (Kuru et al., 2019)*E. coli*; subsequently evidence of 5th residue exchange *in vivo*(gm-AEJAV,gm-AEJIA, gm-AEJAM, gm-AEJAKR and gm-AEJAL)was provided (Patel et al., 2021).

1.5.5 Outer and Inner Membrane Tethering (Diderms)

Outer and inner membrane tethering in Diderms generates connection between the outer membrane and the inner membrane. This tethering is essential in maintaining cell envelope integrity, facilitating transport, coordinating various cellular processes and sensing of damage (Miller and Salama, 2018).

Lipoproteins are a major class of molecules involved in membrane tethering. Braun's lipoprotein (Lpp) in *E. coli* is the most studied example. Lpp C-terminal lysine is covalently linked to PG via Ldts (Dramsi et al., 2008; Magnet et al., 2007) and covalently anchored with the outer membrane with N-lysine. Lpp helps to maintain the periplasm by controlling the distance between the outer and inner membrane and providing stiffness to the cell envelope (Hirota et al., 1977; Sanders and Pavelka, 2013).

Interestingly, most diderm bacteria lack Lpp, suggesting an alternative mechanism for outer membrane stabilization, in *B. abortus* outer membrane proteins (OMPs) are covalently attached to PG, like Lpp these OMPs are attached by Ldts (Godessart et al., 2021). Similarly, in *Coxiella burnetii*, *A. tumefaciens*, and *Legionella pneumophila*, certain β -barrel proteins are also covalently attached to PG by Ldts (Sandoz et al., 2021a), highlighting a conserved role of Ldts in outer membrane stabilization.

While outer membrane tethering to peptidoglycan is a well-documented phenomenon, inner membrane tethering has also been observed. In the phytopathogenic bacterium *Dickeya dadantii*, the inner membrane protein OutB is covalently attached to PG (Nicolai et al., 2024). OutB, is a component of the type 2 secretion system (T2SS) (Zhang et al., 2022). OutB is linked to the PG stem peptide via a C-terminal lysine residue, similarly to Lpp. This attachment is catalysed by specific L,D-transpeptidases, Ldt03 and Ldt84. OutB seems to use the PG as a scaffold to better attach itself and the T2SS secretin pore with the cell wall (Nicolai et al., 2024).

As mentioned before (see 1.1.1) in Monoderms, surface proteins are anchored to the cell wall PG via a mechanism involving membrane-bound transpeptidases named sortases. While a sortase itself is membrane-bound, the proteins it anchors are secreted across the membrane and linked to the PG layer (Perry et al., 2002). In addition to these protein-mediated connections, monoderm bacteria also covalently attach other major glycopolymers to the PG, such as Wall Teichoic Acids (WTAs) and Teichuronic acids (Weidenmaier and Peschel, 2008). The inner membrane serves as a

platform for anchoring a wide array of protein complexes and molecules that are essential for cell wall synthesis.

1.5.6 Glycan chain length and its regulation

Glycan chain length is dynamically regulated by the balance between synthesis and remodelling enzymes (Pedro and Cava, 2015). Synthesis is primarily carried out by glycosyltransferases (GTs), such as aPBPs and SEDS proteins (Garde et al., 2021). The natural termination of growing chains during synthesis is mediated by lytic transglycosylases (LTs), which cap the chain in *E. coli* and *B. subtilis* (Pedro and Cava, 2015; Vollmer et al., 2008a). Additionally, PG hydrolases (muramidases), contribute to chain cleavage during overall PG remodelling (Garde et al., 2021). Precise control over the final distribution of glycan chain length results from the interplay between the rates of polymerization, termination, and remodelling cleavage (Singh et al., 2015). This often involves the coordinated action of GTs and LTs within multi-enzyme complexes (see 1.4.4) (Garde et al., 2021; Vollmer et al., 2008b).

Measuring glycan chain length involves breaking down the insoluble PG network into smaller, soluble fragments that retain information about the original chains (Harz et al., 1990). The standard method involves purifying the intact PG sacculus and digesting it with muramidases to release soluble muropeptides (glycan fragments with stem peptides) (Vollmer et al., 2008b). These muropeptides are separated based on their size, commonly by liquid chromatography (LC) (Harz et al., 1990). LC coupled to mass spectrometry (LC-MS) enables the quantitative analysis of the muropeptide profile, including PG composition, chain length distribution, and crosslinking (see 1.6) (Singh et al., 2015; Vollmer et al., 2008a). The average chain length is calculated from the muropeptide profile by quantifying fragments resulting from LT termination (1,6-anhydro MurNAc residues) and determining the ratio of total muramic acid residues to these chain ends (Harz et al., 1990). Analyzing the full-size distribution of released fragments provides insights into the glycan chain length heterogeneity (Garde et al., 2021; Harz et al., 1990; Vollmer et al., 2008a)

Alternative and complementary methods further refine our understanding of glycan chain length and dynamics. One approach involves releasing intact glycan strands using an amidase, purifying them by charge (ion exchange), and separating them by size (HPLC) (Harz et al., 1990). End-group analysis, such as chemical reduction or enzymatic radiolabelling, is used to determine the length of these strands (Harz et al., 1990; Schindler et al., 1976). Polyacrylamide gel electrophoresis (SDS-PAGE) can separate shorter glycan products and study enzyme processivity *in vitro*, but it lacks sensitivity for long chains, does not provide quantitative rates, and struggles to differentiate long polymers from cross-linked chains (Wang et al., 2008). Metabolic probes, such as fluorescent D-amino acids (FDAA), label newly synthesized PG *in vivo*, allowing visualization of new material incorporation and dynamics via microscopy or flow cytometry. These probes can be combined with enzymatic analysis or MS (Garde et al., 2021; Pedro and Cava, 2015);. *In vitro* assays using labelled substrates like Lipid II can measure polymerization rates but do not directly determine the final chain length distribution. Qualitative methods like turbidity assays assess hydrolase activity, while

Transmission Electron Microscopy (TEM) assesses sacculi integrity, though neither directly measures chain length (Fibriansah et al., 2012; Vollmer et al., 2008a).

1.6 PG Structural Analysis Methods

The first PG analyses started in the 1960's with paper chromatography (Weidel et al., 1960) and later were based on acid hydrolysis followed by thin layer chromatography (Schleifer and Kandler, 1972). Since the late 80's, the purification of PG and analysis involves hot SDS extraction, enzymatic digestion of glycan strands, separation of disaccharide-peptides by rp-HPLC followed by analysis of individual peaks by mass spectrometry (MS). Nowadays the bottleneck is not the experimental approach but the data analysis (Alvarez et al., 2016; Pazos and Peters, 2019; Porfírio et al., 2019).

1.6.1 Nuclear Magnetic Resonance (NMR): Solid-state for PG structure and Solution NMR for purified fragments

Nuclear magnetic resonance (NMR) spectroscopy is a crucial tool for muropeptide structural analysis, complementing techniques such as mass spectrometry (MS) and amino acid analysis. NMR excels in identifying structural features, notably amide groups, whose accurate determination can be challenging with MS alone, particularly in larger muropeptides. Two-dimensional NMR techniques (COSY, TOCSY, ROESY) are instrumental in spectral assignment and muropeptide structural elucidation by revealing amino acid connectivity and amide bond linkages. NMR can confirm the presence of anhydro groups and identify modifications, including N-deacetylation. Beyond structural identification, NMR enables guantitative measurements, such as glycine content in PG bridging segments. Combining NMR with LC-MS facilitates investigation of PG tertiary structure, enabling analysis of both intact cell walls and digested fragments while preserving crucial glycan modifications. Although NMR's sensitivity is generally lower than that of MS, its strength lies in detailed structural elucidation. Computational analyses and chemometric software enhance NMR data interpretation. Techniques such as LC-NMR and solid-state NMR further expand its applications, enabling the study of intact cell walls and larger structures. Solid-state NMR is a useful technique for studying insoluble PG and intact bacterial cells (Porfírio et al., 2019; Romaniuk and Cegelski, 2015). It enables the quantification of chemical composition and the mapping of cell-wall architecture (Romaniuk and Cegelski, 2015). However, compared to liquid-state NMR, solid-state NMR exhibits lower sensitivity (Kim et al., 2015b). This, coupled with the spectral overlapping resulting from the complex mixture of cellular components in intact bacterial cells, can make data analysis and interpretation challenging (Patti et al., 2008b). Furthermore, while solidstate NMR provides average structural information, resolving inherent heterogeneity in PG at a high level of detail across the entire cell wall remains difficult (Patti et al., 2008b).

Despite these limitations, NMR has been successfully employed to analyze muropeptides in *Bacillus subtilis, Cyanophora paradoxa*, and *E. faecium*, revealing key structural details and modifications (Kim et al., 2015b; Patti et al., 2008b; Pfanzagl et al., 1996). The unique ability of NMR to provide structural and compositional information about PG in it native environment makes it an invaluable tool for PG analysis (Kim et al., 2015b).

1.6.2 PG Purification and Digestion

The preparation of the PG starts with the isolation of the cell walls using boiling SDS. Cell walls are separated from other cell components by differential centrifugation, made possible due to the high molecular weight of PG and its solubility. Further purification steps involve treatment by nucleases and proteases. Polymers covalently bound to PG are removed by acid treatment (1 N HCl or 48%v/v HF) (Kühner et al., 2014; Porfírio et al., 2019; Seltmann and Holst, 2002).

PG digestion is carried out using muramidases such as lysozyme or mutanolysin, which cleave the glycosidic bonds between MurNAc andGlcNAc units (Porfírio et al., 2019). Sugars with a reducing end exist in solution as a mixture of α and β anomers, due to a process called mutarotation (the formation of a hemi-acetal group). In a chromatogram unreduced muropeptides appear as two peaks. Sodium borohydride is used to reduce sugars so that each fragment will elute as a single peak (Abdek-Akher et al., 1951; Desmarais et al., 2013; Schaub and Dillard, 2017).

1.6.3 Chemometric Analysis (rp-HPLC, UV detection)

Chemometrics is useful to analyze PG structure. This interdisciplinary field leverages mathematics, statistics, and computer science to extract information from experimental data. It simplifies and automates the comparison/quantification of peaks in HPLC chromatograms, crucial for bacterial cell wall composition analyses. Chemometrics facilitates sample classification, outlier identification, and reveals patterns in complex biological data that might otherwise be obscured.

A chemometric approach has been proposed to compare PG chromatograms. PG-Metrics, a chemometric pipeline, enables rapid bacterial classification based on muropeptide profiles (Kumar et al., 2017). Chemometric approaches automate peak detection and quantification in chromatograms, often using Gaussian fitting for peak identification, alignment, and area quantification. Data pre-processing is crucial, involving correction of artifacts such as baseline offset and retention-time drifts. Multidimensional analysis helps correlate complex datasets, such as metabolite and gene data, identifying potential therapeutic targets. Method comparison is essential to mitigate bias arising from specific chemometric methods.

Several techniques and algorithms are employed, including Self-Organizing Map (SOM) analysis for unsupervised data mapping, Correlation Optimized Warping (COW) and Icoshift algorithms for retention-time drift correction, and Principal Component Analysis (PCA) for classifying bacterial samples based on PG composition. Software tools such as Chromanalysis (Desmarais et al., 2015), MZmine, and MetaboAnalyst (van der Aart et al., 2018) aid in peak detection, data deconvolution, and data normalization.

Data pre-processing is paramount for chemometric analyses. Accurate comparisons require baseline correction using MATLAB routines (Desmarais et al., 2015), and retention-time drift correction, using algorithms like COW (Kumar and Cava, 2019). Despite HPLC/UPLC's high resolution, co-eluting muropeptides necessitate additional

methods, such as MS, for structural confirmation. Careful attention to data artifacts and variations is critical for reliable results.

1.6.4 LC-MS/MS Analysis

Reverse-phase HPLC (rp-HPLC) is used to separate soluble PG fragments based on their size and hydrophobic properties (Glauner, 1988). Muropeptide elution is achieved by decreasing the polarity of the mobile phase with organic solvents (Alvarez et al., 2016). Various buffer systems have been described (phosphate or ammonium phosphate buffer/Methanol). When it is used coupled to mass spectrometry, water/acetonitrile is preferred.

Tandem mass spectrometry (MS/MS) is often used in conjunction to rp-HPLC and provides detailed structural information to confirm the identity of PG structural isomers with identical masses., by revealing their unique fragmentation patterns. In MS/MS, ions can be fragmented through collisions with an inert gas (HCD, Higher energy Collision Dissociation) or with electrons (ETD, Electron Transfer Dissociation).

1.6.5 LC-MS/MS Data Analysis Strategies:

Offline analysis: PG fragments can be separated by reversed-phase highperformance liquid chromatography (rp-HPLC) and collected individually before MS analysis. This approach allows to use more resolutive buffers but requires desalting, eventually freeze-drying and resuspension of fractions in smaller volumes. To avoid the need to desalt samples, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) has been employed in several studies. With this method, salt-containing samples can be directly analysed (Porfírio et al., 2019).

Online analysis (LC-MS): in this setup, the liquid chromatography (LC) system is coupled to the MS instrument (LC-MS). This requires an MS-compatible mobile phase devoid of salts (Kühner et al., 2014). The use of a compatible mobile phase reduces the overall analysis time, but the limiting factor becomes data analysis.

1.6.5.1 Manual Analysis

The seminal work published in 1988 by B. Glauner and collaborators was relying on HPLC separation of muropeptides and the analysis of individual peaks. This involved digestion of muropeptides with various PG hydrolases and the characterisation of peptide stem compositions using derivatization methods and Edman degradation. With this approach, 39 muropeptides were identified (Table 1-1) (Glauner et al., 1988).

Muropeptide	Retention time	Amount	Standard deviation	
	min	%	%	
Tri	11.9	7.41	±5	
Tetra(Gly ⁴)	15.0	1.71	±8	
Tetra	25.2	35.90	±4	
Penta(Gly ⁵)	26.1	0.26	±18	
Di	29.6	2.13	±11	
Penta	39.7	0.07	±23	
Tri-Lys-Arg	45.9	4.68	±6	
Tri(Anh)	55.1	0.36	±8	
Tetra-Tri(A ₂ pm,Gly ⁴)	56.1	0.07	±22	
Tri-Tri(A ₂ pm)	57.3	0.28	±13	
Tetra-Tri(A ₂ pm)	62.4	1.55	±11	
Tetra-Tetra(Gly ⁴)	63.0	1.47	±10	
Tetra-Tri	63.6	2.99	±9	
Tetra-Penta(Gly ⁵)	67.3	0.32	±18	
Tetra-Tetra	69.2	27.30	±3	
Tetra(Anh)	71.0	0.60	±11	
Tetra-Penta	73.8	0.17	±24	
Tri-Tri-Lys-Arg(A ₂ pm)	75.4	0.43	±20	
Tetra-Tetra-Tri(A ₂ pm)	80.1	0.24	±14	
Tetra-Tetra-Tri	80.9	0.28	± 10	
Tetra-Tri-Lys-Arg	83.0	2.92	±7	
Tetra-Tetra-Tetra	84.6	2.33	± 5	
Tetra-Tri(A2pm,Anh)I	88.9	0.26	± 15	
Tetra-Tri(A ₂ pm,Anh)II	89.4	0.27	±16	
Tetra-Tri(Anh)I	90.6	0.42	±12	
Tetra-Tri(Anh)II	91.3	0.14	±15	
Tetra-Tetra-Tetra-Tetra	93.1	0.08	±17	
Tetra-Tetra-Tri-Lys-Arg	95.1	0.27	± 10	
Tetra-Tetra(Anh)I	95.9	0.67	±7	
Tetra-Tetra(Anh)II	96.6	0.67	±5	
Tetra-Tetra-Tri(A2pm,Anh)	101.1	0.08	±13	
Tetra-Tetra-Tri(Anh)	101.7	0.18	± 12	
Tetra-Tetra-Tetra-(Anh)	105.4	0.55	±10	
Tetra-Tetra-Tetra-Tetra(Anh)	110.2	0.05	±12	
Tetra-Tri-Lys-Arg(Anh)I	111.9	0.13	±13	
Tetra-Tri-Lys-Arg(Anh)II	113.2	0.16	± 21	
Tetra-Tetra-Tri-Lys-Arg(Anh)I	116.4	0.03	±20	
Tetra-Tetra-Tri-Lys-Arg(Anh)II	116.9	0.05	± 24	
Tetra-Tetra-Tri-Lys-Arg(Anh)III	118.8	0.05	±17	

Table 1-1 Offline LC-MS/MS manual data analysis (Glauner et al., 1988)

1.6.5.2 Automated Software

The large volume of data produced by modern mass spectrometry instruments makes manual analysis unsuitable due to its time-consuming and error-prone nature. Automation significantly accelerates data processing, enabling higher throughput and faster results. Consistent data processing by automated algorithms minimizes bias and variability, thereby enhancing reproducibility. This is particularly crucial in highthroughput applications such as metabolomics, proteomics, and drug discovery, where rapid data interpretation is essential. Recently, automated analysis pipelines have been developed: *in silico* MS/MS prediction tools (PGN_MS2), and high-throughput automated muropeptide analysis (HAMA). Other tools exist such MassHunter, Byonic[™], and PGFinder can be specifically used for database searching for muropeptide identification, and *in silico* MS/MS fragmentation matching.

MassHunter

MassHunter is a proprietary software package developed by Agilent Technologies for the acquisition and analysis of mass spectrometry (MS) data that was recently used to characterize the PG structure of *Pseudomonas aeruginosa* (Anderson et al., 2020; Anderson et al., 2022). Molecular structures of muropeptides, created using ChemDraw, can be incorporated into the MassHunter Personal Compound Database and Library. MassHunter Profinder is then used to identify muropeptides present in the library. This methodology has been employed to analyze compositional changes across various growth conditions of *P. aeruginosa* (Anderson et al., 2020; Anderson et al., 2022).

Byonic™

Byonic[™] is a software package dedicated to the identification of proteins by MS/MS. It is part of the software platform called Byos® developed by Protein Metrics (Dotmatics). Byonic[™] can analyse glycopeptides MS/MS data and is therefore very suitable for disaccharide-peptides monomers. Muropeptide identification is achieved by searching a database made of peptide stems provided as a fasta file. It allows to define possible modifications at their N-terminus with various sugar moieties (GlcNAc-MurNAc, GlcN-MurNAc etc...). Byonic[™] searches for observed monoisotopic masses matching the sequence provided by the user (modified or not) and identifies corresponding predicted MS/MS fragments. Byonic[™] provides a detailed annotation of each MS/MS spectrum containing at least one fragment ion (a,b,c or x,y,z series). The software can also deconvolute MS/MS data into FTRS format, showing more consistent and better mass accuracy compared to MassHunter or MaxQuant (Patel et al., 2021).

Bern et al., 2017 provided the first proof of concept that Byonic[™] can be used for an automated MS/MS data analysis of PG. The unbiased structural analysis of *C. difficile* PG led to the identification of novel muropeptides, as well as previously described monomers and dimers. However, manual determination of crosslink types (3-3 or 3-4) was still required. The key advantages of this methodology include the utilization of available MS/MS fragmentation data for identification and validation.

HAMA

High-throughput Automated Muropeptide Analysis (HAMA) is a platform designed for the identification and analysis of PG structures using mass spectrometry data (Hsu et al., 2023). HAMA consists of three components: 1) DBuilder, which constructs databases containing monomers, dimers, trimers, glycosidic bonds, peptide bonds, 4-3 crosslinks, and lysine-containing peptide bridges. 2) Analyzer, which processes LC-MS data for mass deconvolution and PG identification. 3) Viewer, which visualizes ion chromatograms and MS/MS spectra annotated with *b*- and *y*-ions.

HAMA simplifies muropeptides into a sequence format, employing a "bottom-up" approach like proteomics and glycoproteomics. It utilizes an *in-silico* MS/MS fragmentation database to identify muropeptides. To achieve this, HAMA uses DBuilder to construct specific databases based on known bacterial species. The Analyzer component of HAMA processes MS data, comparing it with *in silico*-generated b- and y-fragment ions for identification. The platform limitations include the database's reliance on previously reported structures, potentially requiring manual analysis for low abundance muropeptides. Additionally, 3-3 crosslinks are not predicted due to the risk of misidentification.

PGN_MS2

PGN_MS2 is an open-source software tool for the identification of muropeptides from MS/MS data (Kwan et al., 2024). This software is specialized for PG analysis, allowing users to select specific features. It simulates MS/MS spectra and integrates these spectra into an MS Library, facilitating the automated identification. The database is generated based on user-defined parameters, including modifications in GlcNAc and MurNAc, and peptide stem composition. The PG muropeptide is built *in silico* and the database containing chemical formulas, m/z adducts, SMILES strings, degree of acetylation or amidation, and step peptide lengths, is saved as an Excel worksheet.

PGN_MS2 predicts MS/MS spectra *in silico* based on spectra data from known muropeptides. The fragmentation patterns are encoded as a chemical reaction in SMARTS language (Ehrt et al., 2020) and then simulated. The fragmentation model considers that precursor ions frequently undergo both glycan (*b-/z-*) and peptide (*b-/y-*) fragmentation. Usually, PGN_MS2 *in silico-spectra* match experimentally acquired MS/MS spectra. Unlike HAMA, PGN_MS2 incorporates fragmentation of the sugar moieties. The software was implemented like HAMA (Hsu et al., 2023) by analysing *Bifidobacterium* bacteria in the context of gut microbiota (Kwan et al., 2024).

PGFinder

Patel et al., 2021 developed an open-source, automated tool for analyzing deconvoluted MS data in PG analysis. This software identifies muropeptides by matching theoretical masses in user databases with observed monoisotopic masses in the deconvoluted data. Initially, a monomer database is constructed. Subsequent PGFinder searches utilize identified monomers to generate a dimer database, which is then used for a second matching operation. Finally, a third search using identified monomers and dimers (4-3 and 3-3 crosslinked) identifies PG modifications (anhydro groups, deacetylated sugars, amidated amino acids, amidase activity). A key limitation is its reliance on MS1 data, necessitating a separate MS/MS validation step for the inferred structures (Patel et al., 2021).

PGFinder has undergone several improvements since its original release. Distinct versions of PGFinder have been used during over the duration of this thesis. Figure 1-11 illustrates the major improvements by year.



Figure 1-11 Timeline showing the major improvements to PGFinder

Version 0.0.2 was the first version made accessible via Jupyter notebook (Patel et al., 2021), relying on server-side computation. It automatically identified PG fragments based on MS1 data and a theoretical muropeptide mass list but required manual, error-prone database construction and separate MS/MS validation. Version 0.1.1a was showcased analysing *C. difficile* datasets (Galley et al., 2024) by performing unbiased muropeptide searches usingdatabases containing deacetylated muropeptides. It offered a visual interface for customized searches, modification identification, and adjustable stringency. However, this version is no longer available. Version 1.0.3 introduced the capability to run on the user side. Version 1.1.0 further upgraded this by enabling dynamic database creation for dimers and trimers with 1-3 crosslinks, validated using *G. oxydans* datasets (Alamán-Zárate et al., 2024).

The most recent version, 1.2.1 offers a web interface, requiring only static website hosting. It includes a mass calculator module, automating residue and database mass calculations, eliminating manual errors. A fragment predictor module expands capabilities to predict modified muropeptides and 1-3 crosslinks, streamlining MS/MS analysis. This version automates muropeptide precursor and fragment mass computation, incorporates PGLang (a novel language for describing muropeptide structures), and automates output consolidation from single datasets, accelerating data analysis. These developments represent significant advancements in automated PG analysis, improving efficiency and accuracy.

1.6.6 Muropeptide Analysis limitations

While powerful for determining the average chemical PG composition from a bacterial population, standard muropeptide analysis techniques have notable limitations regarding structural detail and accuracy. The process of digesting the PG network into small fragments inherently loses information about the overall intact structural arrangement and the true percentage of cross-links within the native cell wall (Hsu et al., 2023); solid-phase NMR is suggested as a more precise method for evaluating native crosslinking (Chang et al., 2018). Furthermore, relying solely on mass spectrometry data can make it difficult to definitively identify and locate specific modifications (such

as amidation, deacetylation, or O-acetylation on disaccharides) (Anderson et al., 2019) or distinguish between structural isomers (like 3-3 vs 4-3 cross-links or stereoisomers) (Hsu et al., 2023), often requiring complementary techniques like NMR or additional manual verification. Insufficient peptide fragment ions in standard MS/MS spectra also hinder accurate peptide sequencing needed to locate modifications or identify multimers (Kwan et al., 2024). The identification of low-abundance or novel muropeptides is challenging due to detection limits, reliance on databases of known structures, and the necessity for manual verification after automated mass matching, which can be prone to misidentification (Hsu et al., 2023; Patel et al., 2021). Additionally, ion intensity serves only as an indication of relative compositional changes, not absolute molar abundance (Patel, 2021).

Other limitation of muropeptide analysis is its inability to fully capture the heterogeneity of PG structure. By analyzing the bulk PG isolated from a population, the technique yields average information (Bern et al., 2017; Kwan et al., 2024; Patel et al., 2021), missing both spatial heterogeneity within a single cell (e.g., structural differences at poles vs. septum) and transient compositional changes occurring during distinct phases of the cell cycle. It also does not provide information about which cellular proteins (like SPOR domains) might have been associated with specific PG structures. While certain muropeptide features, like anhydro-muropeptides, are linked to specific processes (chain end) (Atrih et al., 1999), the analysis itself does not confirm their precise localization within the cell wall. Methodological factors during sample preparation, such as isolation procedures or excessive enzymatic digestion, can also potentially alter the observed muropeptide pattern (Patti et al., 2008a). Consequently, while revealing differences in bulk composition between populations. muropeptide analysis alone cannot fully capture the intricate spatial and temporal variations of PG chemistry, highlighting the necessity of complementary techniques to gain a comprehensive view of this macromolecule.

1.7 Model System: Rhizobium leguminosarum

The following sections of this thesis will focus on the model organisms studied: *Rhizobium leguminosarum, Clostridioides difficile,* and *Gluconobacter oxydans*. The selection of these organisms was primarily driven by their unique biological features, which will be further detailed. The study of their distinctive muropeptides facilitated the rapid development of PGFinder tools, resources, databases, and analytical workflows presented herein, enabling us to address the diversity of peptidoglycan features and to formulate new biological questions for future investigation in other systems.

1.7.1. Importance of *R. leguminosarum* in agriculture

Rhizobium leguminosarum is a Diderm soil bacterium able to engage in a symbiotic relationship with leguminous plants. It can colonise plant roots to form specialised organs called nodules (Young et al., 2021). During the formation of nodules, *R. leguminosarum* differentiates into bacteroids which can fix atmospheric nitrogen, thereby promoting plant growth. In exchange, bacteroids use carbon compounds

produced by the plant (Maróti and Kondorosi, 2014; Schulte et al., 2022). In indeterminate nodules (in pea), host cell proliferation persists and bacteroids are terminally differentiated with enlarged and branched morphology (Figure 1-12) and cannot resume growth after nodule senescence. In determinate nodules (in bean), the nodule meristem is active transiently; bacteroids do not terminally differentiate and can resume free-living growth after nodule senescence (Maróti and Kondorosi, 2014). The changes in cell morphology during bacteroid formation indicate that cell envelope remodelling is taking place during this transition.



Figure 1- 12. Bacterial morphology of *S. melioti* as free-living cells or as bacteroids in *Medicago sativa* nodules.

The panels show fluorescence microscopy of *Sinorrhizobacterium* as free-living cells and as bacteroids (extracted from nodules). Modified from Nicoud et al., 2021

R. leguminosarum species includes several genetically diverse strains that share nodulation and nitrogen fixation genes, enabling them to form nodules in a restricted host range, like pea (*Pisum sativum*) or faba bean (*Vicia faba*) (Young et al., 2021). For example, *R. leguminosarum* USDA 2370 can only generate nodules in *P. sativum* (Kosslak and Bohlool, 1984). While some other strains can also grow nodules in other legumes like *Trifolium repens* (white clover) and *Phaseolus vulgaris* (common bean) (Kosslak and Bohlool, 1984). Cross-nodulation tests, along with the study of LysM receptor-like kinases, are standard methods for investigating host specificity and host range (Humphrey and Vincent, 1965; Oldroyd and Downie, 2008).

Rhizobia can evolve through the acquisition of symbiosis islands, antibiotic resistance, and megaplasmids (pSyms) carrying genes related to symbiosis (Naamala et al., 2016; Young et al., 2021).

The bacterial cell envelope of *R. leguminosarum* has been studied to explore the mechanisms that underpin the interactions of this bacterium with the root cells and most attention has focused on the composition and structure of LPS and exopolysaccharides (Wheatley et al., 2020). Surprisingly, nothing is known about the PG remodelling that takes place during the life cycle of this organism.

1.7.2 R. leguminosarum Life Cycle

R. leguminosarum has a complex life cycle. This organism is usually found in the soil and enriched in the rhizosphere, in the vicinity of plant roots. The sequence of events

leading to the formation of root nodules has been extensively described. They involve a molecular dialog between the bacterium and the plant roots. The major steps leading to the formation of nodules and the fixation of atmospheric nitrogen are described below and in Figure 1-13.



Figure 1-13 *Rhizobium leguminosarum* lifecycle

Legumes attract rhizobia by producing root exudates called flavonoids (1), Sensing of flavonoids triggers the production of Nod factors and the migration of bacteria towards the root hairs (2) to which they adhere (3). Nod factors induce root hair curling and the migration of rhizobia in infection threads that penetrate the root (4). Nod factors also stimulate cell division, leading to nodule development (5). Rhizobia are released into plant cells, remaining membrane-bound and dividing (6). They transform into bacteroids (7), initiating nitrogen fixation within symbiosomes (8). A vascular system facilitates nutrient exchange. Depending on the type of nodule, rhizobia either terminally differentiate and cannot resume growth (senescent nodule) (9) or they can proliferate again. Upon nodule breakdown, bacteria return to the soil (10), infecting new roots or reverting to a free-living state.

• **Root Colonisation.** *R. leguminosarum* present in the soil can detect the production of root exudates containing flavonoids compounds which are key initiators of the symbiotic process. *Rhizobium* are chemoattracted to the legume root, which they colonize to form a biofilm (Oldroyd and Downie, 2008).

•Root infection: the production of flavonoids by plant roots triggers the transcription of a set of bacterial genes that collectively contribute to produce small oligosaccharide called Nod factors by R. leguminosarum. As the bacteria establish contact with the root cells, the plant cells initiate a complex signalling cascade involving protein phosphorylation and calcium oscillations (a process involving plant lectins and bacterial cell surface polysaccharides), which induce root curling. The plant cell membrane invaginates and the bacteria trapped at the tip of the roots can enter the vascular tissue and form infection threads. Once arrived in root cortical cells, R. membrane-bounded leauminosarum is present in compartments called symbiosomes.

At this stage, plant cells produce nodule-cysteine-rich (NCR) peptides displaying antimicrobial activity that induce the bacterial differentiation into nitrogen-fixing bacteroids (van de Velde et al., 2010). This process is associated with significant changes to the *Rhizobium* metabolism (Pan and Wang, 2017; Prell and Poole, 2006).

•**Nitrogen Fixation:** Once fully differentiated, bacteroids can fix atmospheric nitrogen in the form of ammonium, an essential nutrient for the plant. While shifting from free-living bacteria to bacteroids, *R. leguminosarum* lose the ability to synthesize amino acids, which are provided by the host plant (Wheatley et al., 2020).

1.7.3 Evidence for PG remodelling during *R. leguminosarum* life cycle

It seems legitimate to hypothesize that R. leguminosarum cell envelope is undergoing profound changes as this bacterium goes through its life cycle. Soil bacteria can experience various abiotic stresses including pH variations or changes in osmotic conditions. Examples in the literature have shown that L,D-transpeptidase activity is increasing in such conditions in E. coli and A. tumefaciens (Morè et al., 2019). Bacterial differentiation is also associated with remodelling of the cell envelope, including PG hydrolysis, modification or the formation of different crosslinks. This is the case for sporulation in Bacillaceae (Popham and Bernhards, 2015) which involves the hydrolysis of the mother cell (Popham and Bernhards, 2015) and the formation of delta-lactam rings (Gilmore and Cava, 2025). In Coxiella and Legionella, the transition to actively growing variants is associated with the upregulation of L.D-transpeptidation and the tethering of several outer membrane proteins to PG (Kathayat et al., 2025). In *R. leguminosarum*, two β -barrel proteins (RopA and RopB) have been shown to form amyloid fibres and proposed to play a role in nodulation (Kosolapova et al., 2019). The fact that R. leguminosarum experiences a diverse range of environmental conditions and differentiates into bacteroids displaying morphological changes prompted us to investigate PG remodelling during the life cycle of this organism.

1.7.4 R. leguminosarum multigene families encoding L,D and D,D-transpeptidases

R. leguminosarum biovar viciae strain 3841 (*RIv*3841) genome is 7.75Mb. It contains a large chromosome and 6 plasmids, encoding genes associated with symbiosis. The analysis of *RIv*3841 genome in BV-BRC¹ reveals the presence of at least 12 putative PBPs (Table 1-2) and 18 putative LDTs (Table 1-3). The genetic redundancy amongst transpeptidase genes suggests that both families of proteins are contributing to polymerise and modify a complex PG molecule.

¹ https://www.bv-brc.org/view/Genome/216596.11#view_tab=overview

E. coli PBP	Class	Subclass	Putative R. leguminosarum PBP	Proposed activity	Domain organization 1 10 270 300 100
PBP1a			RL1743	Transglycosylase and transpeptidase	Fercilis
		A1	pRL110261	Transglycosylase and transpeptidase	Transfoly Transpectionse 32
			pRL110250	Transglycosylase and transpeptidase	Transporting
PBP1b	Δ		RL1393	Transglycosylase and transpeptidase	Transbigoosylisse
	А	A2	RL0153	Transglycosylase and transpeptidase	Transport Lisse
			pRL110249	Transglycosylase and transpeptidase	
PBP1c		A6	pRL110012	Transglycosylase and transpeptidase	Caractol young Line Transpept Likese Fent clinited
MGT		A7	N/A	Glycosyltransferase	
PBP2	D	B2	RL3313	DD-transpeptidase	Francisco Francisco
PBP3	Б	B3	N/A	DD-transpeptidase	
PBP4		C4	N/A	DD-carboxypeptidase/DD-endopeptidase	
PBP5			RL4363	DD-carboxypeptidase	Security States States
	С		RL2477	DD-carboxypeptidase	Series (1920)
DDD6a		C5	RL1016	DD-carboxypeptidase	Corrowing cospitations: SLA
r Br 0a		C5	RL2656	DD-carboxypeptidase	Brocksong aptildese (SB
			RL2541	DD-carboxypeptidase	Continue ceptificate ISL
PBP6b			N/A	DD-carboxypeptidase	
PBP7		C7	N/A	DD-endopeptidase	
PBP4b		A	N/A	DD-carboxypeptidase	
AmpH		Атрн	N/A	Weak DD-carboxypeptidase/DD-endopeptidase	2
📔 signa	l peptide	rensigned and tran	nsglycosylase domair Transport Less. tra	npeptidase domain 📷 carboxypeptidase domain 🔤	SPOR domain
			🔤 - BA14K domain	catalytic residue	

Table 1-2. List of putative PBPs encoded by Rlv3841

Table 1-3. List of putative Ldts encoded by Rlv3841

Putative R. leguminosarum LDT	Closest <i>E. coli</i> LDT homolog	Proposed activity	Domain organization 1 100 200 300 400 500 600 700
pRL120340	LdtA	Protein anchoring	
RL0055	LdtA	Protein anchoring	
RL0869	LdtA	Protein anchoring	- YXxD -
RL4558	LdtA	Protein anchoring	-YkuD
RL1845	LdtA*	Protein anchoring?	
RL3834	LdtA*	Protein anchoring?	- NkuD
RL1296	LdtA/B	Protein anchoring	- Nku D
RL1297	LdtA/B	Protein anchoring	
RL0870	LdtB	Protein anchoring	
RL1724	LdtB*	Protein anchoring?	- VituD
RL3342	LdtB*	Protein anchoring?	- Viel
pRL90118	LdtC/E	Protein anchoring?	- YkuD
RL2700	LdtD	L,D transpeptidase/ Carboxypeptidase	
RL1618A	LdtD	L,D transpeptidase/ Carboxypeptidase	
RL1458	LdtE	L,D transpeptidase/ Carboxypeptidase	- 1kuD
RL2819	LdtE	L,D transpeptidase/ Carboxypeptidase	i and a second
RL4356	LdtF	Protein anchor cleavage	(YkuDa
pRL110015	LdtF	Protein anchor cleavage	

*Closest homolog by catalytic domain. 🗾 signal peptide 🛑 catalytic domain 🛑 PG binding domain ү catalytic residue 💷 low complexity region

1.7.5 Potential Role of β -Barrel Proteins

Several bacterial species belonging to gammaproteobacteria tether their outer membrane to PG via a small abundant protein related Braun's lipoprotein (Lpp). A recent study also revealed that in the plant pathogen Dickeya dadantii, a component of a secretion system (GspB/OutB) present in the inner membrane can be covalently attached to PG (Nicolai et al., 2024). The covalent link between Lpp or functional homologs and PG is formed by L,D-transpeptidases (Magnet et al., 2007). Interestingly, several Diderms do not encode homologs of Lpp but use β -barrel proteins for outer membrane tethering (Godessart et al., 2021; Sandoz et al., 2021a). Several β-barrel proteins covalently bound to PG have been identified in L. pneumophila, A. tumefaciens, C. burnetii or B. abortus (Godessart et al., 2021; Sandoz et al., 2021a). Beyond a role in maintaining membrane integrity, the covalent anchoring of β -barrel proteins may contribute to regulate their activity, but no experimental data support this hypothesis. It is however tempting to think that some of these proteins could contribute to antimicrobial resistance or the acquisition of nutrients critical for pathogenesis. In *R. leguminosarum*, two β -barrel proteins that could be linked to PG (RopA and RopB) have been proposed to play a role in symbiosis. No evidence is available showing that they are linked to PG or that they contribute to the nodulation process (Kosolapova et al., 2019).

1.8 Model System: Clostridioides difficile

C. difficile is a Monoderm, spore-forming anaerobic bacterium that is a major cause of antibiotic-associated diarrhea, especially in high-income countries (Lawson et al., 2016; Roo and Regenbogen, 2020). It can colonize the intestinal tracts of humans and other animals, leading to life-threatening pseudomembranous colitis (Sekiya et al., 2021). The infections are triggered by broad spectrum antibiotics that disrupt normal intestinal microbiota allowing *C. difficile* to thrive (Kaus et al., 2020). *C. difficile*'s PG structure has been the focus of several studies due the resistance of this organism to β -lactams and the potential role of Ldts and PBPs in this process (Peltier et al., 2011). *Clostridioides difficile* possesses peptidoglycan-bound proteins bound by a sortase (SrtB), which play a crucial role for bacterial colonization and nutrient acquisition (Donahue et al., 2014).

1.8.1 C. difficile from microbiota to pathogen

Clostridioides difficile can be a harmless member of the gut microbiota (Figure 1-14A). However, antibiotics can disrupt this balance by killing beneficial bacteria and creating the opportunity for *C. difficile* to overgrow (Johanesen et al., 2015).

C. difficile can form spores (Figure 1-14B) that can survive in the gut when vegetative cells will be otherwise killed by antibiotics (Sekiya et al., 2021; Wu et al., 2016). These spores can germinate and repopulate the gut once antibiotic treatment stops, especially when other members of the microbiota have not fully recovered (Coullon et al., 2020; Wu et al., 2016). The *C. difficile* spores cortex contains muramic- δ -lactams,

which made up to 24% of all muropeptides (Coullon et al., 2018). Interestingly, unlike *B. subtilis* (Popham et al., 1996b), the *N*-deacetylase responsible for muramic- δ -lactams (PdaA1), contributes to heat resistance and virulence of spores (Coullon et al., 2018).



Figure 1- 14 *Clostridioides difficile* **SEM morphology** Vegetative cells morphology (**A**) and spore structure (**B**) modified from Baloh et al., 2022

C. difficile pathogenic strains can produce toxins (Toxins A and B) that damages the intestinal lining, leading to disease, but even toxin-producing strains are harmless in low numbers within a healthy microbiota. *C. difficile* resistance to β -lactam antibiotics is multifactorial (Turello et al., 2025; Wydau-Dematteis et al., 2018). L,D-transpeptidases (Ldts) and D,D-transpeptidases (PBPs) are present, as well as efflux pumps that may contribute to resistance.

1.8.2 Multigene Families: L,D -Transpeptidases and its role

C. difficile has natural resistance to β -lactam antibiotics, making it a significant clinical challenge (Wickramage et al., 2021). The balance between different PG crosslinking mechanisms likely plays a crucial role in a bacterium response to antibiotics. Its genome has four D,D-transpeptidases (PBP), and three L,D-transpeptidase (Ldts) paralogs: Ldt_{Cd1}, Ldt_{Cd2}, and Ldt_{Cd3} (Peltier et al., 2011). Unlike most bacteria that primarily utilize penicillin-binding proteins (PBPs) for PG crosslinking via D,D-transpeptidation (4-3 crosslinks), *C. difficile* relies heavily on Ldts. These enzymes form 3-3 crosslinks, which constitute a high proportion (73%) of the crosslinks in *C. difficile* PG (Chapot-Chartier and Kulakauskas, 2014; Peltier et al., 2011). Because Ldts employ a different catalytic mechanism than PBPs, they are not susceptible to most β -lactams, except for penems and carbapenems (Sütterlin et al., 2018).

Three *C. difficile* Ldts have been identified and present distinct enzymatic activities *in vitro* (Table 1-4).

· •								
			L,D-	L,D-				
_	Enzyme	Size	carboxypeptidase	transpeptidation	4th residue exchange			
	Ldt_{Cd1}	469 aa	~	×	×			
	Ldt_{Cd2}	644 aa	~	~	~			
	Ldt_{Cd3}	283 aa	~	~	✓			

 Table 1- 4. C. difficile Ldts enzymatic activities reported by Sütterlin, et al.,2018

While all three *C. difficile* exhibit L,D-carboxypeptidase activity, only Ldt_{Cd2} and Ldt_{Cd3} possess the full set of activities (Sütterlin et al., 2018). *C. difficile* has a unique PG structure, featuring both 4-3 and 3-3 crosslinks, along with high level (89-93%) of muramic acid (MurN) (Peltier et al., 2011), Ldts contribute to *C. difficile* ability to thrive in the gut, particularly after antibiotic disruption of the microbiota, conferring a selective advantage (Ho et al., 2025). Further research into the role of *C. difficile's* Ldts is crucial for understanding their mechanisms and developing effective therapies against this challenging pathogen.

1.9 Model System: Gluconobacter oxydans

Glucnobacter oxydans (G. oxydans), belonging to the *Acetobacteraceae* family, is a diderm that lives as a saprophyte on decaying material from plants and insects. This strictly aerobic organism is well adapted to sugar enriched environments and alcoholic solutions with an acidic pH. The metabolism of *G. oxydans* make it useful for industrial applications (Hommel, 2014; Sheng et al., 2014).

1.9.1 G. oxydans, an Acetobactericeae with industrial relevance

G. oxydans is highly adapted to sugar- and alcohol-rich environments like flowers, fruits, and fermented products (vinegar, sake, wine, beer). It is associated with bacterial brown rot of apples and pears and pink disease of pineapples. Its acid tolerance confers a competitive advantage against organisms like yeast and lactic acid bacteria that prefer anaerobic conditions. *G. oxydans* is also a member of the microbiota of bees (*Apis mellifera*) and other insects, including *Drosophila melanogaster* and mosquitoes. *G. oxydans* is the main species found on healthy and, particularly, damaged red grapes, with colonization varying by grape variety, location, and season (Hommel, 2014).

The rapid oxidation of substrates and release of products such as acetic acid, results in acidification of the habitat, making it favourable for this acid-tolerant organism. Furthermore, these metabolic products are often poorly assimilated by other organisms (Hommel, 2014). *G. oxydans* releases its products into the medium via porins (Vergalli et al., 2020). Its membrane-bound glucose dehydrogenase exhibits activities up to three times greater than the cytoplasmic ones (Pronk et al., 1989). Their localization gives them direct access to substrates and the electron transfer chain, likely contributing to these high reaction rates, allowing for rapid, high-throughput processing of the sugars and alcohols in their environment (Da Silva et al., 2022; Muynck et al., 2007). Certain *Gluconobacter* strains can produce polyenic antibiotics active against *Saccharomyces cerevisiae* and monocyclic β -lactams active against other bacteria (Watanabe et al., 1982). Unlike most bacteria, *G. oxydans* possesses incomplete tricarboxylic acid and Embden-Meyerhoff-Parnas pathways, using these pathways primarly for the generation of metabolic precursors (Deppenmeier et al., 2002).

Beyond its physiological peculiarities, *G. oxydans* is of significant industrial importance. Its ability to efficiently catalyze the partial oxidation of sugar and alcohol substrates via membrane-bound oxidoreductases has been exploited commercially since the 1930s(Da Silva et al., 2022). This metabolic capability makes it a valuable biocatalyst for producing a variety of compounds. Currently, *G. oxydans* is used in the production of acetic acid, the antidiabetic agent miglitol, and dihydroxyacetone (used as tanning agent) (Claret et al., 1994; Da Silva et al., 2022). It also oxidizes glucose to gluconate (used in the textile industry) (Ma et al., 2022) and is most notably employed in the industrial synthesis of vitamin C from sorbitol (Da Silva et al., 2022). Genetically engineered strains have been developed to produce 5-ketogluconic acid from glucose, a precursor for the acidulant tartaric acid (Da Silva et al., 2022; Merfort et al., 2006). Its distinctive metabolic pathways have made *G. oxydans* a valuable model system for studying metabolic processes.

1.9.2 *G. oxydans*' PG shows the presence of 1-3 crosslinks (L,D -Transpeptidase)

Members of the *Acetobacteraceae* family, including *G. oxydans*, possess a unique PG structure characterized by the presence of 1-3 crosslinks (Espaillat et al., 2016). This crosslinking pattern is uncommon in other bacteria, which typically exhibit 4-3 or 3-3 crosslinks.

The enzyme responsible for 1-3 crosslink formation in *G. oxydans* was not identified in the initial report, the crosslink is between the first residue (L-Ala) with the third residue (mDAP) from another peptide stem suggest a mechanism more akin to L,D-transpeptidation than to a D,D-transpeptidation. When this work was initiated both the enzyme involved in generating 1-3 crosslinks and the identity of the donor and acceptor molecules remained as open questions.

1.9.3 Possible role of uncommon L,D-transpeptidation

The precise function of these 1-3 crosslinks in *G. oxydans* is unknown. However, it is hypothesized that these unique crosslinks contribute to the bacterium acid tolerance, competence, and adaptation to its specific ecological niche (Espaillat et al., 2016).

Gluconobacter species, are known to establish symbiotic relationships with insect guts and colonize various tissues and organs, including reproductive ones (Hommel, 2014). These bacteria are believed to pass through body barriers to reach different host organs. In *Drosophila*, they are involved in regulating the insect's immune system, potentially through pattern recognition receptors that recognize mDAP-containing muropeptides. These bacteria are considered secondary symbionts of insects, likely playing diverse roles in the insect biology (Hommel, 2014).

The presence of 1-3 crosslinks in *G. oxydans* highlights the diversity of PG structures in bacteria and opens new avenues for research into the function and evolution of these essential cell wall components.

Aims and objectives

Aim

To advance the understanding of bacterial peptidoglycan (PG) structure and function, with a focus on characterizing PG composition, crosslinking patterns, and the enzymatic activities involved in PG modification in uncharacterized model systems.

Objectives

- a) Optimization of data collection by LC-MS/MS.
- b) Detailed analysis of PG in various bacterial species displaying different modes of PG polymerisation (*C. difficile, G. oxydans*, and *R. leguminosarum*).
- c) Characterization of L,D-transpeptidases activities involved in PG biosynthesis and modification.
- d) Generate a unifying strategy for PG analysis using the software tools developed and tested during this work.

Chapter 2

2. Material and Methods

Chapters 4 (Galley et al., 2024), 5 (Alamán-Zárate et al., 2024), and 6 (Alamán-Zárate et al., 2025) correspond to published articles with their respective "Experimental Procedures" section. This chapter provides a summary of the Materials and Methods that correspond to the work I performed for each publication.

2.1 Bacterial strains, plasmids and oligonucleotides

All bacterial strains, plasmids and oligonucleotides used during my thesis are described in Table 2-1.

Strains/plasmids/ oligonucleotides	Relevant properties/sequence	Source
Strains		
Clostridioides difficile		
R20291	Clinical isolate, ribotype 027	1
R20291 Δldt_{Cd123}	R20291 derivative with an in-frame deletion in ldtCd2, ldtCd1 and	2
Gluconobacter oxydans		
B58	G. oxydans reference strain (ATCC NRLL B58)	ATCC
621H	G. oxydans reference strain (ATCC 621H)	ATCC
B58 <i>ldt</i> _{Go1}	B58 with a transposon insertion in Go2094 (GOX2269 in strain 621H); KanR	3
B58 <i>ldt</i> _{Go2}	B58 with a transposon insertion in Go1074 (GOX1074 in strain 621H); KanR	3
B58 <i>ldtco</i> 2 (pBBR-TetR-Go2227)	B58 ldt _{Ge2} mutant complemented: KanR GmR	4
Escherichia coli		
BL21 (DE3) Lemo	Expression strain	NEB
BL21 (DE3) Lemo (pET-Go2227)	BL21 (DE3) Lemo expressing LdtGo2 ; AmpR	4
Rhizobium leguminosarum		
<i>Bv. viciae</i> strain 3841	Strain able to make nodules in pea	5
Plasmids	I I I I I I I I I I I I I I I I I I I	
pET2818	pET28a derivative for recombinant protein expression; AmpR	6
pET-Ldt _{Go2}	pET2818 derivative encoding full length Ldt _{Go2} ; AmpR	4
pBBR1MCS-5	pBBR1-MCS derivative; GmR	7
pBBR1MCS-5-T _{gdhM} -tetR-mNG	pBBR1-MCS5 derivative for inducible expression with	8
pBBR-TetR-Go2227	pBBR1-TetR derivative expressing full length LdtGo2; GmR	4
Oligonucleotides		
SM_0725	ggctacggtctcccgaagtctcgggccgtctcttgggctt	
SM_0726	ggctacggtctcttccttggattcacttttctctatcactgataggg	
SM_0727	ggctacggtctctaggagatatcatatgcgtgatgtttccagactgac	
SM_0728	ggctacggtctcaaaggtaacggtcttttatccgcaatag	
SM_0729	ggctacggtctcaccttaacgcaaaaaaccccgcttcggcgg	
SM 0730	ggctacggtctctttcgggagcgcctgaagcccgtt	
GmR, resistance to gentamycin	KanR, resistance to kanamycin AmpR, resistance to ampicillin	P102

Table 2-1. Bacterial strains, plasmids, and oligonucleotides

(2009). (2009).

2. Galley, N. F. et al. *Clostridioides difficile* canonical L,D-transpeptidases catalyze a novel type of peptidoglycan cross-links and are not required for β-lactam resistance. Journal of Biological Chemistry 300, 105529 (2024).

3.Schmitz, A. M. et al. Generation of a *Gluconobacter oxydans* knockout collection for improved extraction of rare earth elements. Nat Commun 12, 6693 (2021). 4.Alamán-Zárate, M. G. et al. Unusual 1-3 peptidoglycan cross-links in Acetobacteraceae are made by L,D-transpeptidases with a catalytic domain distantly related to YkuD domains.

Journal of Biological Chemistry 300, 105494 (2024).

5.Karunakaran, R. et al. Thiamine is synthesized by a salvage pathway in Rhizobium leguminosarum bv. viciae strain 3841. J Bacteriol 188, 6661–6668 (2006).

6.Eckert, C., Lecerf, M., Dubost, L., Arthur, M. & Mesnage, S. Functional Analysis of AtlA, the Major N -Acetylglucosaminidase of Enterococcus faecalis. J Bacteriol 188, 8513–8519 (2006).

7.Kovach, M. E. et al. Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. Gene 166, 175–176 (1995). 8.Fricke, P. M. et al. Highly tunable TetR-dependent target gene expression in the acetic acid bacterium *Gluconobacter oxydans*. Appl Microbiol Biotechnol 105, 6835–6852 (2021).

2.2 Chemicals and enzymes

All chemicals and enzymes (analytical grade) were purchased from Sigma-Aldrich, Fisher Scientific and MP Biomedical. Enzymes for molecular biology experiments were purchased from New England Biolabs.

2.3 Growth media and buffers

2.3.1 Growth media and conditions

C. difficile was grown in Brain Heart Infusion broth (BHI). Three flasks, each containing 100 mL of BHI medium (Oxoid), were inoculated with *C. difficile* R20291 to an initial $OD_{600}=0.01$. Cultures were incubated at 37 °C in an anaerobic cabinet under an atmosphere containing 80% nitrogen, 10% hydrogen and 10% carbon dioxide. Cells in exponential growth phase ($OD_{600}=0.7$) were harvested by centrifugation at 6,000 rpm for 10 minutes at 25 °C.

G. oxydans B58 and isogenic derivatives were grown in yeast peptone mannitol (YPM; 5 g/L yeast extract, 3 g/L peptone, 25 g/L mannitol) broth or agar at 30 °C under agitation (200 rpm). *G. oxydans* cultures were inoculated with an overnight preculture at an OD₆₀₀=0.05 and grown for 36 h to stationary phase in 500 mL baffled flasks. *G. oxydans* transposon mutants were grown in the presence of kanamycin (100 µg/mL) and gentamicin (10 µg/mL) for complementation experiments Ldt_{Go2} expression in *G. oxydans* was induced by adding 100 ng/mL anhydrotetracycline to the media at an OD₆₀₀=0.5. Cells in stationary growth phase (OD₆₀₀ = 0.9) were harvested by centrifugation at 8,000 rpm for 15 minutes at room temperature.

R. leguminosarum bv. *viciae* strain 3841 was grown at 28°C in TY (5 g/L Tryptone + 3 g/L Yeast Extract + 1.3 g/L CaCl₂.6H₂O) broth or agar (15 g/L) and MM (CaCl₂·2H₂O (0.51 mM), CoCl₂·6H₂O (4.2 μ M), EDTA-Na₂ (1 μ M), and FeSO₄·7H₂O (0.04 mM)). Liquid cultures were grown in 2L flasks under agitation (180 rpm).

For cloning *E. coli* was grown in LB media (10g/L tryptone, 10g/L NaCl and 5g/L Yeast extract) at 30 °C and 200 rpm under agitation supplemented with the 100 μ g/mL ampicillin. For heterologous expression, *E. coli* was grown in auto-induction medium at 30 °C under agitation (200 rpm) supplemented with 100 μ g/mL ampicillin. Cells in exponential growth phase (OD₆₀₀ = 0.7) were harvested by centrifugation at 8,000 rpm for 15 minutes at room temperature.

Auto Induction Medium (AIM)

NZY (Autoclave) NZ amines (10g/L) Yeast Extract- 5g/L

20X NPS (Autoclave) Ammonium sulphate (0.5M) Potassium di-hydrogen phosphate (1M) Di-sodium hydrogen phosphate (1M) **50X 5052** (Autoclave) Glycerol (250g/L) Glucose (25g/L) α - Lactose (100g/L) **1000X trace salts** (Filter sterilize) Ferric chloride (0.1M) Calcium chloride (1M) Manganese chloride (1M) Zinc sulphate (1M) Cobalt chloride (0.2M) Copper chloride (0.1M) Nickel chloride (0.2M) Sodium molybdate (0.1M) Sodium selenite (0.1M) Boric acid (0.1M) 1M Magnesium sulphate (Autoclave)

Mix:

In 900 mL of NZY add 20 mL of 50X 5052, 50 mL of 20X NPS, 1 mL of 1M Magnesium sulphate and 1 mL of 1000X trace salts mix.

2.3.2 Sodium Borate Buffer for muropeptide reduction

2.3.3 HPLC buffers for LC and LC-MS/MS

Buffer A

Formic acid 0.1% (v/v) Water 99.9% (v/v)

Buffer B

Acetonitrile 99.9% (v/v) Formic acid 0.1% (v/v)

2.3.4 Mutanolysin buffer

Mutanolysin was resuspended in $10 \text{mM} \text{NaH}_2\text{PO}_4$ (pH 5.5). PG digestions were made in the same buffer ($10 \text{mM} \text{NaH}_2\text{PO}_4$ at pH 5.5).

2.4 Methods

2.4.1 Extraction and purification of peptidoglycan

A similar protocol was used for the 3 organisms studied but culture volumes and centrifugation conditions were slightly different.

Cell pellets were resuspended in water an immediately cryopreserved in liquid nitrogen. Frozen pellets were resuspended in 10 mL of boiling MilliQ water for cellular

lysis, followed by the addition of an equal volume of boiling 8% w/v sodium dodecyl sulfate (SDS). Following a 30-minute incubation at 100 °C, the samples were cooled to room temperature. Insoluble cell walls were isolated by ultracentrifugation at 150,000×g for 90 minutes (*C. difficile*) or 120 minutes (for *E. coli, G. oxydans and R. leguminasurm*) using a Beckman Coulter Type 50.2 rotor. The supernatant was discarded, and the pellet was resuspended in MilliQ water and homogenized via sonication for 10 seconds. The samples were subjected to five sequential washes by ultracentrifugation at 150,000×g for 90 or 120 minutes, accordingly, using a Beckman Coulter MLA-80 rotor to remove residual SDS. The purified, insoluble peptidoglycan was lyophilized and subsequently reconstituted to a concentration of 10 mg/mL.

2.4.2 Preparation of soluble muropeptides

One milligram of purified PG was digested overnight in 50 mM sodium phosphate buffer (pH 5.5) with 25U of mutanolysin in a final volume of 100μ L. After heat inactivation of the mutanolysin (5 minutes at 95°C), an equal volume of 200 mM sodium borate buffer (pH 9.0) was added with 1% (w/v) sodium borohydride to reduce muropeptides. After 20 minutes at room temperature, the pH was adjusted to pH 4 using phosphoric acid.

2.4.3 Separation of soluble muropeptides by HPLC

Fractionation of material corresponding to the mutanolysin digestion was carried out on a preparative (Hypersil Gold aQ, 1.9 μ m particles, 150 × 2.1 mm; Thermo Fisher Scientific) at a temperature of 50 °C. and separated with at a flow rate of 300 μ L/min using the gradient shown in Figure 2-1. Individual peaks were observed to confirm the muropeptide digestion and qualitative determine the monomer, dimer and trimer abundance.



Figure 2-1. HPLC gradient to separate muropeptides

Buffer A corresponds to (0.1% w/v formic acid in water) and Buffer B corresponds to (0.1% w/v formic acid in acetonitrile). The flow was set at 300 μ L/min.

2.4.4 LC-MS/MS

LC-MS analysis was performed using an Ultimate 3000 HPLC system (Dionex/Thermo Fisher Scientific) coupled to a high-resolution Orbitrap Exploris 240 mass spectrometer (Thermo Fisher Scientific). Muropeptides were separated on a C18 analytical column (Hypersil Gold aQ, 1.9 μ m particles, 150 × 2.1 mm; Thermo Fisher Scientific) maintained at 50 °C, with elution parameters described previously.

The Orbitrap Exploris 240 mass spectrometer was operated in positive electrospray ionization mode (H-ESI high flow) with full scan acquisition (m/z 150–2250) at a resolution of 120,000 (FWHM) at m/z 200. The normalized AGC target was set to 100%, with automated maximum ion injection time (IT). Data-dependent MS/MS acquisition was performed in 'Top 5' mode (except for Chapter 3, where 'Top 25' was also employed) with the following parameters: resolution 30,000; AGC 100%, automated IT, and normalized collision energy of 25%.

2.4.5 Plasmid constructions

To complement the transposon insertion in the Idt_{Go2} gene of *G. oxydans*, the plasmid pBBR-TetR-Go2227 was built. This plasmid is derived from pBBR1MCS-5-TgdhM-tetR-mNG and allows the inducible expression of proteins under the control of the tetracycline promoter.

Golden Gate assembly was employed to construct pBBR-TetR-Go2227. Three PCR fragments were amplified:

- 1. The gentamicin resistance cassette (1632 bp) using primers SM_0729 and SM_0730.
- 2. The pBBR1 origin of replication and the TetR gene (4029 bp) using primers SM_0725 and SM_0726.
- 3. The full-length *ldtGo2* gene (1021 bp) using primers SM_0727 and SM_0728.

pBBR1MCS-5-TgdhM-tetR-mNG and *G. oxydans* chromosomal DNA served as templates for PCR amplification. The purified PCR products were mixed in equimolar ratios and assembled using the NEBridge Golden Gate Assembly Kit (*Bsa*I-HF v2). Candidate plasmids were screened by PCR and verified by full sequencing (Plasmidsaurus.com) to ensure the absence of mutations.

Additionally, pET-Go2227, a pET2818 derivative, was created for expression of the fulllength LdtGo2 enzyme in *E. coli*. This plasmid utilizes a synthetic DNA fragment with codon optimization for *E. coli* (provided by Genewiz). The synthetic open reading frame, including a stop codon, was cloned into pET2818 as *Ncol-Xhol* fragment.

2.4.6 Preparation of *G. oxydans* competent cells and transformation

G. oxydans cells were grown in 100 mL of YPM medium until reaching an OD₆₀₀ of 0.9. The cells were harvested by centrifugation at 4000 x *g* for 10 minutes at 4 °C and washed three times with 1mM HEPES buffer (pH 7.0). Following the washes the cells were resuspended in 250 μ L of the same buffer.

Electroporation was performed using 50 μ L of electrocompetent cells and 100 ng of plasmid DNA in a volume of 1-2 μ L. The electroporation parameters were set at 2 kV, 25 μ F, and 200 Ω , with 1 mm cuvettes used for the procedure.

Immediately after electroporation, 800 μ L of YPM medium supplemented with 0.25% (w/v) MgSO₄ and 0.15% (w/v) CaCl₂ was added to the cells. The cells were then allowed to recover with agitation for 16 hours before being plated on YPM agar supplemented with kanamycin and gentamicin for selection of transformants.

Chapter 3

3. Optimization of LC-MS/MS data acquisition

3.1 Introduction

The ability to perform both quantitation and unbiased analysis was crucial for the investigations conducted in this thesis. Samples were injected on a liquid chromatography coupled to mass spectrometry (LC-MS/MS) system. Muropeptides were identified and analyzed using mass spectrometry (MS) and PGFinder. Structure validation was achieved through fragmentation (MS/MS). Optimizing mass spectrometry data acquisition was essential to enhance data quality and improve our ability to accurately characterize PG composition. Moreover, optimized data acquisition simplifies downstream analysis in PGFinder, ensuring a well-curated dataset for robust conclusions.

Reversed-phase liquid chromatography (LC), a well-established technique for peptide analysis was coupled with Mass Spectrometry (LC-MS/MS) (Steen and Mann, 2004). Our experimental conditions have previously been described in the literature, with acetonitrile as a solvent, preferred over methanol due to its higher solvent strength and lower viscosity (Patel et al., 2021; van der Aart et al., 2018).

PGFinder was used as the primary analytical tool for this project. PGFinder is a software specifically designed for analyzing deconvoluted MS data. It uses a list of muropeptides and their theoretical monoisotopic masses (called a database) to search for matching observed monoisotopic masses in the deconvoluted data. Ion intensities associated with deconvoluted observed masses are used to determine muropeptide abundances (Patel et al., 2021; Patel, 2021).

Alternative analytical methods like Chemometrics, which rely on run-to-run retention time alignment across multiple LC runs, can encounter reproducibility challenges due to variations in temperature, flow rate, and column degradation. These factors can compromise the accuracy and reliability of analyses. Even with tighter instrumental controls, that improve reproducibility, it has been noted that to effectively use LC data with MS and MS-MS strategies for proteomics database searching, it was necessary to align both LC and MS data (Cohen and Schure, 2008).

All our data were collected with an Orbitrap Exploris 240 mass spectrometer. This instrument is a cutting-edge device that offers exceptional sensitivity, resolution, and accuracy. It works by trapping ions in a hyper-conical electrode assembly and measuring their oscillation frequency to determine their mass-to-charge ratio (m/z) (Eliuk and Makarov, 2015; Strupat et al., 2016). Developed in 1996 by Alexander Makarov (Makarov, 2000), Orbitrap technology has rapidly gained popularity due to its versatility and performance. It has found widespread application in various fields, including proteomics, clinical research, forensics, environmental analysis, and now peptidoglycomics (Eliuk and Makarov, 2015; Makarov, 2000). Orbitrap technology provides high-resolution accurate-mass (HRAM) data with sub-1ppm mass accuracy.

It can be used for both targeted and untargeted analysis, enabling the identification and estimation of a wide range of molecules.

Surprisingly, previous MS analyses using PGFinder identified several muropeptides that were not fragmented. Since we transitioned to a new mass spectrometry facility to perform the data acquisition, we investigated whether increasing the fragmentation frequency could generate MS/MS data for these muropeptides to confirm their structures. This part of my work provided an initial opportunity for me to familiarize with PG composition analysis. I began by analyzing *Clostridioides difficile* PG structure using the strategy originally described by Patel (2021) while adjusting MS/MS data acquisition parameters.

We tested two fragmentation conditions (henceforth referred to as "Top5" and "Top25"), where the top 5 or 25 most abundant ions in each MS scan were fragmented, respectively. My objectives were as follows:

- Compare PGFinder (MS) and Byos® (MS/MS) monomer analysis using a Top5 fragmentation setup as a reference.
- Evaluate the impact of Top25 fragmentation setup on data coverage for monomers.
- Confirm the presence of modified monomers in *C. difficile*.

3.2 Results

3.2.1 MS search: Identification of PG monomers using PGFinder with the Top5 fragmentation setup database

Muropeptides from *C. difficile* R20291 were analysed by LC-MS/MS as previously described (Patel, 2021). For each fragmentation cycle, the five most abundant ions were targeted (Top5). After MS data collection, an unbiased PGFinder (v.0.1.1a) search was performed to identify monomers. Given the high level of GlcNAc deacetylation (GlcN) reported in *C. difficile* PG (Peltier et al., 2011), a database named DB_Cd1 (Table 3-1) was created, consisting of two sets of 211 monomers (a total of 422), each with peptide stems containing 2 to 5 residues and either GlcN-MurNAc or GlcNAc-MurNAc as their disaccharide moieties (Fig 3-1A).

The deconvoluted dataset was analyzed with a 10 ppm tolerance, allowing for the consolidation of in-source decay fragmentation within a 30-second window (Figure 3-2). Multiple matches corresponding to the same muropeptide were consolidated as described previously (Patel et al., 2021). The PGFinder search generated 76 matches, of which 14 were mass coincidences (shaded in blue) (Table 3-2). A mass coincidence occurs when different chemical structures exhibit very similar or identical monoisotopic masses.

Α

В

Structure	Number	Structure	Number	Peptide stem	Number
gm-AE	1	gm(-Ac)-AE	1	A	1
gm-AEJ	1	gm(-Ac)-AEJ	1	AE	1
gm-AEJX	19	gm(-Ac)-AEJX	19	AEJ	1
gm-AEJXY	190	gm(-Ac)-AEJXY	190	AEJX	20
Total	211	Total	211	AEJXY	400
Total Mur	opeptides		422	Total	423

Figure 3-1. Summary of the PGFinder (A) and Byos® (B) database compositions used for the initial searches

A. DB_Cd1 database contains 211 structures with the disaccharide GlcNAc-MurNAc (gm) and 211 structures with the GlcN-MurNAc (gm(-Ac)) disaccharide. **B**. DB_Cd10 database contains 423 structures the peptide stem database used for Byos[®] search; for pentapeptide stems, AEJXY and AEJYX sequences were used, allowing modifications of the N-terminal Alanine residue by GlcNAc-MurNAc or GlcN-MurNAc disaccharides.

Table 3-1. Db_Cd1 composition

Structure	Monoisotopicmass	Structure	Monoisotopicmass	Structure	Monoisotopicmass	Structure	Monoisotopicmass
gm(-Ac)-AE 1	656.27529	gm(-Ac)-AEJMI 1	1072.48463	gm-AE 1	698.2859	gm-AEJMM 1	1132.45167
gm(-Ac)-AEJ 1	828.36008	gm(-Ac)-AEJMK 1	1087.49553	gm-AEJ 1	870.37069	gm-AEJMN 1	1115.45411
gm(-Ac)-AEJA 1	899.39722	gm(-Ac)-AEJMM 1	1090.44106	gm-AEJA 1	941.40783	gm-AEJMP 1	1098.46394
gm(-Ac)-AEJAA 1 gm(-Ac)-AEIC 1	970.43436	gm(-Ac)-AEJIVIN 1 gm(-Ac)-AEIMP 1	1073.4435	gm-AEJAA 1	1012.44497	gm-AEJIVIQ 1	1088 44321
gm(-Ac)-AEJCA 1	1002.40641	gm(-Ac)-AEJMQ 1	1087.45915	gm-AEJCA 1	1044.41702	gm-AEJMT 1	1102.45886
gm(-Ac)-AEJCC 1	1034.37846	gm(-Ac)-AEJMS 1	1046.4326	gm-AEJCC 1	1076.38907	gm-AEJMV 1	1100.47959
gm(-Ac)-AEJCE 1	1060.41186	gm(-Ac)-AEJMT 1	1060.44825	gm-AEJCE 1	1102.42247	gm-AEJMW 1	1187.49049
gm(-Ac)-AEJCG 1	988.39073	gm(-Ac)-AEJMV 1	1058.46898	gm-AEJCG 1	1030.40134	gm-AEJMY 1	1164.47451
gm(-Ac)-AEJCH[1 gm(-Ac)-AEJCK[1	1059 46423	gm(-AC)-AEJIVIW 1	1145.47988	gm-AEJCH 1	1110.43879	gm-AEJN 1	984.41362 1055 45076
gm(-Ac)-AEICN 1	1045.4122	gm(-Ac)-AEIN 1	942.40301	gm-AEICN 1	1087.42281	gm-AEINE 1	1113.45621
gm(-Ac)-AEJCP 1	1028.42203	gm(-Ac)-AEJNA 1	1013.44015	gm-AEJCP 1	1070.43264	gm-AEJNG 1	1041.43508
gm(-Ac)-AEJCS 1	1018.4013	gm(-Ac)-AEJNE 1	1071.4456	gm-AEJCS 1	1060.41191	gm-AEJNH 1	1121.47253
gm(-Ac)-AEJD 1	943.38702	gm(-Ac)-AEJNG 1	999.42447	gm-AEJD 1	985.39763	gm-AEJNN 1	1098.45655
gm(-Ac)-AEJDA 1	1014.42416	gm(-Ac)-AEJNH 1 gm(-Ac)-AEJNN 1	1079.46192	gm-AEJDA 1	1056.43477	gm-AEJNP 1	1081.46638
gm(-Ac)-AEIDD11	1058.41396	gm(-Ac)-AEINP[1	1039.45577	gm-AEIDD 1	1100.42457	gm-AEINT[1	1085.4613
gm(-Ac)-AEJDG 1	1000.40848	gm(-Ac)-AEJNQ 1	1070.46159	gm-AEJDG 1	1042.41909	gm-AEJNW 1	1170.49293
gm(-Ac)-AEJDH 1	1080.44593	gm(-Ac)-AEJNT 1	1043.45069	gm-AEJDH 1	1122.45654	gm-AEJP 1	967.42345
gm(-Ac)-AEJDN 1	1057.42995	gm(-Ac)-AEJNW 1	1128.48232	gm-AEJDN 1	1099.44056	gm-AEJPA 1	1038.46059
gm(-Ac)-AEJDP 1	1040.43978	gm(-Ac)-AEJP 1	925.41284	gm-AEJDP 1	1082.45039	gm-AEJPE 1	1096.46604
gm(-Ac)-AEIDU[1	1071.4430	gm(-Ac)-AEIPE 1	1054 45543	gm-AFIDT 1	1086 44531	gm-AFIPO11	1095 48203
gm(-Ac)-AEJDW 1	1129.46633	gm(-Ac)-AEJPP 1	1022.4656	gm-AEJDW 1	1171.47694	gm-AEJPT 1	1068.47113
gm(-Ac)-AEJE 1	957.40267	gm(-Ac)-AEJPQ 1	1053.47142	gm-AEJE 1	999.41328	gm-AEJPW 1	1153.50276
gm(-Ac)-AEJEA 1	1028.43981	gm(-Ac)-AEJPT 1	1026.46052	gm-AEJEA 1	1070.45042	gm-AEJQ 1	998.42927
gm(-Ac)-AEJED 1	1072.42961	gm(-Ac)-AEJPW 1	1111.49215	gm-AEJED 1	1114.44022	gm-AEJQA 1	1069.46641
gm(-Ac)-AEJEE 1 gm(-Ac)-AEJEG 1	1086.44526	gm(-Ac)-AEJQ 1 gm(-Ac)-AEJOA 1	956.41866 1027.4558	gm-AEJEE 1	1128.45587 1056.43474	gm-AEJQC 1	1101.43846
gm(-Ac)-AEJEI 1	1070.48673	gm(-Ac)-AEJOCI1	1059.42785	gm-AEJEII1	1112.49734	gm-AEJQQ11	1126.48785
gm(-Ac)-AEJES 1	1044.4347	gm(-Ac)-AEJQE 1	1085.46125	gm-AEJF 1	1017.4391	gm-AEJQT 1	1099.47695
gm(-Ac)-AEJF 1	975.42849	gm(-Ac)-AEJQQ 1	1084.47724	gm-AEJFA 1	1088.47624	gm-AEJQW 1	1184.50858
gm(-Ac)-AEJFA 1	1046.46563	gm(-Ac)-AEJQT 1	1057.46634	gm-AEJFC 1	1120.44829	gm-AEJR 1	1026.4718
gm(-AC)-AEJEC 1	10/8.43/68	gm(-AC)-AEJQW 1	1142.49/9/ 984 46119	gm-AEJED 1	1132.40004	gm-AEJRA 1 gm-AFIPC 1	1129 48000
gm(-Ac)-AFIFF[1	1104 47108	gm(-Ac)-AFIRA 1	1055 49833	gm-AFIFF[1	1164 50751	gm-AFIRD 1	1141 49874
gm(-Ac)-AEJFF 1	1122.4969	gm(-Ac)-AEJRC 1	1087.47038	gm-AEJFG 1	1074.46056	gm-AEJRE 1	1155.51439
gm(-Ac)-AEJFG 1	1032.44995	gm(-Ac)-AEJRD 1	1099.48813	gm-AEJFH 1	1154.49801	gm-AEJRH 1	1163.53071
gm(-Ac)-AEJFH 1	1112.4874	gm(-Ac)-AEJRE 1	1113.50378	gm-AEJFI 1	1130.52316	gm-AEJRI 1	1139.55586
gm(-Ac)-AEJFI 1	1088.51255	gm(-Ac)-AEJRH 1	1121.5201	gm-AEJFK 1	1145.53406	gm-AEJRM 1	1157.51229
gm(-Ac)-AEIFN 1	1089 47142	gm(-Ac)-AEIRM[1	1115 50168	gm-AFIFP 1	11114 49186	gm-AFIRP 1	1123 52456
gm(-Ac)-AEJFP 1	1072.48125	gm(-Ac)-AEJRN 1	1098.50412	gm-AEJFQ 1	1145.49768	gm-AEJRO 1	1154.53038
gm(-Ac)-AEJFQ 1	1103.48707	gm(-Ac)-AEJRP 1	1081.51395	gm-AEJFR 1	1173.54021	gm-AEJRR 1	1182.57291
gm(-Ac)-AEJFR 1	1131.5296	gm(-Ac)-AEJRQ 1	1112.51977	gm-AEJFS 1	1104.47113	gm-AEJRT 1	1127.51948
gm(-Ac)-AEJFS 1	1062.46052	gm(-Ac)-AEJRR 1	1140.5623	gm-AEJFT 1	1118.48678	gm-AEJRV 1	1125.54021
gm(-Ac)-AEJFT[1 gm(-Ac)-AEJEV[1	1076.47617	gm(-Ac)-AEJRT[1 gm(-Ac)-AEJRV[1	1085.50887	gm-AEJFV 1	1116.50751	gm-AEJKW 1	1212.55111
gm(-Ac)-AEIFW[1	1161.5078	gm(-Ac)-AEIRW[1	1170.5405	gm-AEJFV/1	1180.50243	gm-AEISA 1	1028.43986
gm(-Ac)-AEJFY 1	1138.49182	gm(-Ac)-AEJS 1	915.39211	gm-AEJG 1	927.39215	gm-AEJSD 1	1072.42966
gm(-Ac)-AEJG 1	885.38154	gm(-Ac)-AEJSA 1	986.42925	gm-AEJGA 1	998.42929	gm-AEJSE 1	1086.44531
gm(-Ac)-AEJGA 1	956.41868	gm(-Ac)-AEJSD 1	1030.41905	gm-AEJGG 1	984.41361	gm-AEJSK 1	1085.49768
gm(-Ac)-AEJGG[1 gm(-Ac)-AEJGG[1	942.403	gm(-Ac)-AEJSK 1	1043.48707	gm-AEJGH 1	1064.45106	gm-AEJSN 1	10/1.44565
gm(-Ac)-AEIGM[1	1016.42203	gm(-Ac)-AEISP 1	1023.43304	gm-AEIGP 1	1024.44491	gm-AEISO 1	1085.4613
gm(-Ac)-AEJGP 1	982.4343	gm(-Ac)-AEJSQ 1	1043.45069	gm-AEJGQ 1	1055.45073	gm-AEJSR 1	1113.50383
gm(-Ac)-AEJGQ 1	1013.44012	gm(-Ac)-AEJSR 1	1071.49322	gm-AEJGR 1	1083.49326	gm-AEJSS 1	1044.43475
gm(-Ac)-AEJGR 1	1041.48265	gm(-Ac)-AEJSS 1	1002.42414	gm-AEJGS 1	1014.42418	gm-AEJST 1	1058.4504
gm(-Ac)-AEJGS 1	972.41357	gm(-Ac)-AEJST 1	1016.43979	gm-AEJGV 1	1026.46056	gm-AEJSW 1	1143.48203
gm(-Ac)-AEJGV [1 gm(-Ac)-AEJGW [1	984.44995 1071 46085	gm(-Ac)-AEJSW 1	929 40776	gm-AEJGW [1	1007 4296	gm-AFITA 1	9/1.41837
gm(-Ac)-AEJH 1	965.41899	gm(-Ac)-AEJTA 1	1000.4449	gm-AEJHA 1	1078.46674	gm-AEJTC 1	1074.42756
gm(-Ac)-AEJHA 1	1036.45613	gm(-Ac)-AEJTC 1	1032.41695	gm-AEJHE 1	1136.47219	gm-AEJTE 1	1100.46096
gm(-Ac)-AEJHE 1	1094.46158	gm(-Ac)-AEJTE 1	1058.45035	gm-AEJHH 1	1144.48851	gm-AEJTG 1	1028.43983
gm(-Ac)-AEJHH 1	1102.4779	gm(-Ac)-AEJTG 1	986.42922	gm-AEJHP 1	1104.48236	gm-AEJTK 1	1099.51333
gm(-Ac)-AEJHP [1 gm(-Ac)-AEJHO [1	1002.47175	gm(-AC)-AEJTTT	977 42849	gm-AFIHS 1	1094 46163	gm_AFIV/1	969 4391
gm(-Ac)-AEJHS 1	1052.45102	gm(-Ac)-AEJVA 1	998.46563	gm-AEJHT 1	1108.47728	gm-AEJVA 1	1040.47624
gm(-Ac)-AEJHT 1	1066.46667	gm(-Ac)-AEJVC 1	1030.43768	gm-AEJHW 1	1193.50891	gm-AEJVC 1	1072.44829
gm(-Ac)-AEJHW 1	1151.4983	gm(-Ac)-AEJVD 1	1042.45543	gm-AEJI 1	983.45475	gm-AEJVD 1	1084.46604
gm(-Ac)-AEJI 1	941.44414	gm(-Ac)-AEJVE 1	1056.47108	gm-AEJIA 1	1054.49189	gm-AEJVE 1	1098.48169
gm(-Ac)-ACIA II	1044.45333	gm(-Ac)-AEIVEI1	1055.52345	gm-AEIID 1	1098,48169	gm-AEIVK11	1097.53406
gm(-Ac)-AEJID 1	1056.47108	gm(-Ac)-AEJVN 1	1041.47142	gm-AEJIG 1	1040.47621	gm-AEJVN 1	1083.48203
gm(-Ac)-AEJIG 1	998.4656	gm(-Ac)-AEJVP 1	1024.48125	gm-AEJIH 1	1120.51366	gm-AEJVP 1	1066.49186
gm(-Ac)-AEJIH 1	1078.50305	gm(-Ac)-AEJVQ 1	1055.48707	gm-AEJII 1	1096.53881	gm-AEJVQ 1	1097.49768
gm(-AC)-AEJII 1 gm(-AC)-AEJINI 1	1054.5282	gm(-AC)-AEJVS 1 gm(-AC)-AEJVT 1	1014.46052 1028.47617	gm-AEJIN 1	1097.49768	gm-AEJVS 1 gm-AFIV/T 1	1056.4/113
gm(-Ac)-AEIIP11	1038.4969	gm(-Ac)-AEIVV11	1026.4969	gm-AEIIOI1	1111.51333	gm-AEJVV11	1068.50751
gm(-Ac)-AEJIQ 1	1069.50272	gm(-Ac)-AEJVW 1	1113.5078	gm-AEJIS 1	1070.48678	gm-AEJVW 1	1155.51841
gm(-Ac)-AEJIS 1	1028.47617	gm(-Ac)-AEJW 1	1014.43939	gm-AEJIT 1	1084.50243	gm-AEJW 1	1056.45
gm(-Ac)-AEJIT 1	1042.49182	gm(-Ac)-AEJWA 1	1085.47653	gm-AEJIV 1	1082.52316	gm-AEJWA 1	1127.48714
gm(-AC)-AEJIV 1	1040.51255 1127 52345	gm(-AC)-AEJWC 1	1117.44858	gm-AEJIW 1	1109.53406 998.46565	gm-AEJWC 1	1185 /0250
gm(-Ac)-AEIK 1	956.45504	gm(-Ac)-AEIWTI1	1115.48707	gm-AEIKA 1	1069.50279	gm-AEIWT11	1157,49768
gm(-Ac)-AEJKA 1	1027.49218	gm(-Ac)-AEJWW 1	1200.5187	gm-AEJKD 1	1113.49259	gm-AEJWW 1	1242.52931
gm(-Ac)-AEJKD 1	1071.48198	gm(-Ac)-AEJY 1	991.42341	gm-AEJKE 1	1127.50824	gm-AEJY 1	1033.43402
gm(-Ac)-AEJKE 1	1085.49763	gm(-Ac)-AEJYA 1	1062.46055	gm-AEJKG 1	1055.48711	gm-AEJYA 1	1104.47116
gm(-Ac)-AEJKG 1	1013.4705	gm(-Ac)-AEJYC[1 gm(-Ac)-AEJYD[1	1094.4320	gm-AEJKH 1	1133.32450 1111 5/1071	gm-AEJIC 1	1120.44321
gm(-Ac)-AEIKI11	1069.5391	gm(-Ac)-AEIYEI1	1120.466	gm-AEIKK11	1126.56061	gm-AEJYEI1	1162.47661
gm(-Ac)-AEJKK 1	1084.55	gm(-Ac)-AEJYG 1	1048.44487	gm-AEJKN 1	1112.50858	gm-AEJYG 1	1090.45548
gm(-Ac)-AEJKN 1	1070.49797	gm(-Ac)-AEJYH 1	1128.48232	gm-AEJKP 1	1095.51841	gm-AEJYH 1	1170.49293
gm(-Ac)-AEJKP 1	1053.5078	gm(-Ac)-AEJYI 1	1104.50747	gm-AEJKQ 1	1126.52423	gm-AEJYI 1	1146.51808
gm(-AC)-AEJKQ[1	1004.01302	gm(-AC)-AEJYK [1 gm(-AC)-AEJYK [1	1105 46624	gin-AEJKK 1	1134.500/0	gm-AEJYK 1	1101.52898
gm(-Ac)-AEIKTI1	1057.50272	gm(-Ac)-AEIYP11	1088.47617	gm-AEIM 1	1001.41118	gm-AEIYP11	1130,48678
gm(-Ac)-AEJKW 1	1142.53435	gm(-Ac)-AEJYQ 1	1119.48199	gm-AEJMA 1	1072.44832	gm-AEJYQ 1	1161.4926
gm(-Ac)-AEJM 1	959.40057	gm(-Ac)-AEJYR 1	1147.52452	gm-AEJMC 1	1104.42037	gm-AEJYR 1	1189.53513
gm(-Ac)-AEJMA 1	1030.43771	gm(-Ac)-AEJYS 1	1078.45544	gm-AEJMD 1	1116.43812	gm-AEJYS 1	1120.46605
gm(-Ac)-AEJIVIC 1	1002.40976	gm(-Ac)-AEJY1 1 gm(-Ac)-AEJY1 1	1092.47109	gm_AFIME 1	1130.43377	gm-AEJYIJ1 gm-AFIVV/11	1132 20242
gm(-Ac)-AEJMEI1	1088.44316	gm(-Ac)-AEJYW11	1177.50272	gm-AEJMH11	1138.47009	gm-AEJYW 1	1219.51333
gm(-Ac)-AEJMF 1	1106.46898	gm(-Ac)-AEJYY 1	1154.48674	gm-AEJMI 1	1114.49524	gm-AEJYY 1	1196.49735
gm(-Ac)-AEJMH 1	1096.45948			gm-AEJMK 1	1129.50614		
Step 1: Upload Deconvoluted Data

Click Upload to upload a .txt file output by MaxQuant (example file), or an .ftrs file.

Upload Deconvoluted Data (0)

Step 2: Select Modifications

Select modifications (Hold down control / command and click to select mulitple items.)

Modification	Sodium Potassium Anh	•
	DeAc DeAc_Anh	+

Step 3: Select or Upload Mass Library

Select

Mass List	0	e_coli_monomer_masses.csv
	0	c_diff_monomer_masses.csv
		Upload Custom

(Optional) Upload Custom Mass Library

Example mass library file.

🏖 Upload Mass Library (0)

Step 4: Set PPM tolerance

Set ppm tolerance 10

Step 5: Set time window for in-source decay and salt adduct clean up

Set time delta value	0.5	
----------------------	-----	--

Step 6: Run Analysis

Click run analysis. After the analysis is complete, a download button will appear.

Figure 3- 2. PGFinder v. 0.1.1a Jupyter notebook visual interface (no longer available)

PGFinder's visual interface allows us to customize the search by selecting our own datasets (Step 1) and to look for a wide range of modifications (Step 2). The software allows us to search using either predesigned or customized datasets (Step 3). We can fine-tune the stringency of muropeptide identification by modifying the ppm tolerance (Step 4) and the time window to look for muropeptide adducts (Step 5). With a total of 92 individual monomers, the PGFinder search revealed a highly deacetylated PG composition (61 deacetylated muropeptides compared to 31 fully acetylated ones) (Table 3-3). Notably, the nine most abundant muropeptides are deacetylated, gm(-Ac)-AEJA being the most abundant.

We observed that most of the structures identified corresponded to muropeptides with pentapeptide stems, some of which exhibiting uncommon composition, like gm(-Ac)-AEJKD, gm(-Ac)-AEJGV or gm(-Ac)-AEJGG (muropeptides 8, 14 and 15 in Table 3-2). Regarding the mass coincidences, we noted that, aside from the perfect mass coincidences (Table 3-2 muropeptides 15, 22, 27, 56 and 61), one of the multiple structures displayed a lower Δ ppm, indicating a closer match between the theoretical and observed masses. This suggests that one structure is more likely to be the correct one.

Individual structures	92
	(14 mass coincidences)
Acetylated (GlcNAc-MurNAc)	31
Dipeptide	1
Tripeptide	1
Tetrapeptide	11
Pentapeptide	18
Deacetylated (GlcN-MurNAc)	61
Dipeptide	1
Tripeptide	1
Tetrapeptide	19
Pentapeptide	40
Intensity collected	1.94E+09

Table 3-2 Top5 Db_Cd1 PGFinder Search summary

1 able 5-5. Top5 PGFInder search output snows 14 mass coincidences (nightighted in blue	Table 3- 3. To	p5 PGFinder searc	h output shows	s 14 mass coincidences	(highlighted in b	lue)
---	-----------------------	-------------------	----------------	------------------------	-------------------	------

	Structure	Intensity	Abundance	RT	TheoMw	Obs	Δppm
1	gm(-Ac)-AEJA 1	1.15E+09	59.29%	7.82	899.3972	899.3946	2.9
2	gm(-Ac)-AEJG 1	3.05E+08	15.75%	3.66	885.3815	885.3800	1.7
3	gm(-Ac)-AEJ 1	1.71E+08	8.85%	3.26	828.3601	828.3583	2.2
4	gm(-Ac)-AEJF 1	7.53E+07	3.89%	22.18	975.4285	975.4267	1.9
5	gm(-Ac)-AE 1	6.49E+07	3.35%	5.58	656.2753	656.2738	2.2
5	gm(-Ac)-AEJV I	4.02E+07	2.08%	16.79	927.4285	927.4260	2.7
, s	gm(-Ac)-AEIF[1 gm(-Ac)-AEIKD[1	1.95E+07	0.74%	10.40 8 17	991.4254 1071 <i>4</i> 82	991.4215 1071 //708	2.0
9	gm(-Ac)-AEIII1	1 21F+07	0.74%	18 17	941 4441	941 4472	2.0
10	gm-AEJA 1	1.00E+07	0.52%	8.71	941.4078	941.4058	2.1
11	gm(-Ac)-AEJM 1	7.99E+06	0.41%	15.22	959.4006	959.3987	2.0
12	gm(-Ac)-AEJK 1	7.77E+06	0.40%	2.82	956.455	956.4530	2.1
13	gm(-Ac)-AEJS 1	6.32E+06	0.33%	6.70	915.3921	915.3908	1.4
14	gm(-Ac)-AEJGV 1,gm(-Ac)-AEJR 1	5.87E+06	0.30%	3.20	984.4499,984.4612	984.4596	9.8,1.6
15	gm(-Ac)-AEJGG 1,gm(-Ac)-AEJN 1	4.88E+06	0.25%	4.86	942.403	942.4004	2.7
16	gm(-Ac)-AEJAA 1	4.15E+06	0.21%	9.24	970.4344	970.4328	1.7
17	gm(-Ac)-AEJWW 1	4.10E+06	0.21%	10.32	1200.5187	1200.5227	3.4
18	gm-AEJG 1	3.48E+06	0.18%	6.79	927.3922	927.3901	2.3
19	gm(-Ac)-AEJQT 1	3.06E+06	0.16%	6.43	1057.4663	1057.4636	2.6
20	gm(-Ac)-AEJNP 1	2.59E+06	0.13%	8.25	1039.4558	1039.4537	2.0
21	gm-AEJPE 1	2.3/E+06	0.12%	7.55	1096.466	1096.4748	8.1
22		2.22E+00	0.11%	21 26	1046 4656	10/6 /629	2.4
23	gm(-Ac)-AEITA 1	2.10L+00 1 94E+06	0.11%	6.47	1000 4449	1000 4428	2.1
25	gm-AEI/1	1.67E+06	0.09%	6.00	870.3707	870.3671	4.1
26	gm(-Ac)-AEJH 1	1.58E+06	0.08%	3.00	965.419	965.4168	2.3
27	gm(-Ac)-AEJE 1,gm-AEJS 1	1.14E+06	0.06%	8.06	957.4027	957.4009	1.9
28	gm(-Ac)-AEJD 1	9.40E+05	0.05%	6.61	943.387	943.3849	2.2
29	gm-AE 1	8.65E+05	0.04%	8.53	698.2859	698.2849	1.4
30	gm(-Ac)-AEJHH 1	8.56E+05	0.04%	8.45	1102.4779	1102.4754	2.2
31	gm(-Ac)-AEJKA 1	8.42E+05	0.04%	7.54	1027.4922	1027.4892	2.9
32	gm(-Ac)-AEJIA 1	8.24E+05	0.04%	19.38	1012.4813	1012.4798	1.5
33	gm(-Ac)-AEJFS 1,gm(-Ac)-AEJYA 1	8.08E+05	0.04%	16.34	1062.4605,1062.4606	1062.4599	0.5,0.6
34	gm(-Ac)-AEJW 1	6.75E+05	0.03%	24.22	1014.4394	1014.4370	2.3
35	gm(-Ac)-AEJP 1	5.31E+05	0.03%	6.59	925.4128	925.4109	2.1
36	gm-AEJF 1	4.45E+05	0.02%	21.78	1017.4391	1017.4367	2.4
3/	gm(-AC)-AEJGV I	4.18E+05	0.022%	14.72	984.4499	984.4475	2.5
30	gm-AEIRM11	3.09L+05	0.019%	11 27	1157 5123	1157 5166	3.7
40	gm(-Ac)-AFIT[1	3 30F+05	0.010%	6 39	929 4078	929 4057	2.7
41	gm-AEJKD 1	2.91E+05	0.015%	9.12	1113.4926	1113.4908	1.6
42	gm-AEJHH 1	2.61E+05	0.013%	10.75	1144.4885	1144.4842	3.7
43	gm(-Ac)-AEJKW 1	2.53E+05	0.013%	7.29	1142.5343	1142.5281	5.4
44	gm-AEJH 1	2.52E+05	0.013%	6.21	1007.4296	1007.4276	1.9
45	gm(-Ac)-AEJPT 1,gm-AEJGV 1,gm-AEJR 1	2.48E+05	0.013%	7.23	1026.4605,1026.4606,1026.4718	1026.4691	8.3,8.2,2.6
46	gm-AEJWW 1	2.19E+05	0.011%	9.05	1242.5293	1242.5326	2.6
47	gm-AEJRC 1	2.18E+05	0.011%	6.60	1129.481	1129.4848	3.4
48	gm(-Ac)-AEJCA 1	1.95E+05	0.010%	10.12	1002.4064	1002.4115	5.1
49	gm(-Ac)-AEJSP 1,gm-AEJAA 1	1.95E+05	0.010%	10.07	1012.4449,1012.445	1012.4427	2.2,2.3
50	gm-AEJFP 1,gm-AEJMI 1	1.92E+05	0.010%	7.38	1114.4919,1114.4952	1114.4848	6.3,9.2
51	gm-AEJI 1	1.42E+05	0.007%	9.45	9/1.4184	9/1.416/	1./
52	gm AEIII1	1.410+05	0.007%	9.02 10.22	1035.4714	1055.4095	2.0
54	gm(-Ac)-AFIFC 1	1.39L+05	0.007%	5 04	1078 4377	1078 4299	7.2
55	gm(-Ac)-AEJC 1	1.14E+05	0.006%	8.86	931.3693	931.3677	1.7
56	gm(-Ac)-AEJIG 1,gm(-Ac)-AEJVA 1,gm-AEJK 1	1.08E+05	0.006%	5.76	998.4656	998.4629	2.7
57	gm(-Ac)-AEJVS 1	1.01E+05	0.005%	8.43	1014.4605	1014.4594	1.0
58	gm-AEJTT 1	9.96E+04	0.005%	10.17	1072.466	1072.4639	2.0
59	gm(-Ac)-AEJPA 1	9.96E+04	0.005%	6.29	996.45	996.4483	1.7
60	gm-AEJGR 1	9.17E+04	0.005%	19.25	1083.4933	1083.4991	5.4
61	gm(-Ac)-AEJNT 1,gm(-Ac)-AEJSQ 1	8.87E+04	0.005%	7.01	1043.4507	1043.4482	2.4
62	gm(-Ac)-AEJKG 1	8.24E+04	0.004%	5.78	1013.4765	1013.4737	2.7
63	gm(-Ac)-AEJWA 1	8.19E+04	0.004%	23.87	1085.4765	1085.4751	1.3
64	gm(-Ac)-AEJHA 1	7.78E+04	0.004%	8.04	1036.4561	1036.4537	2.3
65	gm-AEJMW 1	7.72E+04	0.004%	8.63	1187.4905	1187.4916	0.9
66	gm(-Ac)-AEJDP 1	7.31E+04	0.004%	8.60	1040.4398	1040.4384	1.3
67	gm(-Ac)-AEJGP 1 gm(-Ac)-AEJTC 1	7.14E+04	0.004%	8.1/ 5.04	982.4343 1022 117	982.4323	2.0
50	gm(-AC)-AEJIC 1 gm(-AC)-AEJIC 1 gm(-AC)-AEJIC 1	0.72E+04	0.003%	3.04	1032.41/	1084 5051	1./
70	gm(-Ac)-AFISA 1 gm(-Ac)-AFITG 1	6 34F±04	0.003%	7 41	986 4793 986 4793	986 4270	7.8,2.4
71	gm-AEIPW11	5.24F+04	0.003%	6 96	1153 5028	1153 4950	5.9
72	gm(-Ac)-AEJKE 1,gm-AEJSK 1	4.61E+04	0.002%	10.48	1085.4976,1085.4977	1085.4953	2.0,2.1
73	gm(-Ac)-AEJGQ 1,gm(-Ac)-AEJNA 1	4.33E+04	0.002%	6.32	1013.4401,1013.4402	1013.4378	2.2,2.3
74	gm-AEJFF 1	4.09E+04	0.002%	9.36	1164.5075	1164.5007	5.8
75	gm-AEJPT 1	3.83E+04	0.002%	6.90	1068.4711	1068.4808	9.1
76	gm-AEJP 1	3.71E+04	0.002%	5.86	967.4235	967.4319	8.6
	Total	1.94E+09					

3.2.2 MS/MS search: Identification of PG monomers using Byos® and a Top5 framentation setup

We next investigated the number of muropeptide identifications based on MS/MS data using the Byos® software. To rigorously validate muropeptide identifications, we established a stringent filtering criterion. The Top5 dataset was analysed in Byos® using a database called DB_Cd10, composed of peptide stems with a length between 1 to 5 residues (Fig 3-1B). Two possible disaccharide modifications were allowed at the *N*-terminal of the stem peptide (MurN-GlcNAc or MurNAc-GlcNAc).

The Byos® search output was manually inspected. Half of the 40 muropeptides identified by Byos® exhibited a complete series of *b*- and *y*- ions in their MS/MS spectra (Fig. 3-3). To validate identifications with incomplete *b*- or *y*- series, we established a rule requiring at least half of the expected *b*- or *y*- ions should be present.

For instance, a muropeptide with a pentapeptide stem (gm(-Ac)-AEJAA) is expected to generate an ideal MS/MS spectrum showing 4 *b*- and 4 *y*- ions. To validate such a structure, we required the presence of more than two *b*-ions and more than two *y*ions (Fig 3-3A). This criterion meant that muropeptides were not validated when only one *b*- or *y*- series had more than half of the expected ions whilst the other had not (Fig 3-3B) or when none of the *b*- and *y*- series was complete (Fig 3-3C). A summary of this analysis is presented in Fig 3-3D. Ion series meeting the criteria are highlighted in green and those not meeting the criteria are highlighted in red. Notably, all the validated pentapeptides contain alanine as the fourth residue (shown in red font).

Out of the 40 muropeptides identified with Byos®, 32 were validated and the majority (27) of them were deacetylated. No mass coincidences were identified among the validated muropeptides, suggesting that Byos® methodology effectively clarifies the mass coincidences observed during the initial PGFinder search (Table 3-4).

3.2.3 Merging PGFinder and Byos® analyses using a Top5 fragmentation setup

To comprehensively analyze the data, as a final analysis we integrated the findings from both Byos® and PGFinder searches. Table 3-5 summarizes the muropeptides identified and validated by each method, with Byos® and the PGFinder search results. A correlation was observed between muropeptide abundance and the likelihood of validation. The least abundant muropeptide identified with Byos® has an intensity of 1.3E+05, which was not validated (shown in red in the last column). On the other hand, most validated muropeptides identified by Byos® (in green) clustered with most abundant muropeptides identified by PGFinder. This suggest that more abundant precursor ions generally yield higher quality MS/MS data collection, leading to a greater likelihood of meeting our criteria.





Annotated chromatograms generated by Byos[®] for gm(-Ac)-AEJAA (**A**), gm-AEJH (**B**) and gm-AEJR (**C**), illustrating the output obtained from Byos[®]. The number of *b*- and *y*- ions was collected (**D**) and highlighted in green for series with more than half of the expected ions and in red the ones that did not meet the criteria.

Table 3- 4. Top5 Db_Cd10 PGFinder Se	earch summa	ry
Identifications	40	
(No mass coinc	cidences)	
Validated	32	
Acetylated (GlcNAc-MurNAc)	5	
Dipeptide	1	
Tripeptide	1	
Tetrapeptide	4	
Pentapeptide	0	
Deacetylated (GlcN-MurNAc)	27	
Dipeptide	1	
Tripeptide	1	
Tetrapeptide	18	
Pentapeptide	7	
Intensity collected	N/A	

By combining both searches we were able to solve 8 of the 14 mass coincidences present in the PGFinder search (highlighted in blue in Table 3-3). We observed that the mass coincidence with the lowest Δ ppm were systematically validated in the Byos® search (Table 3-5, muropeptides 14a-b and 45a-c), confirming our hypothesis from the PGFinder search step. In fact, we did not find any validated monomer with a Δ ppm higher than 4.1 (gm-AEJ), suggesting that a more stringent search, perhaps with a 5-ppm tolerance, could have yielded most of the validated monomers.

Interestingly, only the pentapeptides having D-Alanine (AEJAX) as the fourth residue were validated. Most of these muropeptides (7/8; see muropeptides 16, 22a, 23, 24, 31, 32, 33b and 49b in Table 3-5) have MS/MS data available, unlike other peptide stem composition (AEJXY) which either lack fragmentation data (muropeptides 39, 47 and 48 in Table 3-5) or some *b*- and *y*- ions (muropeptides 14a, 15a or 20 in Table 3-5).

Some clarification is needed for specific muropeptide identifications. For example, validated monomers like gm(-Ac)-AEJDK, gm(-Ac)-AEJWW, which are mass coincidences with crosslinked peptide stems gm(-Ac)-AEJ-JA and gm(-Ac)-AEJ-EJA, respectively (Fig 3-3D, muropeptide 8 and 17 in Table 3-5). Similarly, muropeptides gm(-Ac)-AEJQT, gm-AEJKD and g(-Ac)-AEJWK which are also mass coincidences with gm(-Ac)-AEJ-JA and g(-Ac)-AEJ-gm-A(-H₂O₂) respectively, but were not validated (muropeptide 19, 41 and 43 in Table 3-5).

Given that MS/MS data is a critical information to validate PGFinder matches, we needed to know whether increasing fragmentation frequency (*i.e.* amount of MS/MS data collected) would enable us to validate structures like gm-AEJH (Figure 3-3B), which shows a complete *y*- ion series but only 1*b*- ion. Potentially, by fragmenting more, we could collect enough information to validate the remaining *b*- ions. Other structures like gm(-Ac)-AEJAT (Table 3-4, muropeptide 24) seemed promising, as we know that other Byos[®] identified pentapeptides with low Δ ppm and Alanine as a fourth residue were validated in the Top5 dataset. Therefore, we decided to try another fragmentation setup, fragmenting the 25 most abundant muropeptides, now referred to as Top25.

Table 3-5.	Consolidated	PGFinder and B	yos® results f	or Top5 dataset
			•/	

1401	Structure	Intensity	TheoMw	<u>δerest</u> Annm	ChargeOrder	MS/MS?	Byonic ID	Validated
1		1 15E+09	899 3972	2 9	2.1.3		J	Vandated
1	giii(-AC)-AEJA I	2 05E+02	QQ5 2Q15	17	2,1,5	•	•	
2	gm(-Ac)-AEJG[1	1 71F±08	828 3601	2.7	2,3	•	•	
5	gm(-AC)-AEJ 1	7 525+07	075 1205	1.0	2,5	•	•	
4	gm(-AC)-AEJF 1	6 40E±07	575.4205 656 7752	1.5	2,1	•	•	
5	gm(-AC)-AE 1	0.49E+07	030.2735	2.2	2,1	•	•	
6	gm(-AC)-AEJV 1	4.02ET07	927.4203	2.7	1,2	•	•	
/	gm(-Ac)-AEJY 1	1.93E+07	991.4234 1071 402	1.9	2 2:1:2	~	•	
8	gm(-Ac)-AEJDK 1	1.44E+07	10/1.482	2.0	2;1;3	~	•	
9	gm(-Ac)-AEJI 1	1.21E+U/	941.4441	2.0	1;2	v	•	
10	gm-AEJA 1	1.00E+07	941.4078	2.1	1;2	~	~	
11	gm(-Ac)-AEJM 1	7.99E+06	959.4006	2.0	2	~	~	
12	gm(-Ac)-AEJK 1	7.77E+06	956.455	2.1	2;3	~	~	
13	gm(-Ac)-AEJS 1	6.32E+06	915.3921	1.4	2	~	v	
14a	gm(-Ac)-AEJGV 1	5.87E+06	984.4499	9.8	2;3	~	X	
14b	gm(-Ac)-AEJR 1		984.4612	1.6	2;3	~	v	
15a	gm(-Ac)-AEJGG 1	4.88E+06	942.403	2.7	1;2	~	X	
15b	gm(-Ac)-AEJN 1		942.403	2.7	1;2	✓	✓	
16	gm(-Ac)-AEJAA 1	4.15E+06	970.4344	1.7	2;1	~	~	
17	gm(-Ac)-AEJWW 1	4.10E+06	1200.5187	3.4	3;2	~	✓	
18	gm-AEJG 1	3.48E+06	927.3922	2.3	2;1	~	✓	
19	gm(-Ac)-AEJQT 1	3.06E+06	1057.4663	2.6	2	~	✓	
20	gm(-Ac)-AEJNP 1	2.59E+06	1039.4558	2.0	2	~	X	
21	gm-AEJPE 1	2.37E+06	1096.466	8.1	2	~	X	
22a	gm(-Ac)-AEJAG 1	2 22E+06	956.4187	2.4	2	~	✓	
22b	gm(-Ac)-AEJQ 1	2.221.00	956.4187	2.4	2	~	✓	
23	gm(-Ac)-AEJAF 1	2.16E+06	1046.4656	1.7	2	~	~	
24	gm(-Ac)-AEJAT 1	1.94E+06	1000.4449	2.1	2;1	✓	✓	
25	gm-AEJ 1	1.67E+06	870.3707	4.1	1;2	✓	✓	
26	gm(-Ac)-AEJH 1	1.58E+06	965.419	2.3	2;3	~	✓	
27a	gm(-Ac)-AEJE 1	1 1/IF±06	957.4027	1.9	2	✓	✓	
27b	gm-AEJS 1	1.146.00	957.4027	1.9	2	✓	X	
28	gm(-Ac)-AEJD 1	9.40E+05	943.387	2.2	2	✓	✓	
29	gm-AE 1	8.65E+05	698.2859	1.4	1	✓	✓	
30	gm(-Ac)-AEJHH 1	8.56E+05	1102.4779	2.2	2	~	×	
31	gm(-Ac)-AEJAK 1	8.42E+05	1027.4922	2.9	2	✓	✓	
32	gm(-Ac)-AEJAI 1	8.24E+05	1012.4813	1.5	2	✓	✓	
33 a	gm(-Ac)-AEJFS 1	8 08F±05	1062.4605	0.5	2	✓	X	
33b	gm(-Ac)-AEJAY 1	0.002103	1062.4606	0.6	2	~	✓	
34	gm(-Ac)-AEJW 1	6.75E+05	1014.4394	2.3	1;2	✓	✓	
35	gm(-Ac)-AEJP 1	5.31E+05	925.4128	2.1	2;1	~	~	
36	gm-AEJF 1	4.45E+05	1017.4391	2.4	2;1	~	~	
37	gm(-Ac)-AEJGV 1	4.18E+05	984.4499	2.5	2;3	~	×	
38	gm-AEJV 1	3.69E+05	969.4391	1.5	1	×	×	No MS/MS
39	gm-AEJRM 1	3.42E+05	1157.5123	3.7	1	×	×	No MS/MS
40	gm(-Ac)-AEJT 1	3.30E+05	929.4078	2.2	2	~	✓	
		Validated (m	ore than 50% of the second s	of b and y	ions)			
		NOT Validate	d (Less than 5	0% of b ar	nd y ions)			
	×	Data availab	le					
	^	NO Data avai	lable					

	Structure	Intensity	TheoMw	∆ppm	ChargeOrder	MS/MS?	Byonic ID	Validated
41	gm-AEJKD 1	2.91E+05	1113.4926	1.6	2	~	~	
42	gm-AEJHH 1	2.61E+05	1144.4885	3.7	2	~	×	
43	gm(-Ac)-AEJWK 1	2.53E+05	1142.5343	5.4	2;3	✓	~	
44	gm-AEJH 1	2.52E+05	1007.4296	1.9	2	✓	✓	
45a	gm(-Ac)-AEJPT 1	2.48E+05	1026.4605	8.3	2	✓	×	
45b	gm-AEJGV 1		1026.4606	8.2	2	✓	×	
45c	gm-AEJR 1		1026.4718	2.6	2	✓	✓	
46	gm-AEJWW 1	2.19E+05	1242.5293	2.6	3	✓	X	
47	gm-AEJRC 1	2.18E+05	1129.481	3.4	3;2	×	×	No MS/MS
48	gm(-Ac)-AEJCA 1	1.95E+05	1002.4064	5.1	2	×	×	No MS/MS
49 a	gm(-Ac)-AEJSP 1	1.95E+05	1012.4449	2.2	2;1	X	×	No MS/MS
49b	gm-AEJAA 1		1012.445	2.3	2;1	X	×	No MS/MS
50a	gm-AEJFP 1	1.92E+05	1114.4919	6.3	2	X	×	No MS/MS
50b	gm-AEJMI 1		1114.4952	9.2	2	X	×	No MS/MS
51	gm-AEJT 1	1.42E+05	971.4184	1.7	2	✓	×	
52	gm(-Ac)-AEJPQ 1	1.41E+05	1053.4714	2.0	2	×	×	No MS/MS
53	gm-AEJI 1	1.39E+05	983.4547	2.1	2	✓	~	
54	gm(-Ac)-AEJFC 1	1.31E+05	1078.4377	7.2	2	×	×	No MS/MS
55	gm(-Ac)-AEJC 1	1.14E+05	931.3693	1.7	2	×	×	No MS/MS
56a	gm(-Ac)-AEJIG 1	1.08E+05	998.4656	2.7	2	X	×	No MS/MS
56b	gm(-Ac)-AEJVA 1		998.4656	2.7	2	×	×	No MS/MS
56c	gm-AEJK 1		998.4656	2.7	2	×	×	No MS/MS
57	gm(-Ac)-AEJVS 1	1.01E+05	1014.4605	1.0	2	×	×	No MS/MS
58	gm-AEJTT 1	9.96E+04	1072.466	2.0	2	✓	×	
59	gm(-Ac)-AEJPA 1	9.96E+04	996.45	1.7	2	×	×	No MS/MS
60	gm-AEJGR 1	9.17E+04	1083.4933	5.4	2	X	×	No MS/MS
61a	gm(-Ac)-AEJNT 1	8.87E+04	1043.4507	2.4	2	×	×	No MS/MS
61b	gm(-Ac)-AEJSQ 1		1043.4507	2.4	2	X	X	No MS/MS
62	gm(-Ac)-AEJKG 1	8.24E+04	1013.4765	2.7	2	×	×	No MS/MS
63	gm(-Ac)-AEJWA 1	8.19E+04	1085.4765	1.3	2	×	×	No MS/MS
64	gm(-Ac)-AEJHA 1	7.78E+04	1036.4561	2.3	2	~	×	
65	gm-AEJMW 1	7.72E+04	1187.4905	0.9	2	X	×	No MS/MS
66	gm(-Ac)-AEJDP 1	7.31E+04	1040.4398	1.3	2	v	×	
67	gm(-Ac)-AEJGP 1	7.14E+04	982.4343	2.0	2	×	×	No MS/MS
68	gm(-Ac)-AEJTC 1	6.72E+04	1032.417	1.7	2	X	X	No MS/MS
69a	gm(-Ac)-AEJKQ 1	6.51E+04	1084.5136	7.8	2	X	X	No MS/MS
69b	gm-AEJIT 1		1084.5024	2.4	2	X	X	No MS/MS
70a	gm(-Ac)-AEJSA 1	6.34E+04	986.4293	2.3	2	X	X	No MS/MS
70b	gm(-Ac)-AEJTG 1		986.4292	2.2	2	X	X	No MS/MS
71	gm-AEJPW 1	5.24E+04	1153.5028	5.9	2	X	X	No MS/MS
72a	gm(-Ac)-AEJKE 1	4.61E+04	1085.4976	2.0	2	X	X	No MS/MS
72b	gm-AEJSK 1		1085.49/7	2.1	2	X	X	NO IVIS/IVIS
73a	gm(-Ac)-AEJGQ 1	4.33E+04	1013.4401	2.2	2	X	X	
73b	gm(-Ac)-AEJNA 1	4.005.04	1013.4402	2.3	2	X	X	
74	gm-AEJFF 1	4.09E+04	1164.50/5	5.8	2	X	X	
75	gm-AEJPT 1	3.83E+04	1068.4/11	9.1	2	X	X	
76	gm-AEJP 1	S./IE+U4	907.4235	0.0		~	~	110 1015/1015
76	gm-AEJP 1	3.71E+04 Validated (m	967.4235	8.6 of <i>b</i> and <i>y</i>	2 ions)	X	X	No MS/MS

NOT Validated (NO b or y ions) NOT Validated (Less than 50% of b and y ions) Data available No Data available

××

3.2.4 MS search, identification of PG monomers using PGFinder and a Top25 fragmentation setup

The same sample of *C. difficile* muropeptides was analyzed using LC-MS/MS, this time fragmenting the 25 most abundant ions (Top25). An unbiased PGFinder (v.0.1.1a) search was then performed using the same parameters as those employed for the Top5 analysis: the same database (DB_Cd1, Table 3-1), 10ppm mass tolerance, and consolidation of in-source decay fragmentation within a 30-second window. Multiple matches corresponding to the same muropeptide were consolidated as previously.

A summary of the Top25 PGFinder search results compared to the Top5 dataset is shown in Figure 3-4. We identified fewer monomers in Top25 (72) than in Top5 (92) and observed only nine mass coincidences in Top25 (Figure 3-4B and Table 3-7 highlighted in blue) compared to 14 in Top5 (Table 3-4B). The overall ratio of deacetylated/acetylated muropeptides remained unchanged between Top25 and Top5, approximately 67% and 33% respectively, in both cases. Additionally, no significant change in the average Δ ppm was observed; in fact, it was slightly higher in Top25 (2.7 vs 2.5) (Figure 3-4B). Notably, a significant reduction in the amount of deacetylated pentapeptides was observed in Top25 (27) (Figure 3-4A) compared to Top5 (40) (Table 3-7). The intensity collected between both datasets was similar, with Top25 (Figure 3-4A) representing approximately 94% of the abundance collected with Top5 (Table 3-7).

All the monomers identified in Top25 were already present in Top5, including the mass coincidences (Table 3-7). Interestingly, no significant change in the average Δ ppm was observed between the two datasets, with a slight increase in Top25, as previously mentioned.





						I	оръ
Structure	Intensity	Abundance RT	TheoMw	Obs	Δppm	Present	∆ppm
gm(-Ac)-AEIA 1	1.09E+09	59.76% 7.84	899.3972	899.3940	3.6	~	2.9
2 gm(-Ac)-AEIG[1	3 02E+08	16.63% 8.2	885 3815	885 3800	17		17
	1.425.00	7.00% 2.2	828,3615	000.0000	1.7		2.7
3 gm(-AC)-AEJ I	1.43E+08	7.88% 3.30	828.3601	828.3580	2.5	×	2.2
4 gm(-Ac)-AEJF 1	7.39E+07	4.06% 22.2	2 975.4285	975.4267	1.9	×	1.9
5 gm(-Ac)-AE 1	5.57E+07	3.06% 7.7	656.2753	656.2732	3.2	×	2.2
6 gm(-Ac)-AEJV 1	3.97E+07	2.18% 13.8	3 927.4285	927.4259	2.8	V	2.7
7 gm(-Ac)-AEIV[1	1 84F+07	1 01% 16 4	9 991 4234	991 4215	19		19
	1 205 07	0.70% 8.0	1071 492	1071 4902	1.5		2.0
8 gm(-AC)-AEJKD[1	1.285+07	0.70% 8.9	10/1.482	1071.4803	1.0	×	2.0
9 gm(-Ac)-AEJI 1	1.18E+07	0.65% 18.1	9 941.4441	941.4421	2.2	×	2.0
10 gm-AEJA 1	9.11E+06	0.50% 8.74	941.4078	941.4052	2.8	×	2.1
11 gm(-Ac)-AEJM 1	7.25E+06	0.40% 13.7	4 959.4006	959.3980	2.7	v	2.0
12 gm(-Ac)-AEIGV 1.gm(-Ac)-AEIR 1	5.95E+06	0.33% 5.7	984,4499,984,4612	984,4587	8.9.2.5	<u> </u>	9.8.1.6
13 gm(-Ac)-AEIK[1	5 2//E+06	0.29% ///	956.455	956 4525	2.6		21
	3.240100	0.25% 7.4	045 2024	015 0000	2.0		2.1
14 gm(-Ac)-AEJS 1	4.59E+06	0.25% 7.2	915.3921	915.3896	2.8	×	1.4
15 gm(-Ac)-AEJGG 1,gm(-Ac)-AEJN 1	4.39E+06	0.24% 4.93	942.403	942.4010	2.1	×	2.7
16 gm(-Ac)-AEJWW 1	3.59E+06	0.20% 9.35	5 1200.5187	1200.5219	2.7	×	3.4
17 gm(-Ac)-AEJAA 1	3.20E+06	0.18% 9.59	970.4344	970.4318	2.7	v	1.7
18 gm-AEIG 1	2.80E+06	0.15% 9.46	927.3922	927.3899	2.5	<u> </u>	2.3
19 gm(-Ac)-AFIOT 1	2 705+04	0.15% 6.4	1057 4662	1057 4640	2.2		2.6
	2.735100	0.13% 0.4		1030 4530	2.2	, in the second s	2.0
ZU gm(-AC)-AEJNP 1	2.52E+06	0.14% 10.3	/ 1039.4558	1039.4536	2.1	~	2.0
21 gm-AEJPE 1	2.36E+06	0.13% 8.0	5 1096.466	1096.4745	7.7	×	8.1
22 gm(-Ac)-AEJFA 1	2.15E+06	0.12% 24.4	3 1046.4656	1046.4635	2.0	×	1.7
23 gm(-Ac)-AEJGA 1,gm(-Ac)-AEJQ 1	1.89E+06	0.10% 7.68	956.4187	956.4157	3.1	v	2.4
24 gm(-Ac)-AFITA 1	1 76F+06	0.10% 6.53	1000 4449	1000 4430	19	<u> </u>	21
25 gm_AEU1	1 605+06	0.00% 6.0	870 2707	970 2677	2.5		1 1
	1.091+00	0.09% 0.0	870.3707	870.3077	3.5	Ť	4.1
26 gm(-Ac)-AEJH 1	1.48E+06	0.08% 3.0	965.419	965.4169	2.2	×	2.3
27 gm(-Ac)-AEJE 1,gm-AEJS 1	8.63E+05	0.05% 8.08	957.4027	957.4000	2.8	×	1.9
28 gm(-Ac)-AEJFS 1,gm(-Ac)-AEJYA 1	8.29E+05	0.05% 17.1	0 1062.4605,1062.4606	1062.4590	1.4,1.5	×	0.5,0.6
29 gm(-Ac)-AEJD 1	8.05E+05	0.04% 6.1	943.387	943.3844	2.7	V	2.2
30 gm(-Ac)-AEIKA 1	7.98E+05	0.04% 7.58	3 1027.4922	1027.4895	2.7	<u> </u>	2.9
31 gm(-Ac)-AEIHH 1	7 065+05	0.04% 8.4	1102 4779	1102 4747	2.0		2.2
	7.905+05	0.04% 8.4	CO2 2050	1102.4747	2.9	.	2.2
32 gm-AE 1	7.8/E+05	0.04% 8.5	698.2859	698.2845	2.0	×	1.4
33 gm(-Ac)-AEJIA 1	7.36E+05	0.04% 20.3	1 1012.4813	1012.4792	2.1	×	1.5
34 gm(-Ac)-AEJW 1	6.72E+05	0.04% 25.0	4 1014.4394	1014.4380	1.4	v	2.3
35 gm(-Ac)-AEJP 1	4.62E+05	0.03% 6.64	925.4128	925.4115	1.4	V	2.1
36 gm-AFIFI1	4 25F+05	0.02% 21.8	0 1017 4391	1017 4370	2.0	<u> </u>	24
27 gm_AEI//1	2 446±05	0.02% 14.7	2 969 4291	060 4276	1.6		15
	3.44L+03	0.02/8 14.7	000.4070	909.4370	1.0	Ť	1.5
38 gm(-Ac)-AEJT[1	3.07E+05	0.02% 6.44	929.4078	929.4056	2.4	×	2.2
39 gm(-Ac)-AEJPT 1,gm-AEJGV 1,gm-AEJR 1	2.84E+05	0.02% 7.26	5 1026.4605,1026.4606,1026.4718	1026.4681	7.4,7.3,3.5	×	8.3,8.2,2.6
40 gm-AEJKD 1	2.41E+05	0.01% 8.44	1113.4926	1113.4903	2.0	×	1.6
41 gm-AEJRC 1	2.32E+05	0.01% 8.09	1129.481	1129.4845	3.1	×	3.4
42 gm-AEJHH 1	2.25E+05	0.01% 10.8	0 1144.4885	1144.4841	3.8	V	3.7
43 gm(-Ac)-AEJKW 1	2.17F+05	0.01% 7.2	1142 5343	1142 5286	5.0	~	54
14 gm_AEIH 1	2 1/5:05	0.010/ 6.01	1007 4206	1007 4274	2.0		1.0
	2.14ETUD	0.01% 0.2		1114 4040	2.2	, in the second s	1.9
45 gm-AEJFP 1,gm-AEJIVII 1	1.68E+05	0.01% 7.42	1114.4919,1114.4952	1114.4848	6.3,9.2	~	0.3,9.2
46 gm(-Ac)-AEJSP 1,gm-AEJAA 1	1.67E+05	0.01% 10.1	1 1012.4449,1012.445	1012.4426	2.2,2.3	×	2.2,2.3
47 gm-AEJWW 1	1.52E+05	0.01% 10.2	3 1242.5293	1242.5321	2.2	×	2.6
48 gm-AEJRM 1	1.44E+05	0.01% 11.7	4 1157.5123	1157.5176	4.5	¥	3.7
49 gm-AEJT 1	1.36E+05	0.007% 9.5	971.4184	971.4161	2.4	¥	1.7
50 gm(-Ac)-AFIC 1	1 06F+05	0.006% 8.9	931 3693	931 3673	22		17
E1 $gm(Ac)$ AEIIC[1 $gm(Ac)$ AEII/A[1 gm AEII/A]	1.010.00			000 4620	2.2		2.7
ST BUIL-ACI-ACIG 1, BUIL-ACI-ACIACIACIACIK 1	1.01E+05	0.000% 5.79	998.4000	998.4629	2.0		2.7
52 gm(-AC)-AEJVS 1	9.20E+04	0.005% 8.45	1014.4605	1014.4583	2.2	×	1.0
53 gm-AEJTT 1	8.63E+04	0.005% 10.2	0 1072.466	1072.4646	1.3	×	2.0
54 gm(-Ac)-AEJFC 1	8.59E+04	0.005% 4.60	1078.4377	1078.4295	7.6	×	7.2
55 gm(-Ac)-AEJTC 1	8.36E+04	0.005% 5.10	1032.417	1032.4143	2.6	¥	1.7
56 gm-AEII11	8.30F+04	0.005% 19 1	2 983 4547	983,4533	1.4	~	2.1
57 gm(-Ac)-AFIPA 1	7 66F±04	0.00/1% 6.2	L 996 45	996 117F	2		17
	7.001+04	0.004/0 0.54		1005 4727	2.5	, The second sec	1.7
58 gm(-AC)-AEJWA 1	7.49E+04	0.004% 23.9	0 1085.4765	1085.4/37	2.6	×	1.3
59 gm(-Ac)-AEJGP 1	6.96E+04	0.004% 8.1	982.4343	982.4316	2.8	~	2.0
60 gm(-Ac)-AEJKG 1	5.34E+04	0.003% 5.83	3 1013.4765	1013.4734	3.0	×	2.7
61 gm-AEJPW 1	5.33E+04	0.003% 7.02	1153.5028	1153.4959	5.9	¥	5.9
Total	1.82E+09					1.9E+09	

Table 3- 6. All Top25 PGFinder search are present in Top5. 9 mass coincidences highlighted in blue

3.2.5 MS/MS search using Byos® and a Top25 fragmentation setup

We next investigated the number of muropeptide identifications based on MS/MS data using the Byos® software. The Top25 dataset was analysed using a database called DB_Cd10 (Fig 3-1B), and the same conditions as the Top5 Byos® search.

We noticed three more identifications in Top25 compared to Top5. Interestingly, contrary to what this could suggest, we validated one fewer identification. Unexpectedly, the validations were also different, as we managed to validate an acetylated pentapeptide, but we lost the validation of a deacetylated tripeptide and two deacetylated tetrapeptides in Top25 (Figure 3-5C).

When analyzing each identification in detail, we noticed that most of the identifications present in Top25 were also present in Top5. We also observed six new identifications (Figure 3-4D), but only one of them was validated (gm-AEJAA).

We also see the inverse situation, where three identifications in Top5 are not present inTop25 (gm(-Ac)-AEJQ, gm-AEJI and gm-AEJKD). Interestingly, only one of them was validated in Top5 (gm(-Ac)-AEJQ). Regarding the shared identifications, as we noticed in the summary table (Figure 3-5A), we lost the validation of the deacetylated tripeptide (gm(- Ac)-AEJ) and a deacetylated tetrapeptide (gm(-Ac)-AEJR), but we also gained the validation of gm-AEJH (Figure 3-6B and D).

In general terms, we see confirmation of the results for the most abundant monomers (Figure 3-6A). We see as well that our premise was correct to some extent, as we observed an increase in the amount of *-b* and *-y* ions for gm-AEJR (Figure 3-6C and Figure 3-3C) and gm-AEJH, enabling us to validate gm-AEJH (Figure 3-4B and Figure 3-3B).





3.2.6 Merging PGF inder and Byos® analyses for Top25 and Top5 fragmentation setups

As we did with Top5, we merged the Top25 results from both PGFinder and Byos® analysis. Table 3-7 shows the merged results. The most remarkable result is a significant decrease in the number of ions lacking MS/MS data, with 13 missing in Top25 compared to 36 in Top5. Regarding the MS/MS data collected, we noticed that the validated monomer with the lowest intensity for Top5 was gm(-Ac)-AEJT (3.30E5, Table 3-5 muropeptide 40), while for the Top25 it was gm-AEJAA (1.67E5, Table 3-7 muropeptide 50a), corresponding to almost half the intensity of Top5.

In summary, fragmenting the 25 most abundant ions yielded more MS/MS data, but not a clear improvement in the quality, as the number of validations remains almost unchanged at 31 and 32 for Top25 and Top5, respectively. We lost the validation of monomers with high intensity, such as gm(-Ac)-AEJR (Table 3-9 muropeptide 14b) but gained validations for low-intensity monomers like gm-AEJAA (Table 3-9 muropeptide 50a).



C	Expected	Obs	erved
Sequence	ions	b	у
gm(-Ac)-AE 1	1	1	1
gm(-Ac)-AEJAA 1	4	4	4
gm(-Ac)-AEJ <mark>AF</mark> 1	4	4	4
gm(-Ac)-AEJG 1	3	3	3
gm(-Ac)-AEJAG 1	4	4	4
gm(-Ac)-AEJE 1	3	3	3
gm(-Ac)-AEJK 1	3	3	3
gm(-Ac)-AEJM 1	3	3	3
gm(-Ac)-AEJN 1	3	3	3
gm(-Ac)-AEIW 1	3	3	3
gm-AFI1	1	1	1
gm-AFI 1	2	2	2
gm_AFIA 1	3	3	2
gm_AEIE 1	2	2	2
gm-AEIG 1	2	2	2
gm(Ac) AEIA 1	2	2	2
gm(Ac) AEU1	2	2 2	1
gm(Ac) AEIDI1	2	2	2
giii(-Ac)-AEJD[1	2	2	2
gm(-Ac)-AEJT 1	3	3	2
gm(-Ac)-AEJI 1	3	2	3
gm(-Ac)-AEJV 1	3	3	2
gm(-Ac)-AEJY 1	3	3	2
gm(-Ac)-AEJF 1	3	3	3
gm(-Ac)-AEJAI 1	4	3	4
gm(-Ac)-AEJ <mark>AK</mark> 1	4	3	4
gm(-Ac)-AEJH 1	3	2	3
gm(-Ac)-AEJDK* 1	3	4	3
gm(-Ac)-AEJR 1	3	1	3
gm(-Ac)-AEJS 1	3	2	2
gm(-Ac)-AEJ <mark>AY</mark> 1	3	3	4
gm-AEJH 1	3	2	3
gm(-Ac)-AEJQ 1	3		
gm(-Ac)-AEJGV 1	4	2	4
gm(-Ac)-AEJ <mark>AT</mark> 1	4	3	2
gm(-Ac)-AEJVG 1	4	2	3
gm(-Ac)-AEJWW** 1	4	4	3
gm-AEJAA 1	4	3	3
gm(-Ac)-AEJWK 1	4	1	3
gm(-Ac)-AEJP 1	3	2	1
gm(-Ac)-AEJQT 1	4	1	3
gm(-Ac)-AEJCT 1	4	1	3
gm(-Ac)-AEJ <mark>AV</mark> 1	4	2	2
gm-AEJI 1	3		
gm-AEJK 1	3	1	1
gm-AEJKD 1	4		
am-AEIRI1	з	1	2

Figure 3- 6. Top25 Byos® search analysis results.

Annotated chromatograms generated by Byos[®] for gm(-Ac)-AEJAA **(A)**, gm-AEJH **(B)** and gm-AEJR **(C)**. The number of *b*- and *y*- ions was collected **(D)** and highlighted in green for series with more than half of the expected ions and in red the ones that did not meet the criteria. The validated monomers from Top5 Byos[®] search is shown in the last column.

3.2.7 Top 5 and Top 25 QualBrowser analysis

To understand the differences between Top5 and Top25 fragmentation setups, we performed a comparison of the total ion chromatograms (TICs) from both strategies. This revealed a slight time shift in the retention time (RT) of 0.032 ± 0.024 minutes for Top25 compared to Top5. Aside from this minor discrepancy, no other significant difference was observed when we visually examined the chromatograms in Xcalibur 3.0.63 Qual Browser (Thermo fisher Inc, 2013) (Fig 3-7A and 3-7B).

However, upon zooming-in on a two-minute timeframe near the peak 2 at RT=20.7 min, a significant difference in the amount of information collected between Top5 (Fig 3-7A, red square) and Top25 (Fig 3-7B, blue square) is evident. This difference suggests that the two fragmentation strategies, despite yielding seemingly identical TICs, capture different amount of information.

Table 3-7. Consolidated PGFinder and Byos® results for Top25 dataset and Top5 Byos® validated monomers

								Top25	Top5
	Structure	Intensity	TheoMw	Δppm	ChargeOrder	MS/MS?	Byonic ID	Validated	Validated
1	gm(-Ac)-AEJA 1	1.09E+09	899.3972	3.6	2;1				
2	gm(-AC)-AEJG 1 gm(-AC)-AEJJ1	3.02E+08 1.43F±09	303.3615	1./ 25	2.7		ž		
4	gm(-Ac)-AEJF1	7 39F+07	975 4285	1.9	3,2 2·1	Ĵ	Ĵ		
5	gm(-Ac)-AFI1	5.57E+07	656.2753	3.2	1	-			
6	gm(-Ac)-AEJV 1	3.97E+07	927.4285	2.8	2;1	-	2		
7	gm(-Ac)-AEJY 1	1.84E+07	991.4234	1.9	2	~	~		
8	gm(-Ac)-AEJDK 1	1.28E+07	1071.482	1.6	2	~	×		
9	gm(-Ac)-AEJI 1	1.18E+07	941.4441	2.2	1;2	~	~		
10	gm-AEJA 1	9.11E+06	941.4078	2.8	1;2	~	~		
11	gm(-Ac)-AEJM 1	7.25E+06	959.4006	2.7	2;1	~	~		
12	gm(-Ac)-AEJK 1	5.24E+06	956.455	2.6	3;2;1	~	~		
13	gm(-Ac)-AEJS 1	4.59E+06	915.3921	2.8	2		č		
14a 14b	gm(-Ac)-AEJGV 1,	5.95E+06	984.4499	2.5	2;3;1	ž	Ĵ		
15a	gm(-Ac)-AEIGG[1		942.403	2.5	2,3,1	Ĵ	×		
15b	gm(-Ac)-AEJN 1	4.39E+06	942.403	2.1	2;1	-	Ĵ		
16	gm(-Ac)-AEJAA 1	3.20E+06	970.4344	2.7	2;1	~	~		
17	gm(-Ac)-AEJWW 1	3.59E+06	1200.519	2.7	2;3	~	×		
18	gm-AEJG 1	2.80E+06	927.3922	2.5	2;1	~	~		
19	gm(-Ac)-AEJQT 1	2.79E+06	1057.466	2.2	2	~	~		
20	gm(-Ac)-AEJNP 1	2.52E+06	1039.456	2.1	2	~	×		
21	gm-AEJPE 1	2.36E+06	1096.466	7.7	2	~	×		
22a	gm(-Ac)-AEJAG 1	1.89E+06	956.4187	3.1	2	~	ž		
220	gm(-Ac)-AEJQ[1	3.155.00	956.4187	3.1	2	•	×		
23 24	gm(-AC)-AEJAF 1 gm(-AC)-AEJAF 1	2.13E+Ub	1000 445	2.U 1 0	2;1 2·1		ž		
24 25	em-AEJ 1	1.69F+06	870.3707	3.5	2,± 1:7	Ĵ			
26	gm(-Ac)-AEJHI1	1.48E+06	965.419	2.2	2	-	-		
27a	gm(-Ac)-AEJE 1	0.000	957.4027	2.8	2	~	~		
27b	gm-AEJS 1	8.63E+05	957.4027	2.8	2	~	×		
28	gm(-Ac)-AEJD 1	8.05E+05	943.387	2.7	2	~	~		
29	gm-AE 1	7.87E+05	698.2859	2.0	1	~	~		
30	gm(-Ac)-AEJHH 1	7.96E+05	1102.478	2.9	2	~	×		
31	gm(-Ac)-AEJAK 1	7.98E+05	1027.492	2.7	2	~	~		
32	gm(-Ac)-AEJAI 1	7.36E+05	1012.481	2.1	2	~	~		
33a	gm(-Ac)-AEJFS 1	8.29E+05	1062.461	1.4	2	~	×		
33b	gm(-Ac)-AEJAY 1	6 725 . 05	1062.461	1.5	2	~	~		
34	gm(-Ac)-AEJW 1	6.72E+05	1014.439	1.4	2				
35	gm(-ACJ-AEJP 1	4.02E+05	925.4128	1.4	2;1	,			
30	gm(-Ac)-AFIVG[1	4.23L+03 N/Δ	984 4499	2.0 N/A	2,1 N/Δ	Ĵ	Ĵ		
38	gm-AFIV 1	3.44F+05	969.4391	1.6	2:1	Ĵ	×		No MS/MS
39	gm-AEJRM 1	1.44E+05	1157.512	4.5	2	-	×		No MS/MS
40	gm(-Ac)-AEJT 1	3.07E+05	929.4078	2.4	2	~	~		
41	gm-AEJKD 1	2.41E+05	1113.493	2.0	2	~	×		
42	gm-AEJHH 1	2.25E+05	1144.489	3.8	2	~	×		
43	gm(-Ac)-AEJWK 1	2.17E+05	1142.534	5.0	3;2	~	~		
44	gm-AEJH 1	2.14E+05	1007.43	2.2	2	~	~		
45a	gm(-Ac)-AEJPT 1		1026.461	7.4	2	~	×		
45b	gm-AEJGV 1	2.84E+05	1026.461	7.3	2	~	×		
45c	gm-AEJR 1	1.525.05	1026.472	3.5	2	~	ž		
40	gm-AEIRC 1	1.52E+05 2.32E±05	11242.529	2.2	2	ž	÷		No MS/MS
48	gm(-Ac)-AFICA 1	N/A	1002.406	N/A	2,5 N/A	Ŷ	ŝ		No MS/MS
49a	gm(-Ac)-AEJSP 1		1012.445	2.2	2:1	~	x		No MS/MS
49b	gm-AEJAA 1	1.67E+05	1012.445	2.3	2;1	~	~		No MS/MS
50a	gm-AEJFP 1	1 695-05	1114.492	6.3	2	~	×		No MS/MS
50b	gm-AEJMI 1	1.086+05	1114.495	9.2	2	~	×		No MS/MS
51	gm-AEJT 1	1.36E+05	971.4184	2.4	2	~	×		
52	gm(-Ac)-AEJPQ 1	N/A	1053.471	N/A	N/A	×	×	No MS/MS	No MS/MS
53	gm-AEJI 1	8.30E+04	983.4547	1.4	2	~	×		N
54	gm(-Ac)-AEJFC 1	8.59E+04	10/8.438	7.6	2	~	×		No MS/MS
55	gin(-Ac)-AEJC 1	1.U6E+05	931.3693	2.2	2	•	×		NO MS/MS
200 56h	gm(-AC)-AEJIG[1 gm(-AC)-AEJAV[1	1.015+05	998 /656	2.0	2		Ŷ		No MS/MS
56c	gm-AEJK 1	1.012703	998.4656	2.6	2	2	0		No MS/MS
57	gm(-Ac)-AEJVS 1	9.20E+04	1014.461	2.2	2	×	×	No MS/MS	No MS/MS
58	gm-AEJTT 1	8.63E+04	1072.466	1.3	2	~	×		
59	gm(-Ac)-AEJPA 1	7.66E+04	996.45	2.5	2	~	×		No MS/MS
60	gm-AEJGR 1	N/A	1083.493	N/A	N/A	~	×		No MS/MS
61a	gm(-Ac)-AEJNT 1	N/A	1043.451	N/A	N/A	~	×		No MS/MS
61b	gm(-Ac)-AEJSQ 1		1043.451	N/A	N/A	~	×		No MS/MS
62	grn(-Ac)-AEJKG 1	5.34E+04	1095 477	3.0	2	~	×		NO MS/MS
03 64	gm(-AC)-AEJWA 1	7.49E+04 N/A	1036.456	2.0 N /A	2 N/A		Ŷ		INO INIS/INIS
65	gm-AEJMW11	N/A	1187.491	N/A	N/A	×	x	No MS/MS	No MS/MS
66	gm(-Ac)-AEJDP 1	N/A	1040.44	N/A	N/A	~	×	,	,
67	gm(-Ac)-AEJGP 1	6.96E+04	982.4343	2.8	2	×	×	No MS/MS	No MS/MS
68	gm(-Ac)-AEJCT 1	8.36E+04	1032.417	2.6	3;2	~	~		No MS/MS
69a	gm(-Ac)-AEJKQ 1	N/A	1084.514	N/A	N/A	×	×	No MS/MS	No MS/MS
69b	gm-AEJIT 1		1084.502	N/A	N/A	×	×	No MS/MS	No MS/MS
70a	gm(-Ac)-AEJSA 1	N/A	986.4293	N/A	N/A	×	×	No MS/MS	No MS/MS
70b	gm(-Ac)-AEJTG 1	F 335 -	986.4292	N/A	N/A	×	×	No MS/MS	No MS/MS
71	gm-AEJPW 1	5.33E+04	1153.503	5.9	2	ž	×	No Mc / C	No MS/MS
/2a	gm(-AC)-AEJKE 1	N/A	1085.498	N/A	N/A	×	×	NO MS/MS	NO MS/MS
730	gm(-Ac)-AFIGO 1		1013 44	N/A	N/A	Ŷ	Ŷ	No MS/MS	No MS/MS
73b	gm(-Ac)-AEJNA 1	N/A	1013.44	N/A	N/A	×	x	No MS/MS	No MS/MS
74	gm-AEJFF 1	N/A	1164.508	N/A	N/A	~	×	,	No MS/MS
75	gm-AEJPT 1	N/A	1068.471	N/A	N/A	~	×		No MS/MS
76	gm-AEJP 1	N/A	967.4235	N/A	N/A	×	×	No MS/MS	No MS/MS
		Validated (m	ore than 50% o	of b and y ic	ins)				
		NOT Validate	d (NO b or y io	ns)					
		NOT Validate	d (Less than 51	0% of b and	y ions)				
	•	Data availab	le Iable						
	*	. o Data avai	- JIC						



Figure 3- 7. Top5 and Top25 fragmentation setup compared via Total Ion Chromatograms (TIC).

TIC spectra corresponding to Top5 (A) and Top25 (B) displayed in Xcalibur Qual Browser. **1** High abundance peak corresponding to gm(-Ac)-AEJM **2** Low abundance peak corresponding to gm(-Ac)-AEJF.

To further investigate the relationship between visual differences in collection points and data quantity, we compared both a high-abundance (RT ~13.7,) peak and a low abundance peak (RT ~20.7) present in the TICs (Figure 3-5, ① and ②, respectively) for both Top5 and Top25 datasets. These peaks, representing two monomers validated in both fragmentation setups (gm(-Ac)-AEJM and gm(-Ac)-AEJF, respectively), were selected due to their extreme positions within the captured range, making them suitable indicators of data quantity, range, and sensitivity.

QualBrowser quantitative analysis (Table 3-10) revealed that the Top5 TIC contained, on average, 59% more datapoints than the Top25 TIC (*e.g.* 93698/156309 for 2). However, only half of this data points corresponded to precursor ions (actual m/z intensity) in both setups. While the number of precursor ions was slightly higher for Top5, theTop5 TIC consistently collected an average of 2.9% more intensity than theTop25 TIC. This corresponds to an increase of between 2.1% and 4.1% more data points for Top5 compared to the Top25 TIC.

Table 3-8	Qual Browser	MS1 statistics to	highest and low	abundant peaks
	•		d e e e e e e e e e e	

	То	p 5	То	Top 25		
	1	2	1	2		
Scans #	5179-5343	8319-8482	8542-8815	13705-13970		
RT (min)	13.53-13.93	20.45-20.85	13.55-13.95	20.49-20.89		
Datapoints	112310	156309	74042	93698		
Datapoints (Intensity >0)	60625	106995	38450	60288		
	54.0%	68.5%	51.9%	64.3%		
Intensity	5.73E+09	3.30E+08	5.45E+09	3.28E+08		

Given these results, we conclude that the most suitable fragmentation setup for our upcoming experiments should be Top5. This is because Top5 consistently collects

more information in terms of MS intensity (Figure 3-5 and Table 3-8) monomers (Table 3-8), MS/MS quality (Table 3-7) and Byos®-validated monomers (Figure 3-5).

3.2.8 Amidation products

The Byos® results represent nearly 99% of the total abundance in the PGFinder searches. We decided to merge all the validated monomers from Top5 and Top 25 searches and pick the monomers representing the top 95% of the total intensity to generate DB_Cd2 (Table 3-9). The top 95% of abundance aligns with the expectation that modified versions are typically less than one order of magnitude less abundant than the unmodified versions, based on previous results (Patel et al., 2021).. This means that if gm(-Ac)-AEJA represents 59.76% of the total abundance (Table 3-6), the modifications will not exceed 5%.

Structure	Monoisotopicmass
gm(-Ac)-AEJA 1	899.3972
gm(-Ac)-AEJG 1	885.3815
gm(-Ac)-AEJ 1	828.3601
gm(-Ac)-AEJF 1	975.4285
gm(-Ac)-AE 1	656.2753
gm(-Ac)-AEJV 1	927.4285
gm(-Ac)-AEJY 1	991.4234
gm(-Ac)-AEJKD 1	1071.482

Table 3- 9. DB_Cd2: 95% of the total monomers' intensity

A common modification in gram-positive organisms is the amidation of carboxyl groups of aminoacids in the peptide stems of muropeptides (Dajkovic et al., 2017). We searched for amidations using Top5 and Top25 deconvoluted datasets, with DB_Cd2 (Table 3-9) as database and PGFinder configured to look for "Amidation" (Figure 3-2 Step 2). After consolidation of both searches, we found 14 matches, 8 corresponding to the database and 6 corresponding to amidation products (Fig 3-8A).

When comparing the original muropeptide to its modified version (Fig 3-8B), we noticed some interesting features. The amidated products have a slightly lower retention time (RT) than the original version, by less than 1 minute. As a rule, the most abundant modification should correspond to the most abundant muropeptide; In this case, we observed this pattern. Assuming all the modifications could be real, we needed to confirm them by MS/MS fragmentation.

The difference in the monoisotopic masses between the original muropeptide and its modification is less than 1 Dalton (0.984 Da) (Fig 3-8C). We calculated the [M+1] and [M+2] ions for both the original (gm(-Ac)-AEJA) and the amidated muropeptide (gm(-Ac)-AEJA (Amidated)). Then, we looked in XcaliburTM QualBrowser for a known fragmented ion with a Δ ppm of less than 20. We realised that both [M+H]⁺ and [M+2H]²⁺ ions had a Δ ppm lower than 20, so it was possible to distinguish the amidated muropeptide ion from the original muropeptide ion.

Once we found a fragmented precursor ion for validation, we looked for signature product ions that would confirm the presence of amidation. It has been reported that amidation occurs in the ε -carboxyl of *meso*DAP and in the α -carboxyl of D-Glu (only when L-Lys is present in the third position of the peptide stem) (Dajkovic et al., 2017). We looked for a 172.1086Da product ion corresponding to the monoisotopic mass of an amidated *meso*-DAP within Δ ppm =10 (Fig 3-8D), and for amidated D-Glu (129.0664Da).

Unexpectedly, some spectra showed two fragments corresponding to amidated and non-amidated muropeptides (Fig 3-8D). We considered a muropeptide as validly amidated if the amidated *meso*-DAP had higher intensity. Three amidated muropeptides were validated. The muropeptide gm(-Ac)-AE(Amidated) was disregarded because the signature amidated *meso*-DAP was smaller than expected, as no amidated D-Glu was present (Fig3-8E). The strategy for validated amidation is summarized in Figure 3-9.



ł	Structure	Intensity	Abundance	RT ± SD
	gm(-Ac)-AEJA 1	1.12E+09	62.6%	7.83 ± 0.01
	gm(-Ac)-AEJG 1	3.04E+08	17.0%	6.03 ± 0.03
	gm(-Ac)-AEJ 1	1.57E+08	8.8%	5.23 ± 0.02
	gm(-Ac)-AEJF 1	7.46E+07	4.2%	20.68 ± 0.02
	gm(-Ac)-AE 1	6.03E+07	3.4%	7.76 ± 0.01
	gm(-Ac)-AEJV 1	3.99E+07	2.2%	13.81 ± 0.02
	gm(-Ac)-AEJY 1	1.88E+07	1.1%	15.69 ± 0.02
	gm(-Ac)-AEJKD 1	1.36E+07	0.8%	8.18 ± 0.01
	gm(-Ac)-AEJA (Amidated) 1	1.76E+07	0.98%	7.63 ± 0.01
	gm(-Ac)-AEJ (Amidated) 1	5.20E+06	0.29%	4.24 ± 0.02
	gm(-Ac)-AE (Amidated) 1	3.94E+06	0.22%	5.77 ± 0.02
	gm(-Ac)-AEJG (Amidated) 1	8.29E+05	0.05%	5.83 ± 0.02
	gm(-Ac)-AEJF (Amidated) 1	2.05E+05	0.01%	20.14 ± 0.03
	gm(-Ac)-AEJV (Amidated) 1	1.65E+05	0.01%	13.28 ± 0.00

_							
В	Original Structure	Intensity	RT	Modified Structure	Intensity	RT	% modified
	gm(-Ac)-AEJA 1	1.12E+09	7.83	gm(-Ac)-AEJA (Amidated) 1	1.76E+07	7.63	1.57%
	gm(-Ac)-AEJG 1	3.04E+08	6.03	gm(-Ac)-AEJG (Amidated) 1	8.29E+05	5.83	0.27%
	gm(-Ac)-AEJ 1	1.57E+08	5.23	gm(-Ac)-AEJ (Amidated) 1	5.20E+06	4.24	3.31%
	gm(-Ac)-AEJF 1	7.46E+07	20.68	gm(-Ac)-AEJF (Amidated) 1	2.05E+05	20.14	0.28%
	gm(-Ac)-AE 1	6.03E+07	7.76	gm(-Ac)-AE (Amidated) 1	3.94E+06	5.77	6.53%
	gm(-Ac)-AEJV 1	3.99E+07	13.81	gm(-Ac)-AEJV (Amidated) 1	1.65E+05	13.28	0.41%

C	Structure	Monoisotopic Mass	Structure	Monoisotopic Mass	Obs. MS2	Δppm (unmodified)	∆ppm (modified)
	gm(-Ac)-AEJA	899.3972	gm(-Ac)-AEJA (Amidated)	898.4132			
	[M+1]	900.4050	[M+1]	899.4210	899.4177	22.7925	3.67
_	[M+2]	450.7064	[M+2]	450.2144	450.2126	1096.8151	4.00





_		Α	Amidated Amidated				
F	Modified Structure	Fragmented Ion	Е	J	Validated		
	gm(-Ac)-AEJA (Amidated) 1	450.2126	-	172.1077	Yes		
	gm(-Ac)-AEJG (Amidated) 1	443.2047	-	172.1076	Yes		
	gm(-Ac)-AEJF (Amidated) 1	488.2286	-	172.1076	Yes		
	gm(-Ac)-AEJV (Amidated) 1	464.2286	-	172.1078	Yes		
	gm(-Ac)-AEJ (Amidated) 1	414.6942	-	173.0916	No		
	gm(-Ac)-AE (Amidated) 1	-	-	-	No		

Figure 3-8. PGFinder amidation search

A. Consolidated results from the search, the modifications found are shaded in green. **B.** Side by side comparison of the original and modified structures. **C.** Calculated [M+1] ions for gm(-Ac)-AEJA and amidated gm(-Ac)-AEJA. **D.** MS/MS fragmentation chromatogram of amidated gm(-Ac)-AEJA, highlighted in red the signature ion for mDAP amidated (172.1086) and D-Glu amidated (129.0664). **E.** List of amidated muropeptides, mark as "Yes" are the validated ones.



Figure 3-9. Strategy for amidation identification

Two searches were performed using the Byos[®] and PGFinder software. DB_Cd10 and DB_Cd1 contain all possible amino acids in 4th and 5th residue positions. The monomer database DB_Cd2 was built based on the MS/MS analysis carried out with Byos[®] and the results from DB_Cd1 PGFinder search. The identification of the most abundant monomers was used to build the following database (DB_Cd2) to identify 5 amidations.

3.3 Discussion

This chapter describes the initial exploration of the methodology for analysing PG using LC-MS/MS. The work focused on improving the fragmentation coverage for *C. difficile* PG, a bacterium known for causing severe gastrointestinal infections (Mada and Alam, 2025). Two fragmentation strategies were compared:Top5 and Top25, referring to the fragmentation of the five or twenty-five most abundant ions, respectively. The impact of both strategies on the identification of muropeptides, using both MS1 (precursor ion) and MS/MS (fragment ion) was analysed. The effectiveness of each fragmentation strategy was examined by comparing the results obtained with different analysis tools (PGFinder for MS1 and Byos® for MS/MS), with the aim of increasing the number of muropeptides identified.

Our results showed that both fragmentation setups (Top5 and Top25) provide similar amounts of information and quality for the most abundant *C. difficile* monomers. Combined MS and MS/MS allowed us to distinguish between mass coincidences and confidently identify modifications.

Our analyses provided a robust list of structures, allowing us to distinguish between mass coincidences and confidently identify modifications and eventually multimers in any subsequent steps.

Unexpectedly our results also indicate that Top5 is the most suitable fragmentation setup, as it provides more comprehensive data, including higher MS intensity, a greater number of identified monomers in PGFinder and a higher number of monomers validated by Byos®. Even though the Top25 setup increases the frequency of MS/MS acquisition, it did not yield more PGFinder and Byos® identifications. This is because the Top25 configuration leads to increased collection of fragmentation data from a smaller set of muropeptides, compared to Top5, thereby yielding better quality data only for a smaller number of these species.

It is important to note that the higher Orbitrap resolving power requires longer measurement times, leading to fewer data points collected as reported in Table 3-6. This trade-off between resolution and acquisition speed should be considered when selecting the appropriate settings for quantitative analysis.

When we performed the modification analysis (see section 3.8), we noticed that the MS/MS data from either of both datasets was equally useful for validating the modifications. This is likely due to our decision to focus on the top 95% most abundant monomers.

During this analysis, we observed common trends in the modified muropeptides, such as decreased retention time (RT) for amidations and lower abundance compared to the unmodified monomers. We looked for specific product ions to validate amidations, particularly signature ions for *meso*-DAP or D-Glu. We confirmed the presence of *meso*-DAP amidated versions for the 3 most abundant monomers.

3.4 Concluding remarks

Since previous LC-MS/MS experiments (Patel, 2021) were performed in different facilities (Sheffields' Faculty of Sciences Biological Mass Spectrometry Facility and Institut Pasteur) using different mass spectrometers, it was crucial to compare fragmentation setups for our new experiments conducted in the Chemical Engineering at the Life Sciences Interfaces (ChELSI) facility. While Top25 promised improved data acquisition, it ultimately did not outperform Top5.

PGFinder v0.1.1a proved highly effective for identifying monomers, as all Byos® identifications were previously listed in PGFinder searches. However, a significant drawback is the time-consuming manual calculation of database masses, especially for DB_Cd1. This increases the risk of misidentifying muropeptides due to mass calculation errors. This led to the creation of a database builder in the latest version of PGFinder (v1.2.1 as of January 2024).

This work indicated that the size of the database could be simplified for unbiased searches. Keeping only pentapeptides with a D-alanine in the fifth residue seems a reasonable strategy, as we noted that most of the pentapeptides identified have a D-alanine as fourth residue. Another interesting approach is to use a list of monomers validated by Byos® to reduce the initial monomer database.

When we performed the modification searches, the analysis highlighted the usefulness of identifying a single signature ion for validating modifications. However, some modifications, like amidase products (cleavage of the disaccharide moiety), will be challenging to validate. These would require simultaneous analysis for the absence of the disaccharide moiety and the presence of the glycan chain. Therefore, consistent prediction of the product ions for these modifications, particularly for multimers will be crucial. The work done during this chapter help us define some general rules for PG analysis. This is summarized in Figure 3-9 where we show the first draft of a workflow to analyse PG composition, combining Byos® and PGFinder analysis.

Chapter 4

4. Characterization of *Clostridioides difficile's* PG composition and $rLdt_{Cd}s$ activity *in vitro*

As part of my PhD thesis, I am including the following paper:

Galley, N. F., Greetham, D., Alamán-Zárate, M. G., Williamson, M. P., Evans, C. A., Spittal, W. D., Buddle, J. E., Freeman, J., Davis, G. L., Dickman, M. J., Wilcox, M. H., Lovering, A. L., Fagan, R. P., & Mesnage, S. (2024). *Clostridioides difficile* canonical L,D-transpeptidases catalyze a novel type of peptidoglycan cross-links and are not required for β -lactam resistance. The Journal of Biological Chemistry, 300(1), 105529. https://doi.org/10.1016/j.jbc.2023.105529

The work described in this paper was performed in collaboration with the groups of Prof. Andrew Lovering, Prof Mark Dickman, Prof Mark Wilcox and Dr Robert Fagan.

The samples I analysed were generated by Nicola Galley and Darren Greetham from Dr Mesnage Lab. The mass spectrometry data acquisition was performed by Dr Caroline Evans in Prof Dickman's facilities.

The NMR analysis was performed by Prof Mike Williamson, the genome analysis was carried out by Stéphane Mesnage and Jessica Buddle from Dr Fagan Lab and the antimicrobial susceptibility assays by William Spittal, Jane Freeman and Georgina Davis from Prof Wilcox Lab.

My contribution has been the structural analysis of *C. difficile* peptidoglycan and the analysis of *in vitro* transpeptidation products. Both types of experiments involved LC-MS/MS. The original draft was written by Dr Mesnage. The text was edited by Dr Mesnage and me, considering feedback from other co-authors. My major contribution was the figures and the analysis shown in Figure 4-1, Figure 4-2A, Table 1 and Table S4-3, and was acknowledged by a third authorship.

Following the results presented in the previous chapter, we sought to perform a complete PG analysis of *C. difficile* wild type (WT) and *Idt* triple mutant using the optimal fragmentation method identified. Additionally, we aimed to characterise the *in vitro* transpeptidation products generated by the three canonical *C. difficile* recombinant Ldt ($rLdt_{Cd}$) enzymes.

$\begin{array}{c} Clostridioides\ difficile\ {\rm canonical\ L,D-}\\ transpeptidases\ catalyse\ a\ novel\ type\ of\\ peptidoglycan\ cross-links\ and\ are\ not\ required\ for\\ \beta-lactam\ resistance\end{array}$

Nicola F. Galley¹, Darren Greetham¹, Marcel G. Alaman-Zarate¹, Mike P. Williamson¹, Caroline A. Evans², William D. Spittal^{3,4}, Jessica E. Buddle¹, Jane Freeman^{3,4}, Georgina L. Davis^{3,4}, Mark J. Dickman², Andrew L. Lovering⁵, Mark H. Wilcox^{3,4}, Robert P. Fagan¹, Stéphane Mesnage¹

¹School of Biosciences, University of Sheffield, Sheffield, UK

² Department of Chemical and Biological Engineering, University of Sheffield, Sheffield, UK

³ Dept Microbiology, Leeds Teaching Hospitals NHS Trust, Leeds Institute of Medical Research, University of Leeds, Leeds, UK

⁴ Healthcare Associated Infections Research Group, Leeds Institute of Medical Research University of Leeds, Leeds UK

⁵ School of Biosciences, University of Birmingham, Birmingham, UK

* These authors contributed equally

Running title: Novel peptidoglycan remodelling activities in C. difficile

Keywords: *Clostridium difficile*, peptidoglycan, L,D-transpeptidase, cross-link, bacterial cell envelope, antibiotics, beta-lactams

Address correspondence to Stéphane Mesnage, s.mesnage@sheffield.ac.uk

4.1 Abstract

Clostridioides difficile is the leading cause of antibiotic-associated diarrhoea worldwide with significant morbidity and mortality. This organism is naturally resistant to several β -lactam antibiotics that inhibit the polymerisation of peptidoglycan, the essential component of the bacteria cell envelope. Previous work has revealed that *C. difficile* peptidoglycan has an unusual composition. It mostly contains 3-3 cross-links, catalysed by enzymes called L,D-transpeptidases (Ldts) that are poorly inhibited by β -lactams. Unlike PBPs that utilize pentapeptides as donor substrates, Ldts employ tetrapeptides. Because β -lactams are structural analogues of the terminal D-Ala-D-Ala moiety of pentapeptide stems, they do not effectively mimic the Ldt substrate, whose specificity is directed towards tetrapeptides. It was therefore hypothesized that peptidoglycan polymerization by these enzymes could underpin antibiotic resistance. Here, we investigated that catalytic activity of the three canonical Ldts encoded by *C. difficile* (Ldt_{cd1}, Ldt_{cd2} and Ldt_{cd3}) *in vitro* and explored their contribution to growth and antibiotic resistance.

We show that two of these enzymes catalyse the formation of novel types of peptidoglycan cross-links using *meso*-diaminopimelic acid both as donor and an acceptor, also observed in peptidoglycan sacculi. We demonstrated that the simultaneous deletion of these three genes only has a minor impact on both peptidoglycan structure and resistance to β -lactams. This unexpected result therefore implies that the formation of 3-3 peptidoglycan cross-links in *C. difficile* is catalysed by yet unidentified non-canonical Ldt enzymes.

77

4.2 Introduction

Clostridioides difficile is a spore-forming Gram-positive obligate anaerobe that can cause hospital-associated diarrhoea worldwide representing increasing healthcare resource and economic burden. (Wingen-Heimann et al., 2023). Although *C. difficile* has been recognised as a major cause of healthcare associated infection since the 1970s, the more recent increase in morbidity and mortality is linked to the emergence of virulent epidemic strains including those belonging to ribotype 027 (Lessa et al., 2012). *C. difficile* infections (CDIs) are underpinned by the natural resistance of this organism to several antibiotics including broad spectrum β -lactams such as cephalosporins. The dysbiosis caused by an antibiotic treatment creates an environment conducive to the germination of *C. difficile* spores and the production of virulence factors including toxins and several surface proteins (Buddle and Fagan, 2023).

The resistance of C. difficile to B-lactams is poorly understood. These antibiotics covalently bind to D,D-transpeptidases (also known as Penicillin Binding Proteins) and irreversibly inhibit the enzymatic activity of these enzymes (Frère, 1977). In most bacteria, inhibition of D,D-transpeptidation blocks the polymerisation of peptidoglycan, the major and essential component of the bacterial cell wall, and prevents bacterial growth (Cho et al., 2014). The peptidoglycan of C. difficile has an unusual composition. It is mostly polymerised by a class of enzymes called L,D transpeptidases (Ldts) (Peltier et al., 2011). Unlike D,D-transpeptidases, which form bonds between the amino acids in positions 4 and 3 of peptidoglycan peptide stems (3-4 cross-links), Ldts form bonds between two amino acids in positions 3 (3-3 crosslinks). The activity of Ldts involves a catalytic mechanism distinct from the mechanism of D,D-transpeptidases and Ldts are not inhibited by ß lactams, with the exception of carbapenems (Aliashkevich and Cava, 2021). The C. difficile genome encodes 3 enzymes called Ldt_{Cd1}, Ldt_{Cd2} and Ldt_{Cd3} which contain a canonical Ldt domain (YkuD). The contribution of these three enzymes to the peptidoglycan structure was investigated in strain 630 (Peltier et al., 2011). Despite attempts to generate a triple knockout strain, only genes encoding Ldt_{Cd1} and Ldt_{Cd2} could be inactivated simultaneously (Peltier et al., 2011) so it was suggested that at least one Ldt was required for viability. Analysis of the peptidoglycan structure in the double mutant strain revealed only a limited decrease of 3-3 cross-links (Peltier et al., 2011). The mutant remained able to perform 3-3 cross-links in the presence of ampicillin, suggesting that C. difficile Ldts were insensitive to this antibiotic (Peltier et al., 2011). In vitro experiments revealed that these enzymes display distinct enzymatic activities and inhibition by β-lactams (Sütterlin et al., 2018). All enzymes were reported to have L,D carboxypeptidase activity but L.D-transpeptidation and exchange of the amino acid in position 4 could only be detected for Ldt_{Cd2} and Ldt_{Cd3} . Interestingly, Ldt_{Cd3} could not be acylated by any of the β -lactams tested. The acylation efficacy of Ldt_{Cd2} and Ldt_{Cd3} by penicillin and cephalosporin antibiotics was much lower than the acylation by carbapenems and the hydrolysis of these antibiotics was more efficient. It was therefore concluded that Ldt_{Cd} activity could only be inhibited by carbapenems (Sütterlin et al., 2018).

Outstanding questions remain on the individual role of Ldt_{Cd} enzymes in peptidoglycan polymerisation, the essentiality of the L,D-transpeptidation pathway in *C. difficile* and

its contribution to antibiotic resistance. In this work, we further investigate the enzymatic activity of *C. difficile* Ldts, both *in vitro* and during vegetative growth. We show that Ldt_{Cd2} and Ldt_{Cd3} display novel enzymatic activities and that the genes encoding the three canonical Ldts can be deleted simultaneously. High resolution structure of the wild type and triple mutant peptidoglycan only revealed a minor impact on muropeptide composition and no change in resistance to β -lactams could be detected in the mutant strain. This work therefore provides new insights into the catalytic activities of Ldts and implies the existence of another unidentified type of enzyme(s) that does not contain a canonical YkuD domain yet is able to catalyse the formation of 3-3 cross-links in *C. difficile*.

4.3 Results

4.3.1 In vitro assays with recombinant Ldt_{Cd1} and Ldt_{Cd3} reveal distinct activities and a novel type of L,D-transpeptidation

We sought to investigate the activities of the three Ldt_{Cd} enzymes to identify their specific roles in peptidoglycan remodelling. The recombinant enzymes were purified (Fig. S4-1) to test their enzymatic activities using four types of purified substrates: (i) a disaccharide-tetrapeptide (GlcNAc-MurNAc-L-Ala-D-isoGlu-DAP-D-Ala; gm-AEJA) alone to test L,D-carboxypeptidase and L,D-transpeptidase activities (Fig. 4-1A); (ii) the same disaccharide-tetrapeptide (gm-AEJA) in the presence of D-Methionine to test 4th amino acid exchange (Fig. 4-1B); (iii) a 3-4 cross-linked dimer ((GlcNAc-MurNAc-L-Ala-D-isoGlu-DAP-D-Ala)2; gm-AEJA=gm-AEJA, where "=" represents a peptidoglycan cross-link) (Fig. 4-1C) and (iv) a 3-3 cross-linked dimer (GlcN-MurNAc-L-Ala-D-isoGlu-DAP-D-Ala - GlcN-MurNAc-L-Ala-D-isoGlu-DAP (g(-Ac)m-AEJA=g(-Ac)m-AEJA) to test endopeptidase activities ((Fig. 4-1D).



_	
_	
_	

F

Dook Informed structure		Crocolink	Mas	Mass (Da)	
Pear	interred structure	Crossiink	Obs	Theo	∆ppm
1	gm-AEJ	-	870.367	870.370	4.3
2	ğm-AEJA	-	941.404	941.408	4.6
3	gm-AEJ=gm-AEJ	3-3	1722.722	1722.730	5.1
4	gm-AEJA=gm-AEJ	3-3	1793.757	1793.768	5.8
5	gm-AEJM	_	1001.404	1001.411	6.8
6	gm-AEJ=gm-AEJA	3-4	1793.757	1793.768	6.0
7	gm-AEJA=gm-AEJA	3-4	1864.793	1864.805	6.2
а	gm-AEJ=gm-AEJ - H ₂ O	3-3	1704.713	1704.710	2.0
b	gm-AEJ=gm-AEJ=gm-AEJ - H ₂ O	3-3	2557.066	2557.070	1.5
с	gm-AEJ=gm-AEJA - H ₂ O	3-4 + 3-3	1775.747	1775.757	5.8
1*	g(-Ac)m-AEJ	-	828.356	828.360	4.5
2*	g(-Ac)m-AEJA	-	899.394	899.397	4.0
3*	g(-Ac)m-AEJ=g(-Ac)m-AEJ	3-3	1638.702	1638.710	4.6
4*	g(-Ac)m-AEJA=g(-Ac)m-AEJ	3-3	1709.739	1709.747	4.4
a*	$g(-Ac)m-AEJ=g(-Ac)m-AEJ - H_2O$	3-3	1620.691	1620.699	5.0
b*	$g(-Ac)m-AEJ=g(-Ac)m-AEJ=g(-Ac)m-AEJ - H_2O$	3-3	2431.034	2431.049	6.0
d*	$g(-Ac)m-AEJ=g(-Ac)m-AEJ=g(-Ac)m-AEJ=g(-Ac)m-AEJ - H_{3}$	O 3-3	3241.383	3241.398	4.6



Figure 4-1. Ldt in vitro assays.

Recombinant enzymes were incubated in the presence of a disaccharide-tetrapeptide substrate to test carboxypeptidase and transpeptidase activity and exchange (A and B, respectively). Endopeptidase activity was tested using either a 3-4 cross-linked or a 3-3 cross-linked dimer (C and D, respectively). The inferred structures from LC-MS analysis (E) as well as expected structures (F) are described. meso-DAP is abbreviated as J.

The monomer and 4-3 cross-linked dimer were purified from Escherichia coli Δ 6ldt strain and therefore contain fully acetylated sugars (Morè et al., 2019). The 3-3 crosslinked dimer was purified from C. difficile and therefore contained deacetylated

GlcNAc (GlcN). Ldt_{Cd} recombinant enzymes were active against all substrates tested and revealed distinct preferential activities. Ldt_{Cd1} displayed a low carboxypeptidase and transpeptidase activity and only converted half of the substrate during the exchange reaction. No endopeptidase activity was detected with any of the dimers.

Ldt_{Cd2} had a preferential carboxypeptidase activity on the gm-AEJA substrate that was mostly converted into a disaccharide-tripeptide (gm-AEJ). A very weak carboxypeptidase activity was detected with the 3-4 dimer whilst all the 3-3 dimer was completely cleaved, releasing disaccharide-tripeptides (g(-Ac)m-AEJ) as the most abundant products. The distinct activity of Ldt_{Cd2} on 3-4 and 3-3 cross-linked dimers clearly showed a pronounced specificity for 3-3 cross-links. Interestingly, several multimers matching the expected mass for dimers, trimers and tetramers lacking a molecule of water were detected (labelled as a, b, c, a*, b* and d*; Fig. 4-1E). These were further analysed by NMR.

Ldt_{Cd3} had the most pronounced L,D-transpeptidase activity of all enzymes since the 3-3 cross-linked dimer was the most abundant product generated from the gm-AEJA substrate. This enzyme also displayed some carboxypeptidase activity, using the monomer or both dimers. Surprisingly, the carboxypeptidase activity of LdtCd3 was higher on the 3-4 dimer than on the 3-3 dimer. LdtCd3 also produced two transpeptidation products matching the mass of a 3-3 dimer lacking a molecule of water (peak b, also detected with Ldt_{Cd2}) and the mass of a 3-4 dimer lacking a molecule of water (peak c) depending on the substrate used (monomer or dimer, respectively). The structures of all expected and previously described muropeptides produced by Ldt_{Cd1}, Ldt_{Cd2} and Ldt_{Cd3} are described in Fig. 4-1E.

4.3.2 MS/MS and NMR analyses of Ldt_{Cd2} and Ldt_{Cd3} transpeptidation products reveal a novel type of peptidoglycan cross-link

The muropeptide contained in peak 1 (Fig 4-1) was analysed by MS/MS and confirmed the inferred structure for a dimer with doubly cross-linked DAP residues used as both an acceptor and a donor group (Fig. 4-2A; see ions labelled). Several signature ions were found, including a doubly cross-linked DAP-DAP diaminoacid.

Peptidoglycan fragments in peaks 1, a and b (Fig. 1E) were purified and further analysed by NMR. 1D NMR spectra of the peptidoglycan fragments demonstrated a high purity for each. The monomer (Fig. 4-2B) has the amide signals expected for the structure of gm-AEJ (peak 1), namely two sugar N-acetyl signals, and one signal each for the Ala, DAP and iE residues. There is only one signal for DAP because one amine forms an amide with iE, while the other is a free amino group and therefore exchanges too fast with water to be visible. All the other signals are as expected, including the presence of two *N*-acetyl methyl singlets from the two sugars, and two methyl doublets from Ala and the lactyl group on MurNAc.

The dimer has a remarkably simple NMR spectrum. The amide region (Fig. 4-2A) contains only six amide doublets, and there are for example only two N-acetyl methyl singlets and two methyl doublets, as seen in the monomer (Fig. S4-1). This simplicity very strongly suggests a symmetrical dimer, and the similarity of the amide chemical shifts between the monomer and dimer implies a similar covalent structure.



Figure S4- 1. One-dimensional 1H NMR spectra (A)gm-AEJ monomer (peak 1 in Figure 1) and (B) gm-AEJ=gm-AEJ (-H2O) dimer (peak a in Figure 1). The two singlet *N*-acetyl methyl signal are at around 2 ppm, and the two methyl doublets are at around 1.4 ppm.

There are two amide signals from the diaminopimelate, visible in the TOCSY spectrum because they belong to the same spin system (Fig. 4-2C), indicating that both amines in the diaminopimelate take part in amide bonds. Chemical shift assignments for the dimer are listed in Table S4-1. In the NOESY spectrum, there are the expected sequential NOEs present between NH_i and protons in residue (i-1), as indicated in Fig. 4-2D. Crucially, these include NOEs between DAP NH^a and iE CgH₂, and the other "sequential" NOE of DAP NH^b to DAP CaH (Fig. 4-2D and 4-2E). Similarly, the spectrum of the trimer (Fig. 4-2B) is also very similar. The chemical shifts remain very similar to the monomer and dimer, and again there is only one set of signals, indicating a symmetrical trimer. The NMR spectra are thus fully consistent with the structures described in Fig. 4-3, and the simplicity of the spectrum means that no unsymmetrical structure is possible. Based on our NMR data, we conclude that the muropeptides in peaks a, b, c and d all correspond to multiply cross-linked structures (Fig. 4-2 and Fig. 4-3).



Figure 4-2.. MS/MS and NMR analysis of unusual muropeptides a and b.

A, MS/MS analysis of the ion corresponding to peak a (*m/z*=1705.7211). Nine fragment ions containing peptides with doubly bonded m-DAP residues are indicated. **B**, 1D NMR spectra of peptidoglycan fragments: peak 1 (gm-AEJ control), peak a and peak b. The identity of each amide proton is indicated on the spectra. **C**, Part of the TOCSY spectrum of the muropeptide in peak a, showing connectivities between amide protons and sidechains. The signals linked by red lines are the connectivities for the DAP^a and DAP^b amides, showing that they connect to identical sidechain frequencies and are therefore part of the same spin system. **D**, TOCSY (red) and NOESY (black) spectra of the dimer. Significant peaks are marked. **E**, Structures of (left) dimer (with NOEs indicated) and (right) trimer. Only the central part of the trimer is shown, with arrows indicating where the AEJ chains are attached. A, alanine; GlcNAc, *N*-acetylglucosamine; DAP^a, meso-diaminopimelic acid backbone (directly bonded to the isoglutamate), DAP^b, meso-diaminopimelic acid sidechain; MurNAc, *N*-acetylmuramic acid; iE, isoglutamate.



Figure 4-3. Unusual peptidoglycan cross-links catalysed by $Ldt_{\rm Cd2}$ and $Ldt_{\rm Cd3}.$

Based on NMR data, the structure of muropeptides corresponding to peaks a, b, c are shown next to the structure of the canonical 3-3 or 3-4 dimers, 3-3 trimer and 3-3 tetramer. Dimer c contains both types of cross-links (3-3 and 3-4), resulting from D,D- and L,D-transpeptidation. Muropeptides a*, b*, d* display the same cross-links as a, b, d, but contain GlcN instead of GlcNAc.

¹ H signal	Chemical shift (ppm)
GlcNAc amide	8.35
GlcNAc methyl	2.08
GlcNAc anomeric	4.67
GlcNac ring	3.74, 3.58, 3.48, 3.44
MurNAc amide	8.11
MurNac methyl	2.01
MurNAc ring *	4.43, 4.32
Lac CH	4.28
Lac methyl	1.42
A NH	8.41
Α CαH	4.32
A methyl	1.43
iE NH	7.92
iE CaH	4.20
iE CβH ₂	1.90, 2.10
iE CγH ₂	2.26
DAP NaH **	8.11
DAP NbH **	8.04
DAP CaH **	4.43
DAP CbH **	4.26
DAP CaH ₂ **	1.74
DAP CbH ₂ **	1.95
DAP CcH2 **	1.69

Table S4-1. Chemical shift assignments for the dimer peptidoglycan fragment a.

* Note that the muramic acid sugar has been reduced, meaning that there is no anomeric proton for this sugar.

** See structures for the locations of these protons

4.3.3 TraDIS analysis and gene deletion reveal that the three canonical *C. difficile* Ldts are non-essential.

Previous attempts to build a mutant with deletions in all 3 genes encoding the canonical Ldts were unsuccessful (Peltier et al., 2011), suggesting either that one of them is essential or that 3-3 cross-linking is essential. We took advantage of transposon-directed insertion site sequencing (TraDIS) data previously published in a study that identified essential genes in *C. difficile* R20291 (Dembek et al., 2015). Using the number of transposon insertions in each Idt_{Cd} gene as proxy to determine essentiality, our TraDIS analysis indicated that none of these genes was essential (Fig. 4-4), leaving the possibility that the combined deletion of Idt_{Cd1} , Idt_{Cd2} and Idt_{Cd3} could be non-viable. To test this hypothesis, we sought to generate a series of in-frame deletions in Idt_{Cd1} , Idt_{Cd2} and Idt_{Cd3} . All genes were deleted individually, or simultaneously. All the combinations of deletions, including the triple deletion mutant,

could be obtained, showing that these genes are not required for viability. Since this result was unexpected, we performed whole genome sequencing on chosen mutants and confirmed the deletion of the *ldt* genes. All expected deletions were confirmed in the strains sequenced. Single nucleotide polymorphism analysis identified a single variant unique to the triple deletion mutant (Fig. S4-2). The mutation (T>TA) occurred at position 581480 in the intergenic mutation between two genes with no known link to peptidoglycan polymerisation (*CD0482* and *glsA*), encoding a putative phosphoribulokinase/uridine kinase and a putative glutaminase, respectively. This result therefore suggests that the deletion of the *ldt_{Cd}* genes does not lead to genetic mutations likely to compensate for the lack of L,D-transpeptidase activity.



Figure 4-4. TraDIS analysis of the *ldt_{Cd}* loci.

The number of transposition events are shown as histograms depicting the localization (X-axis) and the frequency (Y-axis) of transposon insertion sites (in red are antisense insertions, in blue are same sense insertions).

4.3.4 High-resolution structure of the WT and triple mutant peptidoglycans

Peptidoglycan was extracted from vegetative cells in stationary phase and soluble fragments released after mutanolysin digestion were analysed by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). Surprisingly, the chromatograms of the WT and triple mutant were virtually identical, indicating a very minor contribution of the three Ldts to peptidoglycan structure (Fig. 4-5).

To investigate subtle differences associated with the simultaneous deletion of the 3 Idt_{Cd} genes, we therefore performed a high-resolution analysis of the LC-MS/MS dataset using the Byos® (Protein Metrics) and PGFinder software (Patel et al., 2021). A bespoke search strategy was designed (Fig. S4-3).



Figure S4-2. Genome analysis of *C. difficile* WT and its $\Delta 3ldt$ derivative.

A. Schematic representation of the R20291 chromosome depicting the localisation of the 3 ldt_{Cd} genes and the SNAP detected in the triple mutant in position 581480 (T \rightarrow TA). The sequence in Figure A shows the position of the mutation between the two adjacent genes. **B**. sequence alignments of the contigs assembled from the 3 independent genomes sequenced. Each DNA was purified from the culture that was used for peptidoglycan analysis.



Figure 4-5. HPLC-MS chromatogram of *C. difficile* reduced disaccharide-peptides. Each chromatogram corresponds to a biological replicate of strain R2091 (WT) and its isogenic derivative with in-frame deletions in genes ldt_{Cd1} , ldt_{Cd2} and ldt_{Cd3} (\triangle 3*ldt*).

A first search was performed to identify the monomer search space using the Byonic[™] module from the proprietary software Byos® module (Protein Metrics by Dotmatics). Thirty-four disaccharide-peptides with a fragmentation showing more than half of the expected b and y ions were identified (Fig. S4-3). These included 24 deacetylated monomers containing di-, tri-, tetra- and pentapeptide stems (g(-Ac)m-AE, g(-Ac)m-AEJ, g(-Ac)m-AEJX and g(-Ac)m-AEJAX) and 10 fully acetylated monomers (gm-AE, gm-AEJ, gm-AEJX and gm-AEJAX), where X can be any amino acid. A database called DB 0 made of these 34 monomers was used to perform a PGFinder search (step 2 in Fig. S4-3) to identify the most abundant monomers. 13 disaccharide peptides accounting for more than 98% of the monomers identified were selected to create a second database (DB 1) containing dimers resulting from 3-3 and 3-4 cross-linking. A third PGFinder search (step 3 in Fig. S4-3) was performed to identify the most abundant dimers and generate the next database containing monomers, dimers and trimers (DB 2). The next search with PGFinder and DB 2 (step 4 in Fig. S4-3) identified the most abundant trimers. A final database called DB 3 was created using all the information from sequential searches; it contained all MS2-checked monomers, 26 dimers, 16 trimers and all AnhydroMurNAc derivatives of the 10 most abundant mono-, di- and trimers, as well as 4 unusually cross-linked structures identified during in vitro assays (110 structures in total).





Sequential searches were performed using the Byonic^M and PGFinder software. The monomer database DB_0 was built based on the MS/MS analysis carried out with Byonic^M. The identification of the most abundant monomers was used to build the following databases containing dimers (DB_1), trimers (DB_2) and anhydroMurNAc groups (DB_3). DB_3 was used for final «one off » search.
The result of the PGFinder search using DB_3 and biological replicates from the WT and triple mutant is described in Table 4-1. The search strategy described here combining both MS1 and MS2 analysis allowed us to identify 97 muropeptides, which represent an unprecedentedly detailed analysis. The comparison between the two strains revealed a remarkable similarity between the two peptidoglycan compositions (Table 4-2). No significant difference was found when comparing the proportion of monomers, dimers, trimers, or glycan chain length. Cross-linking index as well as the proportion of 3-3 cross-links were also similar (paired Student *t*-test). The only difference found was a significant decrease in the exchange reaction (23.4 \pm 0.7% in the WT and 16.9 \pm 1.4% in the triple mutant). Overall, our analysis therefore demonstrated that the three canonical L,D-transpeptidases Ldt_{Cd1}, Ldt_{Cd2} and Ldt_{Cd3} only contribute marginally to remodel the peptidoglycan of *C. difficile* vegetative cells.

4.3.5 Comparative phenomics of the parental R20291 strain and its isogenic $3\Delta ldt$ derivative.

A comprehensive set of experiments were carried out to compare the phenotype of the R20291 strain and the triple Idt_{Cd} mutant. As expected, based on the results from peptidoglycan analysis, no significant differences were observed between the two strains in cell size (Fig. S4-4), sporulation (Fig. S4-5) or toxin release (Fig. S4-6). MICs for several β -lactams were also tested for all the mutants generated in this study and did not reveal any difference in the resistance against any of these antibiotics (Table S4-2).

$(\Delta 3Idt)$					
Muropeptide ^a	Ldt activity /	Abundance (%)		RT (Min) °	Average ^c
	crosslink	WT ^b	∆3ldt		Δppm
gm(-Ac)-AEIA 1	1	21.594% ± 1.050%	28.192% ± 4.671%	7.90 ± 0.02	2.71
gm(-Ac)-AEI[1 gm(-Ac)-AEI[1	carboxypeptidase	9.002% ± 1.348%	8.235% ± 1.026%	5.39 ± 0.03	2.85
gm(-Ac)-AEJF11	Exchange	1.820% ± 0.806%	2.083% ± 0.886%	20.72 ± 0.03	2.14
gm(-Ac)-AEI1		3.174% ± 0.185%	2.297% ± 0.520%	7.74 ± 0.05	2.31
gm-AEJA 1 gm (Ac) AEJV 1	Exchange	$0.512\% \pm 0.061\%$	0.716% ± 0.211%	8.84 ± 0.02	1.50
gm(-Ac)-AEV[1 gm(-Ac)-AEV[1	Exchange	$0.944\% \pm 0.111\%$ $0.294\% \pm 0.076\%$	0.389% ± 0.113%	13.83 ± 0.02 18.20 ± 0.04	0.66
gm(-Ac)-AEIY 1	Exchange	0.172%	0.328% 0.072%	15.82 ± 0.04	0.99
gm(-Ac)-AEIAA 1		0.244% ± 0.007%	0.311% ± 0.053%	9.63 ± 0.08	1.68
gm-ADA (Ann) [1 gm(-Ac)-AFIS [1	Exchange	$0.108\% \pm 0.010\%$ $0.302\% \pm 0.036\%$	$0.270\% \pm 0.085\%$ $0.234\% \pm 0.039\%$	7.97 ± 0.11 5.52 ± 0.02	2.24
gm-AE 1	Enteringe	0.197% ± 0.025%	0.207% ± 0.130%	9.41 ± 4.13	1.89
gm(-Ac)-AEIK 1	Exchange	0.324% ± 0.086%	0.184% ± 0.053%	4.55 ± 0.06	2.08
gm(-AC)-AEIQ 1, gm(-AC)-AEJAG 1 gm(-AC)-AEIA (Anh) 1	Exchange	0.107% ± 0.011% 0.033% + 0.002%	0.095% ± 0.022% 0.084% + 0.010%	8.41 ± 1.39 11.35 ± 0.02	0.86
gm-AEIG 1	Exchange	0.115% ± 0.026%	0.071% ± 0.010%	7.74 ± 0.75	1.08
gm-AE 1		0.081% ± 0.016%	0.062% ± 0.019%	8.65 ± 0.03	0.41
gm(-AC)-AEIAEI1 gm(-AC)-AEIAEI1		$0.17/\% \pm 0.021\%$ $0.042\% \pm 0.015\%$	0.096% ± 0.022% 0.055% ± 0.018%	6.63 ± 0.04 24.36 ± 0.02	0.61
gm-AEI 1	Exchange	0.015% ± 0.009%	0.042% ± 0.028%	16.08 ± 4.43	0.88
gm(-Ac)-AEJN 1	Exchange	0.235% ± 0.065%	0.062% ± 0.004%	5.05 ± 0.03	1.49
gm(-Ac)-AEH 1	Exchange	$0.104\% \pm 0.080\%$	ND	5.19 ± 0.20	2.01
gm-ΔEFE11 gm-ΔEFE11	Exchange	0.062% ± 0.010% 0.019% ± 0.013%	$0.041\% \pm 0.003\%$ $0.020\% \pm 0.014\%$	7.51 ± 0.03 21.95 + 0.01	1.37
gm-AEJK 1. gm(-Ac)-AEJAV 1 (mixture)	Exchange	0.000% ± 0.000%	0.000% ± 0.000%	6.69 ± 3.18	1.88
em(-Ac)-AEJAII1		0.021% ± 0.003%	0.023% ± 0.003%	19.36 ± 0.02	0.76
gm(-Ac)-AEJAK 1 gm(-Ac)-AEJD11		0.019% ± 0.001%	0.024% ± 0.006%	7.22 ± 0.12	1.79
gm-AEIH 1	Exchange	0.023% ± 0.027%	0.02470 ± 0.004% ND	6.40 ± 0.02	0.74
gm(-Ac)-AEIT 1	Exchange	0.025% ± 0.002%	0.020% ± 0.002%	6.49 ± 0.14	1.08
gm(-Ac)-AE (Anh) 1		0.029% ± 0.004%	0.015% ± 0.004%	12.00 ± 0.01	1.58
gm-ADAAI1 gm(-Ac)-AFI (Anh) 1		$0.010\% \pm 0.001\%$ $0.025\% \pm 0.005\%$	$0.010\% \pm 0.000\%$ $0.013\% \pm 0.002\%$	10.79 ± 1.54 8.56 ± 0.01	1.10
gm(-Ac)-AEIAY 1		0.005%	0.005% ± 0.004%	17.18 ± 0.03	0.22
gm(-Ac)-AEIG (Anh) 1	Exchange	0.015% ± 0.004%	0.006% ± 0.003%	9.31 ± 0.10	0.46
gm(-Ac)-AEJAA (Anh) 1	Exchange	ND	0.001% ± 0.000%	9.49 ± 0.04	1.74
gm(-AC)-AEIY (Ann) 1 gm_AEIAEI 1	Exchange	ND 0.001% + 0.001%	0.000%	24.42 ± 0.03 24.07 ± 0.01	1.39
gm(-Ac)-AEJA=gm(-Ac)-AEJ12	3-3	16.323% ± 0.074%	19.316% ± 1.031%	13.09 ± 0.08	1.31
gm(-Ac)-AEJA=gm(-Ac)-AEJA 2	3-4	6.806% ± 1.727%	8.963% ± 1.708%	13.93 ± 0.11	1.20
gm(-Ac)-AEI=gm(-Ac)-AEI 2	3-3	5.903% ± 1.156%	5.446% ± 0.636%	11.99 ± 0.03	1.57
gmi-Aci-AEIG=gmi-Aci-AEII2 gm-AEIA=gm(-Ac)-AEII2	3-3 3-3	5./41% ± 0./30% 1.238% + 0.141%	2.088% ± 0.881% 1.677% ± 0.398%	11.54 ± 0.04 13.80 ± 0.03	1.39
gm(-Ac)-AEJF=gm(-Ac)-AEJ12	3-3	0.682% ± 0.213%	0.719% ± 0.283%	22.56 ± 0.01	1.85
gm-AEJA=gm(-Ac)-AEJA 2	3-4	0.560% ± 0.052%	0.713% ± 0.151%	14.78 ± 0.08	0.91
em(-Ac)-AEJG=em(-Ac)-AEJA 2	3-4	$1.112\% \pm 0.202\%$	$0.467\% \pm 0.114\%$	12.40 ± 0.02	0.68
gm-AEA=gm(-AC)-AEJ (Ann) 12 gm(-Ac)-AEIV=gm(-Ac)-AEI 2	3-3	0.244% ± 0.036% 0.308% + 0.047%	0.374% ± 0.131% 0.205% ± 0.057%	19.22 ± 0.10 17.77 ± 0.02	0.83
em(-Ac)-AEJA=em(-Ac)-AEJ (Anh) 2	3-3	0.337% ± 0.040%	0.244% ± 0.023%	15.68 ± 0.32	1.47
gm(-Ac)-AEII=gm(-Ac)-AEI 2	3-3	0.138% ± 0.055%	0.176% ± 0.073%	20.98 ± 0.02	0.88
gm-AEI=gm(-Ac)-AEJ[2 gm(-Ac)-AEIA=gm(-Ac)-AEIA-(Aph)]2	3-3	0.250% ± 0.011%	$0.250\% \pm 0.092\%$ $0.176\% \pm 0.011\%$	12.69 ± 0.03 16.99 \pm 0.02	1.26
gm(-Ac)-AEIF=gm(-Ac)-AEIA 2	3-4	$0.161\% \pm 0.061\%$	$0.130\% \pm 0.011\%$	23.14 ± 0.02	1.43
gm(-Ac)-AEJY=gm(-Ac)-AEJ 2	3-3	0.063% ± 0.084%	0.107% ± 0.034%	18.74 ± 1.18	0.37
gm(-Ac)-AEJS=gm(-Ac)-AEJ 2	3-3	0.207% ± 0.005%	0.133% ± 0.014%	11.41 ± 0.14	0.63
gm-AEIA=gm(-AC)-AEIA (ANN) 2 gm(-AC)-AEIK=gm(-AC)-AEII 2	3-4	0.072% ± 0.019% 0.154% ± 0.033%	$0.125\% \pm 0.012\%$ $0.080\% \pm 0.022\%$	14.02 ± 0.30 10.20 + 0.02	9.40
gm(-Ac)-AEII=gm(-Ac)-AEIA 2	3-4	0.037% ± 0.013%	0.042% ± 0.016%	21.78 ± 0.02	0.42
gm(-Ac)-AEJV=gm(-Ac)-AEJA 2	3-4	0.067% ± 0.006%	0.038% ± 0.011%	18.60 ± 0.02	0.35
gm(-Ac)-AEJS=gm(-Ac)-AEJA 2	3-4	0.053% ± 0.006%	$0.046\% \pm 0.008\%$	12.09 ± 0.02	1.10
gm(-Ac)-AEI=gm(-Ac)-AEI (Anh) 12	3-3	$0.013\% \pm 0.000\%$ $0.110\% \pm 0.021\%$	0.027% ± 0.002%	14.84 + 0.22	0.48
gm(-Ac)-AEIN=gm(-Ac)-AEJ 2	3-3	0.094% ± 0.025%	0.028% ± 0.005%	10.88 ± 0.02	0.64
gm(-Ac)-AEJK=gm(-Ac)-AEJA 2	3-4	0.044% ± 0.008%	0.027% ± 0.002%	10.97 ± 0.02	0.56
gm(-Ac)-AEJY=gm(-Ac)-AEJA 2 gm(-Ac)-AEIA=gm(-Ac)-AEI(-H2O)12	3-4	0.013% ± 0.000%	$0.021\% \pm 0.005\%$	18.92 ± 0.02 14.67 ± 0.54	0.31
gm(-Ac)-AEJN=gm(-Ac)-AEJA [2	3-4	0.021% ± 0.006%	0.009% ± 0.001%	11.62 ± 0.02	1.08
em(-Ac)-AEJG=em(-Ac)-AEJA (Anh)12	3-4	0.004%	0.000%	15.38 ± 0.00	2.26
gm(-Ac)-AEJF=gm(-Ac)-AEJ (Anh) 2 gm(-Ac)-AEJG=gm(-Ac)-AEJ (Anh) 2	3-3	$0.010\% \pm 0.002\%$	0.014% ± 0.014%	20.96 ± 4.59	5.28
$gm(-\Delta c)-\Delta EII=gm(-\Delta c)-\Delta EI (\Delta nh) 2$	3-3	$0.120\% \pm 0.017\%$ $0.002\% \pm 0.001\%$	0.001% + 0.000%	13.30 ± 2.77 23.32 ± 0.03	1.04
gm(-Ac)-AEJV=gm(-Ac)-AEJ (Anh) 2	3-3	0.015% ± 0.011%	0.000%	19.18 ± 3.13	2.32
gm(-Ac)-AEJA=gm(-Ac)-AEJ=gm(-Ac)-AEJ 3	3-3	3.428% ± 0.114%	3.041% ± 0.233%	15.39 ± 0.04	1.42
gm(-Ac)-AEJA=gm(-Ac)-AEJA=gm(-Ac)-AEJ 3	3-3, 3-4	$1.823\% \pm 0.087\%$	2.273% ± 0.327%	15.98 ± 0.02	1.09
gm-AEJA=gm(-AC)-AEJ=gm(-AC)-AEJ[3 gm-AEJA=gm(-AC)-AEJ=gm(-AC)-AEJ[3	3-3, 3-4 3-3	$0.34470 \pm 0.047\%$ $0.505\% \pm 0.052\%$	0.449% ± 0.133%	16.19 ± 0.02	0.27
gm(-Ac)-AEJA=gm(-Ac)-AEJA=gm(-Ac)-AEJA 3	3-4	0.348% ± 0.040%	0.468% ± 0.060%	16.55 ± 0.03	0.30
gm-AEJA=gm(-Ac)-AEJ=gm(-Ac)-AEJ (Anh) 3	3-3	0.259% ± 0.026%	0.342% ± 0.073%	20.34 ± 0.03	1.20
gm(-AC)-AEJ=gm(-AC)-AEJ=gm(-AC)-AEJ[3 gm(-AC)-AEIG=gm(-AC)-AEI=gm(-AC)-AEI[3	5-5 2-2	0.565% ± 0.165% 0.565% ± 0.090%	0.393% ± 0.046% 0.194% + 0.068%	14.40 ± 0.02	0.48
gm-AEJA=gm(-Ac)-AEJA=gm(-Ac)-AEJA 3	3-4	0.095% ± 0.015%	0.142% ± 0.032%	17.30 ± 0.02	0.31
gm(-Ac)-AEJG=gm(-Ac)-AEJA=gm(-Ac)-AEJ 3	3-3, 3-4	0.253% ± 0.052%	0.115% ± 0.038%	14.69 ± 0.01	0.31
gm-AEJA=gm(-Ac)-AEJA=gm(-Ac)-AEJ (Anh) 3	3-3, 3-4	$0.097\% \pm 0.008\%$	$0.136\% \pm 0.017\%$	20.18 ± 1.58	2.32
gm(-Ac)-AEIA=gm(-Ac)-AEI=gm(-Ac)-AEI[3 gm(-Ac)-AEIA=gm(-Ac)-AEI=gm(-Ac)-AEI (Anh)]3	3-3	$0.001\% \pm 0.018\%$ $0.174\% \pm 0.019\%$	$0.041\% \pm 0.019\%$ $0.049\% \pm 0.008\%$	17.55 ± 0.01	0.30
gm(-Ac)-AEJA=gm(-Ac)-AEJA=gm(-Ac)-AEJ (Anh)]3	3-3, 3-4	0.062% ± 0.007%	0.051% ± 0.005%	18.40 ± 0.09	0.40
em(-Ac)-AEJF=em(-Ac)-AEJ=em(-Ac)-AEJA13	3-3.3-4	0.036% ± 0.013%	0.026% ± 0.010%	23.39 ± 0.02	0.40
gm(-Ac)-AEJG=gm(-Ac)-AEJA=gm(-Ac)-AEJA 3	3-4	$0.047\% \pm 0.012\%$	0.021% ± 0.008%	15.23 ± 0.01	0.59
gm(-Ac)-AEJ=gm(-Ac)-AEJ=gm(-Ac)-AEJ (Ann113	3-3 3-3*	0.047% ± 0.007% 0.056% ± 0.015%	ND	16.98 ± 0.02	1.18
em(-Ac)-AEJA=em(-Ac)-AEJA=em(-Ac)-AEJA (Anh) 3	3-4	0.013% ± 0.001%	0.017% ± 0.002%	19.02 ± 0.04	0.40
em(-Ac)-AEJG=em(-Ac)-AEJ=em(-Ac)-AEJA (Anh) 3	3-3.3-4	0.011% ± 0.001%	ND	16.82 ± 0.12	0.25
gm(-Ac)-AEJA=gm(-Ac)-AEJ=gm(-Ac)-AEJ (-H2O) 3	3-3	0.002%	ND	16.83 ± 0.00	2.64
gm(-Ac)-AEI=gm(-Ac)-AEI=gm(-Ac)-AEI (Ann)13	3-4	0.00470 ± 0.006%	0.005% ± 0.007%	16.68 + 0.28	4.89
gm(-Ac)-AEJF=gm(-Ac)-AEJ=gm(-Ac)-AEJ (Anh) 3	3-3	0.002% ± 0.000%	0.005% ± 0.005%	23.51 ± 3.39	3.09
gm(-Ac)-AEJ=gm(-Ac)-AEJ=gm(-Ac)-AEJ (Anh) 3	3-3	0.083% ± 0.006%	0.005% ± 0.003%	16.68 ± 0.28	1.56
gm(-AC)-AEF=gm(-AC)-AEF=gm(-AC)-AEF (Anh) 3	3-3	0.002% ± 0.000%	0.005% ± 0.005%	23.51 ± 3.39	3.09

Table 4- 1. Muropeptide analysis of *C. difficile* R20291 and triple *Idt_{Cd1}, Idt_{Cd2}, Idt_{Cd3}* mutant

 amil-Aci-AEJE=gm(-Aci-AEJ=gm(-Aci-AEJ=amil-Aci-AEJ=Amil-Aci-AEJ=Amil-Aci-AEJ=Amil-Aci-AEJ=Amil-Aci-AEJ=Amil-Aci-AEJ=Amil-Aci-AEJ=Amil-Aci-AEJ=Amil-Aci-AEJ=Amil-Aci-AEJ=A

Muropeptides/properties	WT	$\Delta 3 l d t$
Monomers	49.93% ± 2.82%	49.96% ± 3.13%
Dimers	$41.12\% \pm 2.15\%$	$41.94\% \pm 2.19\%$
Trimers	$8.95\% \pm 1.00\%$	$8.09\% \pm 1.04\%$
Cross-linking index	$23.51\% \pm 1.40\%$	$23.64\% \pm 1.44\%$
Glycan chain length	96.6 ± 1.8	87.3 ± 18.7
gm-AEJX	$23.39\% \pm 0.70\%$	$16.86\% \pm 1.41\%$
3-3 cross-links	81.33% ± 2.54%	$78.35\% \pm 2.68\%$

Table 4-2. Summary of WT and 3Δ*ldt* PG properties



Figure S4- 4. Comparative analysis of WT and $\Delta 3 ldt$ cell size by flow cytometry. Comparison of median forward scattered (FSC) (A) and side scattered (SSC) light values(B) corresponding to WT and $\Delta 3 ldt$ mutant; NS, P>0.05; n=3 (Student unpaired *t*-test with Welch's correction).









Cells were grown overnight and inoculated in TY broth at a 1/100 dilution. Cultures were harvested at different time points to prepare cell lysates and concentrated culture supernatants (SN). Toxin B was detected by Western blot (chemiluminescence) using mouse monoclonal Anti toxin B antibody (MA1-7413) and a secondary rabbit anti-mouse antibody conjugated to horseradish peroxidase. A strain with a deletion in the gene encoding toxin B (\triangle PaLoc) was used as a negative control, M, Molecular weight marker.

		∆l dt _{Cd1}			∆ldt _{Cd1}	∆ldt _{Cd1}		∆ ldt _{Cd1}
	WT		Δldt_{Cd2}		∆ldt _{Cd2}		∆ldt _{Cd2}	Δldt_{Cd2}
				∆ldt Cd3		∆ldt Cd3	∆ldt Cd3	∆ldt Cd3
Penicillins								
Amoxicillin	4	4	4	4	4	4	2	2
Co-amoxicillin	2	2	1	1	2	2	2	2
Piperacillin	16	16	16	16	16	16	16	16
Piperacillin-tazobactam	16	16	16	16	16	16	16	16
Cephalosporins								
Cephalotin ^a	>256	>256	>256	>256	>256	>256	256	256
Cephalexin ^a	>256	>256	>256	>256	>256	>256	>256	>256
Cefuroxime b	>256	>256	>256	>256	>256	>256	>256	>256
Cefoxitin b	>256	>256	>256	>256	>256	>256	>256	>256
Ceftriaxone c	128	128	128	128	128	128	128	128
Cefotaxime ^c	>256	>256	>256	>256	>256	>256	>256	>256
Carbapenems								
Meropenem	8	8	8	8	8	8	8	8
Imipenem	32	32	16	32	16	16	16	16
Glycopeptides								
Vancomycin	1	1	1	1	1	1	1	2

Table S4-2. Minimum inhibitory concentration (MIC) of β -lactams against *C. difficile* R20291 and *ldt*_{Cd} isogenic mutants.

^a first generation cephalosporins

^b second generation cephalosporins

^c third generation cephalosporins

4.4 Discussion

L,D-transpeptidases represent a class of enzymes that are amenable to *in vitro* enzymatic assays since they can use soluble PG fragments as a substrate. *In vitro* assays with distinct PG fragments purified from intact sacculi were used to explore the catalytic activities of the 3 *C. difficile* Ldts. Our comparative analysis based on an LC-MS/MS assay with several substrates provided information about the preferential activity of each Ldt. Previous studies reported that Ldt_{Cd1} only displayed carboxypeptidase activity (Sütterlin et al., 2018). Our data confirmed this result but also revealed that it can also perform transpeptidation and exchange reactions, even though this enzyme was poorly active on all substrates tested. Ldt_{Cd2} was able to perform all reactions but preferentially acted as a carboxypeptidase. Remarkably, Ldt_{Cd2} was the only enzyme with endopeptidase activity, exclusively using 3-3 dimers as a substrate. Unlike other endopeptidases that cleave 3-3 cross-links (MepA, MepM and MepK), Ldt_{Cd2} has a strict substrate specificity for 3-3 cross-links since no activity against 3-4 cross-links can be detected (Chodisetti and Reddy, 2019; Voedts et al., 2021)(Table S4-3). Ldt_{Cd2} activity therefore appears to be unique.

	3-4 dimer	3-4 dimer + Ldt2
gm-AEJA=gm-AEJA	99.963%	0.093%
gm-AEJ=gm-AEJA - H ₂ O	0	86.014%
gm-AEJ=gm-AEJA	0	13.861%
gm-AEJ	0	0.032% ^b
gm-AEJA	0.037%	0
	100.0%	100.0%

Table S4- 3 . Quantification of Ldt_{Cd2} endopeptidase activity against a 3-4 crosslinked dimer $^{\rm a}.$

^a Abundance was calculated based on Ion intensity.

^b The low abundance of gm-AEJ corresponds to carboxypeptidase activity on traces of gm-AEJA present in the substrate

 Ldt_{Cd3} displayed the highest transpeptidase/exchange activity and a relatively weak carboxypeptidase and endopeptidase activity, preferentially against 3-3 cross-linked dimers. Unlike Ldt_{Cd2} , Ldt_{Cd3} could cleave 3-4 dimers with low efficacy.

Beside the exhaustive description of expected Ldt_{Cd} activities, our *in vitro* assays also revealed that Ldt_{Cd2} and Ldt_{Cd3} can generate a novel type of PG cross-link. These result from double transpeptidation reactions that use *m*DAP both as a donor and an acceptor. Interestingly, we identified double cross-linked dimers containing either two 3-3 cross-links or a mixture of 3-4 and 3-3 cross-links. In hindsight, this result is not entirely surprising since the catalytic reaction leading to this type of bond is the same as the reaction leading to the formation of "normal" 3-3 bonds. Double cross-links can be detected in the PG from *C. difficile* and other Gram-negative organisms producing Ldts (*E. coli, R. leguminosarum* and *B. abortus*; S. Mesnage, unpublished). Interestingly, double 3-4 cross-links resulting from D,D-transpeptidation have also been described in *S. aureus* (*Boneca et al., 1997*). The physiological role of this PG cross-link remains unknown and awaits further studies.

Based on the impact of L,D-transpeptidation on antibiotic resistance in *E. faecium*, it is tempting to assume that L,D-transpeptidation in *C. difficile* could underpin β-lactam resistance. This remains an open question since we were unable to generate a mutant devoid of 3-3 cross-links. The mutant harbouring deletions in the genes encoding the three canonical Ldts still contained 78% of 3-3 cross-links, indicating that this organism encodes (an)other enzyme(s) which does not contain a YkuD domain but is (are) able to make 3-3 cross-links. Our findings are surprising and somewhat contrasting with a previous study, where the combined deletion of Idt_{Cd1} and Idt_{Cd2} led to a decrease in the cross-linking index (18.2% as compared to 33.8% for the WT strain). This discrepancy is difficult to explain but could be attributed either to the different strain analysed (C. difficile 630 versus R20291) or to the different strategies followed for PG analysis. The work by Peltier et al. involved offline analysis of individual fractions from single replicates and a quantification of muropeptides based on UV whilst our analysis involved LC-MS (online) analysis of biological triplicates and a quantification using ion intensity. Based on the remarkably similar PG structure of the WT and mutant strain described here (Table 4-1 and Table 4-2), we are confident with the conclusion that the 3 canonical Ldt_{Cd} enzymes have a minor contribution to the formation of 3-3 crosslinks. The identification of alternative (non-canonical) Ldt(s) encoded by C. difficile will therefore be required to investigate (i) whether L,D-transpeptidation is essential in this organism and (ii) whether this activity underpins β -lactam resistance. The interaction of recombinant Ldts with β-lactams has been extensively studied in vitro and the data reported supports the idea that these enzymes play a role in resistance to these antibiotics. Ldts are acylated by β -lactams, but enzyme inactivation only occurs in the presence of carbapenems and penems. Other β-lactams such as cephems (cephalosporin) are poor inhibitors since acylation is slow and the thioester bond formed in the enzyme-antibiotic adduct is prone to hydrolysis (Triboulet et al., 2013). This has been shown for model organisms including E. faecium (Triboulet et al., 2013), M. tuberculosis (Cordillot et al., 2013) and C. difficile Ldts (Sütterlin et al., 2018) and is true for most Ldts studied to date despite some exceptions for Mycobacterium tuberculosis Ldt_{Mt5} (Cordillot et al., 2013), Acintetobacter baumanii Ldt_{Ab} (Toth et al., 2022) and C. difficile Ldt_{Cd3} (Sütterlin et al., 2018) which are not inhibited by carbapenems. The contribution of Ldts to β -lactam resistance has been documented in E. faecium (Mainardi et al., 2000), M. tuberculosis (Gupta et al., 2010), M. smegmatis (Baranowski et al., 2018b) and A. baumanii (Toth et al., 2022). In C. difficile, the inactivation of two of the three Ldts did not lead to a change in β-lactam resistance (Peltier et al., 2011). Our data revealed that the inactivation of all canonical C. difficile ldts has no impact on β-lactam resistance, as expected, based on the results from PG analysis.

4.5 Experimental procedures

4.5.1 Bacterial strains, plasmids and growth conditions

Bacterial strains, plasmids, and oligonucleotides are described in Table S4. *C. difficile* R20291 (ribotype 027) and isogenic derivatives were grown on BHI agar plates or in TY broth. During selection of mutants, strains were grown on *C. difficile* minimal medium (21) supplemented with 5-fluorocytosine (50 μ g/mL) when required. Cultures were incubated at 37 °C in an anaerobic cabinet under an atmosphere containing 80% nitrogen, 10% hydrogen, and 10% carbon dioxide. *E. coli* was grown on Luria Bertani (LB) agar plates or in LB broth at 37 °C. When needed, thiamphenicol was added (30 μ g/mL).

Strains/plasmids/	Relevant properties/sequence	Source
oligonucleotides		
Strains		
Clostridioides difficile		
R20291	Clinical isolate, ribotype 027	(1)
R20291∆ldt1	R20291 derivative with an in-frame deletion in ldt _{Cd1}	This work
R20291Δldt2	R20291 derivative with an in-frame deletion in ldtca2	This work
R20291Δldt3	R20291 derivative with an in-frame deletion in ldt _{Cd3}	This work
$R20291\Delta ldt 2\Delta ldt I$	R20291 derivative with an in-frame deletion in ldt _{Cd2} and ldt _{Cd1}	This work
R20291Aldt3Aldt1	R20291 derivative with an in-frame deletion in ldt _{Cd3} and ldt _{Cd1}	This work
R20291∆ldt3∆ldt2	R20291 derivative with an in-frame deletion in ldt _{Cd3} and ldt _{Cd2}	This work
$R20291\Delta ldt 2\Delta ldt 1\Delta ldt 3$	R20291 derivative with an in-frame deletion in ldtcd2, ldtcd1 and ldtcd3	This work
R20291∆Paloc	R20291 derivative with a deletion in the genes encoding toxins	(2)
Escherichia coli		
NEB5alpha	Cloning strain	NEB
BL21(DE3)	Expression strain	NEB
CA434	HB101 derivative carrying R702	(3)
Plasmids		
pJAK112	Plasmid for gene replacement in C. difficile	(4)
pNG007	pJAK derivative containing homology regions flanking a 1332bp ldt _{Cd1}	This work
pNG008	pJAK derivative containing homology regions flanking a 1695bp ldt _{Cd2}	This work
pNG009	pJAK derivative containing homology regions flanking a 753bp ldt _{Cd3}	This work
pET2818	pET derivative for expression of His-tagged recombinant proteins	(5)
pET-Ldt1	pET2818 derivative expressing Ldt _{Cdl} (residues 3-289; C-terminal His-tag)	This work
pET-Ldt2	pET2818 derivative expressing Ldt _{Cdl} (residues 41-164; N-terminal His-tag)	This work
pET-Ldt3	pET2818 derivative expressing Ldt _{Cd1} (residues 38-469; N-terminal His-tag)	This work
Oligonucleotides		
RF_21 (pJAK112)	GGATTTCACATTTGCCGTTTTGTAAAC	
RF_22 (pJAK112)	GATCTTTTCTACGGGGTCTGAC	
RF1795 (ldt _{Cd1} locus)	GCTTTTACTTTGATACTGTCTGCTG	
RF1796 (ldt _{Cd1} locus)	ATGACAAACTTAAGGAAAGATGGCC	
RF1797 (ldt _{Cd2} locus)	GGATTCAGTTCCTGAATAACTAGGT	
RF1798 (ldt _{Cd2} locus)	GCAGTTGGTGAATCAGTAGAAAAAC	
RF1799 (ldt _{Cd3} locus)	GGAGGAGATATTAGAGACTATGAAG	
RF1800 (ldt _{Cd3} locus)	CGTGTTTATGCACATCCAACTATG	

Table	S4-4	4 Bacteria	al strains	. plasmids	and	oligon	ucleotides	
Lant		I Datuelle	11 SUL allis	, prasinius,	, anu	ungun	uciconues	

- (1) Stabler, R. A., He, M., Dawson, L., Martin, M., Valiente, E., Corton, C., Lawley, T. D., Sebaihia, M., Quail, M. A., Rose, G., Gerding, D. A., Gibert, M., Popoff, M. R., Parkhill, J., Dougan, G. and Wren, B. W. (2009) Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. *Genome Biology* **10** R102
- (2) Buddle, J., Wright, R. C. T., Turner, C. E., Chaudhuri, R. R., Brockhurst, M. A. and Fagan, R. P. (2023) Multiple evolutionary pathways lead to vancomycin resistance in *Clostridioides difficile*. *BioRxiv*
- (3) Woods, C., Humphreys, C. M., Mesquita Rodrigues, R., Ingle, P., Rowe, P., Henstra, A. M., Köpke, M., Simpson, S. D., Winzer, K. and Minton, N. P. (2019) A novel conjugal donor strain for improved DNA transfer into *Clostridium* spp. *Anaerobe* 59:184-191
- (4) Fuchs, M., Lamm-Schmidt, V., Sulzer, J., Ponath, F., Jenniches, L., Kirk, J. A., Fagan, R. P., Barquist, L., Vogel, J. and Faber, F. (2021) An RNA-centric global view of *Clostridioides difficile* reveals broad activity of Hfq in a clinically important gram-positive bacterium. *Proc Natl Acad Sci USA* 118(25): e2103579118
- (5) Eckert, C., Lecerf, M., Dubost, L., Arthur, M. and Mesnage, S. (2006) Functional analysis of AtlA, the major N-acetylglucosaminidase of *Enterococcus faecalis*. J Bacteriol. 188(24):8513-8519

4.5.2 Construction of C. difficile deletion mutants

C. difficile mutant strains were constructed by homologous recombination. Briefly, 1.2 kb upstream and downstream of the region to be deleted was synthesized as a single DNA fragment (Genewiz) and cloned between BamHI and SacI sites in pJAK112, yielding pNG007 (Idt_{Cd1} deletion), pNG008 (Idt_{Cd2} deletion), and pNG009 (Idt_{Cd3} deletion). Plasmids were introduced into *C. difficile* strain R20291 by conjugation, and allelic exchange.

4.5.3 Determination of minimum inhibitory concentrations

MICs were determined according to an agar dilution method using Wilkins Chalgren agar and as recommended by the Clinical and Laboratory Standards Institute guidelines. *C. difficile* isolates were cultured on fresh blood agar plates, prior to inoculation of single colonies into pre-reduced Schaedler Anaerobic Broths and anaerobic culture for 24 h. Cultures were diluted in pre-reduced phosphate-buffered saline to achieve a 1 McFarland standard equivalent, and 10^5 colony-forming units were spotted on Wilkins Chalgren agar containing doubling antibiotic dilutions and non-antibiotic-containing controls. Agar plates containing amoxicillin clavulanate were prepared with a fixed concentration of 2 mg/l clavulanate, and those containing piperacillin tazobactam were prepared with a fixed concentration of a mg/l clavulanate, and those containing modeling antibiotic billy Testing guidelines. Agar plates were incubated anaerobically for 48 h before reading. The MIC was defined as the lowest concentration of antibiotic completely preventing growth, significantly reducing it to a haze or one to three discrete colonies.

4.5.4 Chromosomal DNA extraction, sequencing, and genome analysis

Genomic DNA was purified using phenol–chloroform extraction, and whole genome sequencing was performed by MicrobesNG using their standard Illumina service. Sequence analysis was performed using a custom script. In brief, reads were aligned to the *C. difficile* R20291 reference (accession number: FN545816) using BWA-mem (v0.7.17) and sorted using SAMtools (v1.43). PCR duplicates were removed *via* Picard (v2.25.2) (http://broadinstitute.github.io/picard/). SAMtools (v1.43) mpileup was used to generate the mpileup prerequisite for Varscan. Varscan (v2.4.3-1) was then used to call variants, and snpEff (v5.0) was used to annotate variants. Variants that co-occurred in the WT were removed to generate a list of mutations unique to mutant strains. Mutations were visualized on the genome using a previously published custom script in RStudio (v4.1.0) using the Plotrix package.

4.5.5 TraDIS analysis

The construction of the transposon library, the sequencing of insertion sites, and the mapping to their corresponding reference sequences were done. Visualization of insertion sites was done using the Artemis genome browser.

4.5.6 Peptidoglycan extraction

C. difficile strains were grown overnight in 10 mL of TY broth from a single colony. The starter cultures were used to inoculate 100 mL TY medium (1/100 dilution). After 48 h at 37 °C, cells were spun, supernatant was discarded, and cell pellet was snap frozen in liquid nitrogen; the cell pellet was then resuspended in 20 mL of boiling MilliQ water (MQ) before the addition of 5 mL warm 20% (w/v) SDS (4% SDS final concentration). After 30 min at 100 °C, the cells were allowed to cool down to room temperature. Insoluble cell walls were pelleted at 45,000g for 20 min and washed 5 times using warm MQ water. Proteins covalently bound to PG were removed by pronase treatment

(final concentration of 2 mg/mL, 4 h at 60 °C). Protease-treated cell walls were washed 6 times with 30 mL of MQ water before covalently bound polymers were removed by incubation in 1 M HCl for 5 h at 37 °C. Insoluble pure PG was washed 6 times with MQ water, snap frozen in liquid nitrogen, freeze-dried and resuspended at a final concentration of 10 mg/mL.

$4.5.7 Ldt_{Cd}$ production and purification

The plasmids for protein production were designed and Ldt_{Cd1} and Ldt_{Cd3} were expressed as full-length His-tagged proteins. Ldt_{Cd2} could not be produced as a stable full-length protein, so the catalytic domain was purified. Recombinant Ldt_{Cd} were produced in E. coli BL21(DE3) grown in LB broth. One-liter cultures were inoculated at an OD_{600} of 0.05, and protein expression was induced with 1 mM isopropyl β -D-1thiogalactopyranoside when the cultures reached an OD₆₀₀ of 0.7. They were then cooled down to 20 °C and incubated for 16 h at this temperature. Cells were harvested, resuspended in a buffer containing 50 mM Tris-HCI (pH8.0) + 500 mM NaCl, and mechanically broken using a French press (2 passages at 1250 psi). Cell debris were removed by centrifuging the crude cell extract at 45,000g for 30 min at 4 °C. The entire soluble fraction was loaded on a 5-mL HiTrap column equilibrated in buffer A at a flow rate of 5 mL/min. Elution was performed using a 10 column volume gradient to 250 mM imidazole in buffer A. Fractions containing the Ldt_{Cd} proteins were pooled, concentrated to 2 mg/mL and further purified by gel filtration chromatography using a Hiload 16/600 superdex 75 column equilibrated in 50 mM Tris-HCl pH 8.5 + 250 mM NaCl). Ldt_{Cd} proteins were concentrated on an Amicon centrifugal filter to a final concentration of 2 mg/mL and stored at -80 °C until further use.

4.5.8 Purification of substrates for in vitro assays

Peptidoglycan fragments used as substrates were purified from *E. coli* or *C. difficile* sacculi digested with mutanolysin and reduced with sodium borohydride. Digestion products were separated by reversed-phase HPLC using a Hypersil column (4.6 mm × 250 mm, 5 μ m particle size) using a water–acetonitrile–0.1% (v/v) formic acid gradient. Fractions containing the muropeptides of interest were freeze-dried and quantified by NMR using trimethylsilyl propionate as a standard.

4.5.9 In vitro Ldt assays

Each *in vitro* assay was carried out in triplicate, and average chromatograms are shown in Figure 4-1A. Each reaction was carried out in a phosphate saline buffer (pH 8.0) in a final volume of 50 μ L and contained 100 μ M substrate and 10 μ M enzyme. For exchange reactions, D-methionine was added at a concentration of 1 mM. Reactions were incubated at 37 °C for 4 h.

4.5.10 Preparation of soluble muropeptides for PG structural analysis

Purified PG (1 mg) was digested overnight in 50 mM phosphate buffer (pH 5.5) supplemented with 200 U of mutanolysin (Sigma) in a final volume of 125 μ l. Following

heat inactivation of mutanolysin (5 min at 100 °C), soluble disaccharide peptides were mixed with an equal volume of 250 mM borate buffer (pH 9.25) and reduced with 0.2% (w/v) sodium borohydride. After 20 min at room temperature, the pH was adjusted to 4.5 to 5.5 using phosphoric acid.

4.5.11 Ultrahigh-Performance chromatography coupled to tandem mass spectrometry

An Ultimate 3000 UHPLC (Dionex/Thermo Fisher Scientific) system coupled with a high-resolution Q Exactive Focus mass spectrometer (Thermo Fisher Scientific) was used for LC-MS analysis. Muropeptides were separated using a C18 analytical column (Hypersil Gold aQ, 1.9- μ m particles, 150 mm × 2.1 mm; Thermo Fisher Scientific) at a temperature of 50 °C for PG analysis or on a smaller C18 column for *in vitro* assays (Hypersil Gold aQ, 1.9- μ m particles, 50 mm × 2.1 mm; Thermo Fisher Scientific). For PG analysis, muropeptide elution was performed at 0.25 mL/min by applying a mixture of solvent A (water, 0.1% [v/v] formic acid) and solvent B (acetonitrile, 0.1% [v/v] formic acid). Liquid chromatography conditions were 0 to 12.5% B for 25 min increasing to 20% B for 10 min. After 5 min at 95%, the column was re-equilibrated for 10 min with 100% buffer A. For *in vitro* assays, a flow rate of 0.4 mL/min was used. PG fragments were eluted with a 5-min gradient to 15% B followed by 2 min at 95% B. The column was re-equilibrated for 6 min with 100% buffer A.

The Orbitrap Exploris 240 was operated under electrospray ionization (H-ESI high flow)-positive mode, full scan (m/z 150–2250) at resolution 120,000 (full width at half maximum) at m/z 200, with normalized AGC Target 100%, and automated maximum ion injection time. Data-dependent MS/MS were acquired on a "Top 5" data-dependent mode using the following parameters: resolution 30,000; AGC 100%, automated injection time, with normalized collision energy 25%.

4.5.12 Nuclear magnetic resonance

Purified PG fragments were dissolved in 90% $H_2O/10\%$ D_2O . They were analyzed by NMR at 298 K on a Bruker DRX-600 equipped with a cryoprobe. TOCSY spectra were acquired using a 60-ms spin-lock with a field strength of 10 kHz. NOESY spectra used a 200-ms mixing time. All data were analyzed using Topspin 4.0.5.

4.5.13 Analysis of PG structure

LC-MS datasets were deconvoluted with the Byos® software v3.11 (Protein Metrics). Sequential searches were carried out with PGFinder v1.0.3, with default settings (10 ppm tolerance, 0.5 min cleanup window) following the strategy described in Figure S4-3. Data from individual matching output was consolidated as previously described to calculate average intensities, retention times, observed monoisotopic masses, and ppm differences. The output from individual searches and consolidated data are described. Cross-linking index and glycan chain length were determined: The cross-linking index is defined as 0.5 * (% of dimers) + 0.66 * (% of trimers); glycan chain

length was inferred from the abundance of anhydroMurNAc groups, which are found at the end of glycan chains. It is defined as 1/(% of AnhydroMurNAc monomers + 0.5 * (% of AnhydroMurNAc dimers) + 0.33 * (% of AnhydroMurNAc trimers).

4.5.14 Flow cytometry

Cells corresponding to biological replicates were grown overnight, diluted 1:100 into fresh broth ($OD_{600} \sim 0.02$), and grown to mid-exponential phase ($OD_{600} \sim 0.5$). Bacteria were diluted 1:100 in filtered phosphate-buffered saline and analyzed by flow cytometry using Millipore Guava Easy Cyte system. Light scatter data were obtained with logarithmic amplifiers for 2500 events. Forward scattered and side-scattered light values were compared using Student *t* test with Welch's correction using GraphPad Prism.

4.5.15 Sporulation

Briefly, stationary phase cultures of *C. difficile* were incubated anaerobically for 5 days and the total and heat-resistant (spore) colony-forming units (65 °C for 30 min) were determined every 24 h. Strains were assayed in technical triplicate and the data presented as the mean and standard deviation.

4.5.16 Toxin release assays

Toxin production was detected in whole cell lysates or concentrated culture supernatants by Western blot. For both fractions, material corresponding to the equivalent of 20 mL of culture at OD_{600} = 1 was loaded onto a 6% SDS PAGE, transferred on a polyvinylidene fluoride membrane, and probed with a mouse monoclonal antibody (MA1-7413, Thermo Fisher) against toxin B at a 1/1000 dilution. A secondary rabbit anti-mouse antibody coupled to horseradish peroxidase (#31450, Thermo Fisher) was used at a 1/10,000 dilution. Blots were revealed by chemiluminescence using a BioRad Chemidoc system.

4.6 Concluding remarks

My contribution to this paper shed light on the function of the canonical Ldt_{Cd} during cell growth and *in vitro*. Our results revealed that two of the three canonical Ldts (Ldt_{Cd2} and Ldt_{Cd3}) catalyse a novel type of PG crosslink. Unexpectedly, deleting all three *ldt* genes had only a minor effect on PG structure and no impact on β -lactam resistance. This suggested the presence of other, non-canonical Ldts responsible for forming 3-3 crosslinks in *C. difficile*.

Most of my work involved MS/MS analysis of muropeptides to confirm their structures. *In vitro* assays with recombinant Ldts shown transpeptidation products with masses that did not match any expected value. Our classical MS/MS approach, which relies on identifying theoretical fragment ions, was not applicable. Instead, we had to look for fragment ions from the closest structural analogue (based on retention time or mass). This approach led to the identification of product ions that corresponded to a doubly crosslinked mDAP residues, lacking the mass of a water molecule compared to the normal transpeptidation product. This strategy represents a valuable precedent for identifying novel PG fragments. By analyzing extra or missing masses compared to known muropeptides, we can infer novel structures. I also contributed to confirming the structure of dimers by MS/MS. This process involved using a newly developed fragment predictor module, which I helped beta-test. Since the software could not predict fragmentation of modified muropeptides at the time, I had to adapt the fragment lists to include *C. difficile* deacetylated GlcNAc residues, by subtracting the mass of an acetyl group (42.0106 Da).

After the publication of our manuscript, two other enzymes with L,D transpeptidase activity in the genome of *C. difficile* were identified by bioinformatic analyses and described (Bollinger et al., 2024). These enzymes can perform 3-3 crosslinking but have a VanW domain (PF04294), distinct from the canonical YkuD domain (PF03734) found in the previously studied Ldt_{Cd1}, Ldt_{Cd2} and Ldt_{Cd3} (Kaus et al., 2020; Kirk et al., 2017; Peltier et al., 2011; Sütterlin et al., 2018). Genetic experiments revealed that at least one Ldt is required for *C. difficile* viability (either Ldt_{Cd1}, Ldt_{Cd4} or Ldt_{Cd5}).

We now know that the *C. difficile* genome encodes up to five L,D-transpeptidases (Ldts): three with YkuD catalytic domain and two with VanW catalytic domains. All five contribute to the formation of 3-3 crosslinks in *C. difficile* PG. These Ldts exhibit functional redundancy; deleting one or two does not have a significant impact on the viability or the amount of 3-3 crosslinks. However, deleting all five Ldts leads to cell death, confirming the essential role of L,D-transpeptidation in *C. difficile* survival. Nevertheless, their absence does not significantly impact β -lactam resistance because β -lactams primarily target D,D-carboxypeptidases, which are not essential for 3-3 crosslink formation in the absence of canonical Ldts (Paiva et al., 2024).

Ldts play diverse roles beyond 3-3 crosslink formation, including the incorporation of non-canonical D-amino acids into PG, contributing to β -lactam resistance, maintaining cell envelope integrity through L,D transpeptidation, and tethering outer membrane proteins to PG (Aliashkevich and Cava, 2021; Alvarez et al., 2024). The essentiality of Ldts appears to be species-specific and influenced by factors such as the number of *ldt* genes and growth conditions (stressors). 3-3 crosslinks are essential for the

survival of *C. difficile* and *A. tumefaciens* which encode 5 and 14 *ldt* genes, respectively (Aliashkevich et al., 2024; Bollinger et al., 2024).

Regarding the activities of individual *C. difficile* Ldts, we showed that Ldts with YkuD domains have distinct *in vitro* activities: Ldt_{Cd1} and Ldt_{Cd2} primarily act as carboxypeptidases, while Ldt_{Cd3} exhibits stronger transpeptidase/exchange activity. Ldt_{Cd4} and Ldt_{Cd5} , with VanW domains, are reported to be essential for viability and exclusively catalyze 3-3 crosslinks formation *in vivo* (Bollinger et al., 2024). We also described a novel type of PG crosslink catalyzed by Ldt_{Cd2} and Ldt_{Cd3} , involving double transpeptidation reactions using *m*DAP as both donor and acceptor. We observed these products both *in vivo* (as dimers and trimers in Table 4-1) and *in vitro* (as a dimer, a trimer and a tetramer in Figure 4-1E). As mentioned during the discussion, double transpeptidation products have already been reported for 3-4 transpeptidation (Boneca et al., 1997). Since we could detect traces of doubly crosslinked dimers in the *C. difficile* triple mutant (gm(-Ac)-AEJA=gm(-Ac)-AEJ(-H2O) in Table 4-1), it suggests that Ldts with VanW domains could also have this unusual activity.

Chapter 5

5. Identification and Characterization of an enzyme catalyzing the formation of 1-3 crosslinks in *Gluconobacter oxydans* PG

As part of my PhD thesis, I am including the following paper:

Alamán-Zárate, M. G., Rady, B. J., Evans, C. A., Pian, B., Greetham, D., Marecos-Ortiz, S., Dickman, M. J., Lidbury, I. D. E. A., Lovering, A. L., Barstow, B. M., & Mesnage, S. (2024). Unusual 1-3 peptidoglycan crosslinks in *Acetobacteraceae* are made by *L,D*transpeptidases with a catalytic domain distantly related to YkuD domains. The Journal of Biological Chemistry, 300(1), 105494. <u>https://doi.org/10.1016/j.jbc.2023.105494</u>

The work was carried out in collaboration with the research groups of Prof Mark Dickman, Prof Andrew Lovering, Dr Ian Lidbury and Dr Buz Barstow. The *Gluconobacter oxydans* mutant strains were generated by Brooke Pian and Sabrina Marecos-Ortiz in Dr Buz Barstow Iab. The PG samples I analysed were prepared by myself and Darren Greetham from Dr Mesnage Lab. The mass spectrometry data acquisition was performed by Caroline Evans in Prof Dickman's facilities.

G. oxydans PG structure was analysed by me and Dr Mesnage, while the genomic and structural analyses was a collaborative effort among Brooks Rady, Dr Ian Lidbury, and Prof. Andrew Lovering.

My contribution has been the structural analysis of *G. oxydans* and *E. coli* peptidoglycan, the heterologous expression of Ldt_{Go2} , and the complementation of the Δldt_{Go2} transposon insertion mutant. The initial draft was written by me, Brooks Rady and Dr Mesnage and we also edited the text considering feedback from other co-authors. My contributions are displayed in the analysis and figures shown in Table 5-1, Figure 5-4, Figure 5-5, Table S5-1, Figure S5-2, Table S5-2, Fig S5-3, Table S5-3, and Table S5-4.

Following the PG analysis abilities developed for the previous chapter to identify novel crosslinks, we decided to test and develop new PGFinder capabilities for the analysis of uncommon 1-3 transpeptidation products.

Unusual 1-3 peptidoglycan cross-links in Acetobacteraceae are made by L,Dtranspeptidases with a catalytic domain distantly related to YkuD domains

Marcel G Alaman-Zarate^{1‡}, Brooks J Rady ^{1‡}, Caroline A Evans², Brooke Pian³, Darren Greetham¹, Sabrina Marecos-Ortiz³, Mark J Dickman², Ian DEA Lidbury¹, Andrew L Lovering⁴, Buz M Barstow³, and Stéphane Mesnage^{1*}

- ¹ Molecular Microbiology, Biochemistry to Disease, School of Biosciences, University of Sheffield, Sheffield, UK
- ² Department of Chemical and Biological Engineering, ChELSI Institute, University of Sheffield, Sheffield, UK
- ³ Department of Biological and Environmental Engineering, Cornell University, Ithaca, USA
- ⁴ School of Biosciences, University of Birmingham, Birmingham, UK
- * For correspondence: s.mesnage@sheffield.ac.uk

[†]These authors contributed equally to this work

Running title: 1-3 peptidoglycan cross-links in Acetobacteraceae

Keywords: *Gluconobacter oxydans,* peptidoglycan, L,D-transpeptidase, crosslink, bacterial cell envelope, PGFinder

5.1 Abstract

Peptidoglycan is an essential component of the bacterial cell envelope that contains glycan chains substituted by short peptide stems. Peptide stems are polymerized by D,D-transpeptidases, which make bonds between the amino acid in position 4 of a donor stem and the third residue of an acceptor stem (4-3 cross-links). Some bacterial peptidoglycans also contain 3-3 cross-links that are formed by another class of enzymes called L,D-transpeptidases. In this work, we investigate the formation of unusual bacterial 1-3 peptidoglycan cross-links. We describe a version of the PGFinder software which can identify 1-3 crosslinks and report the high-resolution peptidoglycan structure of Gluconobacter oxydans (a model organism within the Acetobacteraceae family). We reveal that G. oxydans peptidoglycan contains peptide stems made of a single alanine as well as several dipeptide stems with unusual amino acids at their C-terminus. Using a Sudoku transposon library, we identified a G. oxydans mutant with a drastic reduction in 1-3 crosslinks. Through complementation experiments in G. oxydans and recombinant protein production in a heterologous host, we identify an L,D-transpeptidase enzyme with a domain distantly related to the YkuD domain responsible for these non-canonical reactions. This work revisits the enzymatic capabilities of L,D-transpeptidases, a versatile family of enzymes that play a key role in bacterial peptidoglycan remodelling.

5.2 Introduction

Peptidoglycan is an essential component of the bacterial cell envelope that confers cell shape and resistance to a high internal osmotic pressure (Weidel and Pelzer, 1964). This bag-shaped macromolecule surrounding the cytoplasmic membrane is made of disaccharide-peptides as building blocks. Their polymerization forms glycan chains alternating *N*-acetylglucosamine and *N*-acetylmuramic acid residues, substituted by short pentapeptide stems containing L- and *D*-amino acids (Vollmer et al., 2008a). Depending on the bacterial species considered, the composition of peptidoglycan building block can vary (Vollmer et al., 2008a), but in most bacteria (including *Escherichia coli*), pentapeptide stems are made of the sequence L-Ala-iso *D*-Glu-*meso*-DAP-*D*-Ala-*D*-Ala, (where mDAP is diaminopimelic acid).

The polymerization of peptidoglycan has been extensively studied since the late 50's, when it was discovered that this process is inhibited by penicillin, β -lactam antibiotics widely used to combat infections (Hahn and Ciak, 1957; Park and Strominger, 1957). The ubiquitous enzymes that polymerize peptidoglycan, D,D-transpeptidases, are also called Penicillin Binding Proteins (PBPs). They recognize the C-terminal D-Ala-D-Ala extremity of a donor peptide stem, form an acyl-enzyme intermediate with the diamino acid in position 4, and then link this residue to the side-chain amino group of the amino acid in position 3 of an acceptor stem (4-3 crosslink). B-lactams are structural analogues of the D-Ala-D-Ala stems and can be used as suicide substrates (Tipper and Strominger, 1965), leading to growth arrest and cell death (Cho et al., 2014). Alternative 3-3 peptidoglycan crosslinks were originally described in Mycobacteria (Wietzerbin et al., 1974). These types of bonds are prevalent in the peptidoglycan of important pathogens such as *Mycobacterium tuberculosis* (Lavollay et al., 2008), Mycobacterium leprae (Mahapatra et al., 2008) and Clostridium difficile (Peltier et al., 2011). In Enterococcus faecium, resistance to β-lactams and glycopeptides can emerge when 4-3 crosslinks are replaced by 3-3 crosslinks. The complete bypass of the D,D-transpeptidation pathway in E. faecium led to the identification of the enzyme catalyzing the formation of 3-3 bonds (Mainardi et al., 2005b) which is an L,D-transpeptidase. Instead of recognizing the D-Ala-D-Ala extremity of the pentapeptide donor stem, L,D-transpeptidases use a tetrapeptide stem as a substrate, which is generated by the activity of D,D-carboxypeptidases. These enzymes can perform several activities depending on the substrate they use as an acceptor and can act as a carboxypeptidase (cleaving the fourth residue of the donor stem) (Kim et al., 2015a), a transpeptidase (forming 3-3 crosslinked muropeptides or covalently anchoring proteins to peptidoglycan) (Godessart et al., 2021; Sandoz et al., 2021a), or an endopeptidase (cleaving 3-3 crosslinks or the link between peptidoglycan and covalently attached proteins) (Bahadur et al., 2021; Galley et al., 2024; Winkle et al., 2021). Finally, L,D-transpeptidases can also exchange the fourth amino-acid of a peptide stem for another amino-acid (Bern et al., 2017; Sütterlin et al., 2018). The peptidoglycan structural changes catalyzed by *L*,*D*-transpeptidases (called remodelling) plays an important role in cell shape (Kim et al., 2015a), resistance to abiotic stress (More, 2019), pathogenesis, and host immunity (Hernández et al., 2022). A recent study described the existence of peptidoglycan 1-3 crosslinks in *Acetobacteraceae* and proposed that this unusual type of crosslink could play a role for the survival of these organisms in the context of their interaction with the fly immune system and during competition with other organisms (Espaillat et al., 2016). In this work, we describe a version of PGFinder that can automate the analysis of peptidoglycans with 1-3 crosslinks. Using this tool, we determined the high-resolution peptidoglycan structure of *Gluconobacter oxydans* and revealed that it contains a high proportion of previously undescribed disaccharide-dipeptides with non-canonical amino acids at their C-terminus. Using a transposon mutant and its complemented derivative, as well as heterologous expression experiments, we demonstrate that *G. oxydans* 1-3 crosslinks are formed by an enzyme with a domain distantly related to the YkuD domain of canonical *L,D*-transpeptidases. Collectively, our data show that *L,D*-transpeptidases have evolved to carry out enzymatic reactions using either tetrapeptide or dipeptide stems as donors.

5.3 Results

5.3.1 Building a software tool for the structural analysis of 1-3 crosslinked peptidoglycans

Prior to this study, the PGFinder software (v1.0.3; <u>https://mesnage-org.github.io/pgfinder/</u>) could only generate dimers and trimers crosslinked via 3-3 and 4-3 bonds (Patel et al., 2021). To perform the structural analysis of *G. oxydans* peptidoglycan, we modified PGFinder and its graphical user interface to enable the creation of dynamic databases containing dimers and trimers with 1-3 crosslinks (Fig. 5-1). This PGFinder upgrade (v1.1.0) was tested using datasets from *G. oxydans*.

Upload a file or drag and drop Byos (.ftrs) or MaxQuant (.txt)						
Mass Database						
Built-In Custom						
Upload a file or drag and drop PGFinder Mass Library (.csv)						
Advanced Options						
Modifications PPM Tolera	nce					
3-3 and 3-4 Cross-Linked 10 Multimers (=)						
3-1 Cross-Linked Multimers (=) Cleanup Wind	ow 🚯					
Glycosidic Multimers (-) 0.5						
Lactyl Multimers (=Lac) Consolidation	ррм 🚯					
Anhydro-MurNAc (Anh) 1						
Run Analysis						

Figure 5- 1. Analysis of 3-1 crosslinked peptidoglycans using PGFinder v1.1.0. The advanced option command enables of dynamic databases containing 3-1 crosslinked dimers and trimers.

5.3.2 High resolution analysis of *G. oxydans* B58 peptidoglycan

Peptidoglycan was purified from *G. oxydans* B58 cells harvested during both exponential (Fig. 5-2A) and stationary phase (Fig. 5-2B). As expected, the muropeptide profiles revealed changes indicative of major peptidoglycan remodelling during stationary phase. We used a combination of automated tools previously described to determine the high-resolution structure of *G. oxydans* peptidoglycan (Bern et al., 2017; Patel et al., 2021). A two-step custom search strategy was followed (Fig. S5-1). We first used the proprietary ByonicTM software to identify monomers based on tandem mass spectrometry data. The search space contained mono-, di-, tri-, tetra- and pentapeptides containing Alanine in position 1, glutamic acid (E) or glutamine (Q) in position 2, meso-diaminopimelic acid (J) or amidated meso-diaminopimelic acid (Z) in position 3, any possible amino acids (X) in position 4, and pentapeptides containing AX dipeptides at their C-terminus (Table S5-1).



Figure 5- 2. HPLC-MS chromatogram of *Gluconobacter oxydans* reduced disaccharide-peptides.

Strain B58 was grown in YPM media to exponential (**A**) or stationary phase (**B**). The numbers refer to the muropeptide structures described in Table 5-1.



Figure S5-1. Strategy for G. oxydans peptidoglycan structural analysis.

A first search was performed using the Byonic[™] module from the Byos[®] software to identify monomers based on MS/MS data. The search space contained disaccharide substituted by mono-, di-, tri-, tetra- and pentapeptides stems containing glutamic acid (E) or glutamine (Q) in position 2, meso-diaminopimelic acid (J) or amidated meso-diaminopimelic acid (Z) in position 3, any possible aminoacids (X) in position 4 or the AX dipeptides in positions 4 and 5. The 17 monomers identified by MS/MS were combined to generate the database DB_0. A second search was performed with PGFinder, enabling the formation of dynamic libraries containing dimers and trimers resulting from 4-3, 3-3 and 1-3 crosslinks as well as their derivatives containing deacetylated or anhydroMurNAc sugar moeities. The corresponding options are boxes in red in the graphic user interface

Seventeen monomers showing more than half of the expected b and y ions in their fragmentation spectra were identified by Byonic[™] (Fig. S5-2). Interestingly, these included several muropeptides with a dipeptide stem other than AE that were not previously identified and a lack of tetrapeptide stems with unusual amino acids formed by canonical L.D-transpeptidases. In addition to amidated meso-DAP residues, we also found the presence of deacetylated GlcNAc residues (glucosamine) which were not previously reported. The disaccharide-peptides corresponding to these validated monomers were combined to create the database called DB0 Go (Table S5-2). We next performed a PGFinder search, enabling the identification of dimers and trimers with 4-3, 3-3, and 1-3 crosslinks as well as modified disaccharides (deacetylated and containing MurNAc residues). A total of 61 masses matching the monoisotopic mass of theoretical structures were identified (Table 5-1), revealing a far more complex structure than previously reported. A direct comparison of the peptidoglycan from cells harvested during exponential and stationary phase showed an increased crosslinking index (19.9 % versus 15.9 %), partly due to an increased proportion of 1-3 crosslinks in the stationary phase (16.6% versus 5.8%). The higher proportion of 1-3 crosslinks was concomitant with the higher proportion of disaccharide-dipeptide structures detected in stationary phase (13.5% versus 5.7%). Very little variation was

observed in the glycan chain length between exponential and stationary phase, with the average length being equal to 24 and 22 disaccharides, respectively.

By analogy with the transpeptidation reaction leading to the formation of 3-3 bonds (Fig. 5-3A), we hypothesized that the formation of 1-3 bonds uses muropeptides with a dipeptide stem as donor substrates (Fig. 5-3B). According to this hypothesis, the enzyme is predicted to form an acyl enzyme intermediate with a disaccharide-alanine.



Figure 5-3. Schematic representation of the *L,D*-transpeptidation reactions leading to the formation of 3-3 and 1-3 crosslinks.

The enzymatic reactions carried out by *L*,*D*-transpeptidases in organisms with 3-3 crosslinks are described (**A**). By analogy wit these *L*,*D* transpeptidation reactions, we propose a model that leads to distinct reactions in *G. oxydans* (**B**). We hypothesize that an unidentified endopeptidase generates disaccharide dipeptides. These muropeptides are used as substrates to form an acyl-enzyme intermediate. Depending on the acceptor group, the reaction can lead to a carboxypeptidase reaction or a transpeptidation reaction that generates either a dimer or a disaccharide-dipeptide. GlcNAc, *N*-acetylglucosamine; MurNAc, *N*-acetylmuramic acid; LDT, *L*,*D*-transpeptidase; DS, disaccharide (GlcNAc-MurNAc); X, any D amino acid; DS-Tri, disaccharide-tripeptide; DS-Tetra, disaccharide-tetrapeptide.

Table S5-1. List of peptide sequences used for *G. oxydans* MS/MS searches

Sequence	Sequence	Sequence
A	AFIAY	AOIY
AF	AF7	
AC	Δ Ε 7 C	
AR		AQIAE
	ΔE7E	
	A E 7 E	
AD AO		
AQ	AEZG	
AU AU		
	AEZP	DALDA
AF	AEZQ	AQJAR
AP	AEZR	AQJAS
AS	AEZS	
AI	AEZI	AQJAV
AW	AEZV	AQJAW
AY	AEZW	AQJAY
AV	AEZY	AQZ
AEJ	AEZAA	AQZA
AEJA	AEZAC	AQZC
AEJC	AEZAD	AQZD
AEJD	AEZAE	AQZE
AEJE	AEZAF	AQZF
AEJF	AEZAG	AQZG
AEJG	AEZAH	AQZH
AEJH	AEZAI	AQZI
AEJI	AEZAK	AQZK
AEJK	AEZAM	AQZM
AEJM	AEZAN	AQZN
AEJN	AEZAP	AQZP
AEJP	AEZAQ	AQZQ
AEJQ	AEZAR	AQZR
AEJR	AEZAS	AQZS
AEJS	AEZAT	AQZT
AEJT	AEZAV	AQZV
AEJV	AEZAW	AQZW
AEJW	AEZAY	AQZY
AEJY	AQJ	AQZAA
AEJAA	AQJA	AQZAC
AEJAC	AQJC	AQZAD
AEJAD	AQJD	AQZAQ
AEJAE	AQJE	AQZAF
AEJAF	AQJF	AQZAG
AEJAG	AQJG	AQZAH
AEJAH	AQJH	AQZAI
AEJAI	AQJI	AQZAK
AEJAK	AQJK	AQZAM
AEJAM	AQJM	AQZAN
AEJAN	AQJN	AQZAP
AEJAP	AQJP	AQZAQ
AEJAQ	QLQA	AQZAR
AEJAR	AQJR	AQZAS
AEJAS	AQJS	AQZAT
AEJAT	AQJT	AQZAV
AEJAV	VLDA	AQZAW
AEJAW	WLDA	AQZAY

mDAP (J=172.0848) amidated mDAP (Z=171.1008)





Figure S5-2. Tandem mass spectrometry analysis of *G. oxydans* monomers using the Byonic[™] module.

Table 5-1. <i>C</i>	<i>i. oxydans</i>	peptidoglycan	composition
----------------------------	-------------------	---------------	-------------

	Numer entide à	v	/т	عام ا	عام ا	BT (min)b		Theoretical	A	
	Muropeptide "	Ехро	Stat	Lat _{Go1}	Lat _{Go2}	KI	(mi	nj°	mass (Da)	∆ppm
1	gm-AEJ _{NH2}	36.519%	41.440%	48.001%	34.926%	5.32	±	0.05	869.3867	4.01
2	gm-AE	4.239%	10.518%	9.628%	6.887%	6.79	±	0.03	698.2859	2.77
3	gm-AEJA	22.572%	3.270%	4.019%	28.131%	6.80	±	0.01	941.4078	3.38
4	gm-AE (Anh)	0.880%	1.158%	0.656%	0.505%	11.06	±	0.00	678.2597	1.52
5	gm-AEJ _{NH2} (Anh)	0.522%	0.682%	0.472%	0.161%	8.61	±	0.01	849.3605	2.29
6	gm-AEJ _{NH} 2A	0.254%	0.290%	0.343%	0.012%	6.37	±	0.27	940.4238	1.39
7	gm-AEJ	0.711%	0.356%	0.193%	0.014%	5.66	±	0.04	870.3707	2.14
8	gm-AEJAG	0.494%	0.064%	0.130%	0.247%	6.49	±	0.02	998.4293	1.46
9	gm-AF	0.072%	0.171%	0.096%	ND	14.54	±	0.01	716.3117	0.89
10	gm-AEJAA	0.353%	0.074%	0.094%	0.067%	8.08	±	0.39	1012.4450	2.97
11	gm-A	0.335%	1.385%	0.534%	0.024%	6.26	±	0.03	569.2433	1.8
12	gm-AEJA (Anh)	0.425%	0.093%	0.079%	0.131%	10.42	±	0.01	921.3816	1.45
13	gm-AEJAK	0.114%	0.021%	0.042%	0.057%	5.86	±	0.03	1069.5028	1.07
14	gm-AY	0.032%	0.100%	0.042%	ND	10.72	±	0.01	732.3066	0.52
15	gm-Al	0.027%	0.070%	0.038%	ND	12.89	±	0.00	682.3274	0.81
16	gm-Al (Anh)	ND	0.029%	ND	ND	12.75			662.3012	0.99
17	gm-AEJAR	0.121%	0.028%	0.037%	0.062%	6.42	±	0.03	1097.5089	1.57
18	gm-AEJAH	0.100%	0.035%	0.035%	0.050%	5.99	±	0.03	1078.4667	1.41
19	gm-AEJAN	0.022%	0.069%	0.033%	0.038%	10.37	±	4.33	1055.4508	1.72
20	gm-AI (-Ac)	0.055%	ND	ND	ND	6.29			640.3169	0.24
21	gm-AEJ _{NH2} (-Ac)	0.036%	0.015%	0.023%	0.022%	5.05	±	0.06	827.3762	1.15
22	gm-AQ	0.057%	0.026%	0.022%	0.047%	5.77	±	0.04	697.3019	0.97
23	gm-AEJAE	0.029%	0.019%	0.018%	0.031%	7.29	±	0.01	1070.4504	0.50
24	gm-AEJ (Anh)	0.011%	ND	0.017%	ND	9.03	±	0.00	850.3445	1.15
25	gm-AEJA (-Ac)	0.021%	0.002%	0.007%	0.027%	6.43	±	0.02	899.3973	1.11
26	gm-AEJ _{NH2} A (Anh)	ND	ND	0.006%	ND	9.93			920.3976	1.17
27	gm-AE (-Ac)	ND	0.030%	0.003%	ND	5.62	±	0.01	656.2754	0.14
28	gm-AF (-Ac)	ND	ND	0.003%	ND	16.28			674.3012	1.24
29	gm-AEJ _{NH2} =gm-AEJA	13.782%	15.265%	18.938%	15.714%	8.43	±	0.00	1792.7836	3.35
30	gm-AEJ _{NH2} =gm-A	2.592%	13.391%	7.317%	0.037%	8.60	±	0.00	1420.6194	1.75
31	gm-AEJ _{NH2} =gm-AEJA (Ann)	3.741%	4.098%	4.220%	4.109%	10.87	±	0.00	1//2./5/4	2.38
32	gm-AEJA=gm-AEJA	6.802%	2.822%	2.318%	6.976%	9.59	±	0.00	1864.8047	2.51
33	gm-AEJ _{NH2} =gm-A (Ann)	0.243%	0.794%	0.609%	ND 0.0110/	11.43	±	0.01	1400.5932	0.85
34	gm-AEJA=gm-A	2.520%	2.028%	0.512%	0.011%	10.06	±	0.00	1492.6405	1.02
35	gm AEl-gm AElA	0.729%	0.366%	0.199%	0.001%	12.01	Ξ -	0.00	1844.7785	1.57
30 27	gm AELA-gm A (Anh)	0.100%	0.150%	0.127%		9.05		0.03	1/95.7070	0.00
20	gm-AEIA-giii-A (Aiii)	0.317%	0.197%	0.074%	0.078%	0.19	⊥ +	0.02	1472.0143	0.41
20	gm-AEJ-gm-AEJA (Apb)	0.134%	0.033%	0.049%	0.078%	11.46	- +	0.00	1772 7/1/	1 / 5
40	gm-AEJAG=gm-A	0.029%	0.023%	0.034%	0.013%	10.40	+	1 48	1549 6617	3 24
41	gm-AFI=gm-A	0.048%	0.145%	0.022%	ND	9 38	+	0.00	1421 6034	0.94
42	gm-AFIA=gm-AFIA (Anh)	ND	0.014%	ND	ND	9.84	-	0.00	1844 7786	0.09
43	gm-AEJAA=gm-AEJA	0.034%	0.009%	0.016%	0.023%	9.84	±	0.01	1935.8419	0.67
44	gm-AEJ _{NH2} A=gm-AEJA	0.007%	0.008%	0.012%	ND	9.32	±	0.04	1863.8207	0.70
45	gm-AEJAG=gm-AEJA (Anh)	0.032%	0.014%	0.010%	0.016%	11.53	±	0.01	1901.8000	1.14
46	gm-AEJAR=gm-AEJA	0.017%	ND	0.006%	0.007%	9.00	±	0.00	2020.9058	0.32
47	gm-AEJAN=gm-AEJA	ND	ND	0.005%	0.004%	8.76	±	0.01	1978.8477	0.81
48	gm-AEJ _{NH2} A=gm-AEJA (Anh)	ND	ND	0.004%	ND	11.56			1843.7945	0.37
49	gm-AEJ _{NH2} =gm-AEJA (-Ac)	0.005%	ND	0.003%	0.004%	8.18	±	0.03	1750.7731	2.03
50	gm-AEJAH=gm-AEJA	0.013%	ND	ND	0.006%	8.64	±	0.00	2001.8636	0.74
51	gm-AEJ _{NH2} A=gm-A		ND	ND	0.005%	16.44			1491.6565	0.55
52	gm-AEJAA=gm-AEJA (Anh)	0.006%	ND	0.002%	0.004%	12.12	±	0.01	1915.8157	0.63
53	gm-AEJAG=gm-A (Anh)	0.005%	ND	0.002%	ND	9.71	±	2.72	1529.6355	1.74
54	gm-AEJAR=gm-AEJA (Anh)	0.004%	ND	ND	ND	11.24			2000.8796	0.61
55	gm-AEJ _{NH2} =gm-AEJA=gm-AEJA	0.370%	0.384%	0.551%	0.554%	9.97	±	0.00	2716.1805	0.96
56	gm-AEJ _{NH2} =gm-AEJA=gm-AEJA (Anh)	0.180%	0.196%	0.210%	0.229%	11.84	±	0.00	2696.1543	0.78
57	gm-AEJA=gm-AEJA=gm-AEJA	0.099%	0.045%	0.119%	0.136%	10.84	±	0.00	2788.2016	1.89
58	gm-AEJA=gm-AEJA=gm-AEJA (Anh)	0.032%	0.015%	0.024%	0.033%	12.67	±	0.03	2768.1754	0.28
59	gm-AEJ=gm-AEJA=gm-AEJA	ND	ND	0.019%	ND	10.59	±	0.00	2717.1645	1.18
60	gm-AEJ=gm-AEJA=gm-AEJA (Anh)	ND	ND	0.005%	ND	12.51			2697.1383	0.52
61	gm-AEJAG=gm-AEJA=gm-AEJA	0.005%	ND	0.004%	0.003%	10.49	±	0.01	2845.2231	0.53

^a g, GlcNAc; m, MurNAc; A, Alanine; E, isoglutamic acid; J, *meso*-diaminopimelic acid; J_{NH2}, amidated meso-diaminopimelic acid
 ^b standard deviations in bold are determined from 2 values only

Structure	Monoisotopicmass
gm-A 1	569.24331
gm-AE 1	698.2859
gm-AF 1	716.31172
gm-Al 1	682.32737
gm-AQ 1	697.30189
gm-AY 1	732.30664
gm-AEJ 1	870.37069
gm-AEZ 1	869.38668
gm-AEJA 1	941.40783
gm-AEZA 1	940.42382
gm-AEJAA 1	1012.44497
gm-AEJAE 1	1070.45042
gm-AEJAG 1	998.42929
gm-AEJAH 1	1078.46674
gm-AEJAK 1	1069.50279
gm-AEJAN 1	1055.45076
gm-AEJAR 1	1097.50894

 Table S5- 2. Databases for G. oxydans PGFinder searches

5.3.3 Identification of the L,D-transpeptidase catalyzing the formation of 1-3 crosslinks in G. oxydans

Interestingly, G. oxydans peptidoglycan does not contain any tetrapeptide stems with unusual amino acids at their C-terminus, which are characteristic of canonical L,Dtranspeptidase enzymatic activity (Fig. 5-3A). We hypothesized that in G. oxydans, L,Dtranspeptidases could perform a similar enzymatic reaction using disaccharidedipeptides as substrates instead (Fig. 5-3B). We therefore searched the genome to identify genes encoding homologs of the L,D-transpeptidases. We found that the G. oxydans genome encodes two putative L,D-transpeptidases (labelled GOX1074, 337 residues and GOX2269, 171 residues in G. oxydans 621H) that are related to the YkuD catalytic domain (PF03734). We hypothesized that one or both enzymes could catalyze the formation of 1-3 crosslinks and renamed these putative L,D-transpeptidases Ldt_{Go1} and Ldt_{Go2}. To test this hypothesis, we took advantage of the *G. oxydans* B58 Sudoku library previously described (Schmitz et al., 2021) and analyzed the peptidoglycan structure of the two transposon mutants with an insertion in each of the Idt_{Go} genes by LC-MS. Comparison of the TIC profiles indicated the presence of a peak corresponding to the major 1-3 dimer (gm-AEJNH₂=gm-A) in the wild type (Fig. 5-4A) and Idt_{Go1} mutant (Fig. 5-4B), whilst no equivalent peak was detected in the Idt_{Go2} mutant (Fig. 5-4C).





TIC of *G. oxydans* B58 (**A**) and mutants with a transposon insertion in the ldtGo1 (**B**) and ldtGo2 genes (**C**) are shown on the left-hand side. Extracted ions corresponding to the muropeptides eluted between 7.5 and 11.5 min are shown on the right-hand side. The major dimer with a 1-3 crosslink (shown with an arrow on the TIC and on the top MS spectrum) is associated with two major protonated ions: a singly charged ion with an *m/z* at 1421.32 and a doubly charged ion with an *m/z* at 711.32. None of these ions were detected in the peptidoglycan from the *ldt*_{Go2} mutant, demonstrating that this gene is essential for the formation of 1-3 crosslinks.

Analysis of the extracted ion chromatograms for all molecules eluted between 7.5 and 11.5 min revealed a drastic reduction of 1-3 crosslinks in the Idt_{Go2} peptidoglycan sample (0.035% versus 7.3% in the Idt_{Go1} mutant). The Idt_{Go2} mutation was also associated with a reduction of disaccharide-dipeptides and 1-3 crosslinked dimers as compared to the parental strain and the Idt_{Go1} mutant (Table 5-1). Collectively, our LC-MS data showed that Ldt_{Go2} plays a major role in the unusual *L*,*D*-transpeptidation reactions in *G. oxydans*, including the formation of 1-3 crosslinks.

5.3.4 Complementation and heterologous expression experiments show that the LdtGo2 enzyme is sufficient to catalyze peptidoglycan 1-3 crosslinks

To verify that the drastic reduction of 1-3 crosslinks was associated with the disruption of Idt_{Go2} and not a secondary mutation, we built a complementation strain expressing Ldt_{Go2} under the anhydrotetracycline-inducible promoter (Fricke et al., 2021). The production of Ldt_{Go2} in the Idt_{Go2} transposon mutant background clearly restored the presence of a peak corresponding to the major 1-3 crosslinked dimers (Fig. S5-3).



Figure S5- 3. Complementation of the Idt_{Go2} **transposon insertion restores 1-3 crosslinks.** Wild-type train B58 (A), the Idt_{Go2} insertion mutant (B) and the complemented mutant (C) were grown in YPM media. The expression of Idt_{Go2} in the complemented strain was induced at $0D_{600}=0.5$ with 100 ng/mL anhydrotetracycline. The peaks corresponding to the major species are labelled. Numbers refer to the muropeptides described in Table 5-1.

We further confirmed the enzymatic activity of Ldt_{Go2} by producing the full-length protein in *E. coli*. Since *E. coli* peptidoglycan contains disaccharide-dipeptides (gm-AE) that represent the proposed substrate for the 1-3 transpeptidation reaction, we anticipated that recombinant Ldt_{Go2} could generate the expected products found in *G. oxydans* in this heterologous host. Peptidoglycan was purified from *E. coli* transformed with either the empty pET expression vector or a recombinant derivative expressing Idt_{Go2} , digested with mutanolysin, and analysed by reversed-phase HPLC (Fig. 5-5A, top and bottom trace, respectively). A simple search strategy was followed to identify and quantify muropeptides resulting from unusual *L,D*-transpeptidation reactions (Fig. S5-4).

First, we identified monomers based on MS/MS data using the Byonic[™] module from Byos[®]. We searched for all possible disaccharide-peptides containing one to five amino acids (A, AX, AEJ, AEJX and AEJAX, where J is meso-diaminopimelic acid and X any amino acid) adding sugar deacetylation previously identified in E. coli (Patel et al., 2021) as a potential glycan modification (Table S5-3). Twenty-eight monomers validated by MS/MS analysis were selected to create a database called DB0 Ec (Table S5-4). This monomer database was then run through PGFinder to identify, compare, and quantify muropeptides in the E. coli expression strain and its derivative expressing Ldt_{Go2}. To focus on 1-3 L,D transpeptidation products, we only enabled the search for 1-3 dimers which contain the gm-A moiety. Interestingly, the PGF inder search revealed a very low amount of 1-3 transpeptidation products in E. coli (gm-A, 0.44% and gm-AEJA=gm-A, 0.056%) (Fig. 5-5B). A striking increase in gm-A (9.0%), gm-AX (4.3%) monomers and dimers resulting from 1-3 crosslinking (9.5%) was detected in the peptidoglycan of the strain expressing Ldt_{Go2} . demonstrating that this enzyme is an L,Dtranspeptidase that can perform all the reactions described in Fig. 5-3 (carboxypeptidation, exchange and 1-3 transpeptidation).



Figure S5-4. Strategy for peptidoglycan structural analysis of *E. coli* producing Ldt_{Go2}.

A first search was performed using the Byonic[™] module from Byos[®] to identify monomers based on MS/MS data. The search space contained disaccharide substituted by mono-, di-, tri-, tetra- and pentapeptides stems containing glutamic acid (E) or glutamine (Q) in position 2, *meso*-diaminopimelic acid (J) or amidated *meso*-diaminopimelic acid (Z) in position 3, any possible aminoacids (X) in position 4 or the AX dipeptides in positions 4 and 5. The 28 monomers identified by MS/MS were combined to generate the database DB_0_Ec. A second search was performed with PGFinder, enabling the formation of dynamic libraries containing dimers and trimers resulting from 1-3 crosslinks only (boxed in red).

Table S5- 3. List of peptide sequences used for <i>E. coli</i> MS/MS sear					
	Sequence	Sequence			
	Δ	AEIK			

Sequence	Sequence	
A	AEJK	
AE	AEJM	
AC	AEJN	
AD	AEJP	
AF	AEJQ	
AG	AEJR	
AH	AEJS	
AI	AEJT	
AJ	AEJV	
AK	AEJW	
AL	AEJY	
AM	AEJAA	
AN	AEJAC	
AP	AEJAD	
AQ	AEJAE	
AR	AEJAF	
AS	AEJAG	
AT	AEJAH	
AV	AEJAI	
AW	AEJAK	
AY	AEJAM	
AZ	AEJAN	
AEJ	AEJAP	
AEJA	AEJAQ	
AEJC	AEJAR	
AEJD	AEJAS	
AEJE	AEJAT	
AEJF	AEJAV	
AEJG	AEJAW	
AEJH	AEJAY	
AEJI		

Table S5- 4. DB0_Ec

Structure	Monoisotopic mass
gm-AE 1	698.2858
gm-AF 1	716.3116
gm-Al 1	682.3273
gm-AK 1	697.3382
gm-AQ 1	697.3018
gm-AW 1	755.3225
gm-AY 1	732.3065
gm-AEJ 1	870.3706
gm-AEJA 1	941.4077
gm-AEJD 1	985.3975
gm-AEJF 1	1017.439
gm-AEJG 1	927.392
gm-AEJH 1	1007.4295
gm-AEJI 1	983.4546
gm-AEJK 1	998.4655
gm-AEJM 1	1001.4111
gm-AEJN 1	984.4135
gm-AEJT 1	971.4183
gm-AEJV 1	969.439
gm-AEJW 1	1056.4499
gm-AEJY 1	1033.4339
gm-AEJAA 1	1012.4448
gm-AEJAG 1	998.4292
gm-AEJAI 1	1054.4918
gm-AEJAK 1	1069.5027
gm(-Ac)-AEJ 1	828.36008
gm(-Ac)-AEJA 1	899.39722
gm-A 1	569.24331



	Manual day 3	E. coli		DT (min) b	Anne
	Muropeptides	pET2818	pET-Ldt _{Go2}	RI (min)	дррт
1	gm-AEJA	44.973%	28.108%	7.82 ± 0.12	2.4
2	gm-AEJ	17.655%	13.974%	6.57 ± 0.19	3.1
3	gm(-Ac)-AEJA	11.164%	22.256%	7.37 ± 0.14	2.2
4	gm-AEJM	5.849%	0.063%	11.08 ± 0.00	1.0
5	gm-AE	4.219%	2.666%	7.76 ± 0.12	2.1
6	gm-AEJG	3.566%	1.513%	6.91 ± 0.16	1.7
7	gm-Al	ND	0.785%	14.74	1.0
8	gm-AY	ND	0.675%	12.41	0.3
9	gm-AW	ND	0.655%	18.84	0.8
10	gm-AEJN	2.505%	0.235%	6.42 ± 0.21	1.5
11	gm(-Ac)-AEJ	2.288%	5.624%	6.27 ± 0.21	1.9
12	gm-AEJF	2.263%	0.229%	15.10 ± 0.02	1.3
13	gm-AEJK	1.563%	0.196%	6.33 ± 0.22	1.4
14	gm-AEJH	0.786%	0.084%	6.53 ± 0.23	1.4
15	gm-AEJD	0.526%	0.055%	6.93 ± 0.16	1.8
16	gm-A	0.437%	9.038%	6.99 ± 0.15	1.6
17	gm-AEJT	0.364%	0.058%	7.06 ± 0.14	1.1
18	gm-AEJAK	0.341%	0.146%	7.53 ± 0.12	0.6
19	gm-AEJY	0.306%	0.016%	12.40 ± 0.01	0.4
20	gm-AQ	0.273%	0.123%	6.66 ± 0.17	2.0
21	gm-AEJV	0.222%	ND	11.08	0.5
22	gm-AEJAA	0.188%	1.465%	9.07 ± 0.31	1.3
23	gm-AEJI	0.176%	0.008%	14.21 ± 0.65	0.7
24	gm-AEJW	0.173%	0.017%	17.39 ± 0.20	0.6
25	gm-AK	ND	0.052%	5.91	0.9
26	gm-AEJAI	0.051%	0.014%	15.53 ± 0.04	0.4
27	gm-AEJAG	0.035%	0.198%	7.53 ± 0.14	0.7
28	gm-AF	0.019%	2.177%	16.76	1.7
29	gm-AEJA=gm-A	0.056%	6.437%	11.96 ± 0.07	0.4
30	gm(-Ac)-AEJA=gm-A	ND	2.445%	11.68	1.0
31	gm-AEJ=gm-A	ND	0.422%	11.23	1.3
32	gm-AEJG=gm-A	ND	0.154%	10.89	0.2
33	gm(-Ac)-AEJ=gm-A	ND	0.077%	10.89	1.3
34	gm-AEJN=gm-A	ND	0.015%	10.17	0.5
35	gm-AEJF=gm-A	ND	0.012%	18.39	1.5
36	gm-AEJI=gm-A	ND	0.006%	17.14	1.1

^a g. GlcNAc; m. MurNAc; A. Alanine; E. isoglutamic acid; J. meso-diaminopimelic acid; J_{NH2}, amidated meso-diaminopimelic acid ^b standard deviations are determined from 2 values only

Figure 5- 5. Heterologous protein synthesis of Ldt_{Go2} in *E. coli* BL21(DE3) increases the proportion of 1-3 *L*,*D*-transpeptidation products.

E. coli BL21(DE3) transformed with the control pET2818 plasmid or pET2818 encoding Ldt_{Go2} was grown in auto-induction medium overnight and peptidoglycan from both cultures were purified. The muropeptide profile are shown in (**A**); bottom profile is from the control strain; bottom panel is from *E.* coli expressing recombinant Ldt_{Go2}. Two major peaks containing muropeptides of interest resulting from 1-3 *L*,*D*-transpeptidation are indicated. (**B**) PGFinder analysis of *E.* coli control strain (transformed with the empty plasmid, *E.* coli) and expressing Ldt_{Go2} (*E.* coli + Ldt_{Go2}). Only monomers validated by Byonic[™] based on MS/MS data were search as well as their 1-3 transpeptidation products. The monomers and dimers resulting from 1-3 *L*,*D*-transpeptidation are indicated in red.

$5.3.5 \text{ Ldt}_{Go2}$ is characterised by an atypical Yku*D*-like catalytic domain that can be found in distant families of bacteria with 1-3 peptidoglycan crosslinks

To place *G. oxydans' L,D*-transpeptidases in a broader evolutionary context, homologues were extracted from genomes of numerous alphaproteobacterial species (Table S5-5), including those previously shown to contain 1-3 peptidoglycan crosslinks (Espaillat et al., 2016). Four other organisms with characterized *L,D*-transpeptidases (*E. coli, C. difficile, M. tuberculosis* and *E. faecium*) were added. Phylogenetic reconstruction of all putative alphaproteobacterial *L,D*-transpeptidases revealed that both Ldt_{Go1} and Ldt_{Go2} homologues form distinct clades representing previously uncharacterised transpeptidase subfamilies (Fig. 5-6A).

When annotated using InterProScan and its default significance thresholds, Ldt_{Go1} homologues are shown to contain a canonical YkuD (Pfam: PF03734) domain, but the more distantly related Ldt_{Go2} homologues typically lack this annotation. Instead, most are annotated with a YkuD-like (CDD: cd16913) domain, and others contain no domain annotations at all (Fig. 5-6B). Although Ldt_{Go2} is annotated with a canonical YkuD domain, it should be noted that the E-value for this annotation is high (2.0e-1), indicative of a significant divergence from the canonical YkuD domain.



Figure 5- 6. LdtGo2 represents a distinct *L*,*D*-transpeptidase subclade with a divergent catalytic domain that is found primarily in the *Acetobacteraceae* and *Burkholderiaceae* (A) An unrooted phylogenetic tree of putative *L*,*D*-transpeptidases throughout *Alphaproteobacteria* (Table S5-5) reveals that Ldt_{Go1} and Ldt_{Go2} homologues form distinct transpeptidase subfamilies. Ldt_{Go1} and Ldt_{Go2} are labelled with asterisks (*), and previously characterised *L*,*D*-transpeptidases from *Escherichia coli* (LdtA-F), *Clostridioides difficile* (Ldt_{Cd1-3}), *Mycobacterium tuberculosis* (Ldt_{Mt1-5}) and *Enterococcus faecium* (Ldt_{Efm}) have also been labelled. (B) A phylogram of Ldt_{Go1}, Ldt_{Go2}, and their homologues reveals that whilst all Ldt_{Go1} homologues were annotated with the canonical *L*,*D*-transpeptidase Pfam domain (YkuD) (Mistry et al., 2021), most Ldt_{Go2} homologues were annotated only with the CDD YkuD_like domain (Wang et al., 2023), and the rest lacked domain annotation entirely. The Ldt_{Go2} homologues highlighted in orange are found in bacterial species where no 1-3 crosslinks could be detected (Espaillat et al., 2016). (C) An expanded search for Ldt_{Go2} homologues beyond the *Alphaproteobacteria* reveals the presence of this subfamily throughout the *Burkholderiales* and *Desulfovibrionaceae*. Structural homologues located using Flodseek (van Kempen et al., 2024) show a similar evolutionary distribution as those located using BlastP (Camacho et al., 2009).

Table S5-5, Specie	s. TaxIds. and	genome accessions used	l for comparative genomics.
	<i>5, 10/10/5, 0110</i>	Some accessions used	i tor comparative genomes.

Species	Taxid	Accession
Escherichia coli	562	GCF_000005845.2
Clostridioides difficile	1496	GCF_018885085.1
Mycobacterium tuberculosis	1773	GCF_000195955.2
Enterococcus faecium	1352	GCF_009734005.1
Sphingomonas paucimobilis	13689	GCF_016027095.1
Erythrobacter litoralis	39960	GCF_001719165.1
Caulobacter vibrioides	155892	GCF_000022005.1
Asticcacaulis biprosthecum	76891	GCF_000204015.1
Roseobacter denitrificans	2434	GCF_002983865.1
Rhodobacter sphaeroides	1063	GCF_000021005.1
Bartonella grahamii	33045	GCF_000022725.1
Labrys okinawaensis	346911	GCF_002982075.1
Aquamicrobium aerolatum	561088	GCF_900113935.1
Mesorhizobium mediterraneum	43617	GCF_002284565.1
Agrobacterium tumefaciens	358	GCF_003667905.1
Sinorhizobium meliloti	110321	GCF_007827695.1
Angulomicrobium tetraedale	217068	GCF_014195655.1
Hyphomicrobium denitrificans	53399	GCF_000143145.1
Hirschia baltica	2724	GCF_000023785.1
Hyphomonas sediminis	2866160	GCF_019679475.1
Magnetospirillum gryphiswaldense	55518	GCF_000513295.1
Thalassospira lucentensis	168935	GCF_000421265.1
Roseomonas gilardii	257708	GCF_001941945.1
Acidiphilium facilis	525	GCF_000687875.1
Acidomonas methanolica	437	GCF_004346035.1
Komagataeibacter xylinus	28448	GCF_004006375.1
Gluconobacter oxydans	442	GCF_000583855.1
Gluconobacter frateurii	38308	GCF_002723955.1
Acetobacter pasteurianus	438	GCF_009914215.2
Acetobacter tropicalis	104102	GCF_001580945.1
Acetobacter pomorum	65959	GCF_002738225.1
Acetobacter aceti	435	GCF_000379545.1

Finally, to better understand the distribution of this unusual L,D-transpeptidase subfamily beyond the Alphaproteobacteria, the catalytic domain of Ldt_{Go2} was searched against the entirety of the NCBI RefSeq Select database (Camacho, 2009). Out of the 307 hits returned from unique bacterial species, roughly 37% could be attributed to Acetobacteraceae like G. oxydans, but an even greater percentage of hits (45%) came from the Burkholderiaceae (Fig. 5-6C). Though the Acetobacteraceae and Burkholderiaceae encompass the majority of Ldt_{Go2} homologues, others are found sprinkled throughout the broader Burkholderiales and even beyond the Pseudomonadota, with homologues in the Desulfovibrionaceae. Since active site geometry is thought to be a key determinant of L,D-transpeptidase substrate preference and activity, a further search for structural homologues was conducted using Foldseek and an AlphaFold model of the Ldt_{Go2} catalytic domain (Jumper et al., 2021; Sacco et al., 2010). Setting an E-value threshold of <=2e-2 (selecting for matches better than Bacillus subtilis's prototypical YkuD domain), led to 147 hits. The results of this structural search largely validated the results of the sequence-based BlastP search, with 22% of hits coming from the Acetobacteraceae and 39% from the Burkholderiaceae, but a much larger number of hits (23%) now fell outside of the families found by BLAST.

Overall, these analyses establish that *L*,*D*-transpeptidases associated with 1-3 crosslinking contain a catalytic domain related to the canonical YkuD transpeptidase domain, but form a distinct enzymatic subfamily.

5.4 Discussion

In this study, we determined the high-resolution structure of *G. oxydans* peptidoglycan using a version of PGFinder that can generate dynamic databases containing 1-3 crosslinked multimers. We show that *G. oxydans* peptidoglycan contains a high proportion of dipeptide stems with unusual amino acids at their C-terminus, leading us to propose that the enzyme forming 1-3 crosslinks uses dipeptide stems as a donor substrate. We identify two enzymes distantly related to *L,D*-transpeptidases making 3-3 crosslinks. Based on the characterization of a transposon mutant and heterologous expression experiments, we demonstrate that one of these two candidates (Ldt_{Go2}) catalyzes the formation of 1-3 crosslinks.

This work demonstrated that Ldt_{Go2} plays a predominant role in the formation of 1-3 crosslinks in *G. oxydans*. The role of Ldt_{Go1} remains unclear since the inactivation of the corresponding gene is associated with only marginal changes in the peptidoglycan composition (Table 5-1). Our attempts to express recombinant Ldt_{Go1} and Ldt_{Go2} in *E. coli* as His-tagged or maltose-binding fusion proteins remained unsuccessful and both proteins were systematically found in the insoluble fraction, irrespective of the expression strains and conditions tested. Further experiments are therefore required to produce and purify these recombinant proteins to more closely examine their activity *in vitro*.

The formation of 3-3 crosslinks in *Enterococci* is controlled by the availability of disaccharide-tetrapeptides used as donor substrate. In *E. faecium*, *L,D*-transpeptidation can bypass the *D,D*-transpeptidation following the activation of a cryptic *D,D*-carboxypeptidase (Sacco et al., 2010). How the disaccharide-dipeptide substrates are generated in *G. oxydans* remains unknown. *G. oxydans* encodes two potential endopeptidases containing a CHAP domain that could generate Ldt_{Go2} substrates (GOX_RS06930 and GOX_RS07380 in *G. oxydans* 621H). The CHAP (Cysteine-Histidine dependent Amidohydrolases/Peptidases) domain is associated with families of amidase domains and mainly function in peptidoglycan hydrolysis (Bateman and Rawlings, 2003). The transposon inactivation of each gene was tested, but did not abolish the production of 1-3 crosslinks (data not shown), indicating that these genes do not play a predominant role in the formation of dipeptide stems or are functionally redundant. The simultaneous inactivation of both genes will be required to further investigate the Ldt_{Go2} partners that contribute to the formation of unusual crosslinks.

Interestingly, only three Ldt_{Go2} homologs from *Roseomonas gilardii*, *Acidocella facilis*, *and Acidomonas methanolica* did not contain any YkuD-like (CDD: cd16913) catalytic domains. Although 1-3 crosslinks have only been reported in *Acidomonas methanolica* (Espaillat et al., 2016), it would be worth revisiting the peptidoglycan in the two other species to confirm the absence of 1-3 crosslinks using PGFinder.

Moving from sequence to structural analysis, the predicted fold of Ldt_{Go2} revealed the presence of a much more open, bowl-like active site (Fig. 5-7). Given that this enzyme
uses a shorter peptide stem as a donor substrate, it is likely that the catalytic site does not require the canonical cleft or trapping loops to accommodate the substrate. Instead, the open conformation of the catalytic site could ensure that the bulky sugar moieties of a dipeptide substrate don't limit access to the catalytic cysteine residue responsible for the formation of 1-3 crosslinks.



Figure 5-7. Structural analysis of *G. oxydans* Ldt_{Go2}.

(A) Predicted fold of Ldt_{Go2}, inclusive of well-modelled residues 79 to 336, taken from EBI Alphafold repository (Jumper et al., 2021). The catalytic residue (C264) shown in stick form with SH sidechain coloured yellow. (B) Surface representation of Ldt_{Go2}, demonstrating flat bowl-like active site surrounding C264 (yellow). (C) Superimposition of Ldt_{Go2} (grey) with *Vibrio cholerae* LdtA (RCSB entry 7AJO, unreleased, blue; bound reaction intermediate at C444 shown in stick form), reveals the relatively more closed/capped cleft of 3-3 crosslink forming enzymes and outlines that potential donor and acceptor substrates of Ldt_{Go2} will likely be less constrained.

The discovery of enzymes forming 1-3 crosslinks reaffirmed that the catalytic reactions carried out by domains belonging to the YkuD family are very diverse. This work expands our knowledge on peptidoglycan polymerization and opens new avenues to study how remodelling contributes to the maintenance of cell envelope integrity (Morè et al., 2019). *Acetobacteraceae* (also called acetic acid bacteria) are important for the food industry and are key organisms involved in the production of vinegar (Gomes et al., 2018). These organisms have a high capacity to oxidize ethanol as well as various sugars to form acetic acid and display resistance to high concentrations of acetic acid released into the fermentative medium. It is tempting to speculate that the formation of 1-3 crosslinks contributes to the maintenance of cell envelope integrity in these harsh conditions.

5.5 Experimental Procedures

5.5.1 Bacterial strains, plasmids, oligonucleotides, and growth conditions

Bacterial strains, plasmids, and oligonucleotides are described in Table S5-6. *G. oxydans* B58 (ATCC NRLL-BR8) and isogenic derivatives were grown in yeast peptone mannitol (YPM; 5 g/l yeast extract, 3 g/l peptone, 25 g/l mannitol) broth or agar at 30 °C under agitation (200 rpm). *G. oxydans* cultures were inoculated with an overnight preculture at an OD_{600} =0.05 and grown for 36 h to stationary phase. *G. oxydans* transposon mutants were grown in the presence of kanamycin (100 µg/mL) and gentamicin (10 µg/mL) for complementation experiments Ldt_{Go2} expression in *G. oxydans* was induced by adding 100 ng/mL anhydrotetracycline to the media at an

 $OD_{600}=0.5$. For heterologous expression, *E. coli* was grown in an auto-induction medium at 30 °C under agitation (200 rpm) supplemented with 100 µg/mL ampicillin.

StrainsGluconobacter oxydansB58G. oxydans reference strain (ATCC NRLL B58)ATCC621HG. oxydans reference strain (ATCC 621H)ATCCB58 ldtGolB58 with a transposon insertion in Go2094 (GOX2269 in strain 621H); KanR(1)B58 ldtGo2B58 with a transposon insertion in Go2227 (GOX1074 in strain 621H); KanR(1)B58 ldtGo2B58 ldtGo2B58 ldtGo2(1)B58 ldtGo2B58 ldtGo2B58 ldtGo2(1)B58 ldtGo2B58 ldtGo2B58 ldtGo2(1)B58 ldtGo2B58 ldtGo2B58 ldtGo2(1)B58 ldtGo2B58 ldtGo2B58 ldtGo2(1)B121(DE3) LemoExpression strainNEBBL21(DE3) LemoBL21(DE3) Lemo expressing LdtGo2; AmpRThis workPET2818PET28a derivative for recombinant protein expression; AmpR(2)pET2818pET2818 derivative encoding full length LdtGo2; AmpRThis workpBBR1MCS-5pBBR1-MCS derivative; GmR(3)pBBR1MCS-5pBBR1-MCS derivative for inducible expression with anhydrotetracycline; GmR(4)pBR-TetR-Go2227pBBR1-TetR derivative expressing full length LdtGo2; GmRThis workOligonucleotidesStacggtctcccgaagtctcgggcgtcttggggggggggg	Strains/plasmids/ oligonucleotides	Relevant properties/sequence ^a	Source
Gluconobacter oxydansB58G. oxydans reference strain (ATCC NRLL B58)ATCC621HG. oxydans reference strain (ATCC 621H)ATCCB58 ldtGalB58 with a transposon insertion in Go2094 (GOX2269 in strain 621H); KanR(1)B58 ldtGa2B58 with a transposon insertion in Go2227 (GOX1074 in strain 621H); KanR(1)B58 ldtGa2B58 with a transposon insertion in Go2227 (GOX1074 in strain 621H); KanR(1)B58 ldtGa2pB8R-TetR-Go2227)B58 ldtGa2This workBL21(DE3) LemoExpression strainNEBBL21(DE3) (pET-Go2227)BL21(DE3) Lemo expressing LdtGa2; AmpRThis workPET2818pET28a derivative for recombinant protein expression; AmpR(2)pET2818pET28a derivative encoding full length LdtGa2; AmpRThis workgBBR1MCS-5pBBR1-MCS derivative; GmR(3)pBBR1MCS-5pBBR1-MCS derivative for inducible expression with anhydrotetracycline; GmR(4)pBBR-TetR-Go2227gctacggtctccggacgctcttgggctThis workOligonucleotidesSM_0725ggctacggtctccggacgtctctgggcgtgtttccgacgggKangSM_0726ggctacggtctccggacgtctctgggcgtgtttccgacgggSM_0726SM_0728ggctacggtctccaaggtaacgccaaagacgccggtSM_0730SM_0730gredacggtctcacettaacgcaagaaccccgttSM_0730	Strains		
B58G. oxydans reference strain (ATCC NRLL B58)ATCC621HG. oxydans reference strain (ATCC 621H)ATCCB58 ldtGalB58 with a transposon insertion in Go2094 (GOX2269 in strain 621H); KanR(1)B58 ldtGa2B58 with a transposon insertion in Go2227 (GOX1074 in strain 621H); KanR(1)B58 ldtGa2B58 ldtGa2B58 ldtGa2(1)B58 ldtGa2B58 ldtGa2mutant complemented; KanR GmR(1)B58 ldtGa2B58 ldtGa2mutant complemented; KanR GmR(1)B58 ldtGa2B58 ldtGa2mutant complemented; KanR GmRNEBBL21(DE3) LemoExpression strainNEBNEBBL21(DE3) (pET-Go2227)BL21(DE3) Lemo expressing LdtGa2; AmpRThis workPET2818pET28a derivative for recombinant protein expression; AmpR(2)pET-LdtGa2pET2818 derivative encoding full length LdtGa2; AmpRThis workpBBR1MCS-5pBBR1-MCS derivative; GmR(3)pBBR1MCS-5pBBR1-MCS derivative for inducible expression with anhydrotetracycline; GmR(4)pBBR-TetR-Go2227ggctacggtctccgaggtctgggcgtctcttgggcttThis workOligonucleotidesggctacggtctcccgaggtctgggcgtctttgggcgggM O726ggctacggtctcccgaagtacggtggggggggggggggg	Gluconobacter oxydans		
621HG. oxydans reference strain (ATCC 621H)ATCCB58 ldtGolB58 with a transposon insertion in Go2094 (GOX2269 in strain 621H); KanR(1)B58 ldtGo2B58 with a transposon insertion in Go2227 (GOX1074 in strain 621H); KanR(1)B58 ldtGo2B58 ldtGo2 mutant complemented; KanR GmR(1)BL21(DE3) LemoExpression strainNEBBL21(DE3) LemoExpression strainNEBBL21(DE3) (pET-Go2227)BL21(DE3) Lemo expressing LdtGo2; AmpRThis workPlasmidspET2818pET281 derivative for recombinant protein expression; AmpR(2)pET-LdtGo2pET2818 derivative encoding full length LdtGo2; AmpRThis workgBR1MCS-5pBBR1-MCS derivative; GmR(3)pBBR1MCS-5gBR1-MCS derivative expressing full length LdtGo2; GmRThis workOligonucleotidesSM_0725ggctacggtctcccgaagtctcggccgtctcttgggcttThis workSM_0726ggctacggtctctaggagatactatatgcgtagttttccagactgacSM_0727ggctacggtctcacetaaggtaccggtctttatccgcaagagSM_0729ggctacggtctcacetaagcgagaagacccggtcggSM_0730eyctacggtctcacettaagccagaagaacccgtttSM_0730eyctacggtctcacetaagccaaaaacccgcttgggggSM_0730eyctacggtctcacetaagccaaaaacccgcttcgggcg	B58	G. oxydans reference strain (ATCC NRLL B58)	ATCC
B58 ldtGolB58 with a transposon insertion in Go2094 (GOX2269 in strain 621H); KanR(1)B58 ldtGo2B58 with a transposon insertion in Go2227 (GOX1074 in strain 621H); KanR(1)B58 ldtGo2 (pBBR-TetR-Go2227)B58 ldtGo2 mutant complemented; KanR GmR(1)BL21(DE3) LemoExpression strainNEBBL21(DE3) LemoExpression strainNEBBL21(DE3) (pET-Go2227)BL21(DE3) Lemo expressing LdtGo2; AmpRThis workPlasmidspET2818pET28a derivative for recombinant protein expression; AmpR(2)pET2818pET2818 derivative encoding full length LdtGo2; AmpRThis workgBBR1MCS-5pBBR1-MCS derivative; GmR(3)pBBR1MCS-5pBBR1-MCS5 derivative for inducible expression with anhydrotetracycline; GmR(4)pBBR-TetR-Go2227gctacggtctcccgaagtctgggccgtcttgggcttSM_0725ggctacggtctccccgaagtctgggcggcgtcttgggcgggggggg	621H	G. oxydans reference strain (ATCC 621H)	ATCC
B58 ldt_{Go2}B58 with a transposon insertion in Go2227 (GOX1074 in strain 621H); KanR(1)B58 ldt_{Go2} (pBBR-TetR-Go2227)B58 ldt_{Go2} mutant complemented; KanR GmRThis workEscherichia coliExpression strainNEBBL21(DE3) LemoExpression strainNEBBL21(DE3) (pET-Go2227)BL21(DE3) Lemo expressing Ldt_{Go2}; AmpRThis workPlasmidsPET2818pET28a derivative for recombinant protein expression; AmpR(2)pET-Ldt_{Go2}pET2818 derivative encoding full length Ldt_{Go2}; AmpRThis workpBBR1MCS-5pBBR1-MCS derivative; GmR(3)pBBR1MCS-5-TgdhM-tetR-mNGpBBR1-MCS5 derivative for inducible expression with anhydrotetracycline; GmR(4)pBBR-TetR-Go2227ggctacggtctcccggagtctcgggccgtcttgggcttThis workSM_0725ggctacggtctcccggagtctcgggccgtcttgggcgggggggg	B58 ldt _{Go1}	B58 with a transposon insertion in Go2094 (GOX2269 in strain 621H); KanR	(1)
B58 ldt_{Go2} (pBBR-TetR-Go2227)B58 ldt_{Go2} mutant complemented; KanR GmRThis workEscherichia coliExpression strainNEBBL21(DE3) LemoExpression strainNEBBL21(DE3) (pET-Go2227)BL21(DE3) Lemo expressing Ldt_{Go2}; AmpRThis workPlasmidspET2818pET28a derivative for recombinant protein expression; AmpR(2)pET-Ldt_{Go2}pET2818 derivative encoding full length Ldt_{Go2}; AmpRThis workgBBR1MCS-5pBBR1-MCS derivative; GmR(3)pBBR1MCS-5-TgdmM-tetR-mNGpBBR1-TetR derivative expressing full length Ldt_{Go2}; GmR(4)pBBR-TetR-Go2227gctacggtctcccgaagtctcgggccgtcttgggcttSM_0725SM_0725ggctacggtctcccgaagtctcgggccgtcttgggcttSM_0726SM_0726ggctacggtctccagagtatcatagtggtagtttccagactgacSM_0728SM_0728ggctacggtctcccagagtagtatacgtatggtggggggttttacggctgggggggg	B58 ldt _{Go2}	B58 with a transposon insertion in Go2227 (GOX1074 in strain 621H); KanR	(1)
Escherichia coliExpression strainNEBBL21(DE3) LemoExpression strainNEBBL21(DE3) (pET-Go2227)BL21(DE3) Lemo expressing Ldt _{Go2} ; AmpRThis workPlasmidspET2818pET28a derivative for recombinant protein expression; AmpR(2)pET-Ldt _{Go2} pET2818 derivative encoding full length Ldt _{Go2} ; AmpRThis workpBBR1MCS-5pBBR1-MCS derivative; GmR(3)pBBR1MCS-5-TgdnM-tetR-mNGpBBR1-MCS5 derivative expressing full length Ldt _{Go2} ; GmR(4)pBBR-TetR-Go2227ggctacggtctcccgaagtctcgggccgtcttgggcttSM_0725SM_0725ggctacggtctcccgaagtctcgggccgtcttgggctThis workSM_0726ggctacggtctcccgaagtctcgggcgtgttttccagactgacSM_0727ggctacggtctccagagtacggtgttttaccgcaatagSM_0728ggctacggtctcccaagagtacaggtttttaccgcaatagSM_0729ggctacggtctcacettacgcaaaaaaaccccggtSM_0730ggctacggtctcacettacgcgaaaaaaccccggtSM_0730ggctacggtctcacettacgcaaaaaaccccggt	B58 ldt _{Go2} (pBBR-TetR-Go2227)	B58 ldt _{Go2} mutant complemented; KanR GmR	This work
BL21(DE3) LemoExpression strainNEBBL21(DE3) (pET-Go2227)BL21(DE3) Lemo expressing Ldt _{Go2} ; AmpRThis workPlasmidspET2818pET28a derivative for recombinant protein expression; AmpR(2)pET-Ldt _{Go2} pET2818 derivative encoding full length Ldt _{Go2} ; AmpRThis workpBBR1MCS-5pBBR1-MCS derivative; GmR(3)pBBR1MCS-5-TgdhM-tetR-mNGpBBR1-MCS5 derivative expression with anhydrotetracycline; GmR(4)pBBR-TetR-Go2227gctacggtctcccgaagtctcgggccgtctttgggcttThis workSM_0725ggctacggtctcccgaagtctcgggccgtctttgggcttSM_0726SM_0726ggctacggtctccaagagtacagtattcatagcgtgagtttccagactgacSM_0728SM_0728ggctacggtctcccaagagtacaggtctttaccgcaatagSM_0730grctacggtctcacttatagcgaaaaaaccccgttSM_0730grctacggtctcacttagtcggcgctcttgggcg	Escherichia coli		
BL21(DE3) (pET-Go2227)BL21(DE3) Lemo expressing LdtGo2 ; AmpRThis workPlasmidspET2818pET28a derivative for recombinant protein expression; AmpR(2)pET-LdtGo2pET2818 derivative encoding full length LdtGo2; AmpRThis workpBBR1MCS-5pBBR1-MCS derivative; GmR(3)pBBR1MCS-5-TgdhM-tetR-mNGpBBR1-MCS5 derivative for inducible expression with anhydrotetracycline; GmR(4)pBBR-TetR-Go2227ggctacggtctcccgaagtctcgggccgtctttgggcttThis workOligonucleotidesggctacggtctcccgaagtctcgggccgtctttgggcttswnSM_0725ggctacggtctcctaggagatatcatatgcgtgatgtttccagactgac5M_0727ggctacggtctccaaggtacggtctttaccgaagaaaaaacccggtSM_0728ggctacggtctccaaagataacggtctttaccgaaagaSM_0729ggctacggtctcacttagcgaaaaaaaccccggt5M_0730gwctacggtctcacttagcgaagaacggtcttattecgaagagaaga	BL21(DE3) Lemo	Expression strain	NEB
PlasmidspET2818pET28a derivative for recombinant protein expression; AmpR(2)pET2818pET2818 derivative encoding full length Ldt _{Go2} ; AmpRThis workpBBR1MCS-5pBBR1-MCS derivative; GmR(3)pBBR1MCS-5-TgdhM-tetR-mNGpBBR1-MCS5 derivative for inducible expression with anhydrotetracycline; GmR(4)pBBR-TetR-Go2227pBBR1-TetR derivative expressing full length Ldt _{Go2} ; GmRThis workOligonucleotidesSM_0725ggctacggtctcccgaagtctcgggccgtctttgggcttThis workSM_0726ggctacggtctcctggadtaccatttccttacactgatagggSM_0727SM_0728ggctacggtctcacagagtacaggtatcggtttttaccgacatagSM_0728SM_0730grctacggtctccttgggccgtctttgggctSM_0730	BL21(DE3) (pET-Go2227)	BL21(DE3) Lemo expressing LdtGo2; AmpR	This work
pET2818pET28a derivative for recombinant protein expression; AmpR(2)pET-LdtGo2pET2818 derivative encoding full length LdtGo2; AmpRThis workpBBR1MCS-5pBBR1-MCS derivative; GmR(3)pBBR1MCS-5-TgdhM-tetR-mNGpBBR1-MCS derivative for inducible expression with anhydrotetracycline; GmR(4)pBBR-TetR-Go2227pBBR1-TetR derivative expressing full length LdtGo2; GmRThis workOligonucleotidesSM_0725ggctacggtctcccgaagtccgggccgtcttgggcttThis workSM_0726ggctacggtctctcggadtacattgcgtadgttccactgaaggYSM_0727ggctacggtctcacagagatacatggttttccagactgacYSM_0728ggctacggtctcacaagaaaaaccccggtYSM_0730gretacggtctcacttagcgaagaccesttY	Plasmids		
pET-Ldt _{Go2} pET2818 derivative encoding full length Ldt _{Go2} ; AmpR This work pBBR1MCS-5 pBBR1-MCS derivative; GmR (3) pBBR1MCS-5 pBBR1-MCS derivative for inducible expression with anhydrotetracycline; GmR (4) pBBR-TetR-Go2227 pBBR1-TetR derivative expressing full length Ldt _{Go2} ; GmR This work Oligonucleotides ggctacggtctcccgaagtctcgggccgtctttgggctt This work SM_0725 ggctacggtctcccgaagtctcgggccgtctttgggctt SM_0726 ggctacggtctccagagtatcattgggtagtttccagactgac SM_0727 ggctacggtctcacagagtaacggtatttacggtagtgttccagactgac SM_0728 ggctacggtctcacegtagaaaaaaccccggt SM_0729 ggctacggtctcacettaagcgaaaaaaaccccgttggggg SM_0730 gectacggtctcacttatgcgaagaacccstt SM_0730 gectacggtctcacettaagcgaagacccstt	pET2818	pET28a derivative for recombinant protein expression: AmpR	(2)
pBBRIMCS-5 pBBR1-MCS derivative; GmR (3) pBBR1MCS-5 pBBR1-MCS derivative; GmR (4) pBBR-TetR-Go2227 pBBR1-TetR derivative expressing full length Ldt _{Go2} ; GmR This work Oligonucleotides ggctacggtctcccgaagtctcgggccgtctttgggctt ggctacggtctcccgaagtctcgggcggtggtgggggggg	pET-Ldte-2	pET2818 derivative encoding full length Ldt _{G2} : AmpR	This work
pBBR1MC5-5-TgdmM-tetR-mNG pBBR1-MCS5 derivative for inducible expression with anhydrotetracycline; GmR (4) pBBR-TetR-Go2227 pBBR1-TetR derivative expressing full length Ldt _{Go2} ; GmR This work Oligonucleotides ggctacggtctcccgaagtctcgggccgtctttgggctt This work SM_0725 ggctacggtctcccgaagtctcgggccgtctttgggctt ggctacggtctcccgaagtctcgggccgtcttgggcggcgtctttgggct SM_0726 ggctacggtctctaggagatactaatggtgatgtttccagactgac SM_0727 ggctacggtctccagagataccggtatgtttccagactgac SM_0728 ggctacggtctcacatagcgtaagaaaaaacccggtgggg SM_0729 ggctacggtctcacttacgcgaaaaaaaccccggt SM_0730	pBBR1MCS-5	pBBR1-MCS derivative: GmR	(3)
pBBR-TetR-Go2227 pBBR1-TetR derivative expressing full length Ldt _{Go2} ; GmR This work Oligonucleotides ggctacggtctcccgaagtctcgggccgtctttgggctt ggctacggtctcccgaagtctcgggccgtctttgggctt SM_0725 ggctacggtctcccgaagtctcgggccgtctttgggctt ggctacggtctcccgaagtctcgggccgtctttgggctt SM_0726 ggctacggtctcctgggagtatcatatgcgtgatgtttccagactgac SM_0727 SM_0727 ggctacggtctccagagtatcggtatgtttccagactgac SM_0728 SM_0729 ggctacggtctcactgcaaaaaaaccccggttgggg SM_0730	pBBR1MCS-5-T _{adbM} -tetR-mNG	pBBR1-MCS5 derivative for inducible expression with anhydrotetracycline: GmR	(4)
OligonucleotidesSM_0725ggctacggtctcccgaagtctcgggccgtctttgggcttSM_0726ggctacggtctctcctggattcacttttcttatcactgatagggSM_0727ggctacggtctctaggagatatcatatgcgtgatgtttccagactgacSM_0728ggctacggtctcacaggtcaggtctcaccgaaaaaacccggtSM_0729ggctacggtctcacttagcgaaaaaacccggtSM_0730ggctacggtctcttteggagcgcctgaggcgcgtt	pBBR-TetR-Go2227	pBBR1-TetR derivative expressing full length LdtGo2; GmR	This work
SM_0725 ggctacggtctcccgaagtctcgggccgtctttgggtt SM_0726 ggctacggtctctcgggtccttgggttccatgaggg SM_0727 ggctacggtctctaggagatatcatatgcgtagtttccagactgac SM_0728 ggctacggtctcaaggtaggtaggttttaccgcatag SM_0729 ggctacggtctcacttagcgaaaaaccccgcttcggcg SM_0730 ggctacggtcttttgggagaccgcctgaggcgg	Oligonucleotides		
SM_0726 ggctacgglctcttggattcactttctttctatcactgataggg SM_0727 ggctacgglctctaggagattcattgcgtgattcactggtaggg SM_0728 ggctacgglctcaaggtaacggtatttaccgcaatag SM_0729 ggctacgglctcacttaacgcaaaaaaccccgcttcggcgg SM_0730 ggctacggtcttttcggaggcccctgaaggccgc	SM 0725	ggetacggteteeegaagtetegggeegtetettgggett	
SM_0727 ggctacgglctctaggagatatcatatgcgtgatgtttccagacggc SM_0728 ggctacgglctcaaggtaacggtatgtttaccgcaatag SM_0729 ggctacgglctcaccttaacgcaaaaaaccccgcttcggcgg SM_0730 ggctacgglctctttcggagaccccgat	SM_0726	ggetacggtetetteettggatteaetttetetateaetgataggg	
SM_0728 ggctacgglctcaaggtaacggtatgctgatgctgatg SM_0729 ggctacgglctcaccttaacgcaaaaaaccccgcttcggcgg SM_0730 ggctacggtctctttcgggggcccccgaagccccgtt	SM_0727	goctacggtetetegggatetetetetetetetetetetetetete	
SM_0729 ggctacggtctaccttaacgcaaaaaccccgcttcggcg SM_0730 ggctacggtctaccttagggagcgcccggt	SM 0728	ggctacggtctcaaaggtaacggtcttttatccgcaatag	
SM 0730 getacgetcttttcggagcgcctgaagcccgtt	SM 0729	ggetacggtetcacettaacgcaaaaaaaccccgettcggegg	
	SM 0730	ggctacggtctctttcgggagcgcctgaagcccgtt	

Table S5-6. Strains, plasmids and oligonucleotides

^a KanR, resistant to kanamycin; GmR, resistant to gentamicin; AmpR, resistant to ampicillin

 Schmitz, A. M., Pian, B. Medin, S., Reid, M. C., Wu, M., Gazel, E., Barstow, B. (2021) Generation of a *Gluconobacter* oxydans knockout collection for improved extraction of rare earth elements. Nat Commun 12(1): 6693

(2) Eckert, C., Lecerf, M., Dubost, L., Arthur, M. and Mesnage, S. (2006) Functional analysis of AtlA, the major N-acetylglucosaminidase of *Enterococcus faecalis*. J Bacteriol. 188(24):8513-8519

(3) Kovach, M. E., Elzer, P. H., Hill, D. S., Robertson, G. T., Farris, M. A., Roop, R. M., Peterson, K. M. (1995) Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene*. 166(1):175–6.

(4) Fricke, P. M., Lürkens, M., Hünnefeld, M., Sonntag, C. K., Bott, M., Davari, M. D., Polen, T. (2021) Highly tunable TetR-dependent target gene expression in the acetic acid bacterium *Gluconobacter oxydans*. *Appl Microbiol Biotechnol*. 105(18):6835-6852.

5.5.2 Plasmid constructions

Plasmid pBBR-TetR-Go2227 used to complement the transposon insertion in the *Idt_{Go2}* is a derivative of pBBR1MCS-5-T_{gdhM}-tetR-mNG allowing the inducible expression of proteins in *G. oxydans* under the control of the tetracycline promoter. pBBR-TetR-Go2227 was built using Golden Gate assembly. Three PCR fragments corresponding to (i) the gentamicin cassette (1632 bp), (ii) the pBBR1 origin of replication + the TetR gene (4029 bp), and (iii) the Idt_{Go2} full length sequence (1021 bp) were amplified using oligos SM 0729 + SM 0730, SM 0725 + SM 0726 and SM 0727 + respectively using pBBR1MCS-5-T_{gdhM}-tetR-mNG or G. SM 0728, oxydans chromosomal DNA as templates. The PCR products were purified by gel extraction, mixed in an equimolar ratio, and assembled using the NEBridge Golden Gate Assembly Kit (Bsal-HF v2) according to the manufacturer's instructions. Recombinant plasmids were screened by PCR and plasmid candidates were fully sequenced by Plasmidsaurus (Plsamidsaurus.com) to confirm the absence of mutations.

pET-Go2227, a pET2818 derivative expressing the full-length Ldt_{Go2} enzyme was built using a synthetic DNA fragment with optimized codon usage for *E. coli* provided by Genewiz. The synthetic open reading frame corresponding to the full-length Ldt_{Go2} gene (with a stop codon) was cloned into pET2818 as a Ncol-Xhol fragment.

5.5.3 Preparation of G. oxydans competent cells and transformation

G. oxydans was grown in 100 mL of YPM to an OD_{600} of 0.9 and spun for 10 min at 4000*g* at 4 °C. After three washes 1 mM 4-(2-hydroxyethyl)-1in piperazineethanesulfonic acid (HEPES) buffer (pH7.0), cells were resuspended in 250 µL. Electroporation was carried out in 1 mm cuvettes using 50 µL of electrocompetent cells and 100 ng of plasmid in a volume of 1-2 µL; parameters for electroporation were 2 kV, 25 μ F and 200 Ω . After the pulse, 800 μ L of YPM media supplemented with 0.25% (m/v) MgSO₄ and 0.15% (m/v) CaCl₂ was added to the cells that were left to recover under agitation for 16 h before plating on YPM media supplemented with kanamycin and gentamicin.

5.5.4 Peptidoglycan extraction

G. oxydans and *E. coli* strains were grown until the stationary phase in YPM or autoinduction medium, respectively. Cells were pelleted, supernatant discarded, and cell pellet snap frozen in liquid nitrogen. The cell pellet was resuspended in 20 mL of boiling MilliQ water (MQ) before the addition of sodium dodecyl sulfate (SDS) at a final concentration of 4% (m/v). After 30 min at 100 °C, the cells were cooled down to room temperature. Peptidoglycan was pelleted at 150,000*g* for 1 h, washed five times using warm MQ water, freeze-dried and resuspended at a final concentration of 10 mg/mL.

5.5.5 Preparation of soluble muropeptides

2 mg of purified peptidoglycan was digested for 16 h in 20 mM phosphate buffer (pH 5.5) supplemented with 250 Units of mutanolysin (Sigma) in a final volume of 200 μ L. Following heat inactivation of mutanolysin (5 min at 100 °C), soluble disaccharide peptides were mixed with an equal volume of 250 mM borate buffer (pH 9.25) and reduced with 0.2% (m/v) sodium borohydride. After 20 min at room temperature, the pH was adjusted to 5.0 using phosphoric acid. Reduced muropeptides were analyzed by HPLC using a C18 analytical column (Hypersil Gold aQ, 1.9 μ m particles, 150 × 2.1 mm; Thermo Fisher Scientific) at a temperature of 50 °C. Muropeptide elution was performed at 0.3 mL/min by applying a mixture of solvent A (water, 0.1% [v/v] formic acid) and solvent B (acetonitrile, 0.1% [v/v] formic acid). LC conditions were 0 to 12.5% B for 25 min increasing to 20% B for 10 min. After 5 min at 95%, the column was re-equilibrated for 10 min with 100% buffer A. UV absorbance at 202 nm was used to check the quality of samples and determine the volume to inject for LC-MS. A volume of sample with an intensity of the most abundant monomer of 1500 mAU was used, giving an ion intensity of approximately 5.10⁹.

5.5.6 LC-MS/MS

An Ultimate 3000 High-Performance Liquid Chromatography (HPLC; Dionex/Thermo Fisher Scientific) system coupled with a high-resolution Orbitrap Exploris 240 mass spectrometer (Thermo Fisher Scientific) was used for LC-MS analysis. Muropeptides were separated using a C18 analytical column (Hypersil Gold aQ, 1.9 μ m particles, 150 × 2.1 mm; Thermo Fisher Scientific) at a temperature of 50 °C. Muropeptide elution was performed as described in the previous paragraph. The Orbitrap Exploris

240 was operated under electrospray ionization (H-ESI high flow)-positive mode, full scan (m/z 150–2250) at resolution 120,000 (FWHM) at m/z 200, with normalized AGC Target 100%, and automated maximum ion injection time (IT). Data-dependent MS/MS were acquired on a 'Top 5' data-dependent mode using the following parameters: resolution 30,000; AGC 100%, automated IT, with normalized collision energy 25%.

5.5.7 Analysis of peptidoglycan structure

LC-MS datasets were deconvoluted with the Byos® software v3.11 (Protein Metrics). Sequential searches were carried out with PGFinder v1.1.1, with default settings (10 ppm tolerance, 0.5 min cleanup window) following the strategy described in Figures S5-1 and S5-3. Data from individual matching output was consolidated as previously described to calculate average intensities, retention times, observed monoisotopic masses, and ppm differences. The output from individual searches and consolidated data was described in the text. Cross-linking index and glycan chain length were determined as described: The crosslinking index is defined as 0.5 * (% of dimers) + 0.33 * (% of trimers); glycan chain length was inferred from the abundance of anhydroMurNAc groups, which are found at the end of glycan chains. It is defined as 1/(% of AnhydroMurNAc trimers).

5.5.8 Comparative genomics and bioinformatic analysis

Reference genomes and protein sequences in Table S5-5 were downloaded from NCBI Datasets (v15.25.0), and protein sequences were annotated locally using InterProScan (v5.64-96.0). A custom Julia script was then used to search the produced GFF3 files for YkuD-containing proteins and to extract their catalytic domains. Ldt_{Go1} and Ldt_{Go2} homologues were located by running a PSI-BLAST on the RefSeq Select database restricted to taxa in Table S5-5 and iterating until no new hits were returned. Extracted YkuD proteins and Ldt_{Go1/2} homologues were aligned using Muscle (v5.1), and maximum likelihood trees were constructed using IQ-TREE (v2.2.2.7) with ModelFinder (which selected WAG+R7 for Figure 5-6A and WAG + F + G4 for Fig 5-6B) and 1000 UFBoot replicates enabled. Trees were visualised and annotated using iTOL (v6.8.1) with finishing touches applied in Inkscape (v1.3). ColabFold's AlphaFold2_batch notebook (v1.5.2) was used with the default settings and relaxation enabled to obtain predicted structures for Ldt_{Go1}, Ldt_{Go2}, and their respective catalytic domains. Finally, Foldseek (v8-ef4e960) was used to search the AFDB50 database for structural homologues of Ldt_{Go2}.

5.6 Concluding Remarks

In this collaborative work that involved two PhD students in the lab (Brooks Rady and myself), we identified a novel class of enzyme catalysing the formation of uncommon 1-3 PG crosslinks observed in *G. oxydans*. The enzyme contains a YkuD domain, a structural feature previously associated with *L*,*D*-transpeptidases.

To facilitate the high-resolution PG analysis of *G. oxydans*, a new version of PGFinder was created. This new version (v1.1.0) enables the *in-silico* formation and search of dimers and trimers containing 1-3 crosslinks. Interestingly our analyses revealed a high proportion of dipeptide stems (gm-AX) in the *G. oxydans* PG. The transposon mutant screening identified Ldt_{G02}, an L,D-transpeptidase with a YkuD-like domain, as the primary enzyme responsible for the 1-3 crosslinking in *G. oxydans*. This finding represents the key contribution of this work by expanding our understanding of L,D-transpeptidase versatility and its potential role in maintaining call wall integrity under challenging environments.

The initial challenge in analyzing *G. oxydans* PG composition was the detection of 1-3 crosslinked dimers. Before the development of automatic 1-3 crosslinked multimer prediction in PGFinder, we had to manually calculate the masses of all possible dimers. Once Brooks Rady implemented this new feature, I validated the software by comparing the results to those based on our manual calculations. Previously (Espaillat et al., 2016) used NMR to confirm the 1-3 crosslinking. However, since our project relied on LC-MS/MS analysis of muropeptides, a second challenge arose in confirming the existence of 1-3 crosslinks using this method. At the time, the fragment predictor module was unable to generate product ions specific to this type of crosslink. For instance, the J=J ion serves as definitive marker for 3-3 crosslinks, while 4-3 crosslinks lack a specific signature ion as any tetrapeptide structure can yield the A=J product ion. Fortunately, two potential signature ions (gm-A and m-A=J) were identified for 1-3 crosslinks, corresponding specifically to 1-3 crosslink pattern.

Our analysis of *G. oxydans* PG composition across different growth phases confirmed previous observations (Espaillat et al., 2016). This analysis set the conditions for further elucidating where do 1-3 crosslinks play a significant role. Initially, the possibility that a non-canonical L,D-transpeptidase lacking YkuD domain could be responsible for 1-3 crosslinks was not rejected. Dr Barstow's mutant library facilitated the testing of four candidate genes. The analysis we performed, revealed that amongst the two YkuD domain-containing proteins, Ldt_{Go2} was primarily responsible for the generation of 1-3 crosslinks.

Following the identification of the Idt_{Go2} gene, *in vitro* assays were planned to characterize its enzymatic activities. Given the Ldt activities observed in *C. difficile* (L,D-transpeptidation, carboxypeptidation and exchange) (Aliashkevich and Cava, 2021; Galley et al., 2024; Sütterlin et al., 2018), we anticipated that Ldt_{Go2} would exhibit similar behaviour. The first obstacle we faced was the low yield of the required substrates, disaccharide dipeptide (gm-AE) and disaccharide tripeptide (gm-AEJ(amidated)) as donor and acceptor respectively, from *G. oxydans* cultures. Despite 2 days of growth the cultures reached optical densities of 1.5-2.0. We suspect this low

growth may be due to the obligate aerobic nature of *G. oxydans*, as our culture conditions may not have met its oxygen requirements.

An alternative strategy involved digesting *Bacillus subtilis* muropeptides. *B. subtilis* offered several advantages over *G. oxydans*: a faster growth rate, facultative anaerobic nature (Nakano and Hulett, 1997), and the presence of amidated mDAP, similar to *G. oxydans*. Additionally, gm-AEJ(Amidated) could be used as acceptor substrate and digested to obtain the disaccharide dipeptide (gm-AE) needed as donor substrate (Atrih et al., 1999). Two *D,L*-endopeptidases were considered for digesting gm-AEJ(Amidated) into gm-AE: CwIO (Yamaguchi et al., 2004) and CwIS (Fukushima et al., 2006).

While attempting to heterologously express Idt_{Go2} in *E. coli* as an MBP or His tagged protein, we faced limited success. I presented these preliminary results in the Great Wall Symposium 2023. Following the symposium, we communicated with Dr. Felipe Cava's group and learnt about their progress in identifying the enzyme responsible for 1-3 crosslinks. To avoid potential conflicts of interest, the Cava and Mesnage groups agreed to publish their results separately, each submitting manuscripts within a month after the symposium. Due to time constraints arising from this conflict of interest, we had to cut short the experiments aimed at demonstrating the *in vitro* activity of the recombinant Idt_{Go2} enzyme.

To conclude the project and publish the results, we made a final attempt to co-express CwIO and Ldt_{Go2} in *E. coli*. However, this approach was unsuccessful, and we focused in complementing the mutant strain. A major challenge in complementing the transposon mutant was the limited availability of suitable plasmids for *G. oxydans*. Most plasmids available carry kanamycin resistance markers, which is the same resistance of the transposon mutant (Schmitz et al., 2021). Only a few vectors with alternative selection markers (ampicillin and gentamicin) exist (Liu et al., 2021; Ripoll et al., 2023). We obtained the anhydrotetracycline-inducible expression plasmid pBBR1-MCS5 (gentamycin resistance) from Dr Polen's laboratory (Fricke et al., 2021). This plasmid proved to be crucial for complementing the *ldt*_{Go2} mutant and confirming its role in generating 1-3 crosslinks in *G. oxydans*.

To further validate the activity of Ldt_{Go2} , we transformed a $\Delta 6ldt \ E. \ coli$ strain of $E. \ coli$ devoid of the six Ldts genes present in the genome. Although $E. \ coli$ does not naturally contain amidated mDAP, it produces small amounts of the disaccharide-dipeptide donor (gm-AE). The heterologous expression of Ldt_{Go2} in $E. \ coli \ \Delta 6ldt$ resulted in a high amount of 1-3 crosslinks which are not naturally present in *Escherichia coli*. This experiment directly demonstrated the activity of Ldt_{Go2} in catalyzing 1-3 crosslinks.

Following the publication of Dr Cava's laboratory (Espaillat et al., 2024), our understanding of 1-3 transpeptidation has been refined. Contrary to our initial assumptions, Ldt_{Go2} was proposed to use a disaccharide-tetrapeptide (gm-AEJA) as a substrate rather than a disaccharide-dipeptide (gm-AE). The publication by Espaillat et al. 2024 demonstrates L,D-transpeptidase and endopeptidase activity only against tripeptides and tetrapeptides substrates, challenging the notion that gm-AE would be the preferred donor. The study also confirms the exchange activity we observed in disaccharide-dipeptides, revealing that it is a D,L-exchange unlike canonical Ldts which

perform D,D-exchange. Additionally, the paper shows the canonical features of a YkuD Ldt, such as imipenem and copper inhibition, as well as tolerance to β -lactams (ampicillin).

The discovery of 1-3-Ldts continues expanding our understanding of bacterial cell wall diversity, the complexity of PG remodelling, and novel capabilities of known enzymes. A question that remains open is the essentiality of 1,3-L,D-transpeptidation for *G. oxydans* survival in particular and *Acetobactericiae* in general. We observed a minor growth defect in the single transposon mutant, suggesting that one *ldt* gene might be sufficient for cell survival. To address the essentiality of 1-3 crosslinks a double mutant *G. oxydans* strains, lacking both 1-3-Ldts would be informative.

Another intriguing question is whether the presence of 1-3 crosslinks would confer any novel features to the *E. coli* bacterial cell wall. It has been shown that inactivation of 1-3-Ldts sensitizes *G. oxydans* to stress conditions that *E. coli* LdtD, with its ability to generate 3-3 crosslinks, cannot complement (Espaillat et al., 2024). We detected a small amount of 1-3 crosslink dimers (gm-AEJA=gm-A) in *E. coli* Δ 6ldt (Fig5-5B). Interestingly, some *E. coli* pathogenic strains are known to survive at a pH as low as 1.5. In *E. coli*, PBP6b act as acid-induced carboxypeptidase, while MItA (lytic transglycosylase) and MepS (endopeptidase) exhibit elevated *in vitro* enzymatic activity in acidic conditions (Li et al., 2020). With shortened peptide stems is it possible that low pH could induce any Ldt to generate 1-3 crosslinks? Therefore, it is not unreasonable to hypothesize that *E. coli* Δ 6ldt complemented with *Idt_{Go2}*, might display novel characteristics, such as enhanced survival at lower pH, given the known ability *Acetobacteraceæ* to thrive in acidic environments.

Chapter 6

6. Novel software tools and strategy for PG analysis and its application to explore PG remodelling in *Rhizobium leguminosarum* PG

As part of my PhD thesis, I am including the published paper

Alamán-Zárate, M. G*., Rady, B. J.*, Ledermann, R., Shephard, N., Evans, C.A., Dickman, M.J., Turner, R.D., Rifflet, A., Patel, A.V., Gomperts Boneca, I., Poole, P.S., Bern, M., & Mesnage, S. (2025). A software tool and strategy for peptidoglycomics, the high-resolution analysis of bacterial peptidoglycan via LC-MS/MS. Communications Chemistry, 8 (1), 91. <u>https://doi.org/10.1038/s42004-025-01490-6</u>

The work contained in this paper was performed in collaboration with the groups of Dr Ivo Gomperts-Boneca, Prof Mark Dickman, Prof Philip S. Poole and Dr Marshall Bern.

The biological samples were generated by Raphael Lederman in Prof Philip Poole's lab and extracted by Dr Mesnage. The mass spectrometry data acquisition was performed by Aline Rifflet and Dr Caroline Evans in Dr Gomperts-Boneca's and Prof Dickman's facilities, respectively.

Brooks J. Rady, Ankur Patel, Neil Shephard and Bob Turner performed the software development explained in the first section of results. Brooks J. Rady created the Figures 6-1, 6-2, 6-3, Figures S6-1 and S6-2, and Tables S6-1 and S6-2 and contributed to writing and proofreading the manuscript. I performed all the data analyses included in the manuscript, and prepared all related figures, tables, and supplementary material. The original draft was written by me and Dr Mesnage with the help of B. Rady. The text was edited multiple times by me, Brooks Rady and Dr Mesnage, considering feedback from co-authors. The major contribution of Brooks Rady and me to the work was acknowledged by a joint first authorship.

For this chapter, I aimed to characterize the peptidoglycan composition of *Rhizobium legumininosarum* (*Rlv*3841) grown in Minimum (MM) and Tryptone Yeast (TY) Media. This chapter represents the culmination of my thesis research. Here, I present the most optimized workflow for PG analysis, encompassing the following objectives:

- 1. Identification of the most abundant monomers in *Rlv*3841 grown in TY medium.
- 2. Characterization of the peptidoglycan modifications present in *Rlv*3841.
- 3. Determination of the crosslinking pattern of *Rlv*3841's major monomers.
- 4. Comparison of RIv3841's PG composition grown in TY and MM media.

A software tool and strategy for peptidoglycomics, the high-resolution analysis of bacterial peptidoglycan via LC-MS/MS

Marcel G. Alamán-Zárate^{1*}, Brooks J. Rady^{1*}, Raphael Ledermann², Neil Shephard³, Caroline A Evans⁴, Mark J. Dickman⁴, Robert Turner³, Aline Rifflet^{5,6,7}, Ankur Patel¹, Ivo Gomperts-Boneca^{5,6,7}, Philip S. Poole², Marshall Bern⁸, Stéphane Mesnage¹

¹School of Biosciences, University of Sheffield, Sheffield, UK

² Department of Biology, University of Oxford, Oxford, UK.

³ Research Software Engineer team, University of Sheffield, Sheffield, UK

⁴ Department of Chemical and Biological Engineering, ChELSI Institute, University of Sheffield, Sheffield, UK

⁵ Institut Pasteur, Unité Biologie et Génétique de la Paroi Bactérienne, Paris, France

⁶ INSERM, Équipe Avenir, Paris

⁷ CNRS, UMR 2001 Microbiologie intégrative et moléculaire, Paris, France

⁸ Protein Metrics Inc., 20863 Stevens Creek Boulevard #450, Cupertino, California 95014, United States.

* These authors contributed equally

Running title: Novel tools and strategy for LC-MS/MS peptidoglycan structural analysis

Keywords: cell wall, peptidoglycan, mass spectrometry, software

Address correspondence to Stéphane Mesnage, s.mesnage@sheffield.ac.uk

6.1 Abstract

Peptidoglycan is an essential component of the bacterial cell envelope — a mesh-like macromolecule that protects the bacterium from osmotic stress and its internal turgor pressure. The composition and architecture of peptidoglycan is heterogeneous and changes as bacteria grow, divide, and respond to their environment. Though peptidoglycan has long been studied via LC-MS/MS, the analysis of this data remains challenging as peptidoglycan's unusual composition and branching cannot be handled by proteomics software. Here we describe user-friendly open-source tools and a web interface for building peptidoglycan databases, performing MS searches, and predicting the MS/MS fragmentation of muropeptides. We then use *Rhizobium leguminosarum* to describe a step-by-step strategy for the high-resolution analysis of peptidoglycan. The unprecedented detail of *R. leguminosarum*'s peptidoglycan composition (>250 muropeptides) reveals even the subtlest remodelling between growth conditions. These new and easier to use tools enable more systematic analyses of peptidoglycan dynamics in the future.

6.2 Introduction

Peptidoglycan (PG) is a ubiquitous and essential component of the bacterial cell envelope, which forms a single bag-shaped macromolecule (or sacculus) around the cell (Vollmer et al., 2008a). PG synthesis has been extensively studied since many antibiotics work by disrupting it, including widely used β -lactam antibiotics (like penicillin) and last resort antibiotics such as vancomycin (Poole, 2004; Zapun et al., 2008b). The composition and remodelling dynamics of PG during growth, division, and differentiation can be critical for maintaining cell viability in response to changing environmental conditions. During this remodelling, PG fragments are naturally released into the environment; those released by the microbiota are important Microbe-Associated Molecular Patterns (MAMPs) recognised by the innate immune system (Chu and Mazmanian, 2013). They can contribute to acute or chronic inflammatory diseases and are thought to be key signalling molecules in the gut-brain axis (Gonzalez-Santana and Diaz Heijtz, 2020; Huang et al., 2019). PG fragments have also been shown to mediate more unusual symbiotic relationships, as in the case of the Hawaiian bobtail squid, where bioluminescent Vibrio fischeri provide the host with nocturnal camouflage (Nyholm and McFall-Ngai, 2021). PG's unique role in bacterial adaptation, pathogenesis, and symbiosis make it an essential molecule to study.

Whilst the overall structure of peptidoglycan and its building blocks are well conserved, it is continually restructured and modified as bacteria grow and divide, introducing vast and often subtle complexity. Monitoring PG structural dynamics, therefore, requires automated, robust, and sensitive tools. Most analyses currently involve a biased identification of major peaks in UV absorbance chromatograms (Alvarez et al., 2024) that precludes the identification of low abundance or co-eluting muropeptides. The limited number of muropeptides commonly described this way (usually 10–25, even for so-called "high-resolution analyses") (Rimal et al., 2022; van der Aart et al., 2018) does not provide enough detail to track the variation in important muropeptides like those corresponding to covalent protein anchoring. To achieve this greater level of detail, other studies have made use of LC-MS/MS — even proposing the term "peptidoglycomics" for the discipline in 2013 (Wheeler et al., 2014). Despite this, a lack of software tools and published search strategies means that the LC-MS/MS analysis of PG has remained a tediously manual, error-prone, and inconsistent process.

To address this, several more comprehensive tools for peptidoglycomics have recently been developed (Anderson et al., 2020; Hsu et al., 2023; Kwan et al., 2024; Patel et al., 2021). Whilst these tools have vastly improved the consistency and throughput of LC-MS/MS analysis, they remain either inflexible, incomplete, or difficult to use.

Our previously described tool, PGFinder (Patel et al., 2021; Rady and Mesnage, 2024), focused on ease-of-use and the quantification of muropeptides in LC-MS datasets but left room for improvement. Here, we build on PGFinder in two key ways: (i) by improving the usability and capability of the existing MS tool, and (ii) by including new modules that automate additional analysis steps in the LC-MS/MS pipeline. Highlights include PGFinder's new, user-friendly web interface (https://mesnage-org.github.io/pgfinder/) and PGLang, a formal language for the concise description of

muropeptides that enables both automated mass calculation and MS/MS fragment prediction. Finally, to demonstrate how these improvements fit into a complete analysis pipeline, we describe a step-by-step strategy that we use to characterise the changes in *Rhizobium leguminosarum*'s PG composition when grown on minimal (as opposed to rich) media. Empowered by this approach, we report unprecedented PG complexity (>250 muropeptides) and accurately monitor subtle changes in the PG, laying the groundwork for more systemic analyses of muropeptide composition, cross-linking, and protein anchoring in the future.

6.3 Results

6.3.1 Enhancing PGFinder's existing functionality with an improved MS output and web interface

The first published version of PGFinder (v0.02; (Patel et al., 2021)) offered automated MS analysis but search outputs still required post-processing in Excel to build a final table of muropeptides. Processing involved Δ ppm calculation and consolidation of intensities (sum) across retention times. Now, in version 1.3.2, PGFinder automatically picks the best match according to its Δ ppm and consolidates search results into a table of muropeptides sorted by abundance. Finally, a new metadata column makes it possible to keep track of the data analysed, parameters, and PGFinder version used to generate each output. Taken together, these changes are a major step towards reducing the amount of manual processing required.

To use PGFinder v0.02 without installation, we previously provided an interactive Jupyter notebook hosted on MyBinder (Jupyter et al., 2018). This made PGFinder significantly easier to set up and use than similar tools, but the resource limitations imposed by MyBinder regularly made loading our notebook slow (or even impossible). After loading, users also needed to ensure that all cells were run, in order, exactly once and needed to manually reset the notebook between each search. To circumvent these usability issues, we built a new, intuitive web interface that makes running an MS search as simple as uploading your deconvoluted data, picking a mass database, and clicking "Run Analysis" (Fig. 6-1). Since all computation is now done on the client-side (via WebAssembly), we no longer require hosting services like MyBinder and loading/computation times have been dramatically reduced. Moving to this interface has also allowed for bulk processing and made it trivial to add new modules like the Mass Calculator and Fragment Generator (Fig. 6-1). PGFinder is now easier to pick up than ever, further encouraging its adoption by others in the field.



Figure 6- 1. PGFinder's web interface makes MS analysis easier than ever and enables new functionality.

The interface includes the original MS Analysis module for identifying PG fragments from deconvoluted LC-MS data, as well as two newly developed modules: The Mass Calculator for building PGFinder-compatible mass databases, and the Fragment Generator for predicting MS/MS fragment ions.

6.3.2 Condensing complex muropeptide structures into PGLang, a concise formal language

Before PGFinder could be expanded to handle tasks like mass calculation or MS/MS fragment prediction, we needed a way to model muropeptide chemistry in software. PG building blocks (muropeptides) are made of *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) disaccharides linked to a short (possibly branched) peptide stem containing both *L*- and *D*-amino acids (Vollmer et al., 2008a) (Fig. 6-2A). In PGFinder, these muropeptides are represented as monosaccharide and amino acid residues that can be decorated with various modifications and bonded together to form a directed graph (Fig. 6-2B). Each residue contains distinct functional groups that are either free, modified, or donate/accept a particular bond (Fig. 6-2C). These rules ensure that every muropeptide is chemically valid, and tracking each muropeptide's free groups makes it possible to automatically identify potential modification sites and cross-linking positions.



Figure 6-2. Muropeptides are represented as a set of chemically linked, optionally modified residues.

(A) An example muropeptide showcasing several modifications (yellow) and peptide branches (red). (B) The same muropeptide converted to its chemical graph representation. Free functional groups (those that remain unmodified and unbonded) are shown as black dots on each residue, bonds are shown as labelled arrows pointing from donor to acceptor, and modifications are shown as yellow flags. (C) The five functional groups of *N*-Acetylmuramic acid are shown in detail, including the modifications each can have and the bonds they can donate/accept.

To represent these muropeptide graphs compactly, we developed a language called PGLang with a minimal and straightforward syntax (Fig.6- 3). Each monomer is partitioned into a glycan chain (represented by lowercase letters) and a stem peptide (represented by uppercase letters). Lateral chains can be attached to any diamino or dicarboxylic amino acids using square brackets, and any residue can be modified using round brackets. Monomers can be connected via their glycan chain (~), or via cross-linked stem peptides (=). When monomers are connected via cross-linked peptides, the structure is followed by a bracketed list of cross-link descriptors: 3-3, 4-3, etc. A complete syntax diagram for PGLang is available in Fig. S6-1, and tables detailing the currently available monosaccharides, amino acids, and modifications are provided in Tables S6-1-S6-2. Finally, to close the loop and move backwards from PGLang to a full molecular structure (including stereochemical information), we've included a PGLang to SMILES translator in PGFinder's new Mass Calculator module



Figure 6- 3. PGLang is a simple language for describing potentially modified, branched, or cross-linked muropeptides.

Each cartoon representation of a peptidoglycan fragment is color-coded to match its corresponding PGLang structure. Hexagons depict monosaccharides forming glycan chains, while circles represent amino acids. Modifications are indicated by flags extending from the modified residue (Am, Amidation; Anh, anhydroMurNAc; Glyc, Glycolylation). J represents *meso*-diaminopimelic acid.

6.3.3 Expanding PGF inder to automate mass calculation and MS/MS fragment prediction

Once muropeptides described using PGLang have been translated into their chemical graph representations, implementing a number of new features becomes straightforward. Here that means automating two additional parts of the analysis pipeline that were previously out of scope for PGFinder: monoisotopic mass calculation and MS/MS fragment prediction. The mass of any given muropeptide is simply the sum of its residue, modification, and bond masses, and fragment prediction is a three-step process involving bond-cleavage, ion formation (depending on the bond broken, the acceptor and donor fragments may gain or lose a particular chemical group), and mass-charge ratio (m/z) calculation (all ions are currently [M+H]+ adducts). Exposing this functionality are two new UI modules: the Mass Calculator which generates mass databases that can be fed directly into MS Analysis module, and the Fragment Generator which produces a list of ions with PGLang-like descriptions that make it clear to users what fragment each ion represents. To get users started quickly, the Mass Calculator also includes several PGLang databases for common model organisms that can be easily downloaded and adapted using a text editor.



Table S6-1. Railroad syntax diagram for PGLang

Any path from the top left to the bottom right corner of a PGLang structure, including loops or skipped sections, is valid. However, not all valid structures are chemically possible. PGLang uses lowercase letters 'a' to 'z', uppercase letters 'A' to 'Z', digits '0' to '9', and positions '1' to '5'.

Symbol	Name	Formula	Monoisotopic Mass
g	N-Acetylglucosamine	$C_8H_{15}NO_6$	221.089937
m	N-Acetylmuramic Acid	$C_{11}H_{19}NO_8$	293.111067
х	Unknown Monosaccharide		0
Α	Alanine	$C_3H_7NO_2$	89.047678
В	Diaminobutyric Acid	$C_4H_{10}N_2O_2$	118.074228
С	Cysteine	$C_3H_7NO_2S$	121.019750
D	Aspartic Acid	C ₄ H ₇ NO ₄	133.037508
E	Glutamic Acid	$C_5H_9NO_4$	147.053158
F	Phenylalanine	$C_9H_{11}NO_2$	165.078979
G	Glycine	$C_2H_5NO_2$	75.0320028
н	Histidine	$C_6H_9N_3O_2$	155.069477
1	Isoleucine	$C_6H_{13}NO_2$	131.094629
J	Diaminopimelic Acid	$C_7H_{14}N_2O_4$	190.095357
Κ	Lysine	$C_6H_{14}N_2O_2$	146.105528
L	Leucine	$C_6H_{13}NO_2$	131.094629
Μ	Methionine	$C_5H_{11}NO_2S$	149.051050
N	Asparagine	$C_4H_8N_2O_3$	132.053492
0	Ornithine	$C_5H_{12}N_2O_2$	132.089878
Ρ	Proline	$C_5H_9NO_2$	115.063329
Q	Glutamine	$C_5H_{10}N_2O_3$	146.069142
R	Arginine	$C_6H_{14}N_4O_2$	174.111676
S	Serine	$C_3H_7NO_3$	105.042593
Т	Threonine	C ₄ H ₉ NO ₃	119.058243
U	Homoserine	C ₄ H ₉ NO ₃	119.058243
V	Valine	$C_5H_{11}NO_2$	117.078979
W	Tryptophan	$C_{11}H_{12}N_2O_2$	204.089878
Х	Unknown Amino Acid		0
Υ	Tyrosine	$C_9H_{11}NO_3$	181.073893
Z	Threo-3-Hydroxyglutamic	$C_5H_9NO_5$	163.048072

Table S6-2. Residues built into PGFinder

Both "x" and "X" can be used as wild-cards with the PGLang offset syntax — e.g. "X(+C5H12N4O3)".

Table S6-3. Modifications built into PGFinder.

Symbol	Name	Lost Atoms	Gained Atoms	Net Monoisotopic Mass	Targeted Functional Groups
Ac	O-Acetylation	Н	C_2H_3O	42.010565	"Hydroxyl" at="6-Position"
Am	Amidation	OH	NH_2	-0.984016	"Carboxyl" at="Sidechain"
Anh	1,6-Anhydro	H ₂ O	-	-18.010565	"Hydroxyl" at="Reducing End" of="N-Acetylmuramic Acid"
DeAc	De-N-Acetylation	C_2H_3O	н	-42.010565	"Acetyl" at="Secondary Amide"
Glyc	Glycolylation	CH₃	CH ₂ OH	15.994915	"Acetyl" at="Secondary Amide" of="N-Acetylmuramic Acid"
Poly	Wall Polymer Linkage	н	PO ₃	77.950681	"Hydroxyl" at="6-Position"
Red	Reduced	-	H ₂	2.015650	"Hydroxyl" at="Reducing End"

Note that specific modifications can only target specific residues / functional groups. The groups each modification targets are described in the last column.

$6.3.4\ Rhizobium\ leguminosarum\ as\ a\ model\ system\ for\ describing\ a\ five-step\ PG$ analysis strategy

The *R. leguminosarum* genome encodes many D,D and L,D-transpeptidases, so its PG structure is expected to be complex, making it a good model organism for testing our new PG analysis tools. To prepare some sample datasets, triplicate cultures of *R. leguminosarum* were grown in both minimal and rich media, and their PG was analysed via UHPLC-MS/MS (Fig. 6-4).



Figure 6- 4. *R. leguminosarum's* PG compositon is comples and varies between growth conditions.

Total ion chromatograms (TIC's) show reduced *R. leguminosarum* muropeptides. PG was extracted from cells grown in either rich (**TY**) or minimal (**MM**) media. The TICs corresponding to each triplicate are shown.

The chromatograms confirmed that *R. leguminosarum's* muropeptide profile was complex and revealed differences between the two media conditions. Consequentially, the corresponding LC-MS/MS datasets were ideal for showcasing our comparative PG analysis strategy:

Four sequential searches focus on monomers (step 1), modifications (step 2), PG-anchored proteins (step 3) and multimers (step 4) that inform a final, fifth search producing a comprehensive muropeptide quantification that can be used in statistical comparisons (Fig.6-5).



Figure 6-5. An end-to-end strategy for PG structural analysis via LC-MS/MS.

Sequential searches were performed using PGFinder and Byonic[™]. The monomer database DB_2 was built based on MS/MS analysis. The most abundant monomers were then used to build DB_3, which was used to identify modified muropeptides. DB_4 contained muropeptides with a gm-AEJ stem followed by *N*-terminal porin sequences (with signal peptides removed). DB_2 was used to identify dimers and trimers, then MS/MS data from matching output was manually inspected to differentiate between mass coincidences. A final search was carried out with DB_5, which combines muropeptides from DB_2 (monomers) and the muropeptides from DB_4 corresponding to the MS/MS confirmed RopA1,2,3, RopB and pRL90069 porins. The final PGFinder search, with anhydroMurNAc modifications and 3-3 / 4-3 multimers enabled, was carried out with a 5 ppm tolerance. The final search output was manually inspected and modified to remove any known mass coincidences.

6.3.5 Step1: Identifying PG monomers using MS and MS/MS

An initial, unbiased search of the TY datasets was performed using PGFinder's "MS Analysis" module. The monomer database (DB_1; Table S6-3) contained 223 disaccharide peptides with stem lengths ranging from one to five amino acids. A total of 131 unique matches were found within 20 ppm of the observed masses; 78 of which were found in all three datasets, 22 in only two, and 31 in just one (Table S6-4).

Structure	Monoisotopicmass	Structure	Monoisotopicmass
gm-AE 1	698.28590	gm-AEJHE 1	1136.47219
gm-AEJ 1	870.37069	gm-AEJHH 1	1144.48851
gm-AEJA 1	941.40783	gm-AEJHP 1	1104.48236
gm-AEJAA 1	1012.44497	gm-AEJHQ 1	1135.48818
gm-AEJC 1	973.37988	gm-AEJHS 1	1094.46163
gm-AEJCA 1	1044.41702	gm-AEJHT 1	1108.47728
gm-AEJCC 1	1076.38907	gm-AEJHW 1	1193.50891
gm-AEJCE 1	1102.42247	gm-AEJI 1	983.45475
gm-AEJCG[1	1030.40134	gm-AEJIA 1	1054.49189
gm-AEJCH 1	1110.43879	gm-AEJID 1	1098.48169
gm-AEJCK 1	1101.47484	gm-AEJIE	1112.49734
gm-AEJCN 1	1087.42281	gm-AEJIH 1	1120.51366
gm-AEJCP 1	1070.43264	gm-AEJII 1	1096.53881
gm-AEJCS[1	1060.41191	gm-AEJIK 1	1111.54971
gm-AEJD 1	985.39763	gm-AEJIN 1	1097.49768
gm-AEJDA 1	1056.43477	gm-AEJIP 1	1080.50751
gm-AEJDCl1	1088.40682	gm-AEJIQI1	1111.51333
gm-AEJDD 1	1100.42457	gm-AEJIS 1	1070.48678
gm-AEJDE 1	1114.44022	gm-AEJIT 1	1084.50243
gm-AEJDG 1	1042.41909	gm-AEJIW 1	1169.53406
gm-AEJDH 1	1122.45654	gm-AEJK 1	998.46565
gm-AEJDN 1	1099.44056	gm-AEJKA 1	1069.50279
gm-AEJDP 1	1082.45039	gm-AEJKD 1	1113.49259
gm-AEJDQ 1	1113.45621	gm-AEJKE[1	1127.50824
gm-AEJDT 1	1086.44531	gm-AEJKG 1	1055.48711
gm-AEJDW 1	1171.47694	gm-AEJKH 1	1135.52456
gm-AEJE 1	999.41328	gm-AEJKI 1	1111.54971
gm-AEJEA 1	1070.45042	gm-AEJKK 1	1126.56061
gm-AEJED 1	1114.44022	gm-AEJKN 1	1112.50858
gm-AEJEE 1	1128.45587	gm-AEJKP 1	1095.51841
gm-AEJEG 1	1056.43474	gm-AEJKQ1	1126.52423
gm-AEJEH 1	1136.47219	gm-AEJKR 1	1154.56676
gm-AEJEI 1	1112.49734	gm-AEJKT 1	1099.51333
gm-AEJEN 1	1113.45621	gm-AEJKW 1	1184.54496
gm-AEJES 1	1086.44531	gm-AEJL 1	983.45475
gm-AEJEV 1	1098.48169	gm-AEJLC 1	1086.46394
gm-AEJF 1	1017.43910	gm-AEJLG 1	1040.47621
gm-AEJFA 1	1088.47624	gm-AEJLV 1	1082.52316
gm-AEJFC 1	1120.44829	gm-AEJM 1	1001.41118
gm-AEJFD 1	1132.46604	gm-AEJMA 1	1072.44832
gm-AEJFE 1	1146.48169	gm-AEJMC 1	1104.42037
gm-AEJFF 1	1164.50751	gm-AEJMD 1	1116.43812
gm-AEJFG 1	1074.46056	gm-AEJME 1	1130.45377
gm-AEJFH 1	1154.49801	gm-AEJMF 1	1148.47959
gm-AEJFI 1	1130.52316	gm-AEJMH 1	1138.47009
gm-AEJFK 1	1145.53406	gm-AEJMI 1	1114.49524
gm-AEJFN 1	1131.48203	gm-AEJMK 1	1129.50614
gm-AEJFP 1	1114.49186	gm-AEJMM 1	1132.45167
gm-AEJFQ 1	1145.49768	gm-AEJMN 1	1115.45411
gm-AEJFR 1	1173.54021	gm-AEJMP 1	1098.46394
gm-AEJFS 1	1104.47113	gm-AEJMQ 1	1129.46976
gm-AEJFT 1	1118.48678	gm-AEJMS 1	1088.44321
gm-AEJFV 1	1116.50751	gm-AEJMT 1	1102.45886
gm-AEJFW 1	1203.51841	gm-AEJMV 1	1100.47959
gm-AEJFY 1	1180.50243	gm-AEJMW 1	1187.49049
gm-AEJG 1	927.39215	gm-AEJMY 1	1164.47451

Table S6- 4. Database 1 (DB_1)

Structure	Monoisotopicmass	Structure	Monoisotopicmass
gm-AEJGA 1	998.42917	gm-AEJN 1	984.41362
gm-AEJGG[1	984.41361	gm-AEJNA 1	1055.45076
gm-AEJGH 1	1064.45106	gm-AEJNE 1	1113.45621
gm-AEJGM 1	1058.43264	gm-AEJNG 1	1041.43508
gm-AEJGP 1	1024.44491	gm-AEJNH 1	1121.47253
gm-AEJGQ 1	1055.45073	gm-AEJNN 1	1098.45655
gm-AEJGR 1	1083.49326	gm-AEJNP 1	1081.46638
gm-AEJGS 1	1014.42418	gm-AEJNQ 1	1112.47220
gm-AEJGV 1	1026.46056	gm-AEJNT 1	1085.46130
gm-AEJGW 1	1113.47146	gm-AEJNW 1	1170.49293
gm-AEJH 1	1007.42960	gm-AEJP 1	967.42345
gm-AEJHA 1	1078.46674	gm-AEJPA 1	1038.46059
gm-AEJPE 1	1096.46604	gm-AEJTE 1	1100.46096
gm-AEJPP 1	1064.47621	gm-AEJTG 1	1028.43983
gm-AEJPQ 1	1095.48203	gm-AEJTK 1	1099.51333
gm-AEJPT 1	1068.47113	gm-AEJTT 1	1072.46605
gm-AEJPW 1	1153.50276	gm-AEJTV 1	1070.48678
gm-AEJQ 1	998.42917	gm-AEJV 1	969.43910
gm-AEJQA 1	1069.46641	gm-AEJVA 1	1040.47624
gm-AEJQC 1	1101.43846	gm-AEJVC 1	1072.44829
gm-AEJQE 1	1127.47186	gm-AEJVD 1	1084.46604
gm-AEJQG 1	1055.45073	gm-AEJVE 1	1098.48169
gm-AEJQQ 1	1126.48785	gm-AEJVH 1	1106.49801
gm-AEJQS 1	1085.46130	gm-AEJVK 1	1097.53406
gm-AEJQT 1	1099.47695	gm-AEJVN 1	1083.48203
gm-AEJQW 1	1184.50858	gm-AEJVP 1	1066.49186
gm-AEJR 1	1026.47180	gm-AEJVQ 1	1097.49768
gm-AEJRA 1	1097.50894	gm-AEJVS 1	1056.47113
gm-AEJRC 1	1129.48099	gm-AEJVT 1	1070.48678
gm-AEJRD 1	1141.49874	gm-AEJVV 1	1068.50751
gm-AEJRE 1	1155.51439	gm-AEJVW 1	1155.51841
gm-AEJRH 1	1163.53071	gm-AEJW 1	1056.45000
gm-AEJRI 1	1139.55586	gm-AEJWA 1	1127.48714
gm-AEJRM 1	1157.51229	gm-AEJWC 1	1159.45919
gm-AEJRN 1	1140.51473	gm-AEJWE 1	1185.49259
gm-AEJRP[1	1123.52456	gm-AEJWT 1	1157.49768
gm-AEJRQ 1	1154.53038	gm-AEJWW/1	1242.52931
gm-AEJRR[1	1182.57291	gm-AEJY 1	1033.43402
gm-AEJR1[1	1127.51948	gm-AEJYA 1	1104.47116
gm-AEJRV[1	1125.54021	gm-AEJYC[1	1136.44321
gm-AEJRW[I	1212.55111	gm-AEJYD I	1148.46096
gm-AEJSII	957.40272	gm-AEJYEJI	1000 455 40
gm-AEJSAJI	1028.43986	gm-AEJYGII	1090.45548
gm-AEJSD[1	1072.42966	gm-AEJYH I	1170.49293
gm-AEJSEIT	1086.44531	gm-AEJYIJI	1146.51808
gm-AEJSKII	1085.49768	gm-AEJYK I	1161.52898
gm-AEJSNII	1071.44363	gm-AEJYNJI	1147.47695
gIII-AEJOPII	1004.40048	grii-AEJYP I	113U.48078
gIII-AEJOWI	1000.4013U	grii-AEJYQ I	1101.4920U 1100.59519
giii-AEJOKII	1044 40475		1100 40005
gIII-AEJOOII	1044.43473	giii-AEJISII	1120.40000 1194.40170
gIII-AEJOIII dm AEISIA/11	11/20.40040	giii-AEJIII dm AEIXV/li	1134.40170
gIII-AEJOVVII	1143.48203 071.41097	giii-AEJIVII	1010 51999
dm AEJTAI1	971.41007 104945551	grii-AEJTVVII dm AEIVVII	1213.01000
gm-ALJTALT dm AEITCI1	1042.40001	giii-ALUTTII	1190.49700
gill-ALUI UII	1074.42700		

Several matches had highly unusual compositions (e.g., gm-AEJCC or gm-AEJYS) and relatively high Δ ppm values (8.8 and 9.1, respectively), suggesting that these identifications resulted from mass coincidences. Manual inspection of MS spectra confirmed this hypothesis; as many ions matching the theoretical *m*/*z* of these unusual muropeptides lacked the signature ion corresponding to the loss of GlcNAc loss due to in-source fragmentation. To screen out these mass coincidences and resolve the structure of any isomers, we confirmed each monomer via MS/MS.

	Muropeptide ^a	TY1 ^b	TY2	ТҮЗ	RT	' (m	in)	Theoretical mass (Da)	∆ppm	Present in
1	dm AFIAI1	34 346%	13 678%	45.233 %	95	+	0.0	941.4078	0.2	3
2	dm AFIGI1	0 361%	9016%	0 535%	6.8	- +	0.0	927 3922	07	3
2		8 833%	6 551%	1 017%	12.5	÷ +	0.0	1127 4719	12	3
1	am AE II1	6344%	5 207%	7.01770	5.5	- +	0.0	870 3707	0.5	3
5	dm-AE.IIAI1	5.847%	6.038%	5 729%	0.0 01/	+ +	0.0	1054 4919	0.9	3
6	om-ΔE.IEI1	5 4 2 3 %	5 917%	5 617%	21.4	+	0.0	1017.4391	2.0	3
7	dm-AEIAAl1	1862%	2 801%	9 779%	10.9	+	0.0	1012,4450	1.2	3
, 8	om-ΔF.IFΔI1	2 248%	2.001/0	1456%	15.4	+	0.0	1088.4762	0.7	3
9	$gm_{AE} \Delta G 1 gm_{AE} G 1$	1605%	1.697%	1.400%	8.8	+	0.0	998.4292	1.2	3
10	gm-AF.IKI1	2 314%	1.576%	1371%	77	+	0.0	998.4656	1.6	3
11	om-ΔΕ.IPΔI1	3 017%	1059%	0.695%	29.3	+	0.0	1038.4606	9.1	3
12	gm-AF.IYI1	1 251%	1432%	1202%	16.8	+	0.0	1033.4340	1.4	3
13	gm-AF.IWWI1	2 392%	0.508%	0.898%	9.9	+	01	1242.5293	6.2	3
14	gm-AF.IVCl1 gm-AF.IMAl1	1066%	1 216%	1.017%	16.6	+	01	1072.4483	1.8	3
15	gm-AF.IRCI1	1004%	1288%	0.984%	22.2	+	0.0	1129.4810	7.5	3
16	gm-AEJGGI1 gm-AEJNI1	0.955%	0.961%	1178%	5.0	+	0.0	984.4136	0.8	3
17	gm-AF.IDAl1 gm-AF.IFGl1	1376%	0.807%	0 701%	9.6	+	3.9	1056.4348	1.4	3
18	gm-AF.IMI1	0 790%	0.897%	0.648%	15.6	+	0.0	1001.4112	0.5	3
	8	0110070	0.00170	0.01070	20.	-	0.0	000 45 47	0.0	0
19	gm-AEJI 1	0.745%	0.722%	0.611%	3	±	0.0	983.4547	0.8	3
20	gm-AE 1	0.875%	0.423%	0.680%	8.7	±	0.0	698.2859	0.3	3
21	gm-AEJS 1	0.643%	0.403%	0.753%	5.8	±	5.0	957.4027	0.4	3
22	gm-AEJWA 1	0.541%	0.641%	0.510%	25.2	±	0.0	1127.4871	1.3	3
23	gm-AEJH 1	0.427%	0.343%	0.334%	5.9	±	0.0	1007.4296	1.2	3
24	gm-AEJCC 1	0.619%	0.235%	0.150%	29.3	±	0.0	1076.3891	8.9	3
25	gm-AEJR 1	0.292%	0.270%	0.272%	6.7	±	0.0	1026.4718	0.6	3
26	gm-AEJW 1	0.244%	0.276%	0.217%	23.9	±	0.0	1056.4500	1.6	3
27	gm-AEJV 1	0.230%	0.255%	0.237%	14.6	±	0.1	969.4391	0.9	3
28	gm-AEJQA 1	0.206%	0.198%	0.215%	11.0	±	0.0	1069.4664	1.9	3
29	gm-AEJMY 1	0.403%	0.131%	0.054%	25.9	±	0.0	1164.4745	5.1	3
30	gm-AEJHA 1	0.209%	0.197%	0.169%	7.3	±	0.0	1078.4667	1.9	3
31	gm-AEJFS 1, gm-AEJHP 1,	0.187%	0.179%	0.196%	17.4	±	0.2	1104.4712	1.1	3
32	gm-AFJTAJT gm-AFJTI	0.169%	0.184%	0.165%	7.6	+	8.1	971.4184	1.9	3
33	gm-AF.IFAl1	0 143%	0.147%	0.226%	10.7	+	22	1070.4504	1.9	3
34	gm-AEJKDI1. gm-AEJSRI1	0.235%	0.130%	0.147%	9.1	±	0.1	1113.4926	1.4	3
	8				30.			1005 5104	4.0	0
35	gm-AEJKP 1	0.166%	0.168%	0.148%	3	±	0.0	1095.5184	4.3	3
36	gm-AEJIQ 1	0.064%	0.165%	0.239%	27.3	±	0.6	1111.5133	4.2	3
37	gm-AEJST 1	0.219%	0.065%	0.086%	23.1	±	0.0	1058.4504	2.7	3
38	gm-AEJMS 1	0.125%	0.114%	0.108%	10.1	±	0.1	1088.4432	2.0	3
39	gm-AEJEV 1, gm-AEJID 1	0.109%	0.103%	0.095%	18.7	±	0.0	1098.4817	0.7	3
40	gm-AEJRA 1	0.092%	0.113%	0.099%	9.1	±	6.3	1097.5089	1.9	3
41	gm-AEJKA 1	0.135%	0.082%	0.071%	7.3	±	0.0	1069.5028	2.2	3
42	gm-AEJVP 1	0.123%	0.076%	0.066%	31.9	±	0.0	1066.4919	2.1	3
43	gm-AEJCG 1	0.119%	0.096%	0.050%	10.8	±	14.6	1030.4013	1.3	3

Table S6-5. Rhizobium leguminosarum unbiased search with DB_1.

	Muropeptide ^a	TY1 ^b	TY2	ТҮЗ	RT	۲ (mi	in)	Theoretical mass (Da)	Δppm	Present in
44	gm-AEJD 1	0.104%	0.030%	0.117%	7.0	±	0.1	985.3976	0.3	3
45	gm-AEJTA 1	0.095%	0.098%	0.046%	22.1	±	0.0	1042.4555	2.7	3
46	gm-AEJTC 1	0.082%	0.050%	0.059%	24.2	±	0.0	1074.4276	3.2	3
47	gm-AEJKQ 1	0.058%	0.059%	0.051%	7.8	±	0.0	1126.5242	1.4	3
48	gm-AEJEI 1, gm-AEJKN	1 0.050%	0.058%	0.059%	26.4	±	1.4	1112.4973	2.9	3
49	gm-AEJMC 1	0.078%	0.036%	0.051%	11.4	±	0.0	1104.4204	5.5	3
50	gm-AEJR 1, gm-AEJGV	1 0.044%	0.055%	0.064%	30.1	±	2.6	1026.4718	1.3	3
51	gm-AEJCH 1	0.080%	0.051%	0.026%	28.9	±	0.0	1110.4388	5.6	3
52	gm-Aejin 1, gn AejvQ 1, gm-AejRA 1	n- 0.046%	0.050%	0.061%	15.8	±	7.1	1097.4977	1.1	3
53	gm-AEJKG 1	0.049%	0.069%	0.038%	16.5	±	0.0	1055.4871	4.0	3
54	gm-AEJCK 1	0.062%	0.043%	0.041%	28.5	±	8.1	1101.4748	5.9	3
	gm-AEJDT 1, gn	n-						1086 4453	26	3
55	AEJES 1	0.072%	0.034%	0.036%	22.5	±	0.0	100011100	2.0	
56	gm-AEJHT 1	0.049%	0.054%	0.034%	18.0	±	0.0	1108.4773	2.1	3
57	gm-AEJTE 1	0.093%	0.013%	0.026%	18.8	±	0.0	1100.4610	7.6	3
58	gm-AEJNA 1, gn	n-	0.046%	0.044%	77	+	17	1055.4508	1.3	3
50		0.044%	0.04070	0.04470	10.0	÷.	0.0	1173 5402	14	3
29	gm-AEJFRI	0.052%	0.036%	0.042%	10.9	Ξ	0.0	1194 4017	15	2
60	gm-AEJYT 1	0.075%	0.029%	0.026%	8.6	±	0.0	1134.4017	1.5	3
61	gm-AEJP 1	0.089%	0.016%	0.021%	28.6	±	0.0	967.4235	8.0	3
62	gm-AEJRH 1	0.050%	0.031%	0.029%	21.4	±	0.8	1163.5307	6.8	3
63	gm-AEJYC 1	0.040%	0.040%	0.025%	10.7	±	0.0	1136.4432	1.5	3
64	gm-AEJSA 1, gn AEJTG 1	n- 0.025%	0.030%	0.044%	8.7	±	0.2	1028.4399	0.8	3
65	gm-AEJYI 1	0.068%	0.012%	0.015%	29.2	±	0.0	1146.5181	4.5	3
66	gm-AEJIP 1	0.033%	0.021%	0.039%	16.0	±	0.1	1080.5075	2.6	3
67	gm-AEJQT 1	0.031%	0.024%	0.026%	4.7	±	0.0	1099.4769	1.0	3
68	gm-AEJRV 1	0.035%	0.023%	0.021%	28.0	±	3.4	1125.5402	7.4	3
69	gm-AF,IIHI1	0.045%	0.013%	0.015%	17.4	+	0.0	1120.5137	2.7	3
70	gm_AEJDNI1	0.026%	0.025%	0.021%	16.2	+	0.4	1099.4406	3.0	3
71		0.020%	0.02070	0.021/0	00.5	- -	0.4	1132,4660	2.9	3
71		0.017%	0.020%	0.020%	20.5	<u>+</u>	0.0	1104 4894	8.8	2
72	gm-AEJHP[1	0.031%	0.018%	0.013%	28.5	±	0.0	1146 4017	5.0	0
73	gm-AEJFE 1	0.037%	0.008%	0.014%	24.2	±	0.0	1140.4817	5.1	3
74	gm-AEJYS 1	0.020%	0.023%	0.014%	10.7	±	0.0	1120.4660	9.1	3
75	gm-AEJFY 1	0.023%	0.018%	0.013%	10.3	±	0.0	1180.5024	2.3	3
76	gm-AEJRD 1	0.022%	0.018%	0.009%	21.2	±	0.0	1141.4987	4.9	3
77	gm-AEJVS 1	0.031%	0.009%	0.010%	15.9	±	0.0	1056.4711	2.1	3
78	gm-AEJPW 1	0.019%	0.011%	0.016%	8.2	±	0.0	1153.5028	6.0	3
79	gm-AFJSRI1	ND	0.009%	0.230%	27.1	+	0.0	1113.5038	7.9	2
80	gm-ΔF.IMDI1	0.075%	0.058%		14.3	+	8.8	1116.4381	0.4	2
01		0.01070	0.00070	0.050%	20.0	- -	0.0	1114,4919	8.3	2
01	gill-AEJFF I		0.039%	0.000%	30.9	1	0.0	1130 / 868	0 1	-
82	gm-AEJYP[1	0.053%	ND	0.022%	28.9	±	0.0	1101.4705	2.1	2
83	gm-AEJNH 1	0.041%	ND	0.030%	27.3	±	0.0	1121.4725	1.0	z
84	gm-AEJIW 1	0.051%	ND	0.012%	23.2	±	0.0	1169.5341	8.0	2
85	gm-AEJE 1	0.016%	0.047%	ND	18.3	±	0.0	999.4133	2.0	2
86	gm-AEJLC 1	0.043%	ND	0.017%	25.0	±	0.0	1086.4639	8.0	2
87	gm-AEJVN 1, gn AEJGR 1	n- 0.034%	0.026%	ND	28.7	±	0.0	1083.4820	3.0	2

	Muropeptide ^a	TY1 ^b	TY2	ТҮЗ	R	Г (mi	in)	Theoretical mass (Da)	∆ppm	Present in
88	gm-AEJGR 1	0.042%	0.017%	ND	21.0	±	0.0	1083.4933	4.9	2
89	gm-AEJKD 1	ND	0.043%	0.015%	20.9	±	0.0	1113.4926	6.5	2
90	gm-AEJRW 1	ND	0.015%	0.032%	28.9	±	0.1	1212.5511	9.9	2
91	gm-AEJYD 1	0.031%	ND	0.009%	22.4	±	0.0	1148.4610	3.6	2
92	gm-AEJKE 1,gm-AEJRT 1	ND	0.023%	0.015%	16.2	±	0.0	1127.5082	3.6	2
93	gm-AEJRI 1	0.022%	ND	0.009%	28.9	±	0.0	1139.5559	8.4	2
94	gm-AEJED 1	ND 0.0140/	0.008%	0.023%	10.0	±	0.0	1114.4402	0.2	2
95	gm-AEJKHII	0.014%	0.013%	ND	13.8	±	0.0	110/ 1711	2.3	2
90 07	gm-AEJFS[1,gm-AEJTA]1		0.012%	0.010%	29.3	Ξ +	0.0	1064.4711	5.2 5.5	2
98	dm-AEJMT1	0.010%	0.012 <i>7</i> 0	0.010%	20.2 91.9	+ +	0.0	1102 4589	5.5	2
99	gm-AFJRNI1	ND	0.010%	0.009%	26.8	+	0.0	1140.5147	4.7	2
	gm-AEJMF 1,gm-		0101070	0.00070	2010	-	0.0	1140 4700	6.0	0
100	AEJYD 1	ND	0.005%	0.006%	9.2	±	0.0	1148.4796	6.9	2
101	gm-AEJNP 1	0.064%	ND	ND	31.9	±	9.7	1081.4664	8.8	1
102	gm-AEJSD 1	0.054%	ND	ND	22.1	±	0.0	1072.4297	7.1	1
103	gm-AEJFH 1	0.051%	ND	ND	16.5	±	0.0	1154.4980	7.6	1
104	gm-AEJMP 1	0.041%	ND	ND	13.3	±	0.0	1098.4639	3.8	1
105	gm-AEJFP 1, gm-AEJMI 1	0.032%	ND	ND	21.7	±	0.0	1114.4919	1.7	1
106	gm-AEJQQ 1	0.031%	ND	ND	30.3	±	0.0	1126.4878	0.5	1
107	gm-AEJGS 1	0.030%	ND	ND	20.9	±	9.9	1014.4242	9.8	1
108	gm-AEJRM 1	0.027%	ND	ND	27.9	±	0.0	1157.5123	7.5	1
109	gm-AEJYN 1	0.024%	ND	ND	17.1	±	0.0	1147.4769	5.0	1
110	gm-AEJDG 1	ND	ND	0.023%	7.7	±	0.0	1042.4191	9.3	1
111	gm-AEJYQ 1	ND	ND	0.022%	17.1	±	0.0	1161.4926	1.1	1
112	gm-AEJNN 1	0.022%	ND	ND	16.9	±	11.1	1098.4566	8.5	1
113	gm-AEJMV 1	ND	0.021%	ND	31.4	±	0.0	1100.4796	8.3	1
114	gm-AEJEH 1	ND	ND	0.019%	27.3	±	0.0	1136.4722	5.6	1
115	gm-AEJFT 1	0.016%	ND	ND	24.6	±	0.0	1118.4868	1.8	1
116	gm-AEJGPl1	ND	0.013%	ND	28.4	±	0.0	1024.4449	9.1	1
117	gm-AEJWE 1	0.011%	ND	ND	28.7	±	0.0	1185.4926	3.6	1
118	gm-AEJKR 1	ND	0.011%	ND	23.5	±	0.0	1154.5668	9.5	1
119	gm-AEJHH 1	ND	ND	0.010%	10.2	±	0.0	1144.4885	0.2	1
100	gm-AEJGM 1, gm-	ND	ND	0.0100/	10.0		0.0	1058.4326	4.5	1
120	aejkaji, gm-AEJIN 1,		ND	0.010%	13.8 00.0	±	0.0	1007 5000	0.0	-
121	gm-AEJVQ 1	ND	ND	0.010%	26.8	±	0.0	1097.5089	0.2	I
122	gm-AEJSTIT, gm- AEJGM 1	ND	0.010%	ND	13.8	±	0.0	1058.4504	7.3	1
123	gm-AEJMF 1	0.010%	ND	ND	9.0	±	0.0	1148.4796	5.9	1
124	gm-AEJMN 1	0.010%	ND	ND	21.9	±	0.0	1115.4541	9.7	1
125	gm-AEJHQ 1	ND	ND	0.010%	28.1	±	0.0	1135.4882	5.0	1
126	gm-AEJSN 1	0.009%	ND	ND	23.9	±	0.0	1071.4456	9.6	1
127	gm-AEJKT 1	0.009%	ND	ND	12.8	±	0.0	1099.5133	3.5	1
128	gm-AEJMP 1, gm- AEJNN 1	ND	ND	0.008%	8.0	±	0.0	1098.4639	0.9	1
129	gm-AEJKN 1	ND	ND	0.007%	23.7	±	0.0	1112.5086	8.9	1
130	gm-AEJKW 1	ND	0.007%	ND	11.5	±	0.0	1184.5450	6.5	1
131	gm-AEJRT 1	ND	0.005%	ND	4.2	±	0.0	1127.5195	4.9	1

^ag, GlcNAc; m, MurNAc ^b ND, Not Detected

We carried out a search of the TY datasets using the ByonicTM module of the Byos® software that can automatically analyse and score MS/MS spectra. The list of

monomers to search for contained 423 muropeptides with stem lengths ranging from one to five amino acids, including disaccharide-tetrapeptides with all possible residues in position four (gm-AEJX) and disaccharide-pentapeptides with all possible combinations in positions four and five (gm-AEJXX). Based on the automatic scoring of MS/MS spectra, ByonicTM confirmed 39 monomers (Fig. S6-2).

Monomer	Max ions	h ions	v ions
Wohomen	expected	0 10113	y 10113
gm-AE	1	1	1
gm-AEJ	2	2	2
gm-AEJA gm-AEJD gm-AEJF gm-AEJG gm-AEJH gm-AEJI gm-AEJK gm-AEJM gm-AEJN gm-AEJQ * gm-AEJS gm-AEJT gm-AEJV gm-AEJV gm-AEJW gm-AEJY	3	3 2 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3	3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3
gm-AEJAA gm-AEJAD gm-AEJAF gm-AEJAF gm-AEJAH gm-AEJAH gm-AEJAI gm-AEJAK gm-AEJAM gm-AEJAQ gm-AEJAR gm-AEJAR gm-AEJAW	4	4 4 4 3 4 3 4 3 4 3 4 3 4	4 3 4 3 4 4 4 4 4 4 4 4

* exact mass coincidence

gm-AEJQ, gm-AEJAG = 998.429165 Da







NON Validated MS/MS spectra:

Monomer	Max ions expected	b ions	y ions
gm-AEJR	3	1	3
gm-AEJDK		2	3
gm-AEJQK		2	3
gm-AEJMS		2	3
gm-AEJGA *		3	4
gm-AEJAN *	4	3	3
gm-AEJAY *		4	4
gm-AEJQE **		3	3
gm-AEJEQ **		3	4
gm-AEJWW **		4	3

* only one MS/MS spectrum across 3 biological replicates

** mass coincidence (EQ=QE=ADA; WW=JEA)





Figure S6- 1. Validated and non-validated MS/MS spectra (more than 50% of the expected *b*- and *y*- ions)

Further, manual inspection of the Byonic[™] output allowed us to validate an additional 29 monomers that satisfied two criteria: they were fragmented in at least one replicate and contained at least half of the expected b-ions and half of the expected y-ions. The validated muropeptides contained many tetra- and pentapeptide stems with unusual residues in their final position — indicative of D,D or L,D-transpeptidase exchange activity. Two structures with identical masses, however, could not be differentiated with certainty (gm-AEJAG and gm-AEJQ, both 998.429165 Da). Seven monomers did not meet the criteria for validation (only one MS/MS spectrum across three biological replicates or less than half of the expected b- or y-ions). The lack of b- or y-ions for the three monomers (gm-AEJQE, gm-AEJEQ and gm-AEJWW) prompted us to explore the corresponding MS/MS spectra and revealed that ByonicTM's automatic identification of these monomers was incorrect. The gm-AEJQE and gm-AEJEQ monomers were found to really be gm-AEJADA, whilst gm-AEJWW was in fact gm-AEJ=AEJ (3-3) (a dimer of tripeptides missing a disaccharide molety). The 29 validated monomers that were present in all three TY replicates became the monomer database DB 2 (Table S6-5).

Table S6-6. Monomers identified	by MS/MS in TY datasets.
---------------------------------	--------------------------

Structure	Monoisotopicmass
gm-AE 1	698.28580
gm-AEJ 1	870.37059
gm-AEJA 1	941.40770
gm-AEJAA 1	1012.44482
gm-AEJAD 1	1056.43465
gm-AEJAE 1	1070.45030
gm-AEJAF 1	1088.47612
gm-AEJAG 1	998.42917
gm-AEJAH 1	1078.46661
gm-AEJAI 1	1054.49177
gm-AEJAK 1	1069.50267
gm-AEJAM 1	1072.44819
gm-AEJAQ 1	1069.46628
gm-AEJAR 1	1097.50881
gm-AEJAW 1	1127.48702
gm-AEJD 1	985.39753
gm-AEJF 1	1017.43900
gm-AEJG 1	927.39205
gm-AEJH 1	1007.42950
gm-AEJI 1	983.45465
gm-AEJK 1	998.46555
gm-AEJM 1	1001.41107
gm-AEJN 1	984.41352
gm-AEJQ 1	998.42917
gm-AEJS 1	957.40262
gm-AEJT 1	971.41827
gm-AEJV 1	969.43900
gm-AEJW 1	1056.44990
gm-AEJY 1	1033.43392

6.3.4 Step 2: Identifying PG modifications

Database DB_2 was then used to determine which monomers were the most abundant using PGFinder. A subset of 11 muropeptides, accounting for >90% of the monomer abundance (Table S6-6), were selected to create a third database called DB_3 (Table S6-7). DB_3 was then used to search for six different modifications using PGFinder. The modifications considered are listed in Table 1 and can be sorted into glycan modifications (deacetylation, O-acetylation, and 1,6-anhydroMurNAc), peptide modifications (amidation), and hydrolysis products resulting from Glucosaminidase (loss of GlcNAc) or amidase activity (presence of an extra GlcNAc-MurNAc). For each modification, matches were consolidated (summing the intensities of matches found at different retention times) and matches absent from any of the three replicates were discarded. Three additional criteria were then used to validate the modified muropeptide matches: (i) a retention time consistently higher or lower (depending on the modification considered) than the unmodified muropeptide; (ii) the presence of signature ions corresponding to each modification; (iii) a similar relative abundance of modified and unmodified muropeptides (Table 6-1).

	Christense	Ave	Cumulated				
	Structure	Intensity	Abundance	abundance			
1	gm-AEJA 1	1.07E+09	47.90%	47.90%			
2	gm-AEJG 1	2.43E+08	10.84%	58.74%			
3	gm-AEJ 1	1.64E+08	7.33%	66.07%			
4	gm-AEJA[I/L] 1	1.54E+08	6.85%	72.92%			
5	gm-AEJF 1	1.48E+08	6.59%	79.51%			
6	gm-AEJAA 1	6.55E+07	2.92%	82.43%			
7	gm-AEJAF 1	5.00E+07	5.00E+07 2.23%				
8	gm-AEJAG 1,gm-AEJQ 1	4.61E+07	4.61E+07 2.06%				
9	gm-AEJK 1	4.60E+07	2.05%	88.77%			
10	gm-AEJY 1	3.38E+07	1.51%	90.28%			
11	gm-AEJAM 1	2.87E+07	1.28%	91.56%			
12	gm-AEJN 1	2.70E+07	1.20%	92.77% 93.89%			
13	gm-AEJAD 1	2.52E+07	1.13%				
14	gm-AEJM 1	2.03E+07	0.91%	94.80%			
15	gm-AEJ[I/L] 1	1.81E+07	0.81%	95.61%			
16	gm-AE 1	1.73E+07	0.77%	96.38%			
17	gm-AEJS 1	1.58E+07	0.70%	97.09%			
18	gm-AEJAW 1	1.47E+07	0.66%	97.74%			
19	gm-AEJH 1	9.64E+06	0.43%	98.17%			
20	gm-AEJW 1	6.41E+06	0.29%	98.46%			
21	gm-AEJV 1	6.29E+06	0.28%	98.74%			
22	gm-AEJAQ 1	5.40E+06	0.24%	98.98%			
23	gm-AEJAH 1	5.02E+06	0.22%	99.20%			
24	gm-AEJT 1	4.51E+06	0.20%	99.40%			
25	gm-AEJAE 1	4.51E+06	0.20%	99.61%			
26	gm-AEJAR 1	AR 1 4.11E+06		99.79%			
27	gm-AEJAK 1	2.53E+06	0.11%	99.90%			
28	gm-AEJD 1	2.21E+06	0.10%	100.00%			
	Total	2.24E+09					

 Table S6-7. Cumulated abundance of monomers validated by MS/MS

Table S6- 8. Database 3 (DB_3)____

Structure	Monoisotopicmass
gm-AEJ 1	870.37059
gm-AEJA 1	941.40770
gm-AEJAA 1	1012.44482
gm-AEJAF 1	1088.47612
gm-AEJAI 1	1054.49177
gm-AEJF 1	1017.43900
gm-AEJG 1	927.39205
gm-AEJK 1	998.46555
gm-AEJAG 1, gm-AEJ	IQ 1 998.42917
gm-AEJY 1	1033.43392

Table 6-1. A list of the PG modifications s	earched for, and the strategy used to validate them.
---	--

Modification	Description	Retention	Mass	Signature ion	Relative	Present
		time	change	<i>m/z</i> ([м+н]⁺)	abundance	
Glycan modification AnhydroMurNAc De-acetylation <i>O</i> -acetylation	(Anh) 1,6-anhydroMurNAc (DeAc) Loss of acetyl group (Ac) Gain of acetyl group	Increased Decreased Increased	-20.026 -42.010 +42.010	258.096 (AnhMurNAc) 162.077 (GlcN); 236.113 (MurN ^{red}) 246.098 (GlcNAc+Ac) 320.134 (MurNAc ^{red} +Ac)	1) Similar % of modified monomers for all monomers 2) Belative	YES NO NO
Peptide modification Amidation	ıs (Am) Glu or <i>m</i> Dap amidatio	Glu or <i>m</i> Dap amidation Decreased -0.984 172.109 (mDAP _{NH2}); 1		172.109 (mDAP _{NH2}); 129.066 (Gln)	abundance of modified monomers should	NO
Loss of g	Glucosaminidase product	Decreased	-203.079 278.124 (MurNAc ^{red}) relative 206.087 (MurNAc ^{red} -Lactyl) of unmo +478.180 276.108 (MurNAc) monom	relative abundance of unmodified	NO	
Extra gm	Partial muramidase digestion	Increased		276.108 (MurNAc)	monomers	NO

As an example, Fig. 6-6A shows how nine putative AnhydroMurNAc-containing muropeptides were identified. Most modified muropeptides were present in all three replicates and had a consistently higher retention time than their unmodified counterparts. Next, we manually searched the MS/MS data for signature fragment ions predicted by PGFinder's "Fragment Generator" module. Fig. 6-6B summarizes the MS/MS analysis of the gm(Anh)-AEJA monomer. Out of the 20 predicted fragment ions, 12 were present, including five out of nine possible signature ions (highlighted in red in Fig. 6-6B). The presence of these fragment ions ultimately contributes to validate the gm(Anh)-AEJA monomer.

On average, AnhydroMurNAc-modified muropeptides were 10% as abundant as their unmodified counterparts, with a particularly high proportion of the disaccharide-tripeptide gm-AEJ being modified (48% of its unmodified intensity) (Fig. 6-6A). In the end, only the anhydro versions of the three most abundant monomers (gm-AEJA, gm-AEJG, and gm-AEJ) could be confirmed (Fig. 6-6A), though this was in part due to a lack of MS/MS data for the other matches. Five other modifications (Table 6-1) were searched for using the same strategy, but none of these modifications could be confirmed via MS/MS (Fig. S6-3).

Α

Structure ^a	Intensity	% Anhydro	RT (min)	ΔT (min) ^t	Presence of signature ions		
gm-AEJA 1	1.1E+09	2 50/	9.50 ± 0.03	1 6 62	Vacuualidated with 1 Land 2 Lions (TV1 dataset)		
gm(Anh)-AEJA 1	3.7E+07	3.370	16.12 ± 0.00	+ 0.02	res; vanuated with 1+ and 2+ ions (111 dataset)		
gm-AEJG 1	2.4E+08	7 70/	6.85 ± 0.04	. 7 4 2	Vecture lighted with 1, and 2, ions (TV1 detect)		
gm(Anh)-AEJG 1	1.8E+07	1.270	14.27 ± 0.01	+ 7.42	res, vanuated with 1+ and 2+ lons (141 dataset)		
gm-AEJ 1	1.6E+08	49.00/	5.48 ± 0.03	L Q 10	Vacuualidated with 1, and 2, ions (TV1 dataset)		
gm(Anh)-AEJ 1	7.9E+07	48.0%	13.58 ± 0.00	+ 8.10	res; validated with 1+ and 2+ lons (1Y1 dataset)		
gm-AEJAI 1	1.5E+08	0.6%	21.40 ± 0.00		No for encoderation and the		
gm(Anh)-AEJAI 1	8.8E+05	0.6%	28.39 ± 0.02	+ 6.99	No fragmented fon available		
gm-AEJF 1	1.5E+08	2.0%	21.62 ± 0.00	1 7 08	No fragmented ion available		
gm(Anm)-AEJF 1	4.5E+06	5.0%	28.70 ± 0.00	+ 7.08	No magmenteu ion avanable		
gm-AEJAA 1	6.5E+07	0.0%	10.86 ± 0.02	+ 6.40	Only present in 2 of 3 datasets		
gm(Anh)-AEJAA 1	5.9E+05	0.5%	17.26 ± 0.02	+ 0.40			
gm-AEJK 1	4.6E+07	0.49/	7.71 ± 0.06	41	No frequented ion evolution		
gm(Anh)-AEJK 1	4.3E+06	9.4%	14.12 ± 0.01	+ 0.41	No fragmented fon available		
gm-AEJAG 1, gm-AEJQ 1	4.6E+07	1 /0/	8.80 ± 0.05	+ 11 1E	No signature ions found		
gm(Anh)-AEJAG 1, gm(Anh)-AEJQ 1	6.5E+05	1.470	19.95 ± 6.37	+ 11.15	No signature ions round.		
gm-AEJAD 1	2.5E+07	11 79/	9.61 ± 0.04	6.21	No fragmented ion available		
gm(Anh)-AEJAD 1	2.9E+06	11.7%	15.82 ± 0.01	+ 0.21	No fragmenteu fon avanable		
gm-AEJAF 1	5.0E+07	0.0%	15.4 ± 0.0	-	No modification found		

^a Non modified monomers are sorted by abundance (most abundant first)

^b ΔT is defined as the difference (in min) between the average RT of the unmodified and modified muropeptide



Figure 6-6. The existence of AnhydroMurNAc-modified monomers can be confirmed via MS/MS.

(A)Summary of unmodified monomers and their anhydroMurNAc counterparts identified by PGFinder. For each muropeptide, the intensity and abundance of the AnhydroMurNAc modification is provided, alongside with the retention time. The shift in retention time (ΔT (min)) and presence of signature ions are indicated. (B) Example MS/MS spectrum showing the identified signature ions in red for a singly charged ([M+H]⁺) ion corresponding to the gm(Anh)-AEJA muropeptide. RT: retention time.

6.3.5 Step 3: Identifying outer membrane proteins covalently anchored to the PG

The covalent anchoring of outer membrane β -barrel proteins to the PG helps maintain cell envelope integrity in Alphaproteobacteria (Godessart et al., 2021; Sandoz et al., 2021a). We identified ten putative β -barrel proteins encoded by the genome of *R. leguminosarum* bv. *viciae* (strain 3841) (Fig. S6-4) and investigated if any of them were anchored to the PG.

A fourth database (DB_4, Table S6-8) was created to search for any amino acid scars left behind by β -barrel proteins that had been attached to the PG. Since tetrapeptide stems are thought to act as the donors during L,D-transpeptidase mediated protein anchoring, DB_4 contained muropeptides comprised of disaccharide tripeptides (gm-AEJ) followed by residues corresponding to *N*-terminal of each anchored porin with its signal peptide removed

(Fig S6-5A). Although trypsin digestion is expected to generate porin "scars" with a basic residue at the C-terminal, previous analyses revealed that muropeptides containing non-canonical and missed cleavages are common. To avoid missing any of these non-canonical "scars", the muropeptides in DB_4 contained every N-terminal porin sequence from 2 to 17 amino acids in length.

Α				AnhydroMu	rNAc (Anh)	D			Amidati	on (Am)	
S	itructure	Intensity	% Anh	RT (min)	ΔT (min)*	Comments	Structure	Intensity	% (+NH ₂)	RT (min)	ΔT (min)*	Comments
g g	m-AEJA 1 m-AEJA (Anh) 1	1.07E+09 3.71E+07	3.5%	9.50 ± 0.03 16.12 ± 0.00	6.62	Validated with 1+ and 2+ ions	gm-AEJA 1 gm-AEJA (Am) 1	1.07E+09 1.90E+06	0.2%	9.5 ± 0.0 11.7 ± 0.1	2.2	RT not compatible with amidation (elutes later than unmodified monomer)
g	m-AEJG 1 m-AEIG (Anh) 1	2.43E+08 1.76E+07	7.2%	6.85 ± 0.04 14.27 + 0.01	7.42	Validated with 1+ and 2+ ions	gm-AEJG 1 gm-AEJG (Am) 1	2.43E+08 5.57E+05	0.2%	6.8 ± 0.0 10.7 ± 0.0	3.9	RT not compatible with amidation (elutes later than unmodified
g	m-AEJ 1 m-AEJ(Anh) 1	1.64E+08 7.90E+07	48.0%	5.48 ± 0.03 13.58 ± 0.00	8.09	Validated with 1+ and 2+ ions	gm-AEJ 1 gm-AFI (Am) 1	1.64E+08 7.82E+05	0.5%	5.5 ± 0.0 4.4 + 0.0	-1.0	No MS/MS data available
g	m-AEJAI 1 m-AEJAI (Anh) 1	1.54E+08 8.83E+05	0.6%	21.40 ± 0.00 28.39 ± 0.02	6.99	No MS/MS data available	gm-AEJAI (1 gm-AEJAI (Am) 1	1.54E+08 1.26E+06	0.8%	21.4 ± 0.0 26.3 ± 0.0	4.9	RT not compatible with amidation (elutes later than unmodified
g	m-AEJF 1 m-AEJF (App) 1	1.48E+08	3.0%	21.62 ± 0.00	7.08	No MS/MS data available	gm-AEJF 1 gm-AEJE (Am) 1	1.48E+08 4.10E+05	0.3%	21.6 ± 0.0 23.7 ± 9.1	2.0	RT not compatible with amidation (elutes later than unmodified
g	m-AEJAF (Anh) 1	5.00E+07	0.0%	15.4 ± 0.00	-	Not present	gm-AEJAA 1	6.55E+07	2.2%	10.9 ± 0.0	40.2	RT not compatible with amidation (elutes later than unmodified
g	m-AEJAA 1 m-AEJAA (Anh) 1	6.55E+07 5.87E+05	0.9%	10.86 ± 0.02 17.26 ± 0.02	6.40	Only present in 2 of 3 datasets	gm-AEJAA (Am) 1 gm-AEJAF 1	2.07E+06 5.00E+07	0.0%	30.1 ± 1.6 15.4 ± 0.0	- 19.3	monomer) No modification found
g	m-AEJK 1 m-AEJK (Anh) 1	4.60E+07 4.32E+06	9.4%	7.71 ± 0.06 14.12 ± 0.01	6.41	No MS/MS data available	gm-AEJAG 1,gm-AEJQ 1 gm-AEJAG (Am) 1,gm-AEJQ (Am) 1	4.61E+07 6.81E+05	1.5%	8.8 ± 0.1 23.8 ± 0.0	15.0	RT not compatible with amidation (elutes later than unmodified monomer)
g	m-AEJAG 1,gm-AEJQ 1 m-AEJAG (Anh) 1 gm-AEIO (Anh) 1	4.61E+07 6.53E+05	1.4%	8.80 ± 0.05	11.15	No signature ions found.	gm-AEJK 1 gm-AEJK (Am) 1	4.60E+07 9.45E+05	2.1%	7.7 ± 0.1 20.7 ± 0.0	13.0	RT not compatible with amidation (elutes later than unmodified
g	m-AEJAD (Anh) 1	2.52E+07	11.7%	9.61 ± 0.04	6.21	No MS/MS data available	gm-AEJAD 1	2.52E+07		9.6 ± 0.0		indioner)
g	(m-AEJAD (Ann) 1	2.94E+06		15.82 ± 0.01		-			4.6%		-1.9	No MS/MS data available
R	[™] Δ1 is defined as the difference (in m	in) between tr	ie average K	DeAcetyla	tion (-Ac)	ed muropeptide	gm-AEJAD (Am) 1	1.16E+06		7.7 ± 0.0		
۲.		1-4	0// 0-)	DEALELYIA	AT (min)*	Commonto	* ∆T is defined as the difference (in m	nin) between th	e average RT of	the unmodifie	d and modifie	ed muropeptide
g	m-AEJA 1	1.07E+09	% (-AC)	9.5 ± 0.03	Δ1 (min)*	Comments	E			Extra-g	m (gm-)	
g	m-AEJA (-Ac) 1	3.44E+06	0.3%	7.8 ± 0.06	-1.7	No signature ions	Structure	Intensity	% Extra gm	RT (min)	ΔT (min)*	Comments
g g	m-AEJG 1 m-AEJG (-Ac) 1	2.43E+08 1.24E+06	0.5%	6.8 ± 0.04 6.4 ± 2.16	-0.4	No MS/MS data available	gm-AEJA 1 gm-gm-AEJA 1	1.07E+09 4.17E+06	0.4%	9.5 ± 0.0 12.6 ± 0.0	3.1	No convincing signature ions
g g	m-AEJ 1 m-AEJ (-Ac) 1	1.64E+08 7.78E+05	0.5%	5.5 ± 0.03 4.1 ± 0.03	-1.3	No MS/MS data available	gm-AEJG 1 gm-gm-AEJG 1	2.43E+08 6.03E+05	0.2%	6.8 ± 0.0 11.3 ± 0.0	4.5	No MS/MS data available
g	m-AEJAI 1 m-AEJAI (-Ac) 1	1.54E+08 6.51E+06	4.2%	21.4 ± 0.00 25.9 ± 1.67	4.5	RT not compatible with deacetylation (elutes later than unmodified monomer)	gm-AEJ 1 gm-gm-AEJ 1	1.64E+08 7.59E+05	0.5%	5.5 ± 0.0 10.8 ± 0.0	5.3	No MS/MS data available
g	m-AEJF 1	1.48E+08	0.0%	21.6 ± 0.00	0.0	Not present	gm-AEJAI 1	1.54E+08	13.8%	21.4 ± 0.0	10.1	No MC/MC data quailable
g	m-AEJAA 1	6.55E+07	0.0%	10.9 ± 0.02	0.0	Not present	gm-gm-AEJAI 1	2.12E+07		31.5 ± 0.0	10.1	NO IVIS/IVIS GALA AVAILADIE
g	m-AEJAF I m-AEJAF (-Ac) 1	5.00E+07 4.65E+05	0.9%	15.4 ± 0.00 267 ± 0.00	11.3	Only present in 1 of 3 datasets	gm-AEJF 1	1.48E+08	-	$\frac{21.6 \pm 0.1}{10.0 \pm 0.1}$	-	Not present
g	m-AEJAG 1,gm-AEJQ 1	4.61E+07	0.0%	8.8 ± 0.05	0.0	Not present	gm-AFIAFI1	5.00E+07	12.4%	15.4 + 0.0	-	Not present
g	m-AEJK 1	4.60E+07	0.0%	7.7 ± 0.06	0.0	Not present	gm-gm-AEJAF 1	6.18E+06	12.470	31.4 ± 1.8	16.0	Not validated, too low coverage (13%)
g	m-AEJAD 1	2.52E+07	2 10/	9.6 ± 0.04	11.2	Only present in 1 of 2 detects	gm-AEJAG 1,gm-AEJQ 1	4.61E+07	-	8.8 ± 0.0	-	Not present
g	m-AEJAD (-Ac) 1	7.91E+05	5.1%	20.9 ± 0.00	11.2	Univ present in 1 or 5 datasets	gm-AEJK 1 gm-gm-AEJK 1	4.60E+07 7.23E+05	1.6%	7.7 ± 0.0 17.4 ± 0.0	9.7	No MS/MS data available
<u>`</u>	Δ1 is defined as the difference (in mir	i) between the	e average RT	of the unmodified	and modifie	d muropeptide	gm-AEJAD 1	2.52E+07	-	9.6 ± 0.0	-	Not present
ι_				O-Acetylat	ion (+Ac)		* ΔT is defined as the difference (in m	nin) between th	e average RT of	the unmodifie	d and modifie	ed muropeptide
<u>s</u>	Structure	Intensity	<u>% O-Ac</u>	RT (min)	ΔT (min)*	Comments	F			Loss of G	cNAc (m-)	
g	m-AEIGI1	2.43E+08	0.0%	9.5 ± 0.0	0.0	Not present	Structure	Intensity	oss of GlcN4	RT (min)	AT (min)*	Comments
g	m-AEJG (+Ac) 1	2.36E+06	1.0%	14.3 ± 2.3	7.4	No MS/MS data available	gm-AEJA 1	1.48E+09	2.0%	9.5 ± 0.0		
<u>g</u>	m-AEJ I	1.64E+08	0.0%	5.5 ± 0.0	0.0	Not present	m-AFIA 1	4.38F+07	3.0%	8.6 + 0.1	-0.9	NO MS/MS data available
5 g	m-AEJAI (+Ac) 1	3.36E+05	0.2%	26.5 ± 0.0	5.1	Only present in 2 of 3 datasets	gm-AEJG 1	3.70E+08	-	6.8 ± 0.0	-	Not present
g	m-AEJF 1	1.48E+08	0.0%	21.6 ± 0.0	0.0	Not present	gm-AEJ 1	2.65E+08	0.2%	5.5 ± 0.0	18.2	RT not compatible with loss of GlcNAc (elutes later than unmodified
g	m-AEJAA 1	6.55E+07	0.0%	10.9 ± 0.0	0.0	Not present	m-AEJ 1	5.09E+05	0.270	23.7 ± 0.0	10.2	monomer)
g g	m-AEJAF 1 m-AEJAF (+Ac) 1	5.00E+07 1.00E+06	2.0%	15.4 ± 0.0 28.9 ± 0.0	13.5	Only present in 2 of 3 datasets	m-AEJF 1 m-AEJF 1	2.12E+08 9.09E+06	4.3%	21.6 ± 0.0 28.7 ± 0.0	7.1	monomer)
g g	m-AEJAG 1,gm-AEJQ 1 m-AEJAG (+Ac) 1,gm-AEJQ (+Ac) 1	4.61E+07 5.17E+05	1.1%	8.8 ± 0.1 6.2 ± 0.0	-2.6	Only present in 1 of 3 datasets	gm-AEJAI 1 m-AEJAI 1	1.94E+08 3.16E+06	1.6%	21.4 ± 0.0 16.2 ± 0.0	-5.3	Not validated low coverage (20%)
g	m-AEJK 1	4.60E+07	0.0%	7.7 ± 0.1	0.0	Not present	gm-AEJAA 1	8.89E+07	-	10.9 ± 1.4	-	Not present
g	m-AEJAD 1	2.52E+07	2.3%	9.6 ± 0.0	73	Only present in 1 of 3 datasets	gm-AEJAG 1,gm-AEJQ 1	6.32E+07	-	8.8 ± 0.1	-	Not present
g	m-AEJAD (+Ac) 1	5.73E+05	2.370	16.9 ± 0.0	7.5	ony present in 2 or 5 datasets	m-AFIK 1	5.05E+07 6.91E+05	1.2%	14.1 + 0.1	6.4	monomer)
*	ΔT is defined as the difference (in mir	n) between the	e average RT	of the unmodified	and modifie	d muropeptide	gm-AEJAF 1	5.32E+07	-	18.1 ± 0.0	-	Not present
							gm-AEJAD 1	3.07E+07	-	9.6 ± 0.0	-	Not present

gm-AEJAD 1 * ΔT is defined as the

Figure S6-2. Summary of modifications analyses.

A, AnhydroMurNAc products; B, DeAcetylation; C, O-acetylation; D, Amidation; E, Extra gm-; F, Loss of GlcNAc.

>pRL90069 Q1M8N4 |Q1M8N4 RHIL3 Conserved hypothetical exported protein

MKKILATAFAAVSLTFVGAAAVNÄADLGTRTYEEPDLRNGVKIGYLTCDIGGGTGYVLGSSKEADCIFOSTVGNELSDRYTGEMRKLGIDLGFTTRS. WAVFAFTAGYHRGSLAGLYVGATAEATLGAGVGANLLVGGTSGSIHLOTVSLTGOLGLNVAAGSASMTLTAAN



>RL0868 (LpxQ) Q1MKZ0 RHIL3 Lipid A oxidase

WSASAEDLQFSIYGGYQTAPHSGVDLSDGTSFTAGWEGKSFGSPPYYGARVTWWLENFNKPNWGISLDYTHDKV DDTLAKAGWSHFEFTDGLNLITVNGLYRFQDPTRRWTPYLGAGIGVNIPHVEVIRPEGKTWAYEFGGVTLQAQAGVDFKVTERWSTFVEYKGTYSRI IDSGVDLKTNIFTNAVNVGVSFHW



>RL3165 Q1MEH4_RHIL3 Conserved hypothetical TPR repeat receptor protein

 >R12752 0mp4 family protein 01MF3 RHID
 >R12752 0mp4 family protein 01MF3 RHID

 >RCXP372 0mp4 family protein 01MF3 RHID
 >NIGRWKNICRVCMALAFAVPGGLPAMAPUVPRATVAGSVIARKIGEEVRFIDVSNWRVDINODLIGDVLRTNANGOLAT

 >RCXP372 0mp4 family protein 01MF3 RHID
 >NIGRWKNICRVCMALAFAVPGGLPAMAPUVPRATVAGSVIARKIGEEVRFIDVSNWRVDINODLIGDVLRTNANGOLAT

 >RCXP372 0mp4 family protein 01MF3 RHID
 >NIGRWKNICRVCMALAFAVPGGLPAMAPUVPRATVAGSVIARKIGEEVRFIDVSNWRVDINODLIGDVLRTNANGOLAT

 >RCXP372 0mp4 family protein 01MF3 RHID
 >NIGRWKNICRVCMALAFAVPGGLPAMAPUVPTGLAVRAGVIARKIGEEVRFIDVSNWRVDINODLIGDVLRTNANGARGA

 APPEPPRAEAPAKPEDPOPESKPARKAKSEA0PEAKPEAE0PVT0EKPKKPKKPEA00AEPE0QPAAKEA0PEAE0APEAKPEAGPEAKPEAGVATIGOAFSKIISVNPDIRDGVMLFVLDLRDGFULDFF50LPARKALLGEPERKTIEDWLEAFEAE0

 RCKKARELPAPAPAPAPDAVPTTESSAFFEATPEAKPAAEAPAEKPKAGGTAAPAKAPTUSATUTGOAFSKIISVNPDIRDGVMLFVLDLRDGFULDFF50LPARKARLLGEPERKTILLEGPERKTIEDWLEAFEAE0

 PACA0PAAPDAVPTDIASGQ02VF02FFEKESKRKKIADPAKSTULDFVLDVENGAULDSKKDARGFEAPAFEKPKOTAKFA TSGFUCALAFAFEAFEAPAFEKPKUTSTENTERVATUSTULDSKCDARGFEAPAFEKPKUTSTENTERVATUSTULDSKCDARGFEAPAFEKPKUTSTENTERVATUSTULDSKCDARGFEAPAFEKPKUTSTENTERVATUSTULDSKCDARGFEAPAFEKPKUTSTENTERVATUSTULDSKCDARGFEAPAFEKPKUTSTENTERVATUSTULDSKCDARGFEAPAFEXPVELEAFKKIADPAKSSETVULDSKCDARGFEAPAFEKPKUTSTULDSKCDARGFEAPAFEKPKUTSTULDSKCDARGFEAPAFEKPKUTSTULDSKCDARGFEAPAFEKPKUTSTENTUDIGGSDSNULGUDARGKGEAPAFEKPKUTSTULDSKCDARGFEAPAFEKPKUTSTULDSKCDARGFEAPAFEKPKUTSTULDSKCDARGFEAPAFEKPKUTSTULDSKCDARGFEAPAFEKPKUTSTULDSKCDARGFEAPAFEKPKUTSTULDSKCDARGFEAPAFEKPKUTSTULDSKCDARGFEAPAFEKPKUTSTULDSKCDARGFEAPAFEKPKUTSTULDSKCDARGFEAPAFEKPKUTSTULDSKCDARGFEAPAFEKPKUTSTULDSKCDARGFEAPAFEKPKUTSTULDSKCDARGFEAPAFEKPKUTSTULDSKCDARGFEAPAFEKPKUTSTULDSKCDARGFEAPAFEKPKUTSTULDSKCDARGFEAPAFEKPKUTSTULDSKCDARGFE





>RL3968 (Pal) Q1MC73_RHIL3 Peptidoglycan-associated protein

MSRIHTPAMSRMQNFARNPVMINLIAGLALASCAKKNVPNSAGDLGLGAGAGAATPGSAQDFTVNVGDRIFFDTDSSSIRADASQTLDRQAQWLGRY QITVEGHADERGTREYNLALGARRAAAAKDYLASRGVPAQRLKTISYGKERPVAVCDDISCWSQNRRAVTVLGGAGM*



>RL4133 Q1MBR1 RHIL3 Conserved hypothetical exported protein

MEKQTMIAAAALTAAAWASPAGAENYVTLGRLVCGSDGGQGLIVTSQKNLICTYTPSAGGAKAVYAGKIEKFGLDIGQTGKSVMIWQVLAKTGTDIP>RL (RopA1) Q1MFL0_RHIL3 Porin QFALAGEYYGIGADASIGAGAGAKVIAGGTDKAFMLQPLNVQAQEGLNLAIGVEKMTLVPGET



Figure S6-3. Putative β-barrel proteins encoded by *Rhizobium leguminosarum* bv. *viciæ* strain 3841.
Structure	Monoisotopicmass
gm-AEJADL 1	1169.51871
gm-AEJADLG 1	1226.54017
gm-AEJADLGT 1	1327.58785
gm-AEJADLGTR 1	1483.68896
gm-AEJADLGTRT 1	1584.73664
gm-AEJADLGTRTY 1	1747.79997
gm-AEJADLGTRTYE 1	1876.84256
gm-AEJADLGTRTYEE 1	2005.88516
gm-AEJADLGTRTYEEP 1	2102.93792
gm-AEJADLGTRTYEEPD 1	2217.96486
gm-AEJADLGTRTYEEPDL 1	2331.04893
gm-AEJADLGTRTYEEPDLR1	2487.15004
	2601.19296
gm-AEJADLGTRTTEEPDLRING 1	2028.21443
gm-AEJADLGTRTTEEPDLRINGV/1	2737.20204
gm_AEIED 1	1114 44012
gm-AEIEDI 1	1227 52419
gm-AFIFDI O 1	1355 58277
gm-AFIEDI OF 1	1502 65118
gm-AEJEDLOFS 1	1589.68321
gm-AEJEDLQFSI 1	1702.76727
gm-AEJEDLQFSIY 1	1865.83060
gm-AEJEDLQFSIYG 1	1922.85206
gm-AEJEDLQFSIYGG 1	1979.87353
gm-AEJEDLQFSIYGGY 1	2142.93686
gm-AEJEDLQFSIYGGYQ 1	2270.99543
gm-AEJEDLQFSIYGGYQT 1	2372.04311
gm-AEJEDLQFSIYGGYQTA 1	2443.08023
gm-AEJEDLQFSIYGGYQTAP 1	2540.13299
gm-AEJEDLQFSIYGGYQTAPH 1	2677.19190
gm-AEJEDLQFSIYGGYQTAPHS 1	2764.22393
gm-AEJEDLQFSIYGGYQTAPHSG 1	2821.24539
gm-AEJAV [1	1040.47612
gm-AEJAVR 1	1196.57723
gm-AEJAVRD I	1311.60417
gm-AEIAVRDV 1	1410.07258
gm-AEIAVRDVAT1	1582 75738
gm-AFIAVRDVATO 1	1710 81595
gm-AFIAVRDVATOA 1	1781.85307
gm-AEJAVRDVATOAS 1	1868.88510
gm-AEJAVRDVATQASA 1	1939.92221
gm-AEJAVRDVATQASAV 1	2038.99062
gm-AEJAVRDVATQASAVQ 1	2167.04920
gm-AEJAVRDVATQASAVQQ 1	2295.10778
gm-AEJAVRDVATQASAVQQA 1	2366.14489
gm-AEJAVRDVATQASAVQQAE 1	2495.18749
gm-AEJAVRDVATQASAVQQAEQ 1	2623.24606
gm-AEJAVRDVATQASAVQQAEQG 1	2680.26753
gm-AEJDP 1	1082.45030
gm-AEJDPV 1	1181.51871
gm-AEJDPVP 1	1278.57147
gm-AEJDPVPR 1	1434.67258
gm-AEJDPVPRA 1	1505.70970
gm-AEJDPVPRAT 1	1606.75738
gm-AEJDPVPKATP[1 gm-AEJDDVDPATDV/1	1703.81014
gni-AEJDEVERATEV 1 gm_AEIDD//DRATD//A 1	1072 01567
	1030 0212
	2017 06016
gm-AFIDP//PRATP//AGS//1	2017.30310
gm-AEJDPVPRATPVAGSVI11	2230.12164
gm-AEJDPVPRATPVAGSVIA 1	2301.15875
gm-AEJDPVPRATPVAGSVIAR 1	2457.25986
gm-AEJDPVPRATPVAGSVIARK 1	2585.35483
gm-AEJDPVPRATPVAGSVIARKI 1	2698.43889

Structure	Monoisotopicmass
gm-AEJCA 1	1044.41689
gm-AEJCAK 1	1172.51185
gm-AEJCAKK 1	1300.60681
gm-AEJCAKKN 1	1414.64974
gm-AEJCAKKNV 1	1513.71815
gm-AEJCAKKNVP 1	1610.77092
gm-AEJCAKKNVPN 1	1724.81385
	1811.84587
gm_AEICAKKNVPNSA[1	1032.00233
gm-AFICAKKNVPNSAGD11	2054 93139
gm-AEJCAKKNVPNSAGDL 1	2168.01546
gm-AEJCAKKNVPNSAGDLG 1	2225.03692
gm-AEJCAKKNVPNSAGDLGL 1	2338.12099
gm-AEJCAKKNVPNSAGDLGLG 1	2395.14245
gm-AEJCAKKNVPNSAGDLGLGA 1	2466.17956
gm-AEJCAKKNVPNSAGDLGLGAG 1	2523.20103
gm-AEJEN 1	1113.45611
gm-AEJENY/11	1270.51944
gm-AFIFNYVT1	1476 63553
gm-AEJENYVTL1	1589.71959
gm-AEJENYVTLG11	1646.74106
gm-AEJENYVTLGR 1	1802.84217
gm-AEJENYVTLGRL 1	1915.92623
gm-AEJENYVTLGRLV 1	2014.99465
gm-AEJENYVTLGRLVC 1	2118.00383
gm-AEJENYVTLGRLVCG 1	2175.02530
gm-AEJENYVILGRLVCGS11	2262.05732
gm-AEJENYVILGRLVCGSD[1 gm-AEJENVVILGRLVCGSDG]1	2377.08427
gm-AFIENYVTI GRI VCGSDGG 1	2434.10373
gm-AEJENYVTLGRLVCGSDGGQ11	2619.18577
gm-AEJENYVTLGRLVCGSDGGQG 1	2676.20724
gm-AEJADA 1	1127.47176
gm-AEJADAI 1	1240.55582
gm-AEJADAIV 1	1339.62424
gm-AEJADAIVA 1	1410.66135
gm-AEJADAIVAA 1	1481.69846
gm_AEIADAIVAAEII	1707 79382
gm-AFIADAIVAAFPF11	1836 83641
gm-AEJADAIVAAEPEP 1	1933.88918
gm-AEJADAIVAAEPEPV 1	2032.95759
gm-AEJADAIVAAEPEPVE 1	2162.00019
gm-AEJADAIVAAEPEPVEY 1	2325.06351
gm-AEJADAIVAAEPEPVEYV 1	2424.13193
gm-AEJADAIVAAEPEPVEYVR 1	2580.23304
gm-AEJADAIVAAEPEPVEYVRV 1	2679.30145
gm-AEJADAIVAAEPEPVEYVRVCJ1	2782.31064
gm-AFIADAVD1	1341 56712
gm-AEJADAVDO 1	1469.62569
gm-AEJADAVDQV 1	1568.69411
gm-AEJADAVDQVP 1	1665.74687
gm-AEJADAVDQVPE 1	1794.78946
gm-AEJADAVDQVPEA 1	1865.82658
gm-AEJADAVDQVPEAP 1	1962.87934
gm-AEJADAVDQVPEAPV 1	2061.94776
	2132.98487
gm-AEJADAVDQVPEAPVAQ 1 gm-AEJADAVDQVPEAPVAQE11	2201.04345 2390 08601
gm-AEJADAVDOVPEAPVAOFAI1	2461 12315
gm-AEJADAVDQVPEAPVAQEAP 1	2558.17592
gm-AEJADAVDQVPEAPVAQEAPV 1	2657.24433



Figure S6- 4. The existence of β -barrel anchoring scars is confirmed via MS/MS.

A. N-terminal sequences of five putative peptidoglycan-anchored β -barrel proteins; the arrow shows the predicted signal peptide cleavage site. **B**, Example MS/MS spectra confirm the N-terminal anchoring of these β -barrel proteins. Ions corresponding to internal fragments that were manually annotated are boxed.

A total of 29 masses matching PG-anchored β -barrel proteins were found in all three TY replicates, though MS/MS data was only available to confirm muropeptides associated with RopA1,2,3, RopB, and pRL90069 (Fig. S5B). Collectively, muropeptides corresponding to these proteins accounted a total intensity of 3.51E+08, representing 13.3% of all monomer intensity (2.64E+09) (Table 6-2).

	Structure	Av. Intensity	TY1	TY2	ТҮЗ	Sum
RopA1,2,3, RopB						
or PRL90069	gm-AEJAD	2.52E+07	3.66E+07	2.05E+07	1.86E+07	2.52E+07
RopA1,2,3 or RopB	gm-AEJADA	1.69E+08	2.35E+08	1.66E+08	1.07E+08	1.69E+08
	gm-AEJADAIVA	8.03E+07	7.96E+07	6.78E+07	9.36E+07	
	gm-AEJADAIVAA	1.14E+07	1.87E+07	9.81E+06	5.74E+06	
	gm-AEJADAIVAAEPEPVE	8.19E+06	9.59E+06	4.72E+06	1.03E+07	
PonA1 2 2	gm-AEJADAIV	7.24E+06	8.39E+06	7.55E+06	5.77E+06	
кора1,2,5	gm-AEJADAIVAAEPEPV	2.41E+06	3.28E+06	1.14E+06	2.79E+06	
	gm-AEJADAI	1.24E+06	1.54E+06	1.14E+06	1.04E+06	
	gm-AEJADAIVAAE	9.13E+05	9.10E+05	9.26E+05	9.02E+05	
	gm-AEJADAIVAAEPEP	6.88E+04	6.88E+04	ND	ND	1.12E+08
RopB or pRL90069	gm-AEJADAV / gm-AEJADLG	3.01E+06	3.91E+06	2.84E+06	2.28E+06	3.01E+06
	gm-AEJADAVDQVPEAPVAQ	1.69E+07	1.44E+07	1.29E+07	2.35E+07	
PonP	gm-AEJADAVDQVPEAPVAQE	2.50E+06	2.96E+06	2.31E+06	2.23E+06	
КОРВ	gm-AEJADAVD	2.28E+06	3.18E+06	2.33E+06	1.32E+06	
	gm-AEJADAVDQVPEAP	1.40E+06	1.48E+06	1.61E+06	1.12E+06	2.31E+07
	gm-AEJADLGTRTYEEPDLRNGV	8.12E+06	1.56E+07	4.60E+06	4.17E+06	
pRL90069	gm-AEJADL	5.55E+06	5.43E+06	6.16E+06	5.08E+06	
	gm-AEJADLGTR	5.13E+06	6.56E+06	4.55E+06	4.29E+06	1.88E+07
						3.51E+08

Table 6-2. PGFinder identification of *N*-terminal peptides from RopA1,2,3,RopB and pRL90069 anchored to gm-AEJ.

6.3.6 Step 4: Identifying PG multimers and confirming their structure

PGFinder's "MS Analysis" module was used to identify dimers and trimers resulting from both *D*,*D*- and L,D-transpeptidation using the monomers previously validated (DB_2; Table S6-6) as acceptors (apart from gm-AE). A total of 88 multimeric masses (41 dimers and 47 trimers) were present across all three replicates (Table S6-9).

Out of the 41 dimers, 28 could be unambiguously classified as products of *D*,*D* or L,D-transpeptidation. This was possible because muropeptides like gm-AEJ=gm-AEJX (3-3) could only be formed via L,D-transpeptidation, and gm-AEJA=gm-AEJAX (4-3) structures could only be formed via *D*,*D*-transpeptidation. To differentiate the remaining 13 muropeptides, MS/MS analysis was necessary. They were either (i) isomers — with the same residues but distinct crosslinking, e.g., gm-AEJA=gm-AEJ (4-3) vs gm-AEJ=gm-AEJA (3-3) — or (ii) mass coincidences — with distinct compositions but the same chemical formula, e.g., gm-AEJ=gm-AEJQ (3-3) vs gm-AEJ=gm-AEJAG (3-3). In the Q vs AG case, MS/MS did not allow us to discriminate between the two structures, as a fragmentation between the *C*-terminal alanine and glycine was never observed. It was possible, however, to assign one of the two possible isomeric structures — gm-AEJA=gm-AEJ (4-3) or gm-AEJ=gm-AEJA (3-3) to each of the three peaks in the extracted ion chromatogram (*m*/*z*=897.88; Fig. 6-7A).

As a first step, we predicted a list of fragment ions using PGFinder's "Fragment Generator" for each of the isomeric structures (71 fragments for the 3-3 dimer and 55 for the 4-3 dimer, coming together to form a set of 59 unique m/z values; Table S6-10). As expected, a large proportion of the predicted ions (68%, 59%, and 58%) were detected in each peak. To assign each of the three peaks in the extracted ion chromatogram to a 3-3 or a 4-3 dimer, we computed a list of signature ions for each

structure (8 for the 4-3 dimer and 16 for the 3-3 dimer; Fig. 6-7B) and recorded their presence and intensity in each peak. The intensity associated with each set of signature ions allowed us to conclude that peak 1 was comprised of mostly 4-3 dimer whilst peaks 2 and 3 were mostly 3-3 dimer. The remaining ambiguous dimers were then analysed using the same strategy. Out of 46 total dimers identified, 21 contained 3-3 cross-links, and 25 contained 4-3 cross-links (Table S6-11).



Figure 6-7. Validation of gm-AEJ=gm-AEJA and gm-AEJA=gm-AEJ mass coincidence.

(A)Predicted list of signature ions for gm-AEJ=gm-AEJA (3-4 cross-link) and gm-AEJA=gm-AEJ (3-3 cross-link). (B), MS data corresponding to 897.8918 *m/z* ion (top panel) and extracted ion chromatogram showing 3 peaks at consecutive retention times (bottom).

Structure	Theo mass	Abundance	Cumul.	DT (min)
Structure	(Da)	(%)	%	KI (min)
DIMERS				
1 gm-AEJAA-gm-AEJ 2, gm-AEJA-gm-AEJA 2	1864.805	24.40%	24.40%	15.7 ± 0.0
2 gm-AEJ-gm-AEJA 2, gm-AEJA-gm-AEJ 2	1793.768	23.04%	47.44%	14.9 ± 0.0
3 gm-AEJG-gm-AEJ 2	1779.752	16.18%	63.62%	13.6 ± 0.0
4 gm-AEJ-gm-AEJ 2	1722.731	12.82%	76.44%	13.9 ± 0.0
5 gm-AEJG-gm-AEJA 2, gm-AEJQ 1, gm-AEJAG-gm-AEJ 2	1850.789	7.79%	84.23%	14.4 ± 0.0
6 gm-AEJF-gm-AEJ 2	1869.799	2.40%	86.63%	23.9 ± 0.1
7 gm-AEJN-gm-AEJ 2	1836.773	1.84%	88.47%	13.0 ± 0.0
8 gm-AEJ(AG/Q)-gm-AEJA 2	1921.826	1.57%	90.03%	15.0 ± 0.0
9 gm-AEJS-gm-AEJ 2	1809.763	1.18%	91.22%	13.3 ± 0.0
10 gm-AEJF-gm-AEJA 2, gm-AEJAF-gm-AEJ 2	1940.836	0.94%	92.16%	24.7 ± 0.0
11 gm-AEJAI-gm-AEJ 2, gm-AEJI-gm-AEJA 2	1906.852	0.92%	93.08%	23.4 ± 0.0
12 gm-AEJAI-gm-AEJA 2	1977.889	0.90%	93.98%	24.2 ± 0.0
13 gm-AEJN-gm-AEJA 2	1907.811	0.57%	94.55%	13.6 ± 0.0
14 gm-AEJS-gm-AEJA 2	1880.800	0.49%	95.04%	14.1 ± 0.0
15 gm-AEJY-gm-AEJ 2	1885.794	0.47%	95.51%	19.6 ± 0.0
16 gm-AEJAD-gm-AEJA 2, gm-AEJW-gm-AEJA 2, gm-AEJAW-gm-AEJ 2	1979.832	0.42%	95.93%	16.8 ± 0.0
17 gm-AEJK-gm-AEJ 2	1850.825	0.40%	96.34%	12.6 ± 0.0
18 gm-AEJM-gm-AEJ 2	1853.771	0.37%	96.71%	19.0 ± 0.0
19 gm-AEJAA-gm-AEJA 2	1935.842	0.35%	97.06%	16.3 ± 0.0
20 gm-AEJI-gm-AEJ 2	1835.815	0.34%	97.40%	22.8 ± 0.0
21 gm-AEJH-gm-AEJ 2	1859.789	0.33%	97.74%	12.6 ± 0.0
22 gm-AEJH-gm-AEJA 2, gm-AEJAH-gm-AEJ 2	1930.827	0.21%	97.95%	13.3 ± 0.0
23 gm-AEJT-gm-AEJ 2	1823.778	0.20%	98.15%	13.9 ± 0.0
24 gm-AEJM-gm-AEJA 2, gm-AEJAM-gm-AEJ 2	1924.808	0.19%	98.34%	19.7 ± 0.0
25 gm-AEJY-gm-AEJA 2	1956.831	0.18%	98.52%	20.4 ± 0.0
26 gm-AEJW-gm-AEJA 2, gm-AEJAW-gm-AEJ 2, gm-AEJAD-gm-AEJA 2	1979.847	0.16%	98.68%	26.5 ± 0.0
27 gm-AEJAM-gm-AEJA 2	1995.845	0.15%	98.83%	20.4 ± 0.0
28 gm-AEJAK-gm-AEJ 2, gm-AEJK-gm-AEJA 2	1921.863	0.14%	98.97%	13.2 ± 0.0
29 gm-AEJAD-gm-AEJ 2, gm-AEJD-gm-AEJA 2, gm-AEJW-gm-AEJ 2	1908.795	0.14%	99.11%	14.5 ± 0.0
30 gm-AEJAW-gm-AEJA 2	2050.884	0.14%	99.25%	27.0 ± 0.0
31 gm-AEJV-gm-AEJ 2	1821.799	0.11%	99.36%	19.1 ± 0.0
32 gm-AEJD-gm-AEJ 2	1837.757	0.10%	99.46%	13.7 ± 0.0
33 gm-AEJT-gm-AEJA 2	1894.815	0.10%	99.56%	14.7 ± 0.0
34 gm-AEJAR-gm-AEJ 2	1949.869	0.09%	99.65%	13.9 ± 0.1
35 gm-AEJAF-gm-AEJA 2	2011.873	0.08%	99.73%	25.9 ± 0.0
36 gm-AEJW-gm-AEJ 2, gm-AEJD-gm-AEJA 2, gm-AEJAD-gm-AEJ 2	1908.810	0.08%	99.80%	25.4 ± 0.0
37 gm-AEJAH-gm-AEJA 2	2001.864	0.06%	99.86%	13.9 ± 0.0
38 gm-AEJV-gm-AEJA 2	1892.836	0.05%	99.92%	19.9 ± 0.0
39 gm-AEJAE-gm-AEJA 2	1993.847	0.03%	99.95%	15.8 ± 0.0
40 gm-AEJAK-gm-AEJA 2	1992.900	0.03%	99.98%	13.8 ± 0.0
41 gm-AEJAQ-gm-AEJA 2	1992.863	0.02%	100.00%	16.4 ± 0.0

Table S6-10. PGFinder identification of dimers and trimers

Structure	Theo mass	Abundance	Cumui.	RT (min)
TDIMER	(Da)	(%)	%	
	2788 202	18 26%	18 26%	186 + 00
2 gm-AFIAA-gm-AFIAm-AFIA, gm-AFIA-gm-AFIA[3] gm-AFIAm-AFIA]	2700.202	18.09%	26 25%	18.0 ± 0.0 18.1 + 0.0
	2646 127	15 90%	50.55%	176 ± 0.0
	2622 112	12.08%	52.25/0	165 ± 0.0
5 gm AELam AELam AELam	2575 090	9 3/%	72 66%	16.5 ± 0.0
5 gir ALI'gir ALI'gir ALI'gir ALI'gir ALI'gir AEl an AEl an AEl an AEl an AEl Ali'	2373.030	7 47%	75.00% 01 120/	10.7 ± 0.0
7 gm_AEIAA_gm_AEIam_AEI3gm_AEIA1_gm_AEIA1_gm_AEIAG_gm_AEIA13gm_AEIA_gm_AEIA_gm_AEIA13	2703.145	5 30%	01.15/0	17.0 ± 0.0 175 ± 0.0
2 gm Abrag gm Abrag 13, gm Abrag 14, gm Abrag 16, gm Abrag 16, gm Abrag 16, gm Abrag 16, ann an 15	2774.100	2.02%	00.JZ/0	250 ± 0.0
9 gm AEIAC gm AEIAm AFIAI2 gm AEIAI1 gm AEIAC gm AEIAC gm AEIAI2	2722.133	1 2/1%	00.34/0	180 ± 0.0
	2690 122	1 20%	09.00%	15.0 ± 0.0
	2005.155	1.29%	91.17%	15.8 ± 0.0
11 gir-AEJA-gir-AEJS	2002.122	0.30%	92.08%	10.2 ± 0.0
12 gin-Aciar-gin-Aci-gin-Acialis, gin-Aci-gin-Acialis	2795.190	0.77%	92.85%	25.0 ± 0.0
	2700.170	0.72%	93.57%	10.5 ± 0.0
14 gir-Acin-gir-Acin-gir-Acina (5)	2051.207	0.55%	94.10%	16.9 ± 0.0
15 gin-Aciar-gin-Aci-gin-Acia/5, gin-Acia-gin-Acia/5	2004.233	0.50%	94.60%	20.0 ± 0.0
10 gm-Aels-gm-Ael-gm-AelA[3	2/33.159	0.44%	95.04%	16.7 ± 0.0
1/ gm-AEJAl-gm-AEJ-gm-AEJ-gm-AEJ-gm-AEJ-gm-AEJ-gm-AEJA 3	2759.212	0.40%	95.43%	24.0 ± 0.0
18 gm-AEJA-gm-AEJ-gm-AEJ 3	2/12.149	0.35%	95.79%	17.0 ± 1.1
19 gm-Acial-gm-Acial-gm-Acial 3	2901.280	0.35%	96.14%	25.0 ± 0.0
20 gm-AEJS-gm-AEJA-gm-AEJA 3	2804.197	0.35%	96.49%	17.2 ± 0.0
21 gm-AEJAI-gm-AEJ-gm-AEJA[3, gm-AEJI-gm-AEJA-gm-AEJA]3	2830.249	0.32%	96.81%	25.0 ± 0.0
22 gm-AEJY-gm-AEJ-gm-AEJ-3	2/38.154	0.30%	97.12%	21.3 ± 0.0
23 gm-AEJM-gm-AEJ-gm-AEJ	2706.131	0.28%	97.40%	21.4 ± 0.7
24 gm-AEJK-gm-AEJ-gm-AEJ 3	2703.185	0.24%	97.64%	15.4 ± 0.0
25 gm-AEJAA-gm-AEJA-gm-AEJA 3	2859.239	0.24%	97.88%	19.0 ± 0.0
26 gm-AEJ-gm-AEJ-gm-AEJ 3	2688.174	0.22%	98.10%	24.0 ± 0.0
2/ gm-ALJAD-gm-AEJ-gm-AEJA 3, gm-AEJAW-gm-AEJ-gm-AEJ 3, gm-AEJAW-gm-AEJ-gm-AEJA 3, gm-AEJD-gm-AEJA-gm-AEJA 3	2832.192	0.20%	98.30%	18.9 ± 0.0
28 gm-AEJAM-gm-AEJ-gm-AEJ 3, gm-AEJM-gm-AEJ-gm-AEJA 3	2777.168	0.15%	98.45%	21.5 ± 0.0
29 gm-AEJT-gm-AEJ-gm-AEJ 3	2676.138	0.15%	98.60%	16.7 ± 0.0
30 gm-AEJAH-gm-AEJ-gm-AEJ 3, gm-AEJH-gm-AEJ-gm-AEJA 3	2783.186	0.13%	98.73%	16.0 ± 0.0
31 gm-AEJAM-gm-AEJ-gm-AEJA 3, gm-AEJM-gm-AEJA-gm-AEJA 3	2848.205	0.12%	98.85%	22.0 ± 0.0
32 gm-AEJAW-gm-AEJ-gm-AEJ 3, gm-AEJW-gm-AEJ-gm-AEJA 3, gm-AEJAD-gm-AEJ-gm-AEJA 3, gm-AEJD-gm-AEJA-gm-AEJA 3	2832.207	0.12%	98.97%	27.1 ± 0.0
33 gm-AEJY-gm-AEJ-gm-AEJA 3	2809.191	0.12%	99.09%	21.8 ± 0.0
34 gm-AEJY-gm-AEJA-gm-AEJA 3	2880.228	0.12%	99.21%	22.5 ± 0.0
35 gm-AEJAW-gm-AEJA-gm-AEJA 3	2974.281	0.10%	99.31%	24.5 ± 4.8
36 gm-AEJAH-gm-AEJ-gm-AEJA 3, gm-AEJH-gm-AEJA-gm-AEJA 3	2854.223	0.10%	99.41%	16.5 ± 0.0
37 gm-AEJV-gm-AEJ-gm-AEJ 3	2674.159	0.08%	99.49%	21.4 ± 0.0
38 gm-AEJAM-gm-AEJA-gm-AEJA 3	2919.242	0.07%	99.56%	22.5 ± 0.0
39 gm-AEJAK-gm-AEJ-gm-AEJ 3, gm-AEJK-gm-AEJ-gm-AEJA 3	2774.222	0.07%	99.63%	15.9 ± 0.0
40 gm-AEJD-gm-AEJ-gm-AEJ 3	2690.117	0.06%	99.70%	17.2 ± 0.0
41 gm-AEJW-gm-AEJ-gm-AEJ 3, gm-AEJAD-gm-AEJ-gm-AEJ 3, gm-AEJD-gm-AEJ-gm-AEJA 3	2761.170	0.06%	99.76%	26.1 ± 0.0
42 gm-AEJAW-gm-AEJ-gm-AEJA 3, gm-AEJW-gm-AEJA-gm-AEJA 3, gm-AEJAD-gm-AEJA-gm-AEJA 3	2903.244	0.06%	99.82%	27.4 ± 0.0
43 gm-AEJT-gm-AEJ-gm-AEJA 3	2747.175	0.05%	99.87%	17.2 ± 0.0
44 gm-AEJAD-gm-AEJA-gm-AEJA 3, gm-AEJAW-gm-AEJ-gm-AEJA 3, gm-AEJW-gm-AEJA-gm-AEJA 3	2903.229	0.04%	99.91%	19.3 ± 0.0
45 gm-AEJV-gm-AEJ-gm-AEJA 3	2745.196	0.03%	99.95%	21.7 ± 0.0
46 gm-AEJV-gm-AEJA-gm-AEJA 3	2816.233	0.03%	99.98%	22.1 ± 0.0
47 gm-AEJAD-gm-AEJ-gm-AEJ 3, gm-AEJW-gm-AEJ-gm-AEJ 3, gm-AEJD-gm-AEJ-gm-AEJA 3	2761.154	0.02%	100.00%	17.8 ± 0.0

Ion Nb	Type	<u>m/z</u>	Parts	Ion Nb	Type	m/z	Parts
	.,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	AFIA-gm_A	EL (3-4) fragments		.) 0	n_AEI_gm_AE	(3-3) fragments
1	C-Terminal	871 3770	dm(r)-AEI	1	C-Terminal	90 0550	Δ*
2	C Torminal	0/1.3//9	$A = am(r) A \equiv I (A = 2)$	2	C Torminal	90.0330	
2	C Torminal	1111 / 1000	A = g(1)(1) = A = J(4 = 3)	2	C Torminal	1111 / 1000	gm(r) AEI=IA (2, 2)
3	C Torminal	1114.4990	$\beta = \alpha m(r) \Lambda = 1 (4 - 3)$	3	C Torminal	1114.4990	$\lim_{r \to \infty} \frac{1}{r} \sum_{r \to \infty} $
4	C-Terminal	1114.4990	JA-gIII(1)-AEJ(4-3)	4	C-Terminal	1114.4990	J = gin(1) - AEJA (3-3)
5	C-Terminal	1243.5424	$g_{III}(I)$ -AEJA=EJ (4-3)	5	C-Terminal	1243.5424	$g(\Pi(I)-AEJ=EJA(3-3)$
	C-Terminal	1243.5424	EJA=gIII(I)-AEJ (4-3)	6	C-Terminal	1243.5424	EJ=gIII(I)-AEJA(3-3)
/	C-Terminal	1314.5795	gm(r)-AEJA=AEJ (4-3)	/	C-Terminal	1314.5795	gm(r)-AEJ=AEJA (3-3)
8	C-Terminal	1314.5795	AEJA=gm(r)-AEJ (4-3)	8	C-Terminal	1314.5795	AEJ=gm(r)-AEJA (3-3)
9	C-Terminal	1591.6956	m(r)-AEJA=gm(r)-AEJ(4-3)	9	C-Terminal	1591.6956	m(r)-AEJ=gm(r)-AEJA (3-3)
10	C-Terminal	1591.6956	gm(r)-AEJA=m(r)-AEJ (4-3)	10	C-Terminal	1591.6956	gm(r)-AEJ=m(r)-AEJA (3-3)
11	Internal	72.0444	A	11	Internal	72.0444	A
12	Internal	130.0499	E	12	Internal	130.0499	E
13	Internal	173.0921	J	13	Internal	173.0921	J
14	Internal	191.1026	J*	14	Internal	201.0870	AE
15	Internal	201.0870	AE	15	Internal	262.1397	JA
16	Internal	244.1292	JA	16	Internal	278.1234	m(r)
17	Internal	262.1397	A=J (4-3)	17	Internal	302.1347	EJ
18	Internal	278.1234	m(r)	18	Internal	345.1769	J=J (3-3)
19	Internal	302.1347	EJ	19	Internal	349.1605	m(r)-A
20	Internal	320,1452	FI	20	Internal	373 1718	AFL
21	Internal	349,1605	m(r)-A	21	Internal	391,1823	FIA
22	Internal	373 1718	EIA	22	Internal	434 2245	I = IA(3-3)
23	Internal	373 1718	ΔE1	23	Internal	/62 2195	
24	Internal	301 1823	AEI	24	Internal	402.2100	I=FI (3-3)
24	Internal	201 1023	A=EI(A,2)	24	Internal	474.2195	I = I (3 - 3)
25	Internal	124 2245	A = LJ(4 - 3)	25	Internal	474.2193	$m(r) \Lambda E$
20	Internal	434.2243	JA-J (4-3)	20	Internal	470.2031 E4E 2ECC	
2/	Internal	444.2009		2/	Internal	545.2500	J-AEJ (3-3)
28	Internal	462.2195	A=AEJ (4-3)	28	Internal	545.2566	AEJ=J (3-3)
29	Internal	4/8.2031	m(r)-AE	29	Internal	563.2671	J=EJA (3-3)
30	Internal	563.2671	JA=EJ (4-3)	30	Internal	563.2671	EJ=JA (3-3)
31	Internal	563.2671	EJA=J (4-3)	31	Internal	603.2620	EJ=EJ (3-3)
32	Internal	634.3042	JA=AEJ (4-3)	32	Internal	634.3042	J=AEJA (3-3)
33	Internal	634.3042	AEJA=J (4-3)	33	Internal	634.3042	AEJ=JA (3-3)
34	Internal	650.2879	m(r)-AEJ	34	Internal	650.2879	m(r)-AEJ
35	Internal	668.2985	m(r)-AEJ	35	Internal	674.2992	EJ=AEJ (3-3)
36	Internal	692.3097	EJA=EJ (4-3)	36	Internal	674.2992	AEJ=EJ (3-3)
37	Internal	721.3250	m(r)-AEJA	37	Internal	692.3097	EJ=EJA (3-3)
38	Internal	739.3356	A=m(r)-AEJ (4-3)	38	Internal	739.3356	m(r)-AEJA
39	Internal	763.3468	EJA=AEJ (4-3)	39	Internal	745.3363	AEJ=AEJ (3-3)
40	Internal	763.3468	AEJA=EJ (4-3)	40	Internal	763.3468	EJ=AEJA (3-3)
41	Internal	834.3840	AEJA=AEJ (4-3)	41	Internal	763.3468	AEJ=EJA (3-3)
42	Internal	911.4204	m(r)-AEJA=J (4-3)	42	Internal	822.3727	m(r)-AEJ=J (3-3)
43	Internal	911.4204	JA=m(r)-AEJ (4-3)	43	Internal	822.3727	J=m(r)-AEJ (3-3)
44	Internal	1040.4630	m(r)-AEJA=EJ (4-3)	44	Internal	834.3840	AEJ=AEJA (3-3)
45	Internal	1040.4630	EJA=m(r)-AEJ(4-3)	45	Internal	911,4204	m(r)-AEJ=JA(3-3)
46	Internal	1111.5001	m(r)-AEJA=AEJ (4-3)	46	Internal	911.4204	J=m(r)-AEJA(3-3)
47	Internal	1111.5001	AFIA=m(r)-AFI(4-3)	47	Internal	951,4153	m(r)-AFI=FI(3-3)
48	Internal	1388.6163	m(r)-AFIA= $m(r)$ -AFI(4-3)	48	Internal	951,4153	FI=m(r)-AFI(3-3)
49	N-Terminal	204 0866	g	49	Internal	1022 4524	m(r)-AEI=AEI(3-3)
50	N-Terminal	/81 2028	gm(r)	50	Internal	1022.4024	$\Delta EI = m(r) - \Delta EI (3-3)$
51	N-Terminal	552 2300	$gm(r)_{-}\Lambda$	50	Internal	1025.4524	$dm(r)_AEI=1(3-3)$
51	N Torminal	601 2025	dm(r) AE	51	Internal	1025.4521	$\int dr r (r) \Delta E I (2, 2)$
52	N Terminal	001.2023	gm(r) AEI	52	Internal	1025.4521	$p = g(r) \wedge E = E [A (2, 2)]$
55	N-Terminal	003.3073	gm(r) AEIA	55	Internal	1040.4630	$\prod(I) - AEJ - EJA(3-3)$
54	N-Terminal	924.4044	gm(r)-AEJA	54	Internal	1040.4630	EJ=III(I)-AEJA(3-3)
55		1/94.//50	gm(r)-AEJA=gm(r)-AEJ (4-3)	55	Internal	1111.5001	m(r)-AEJ=AEJA (3-3)
C-term	inal residue (+	Π ₂ Ο)		50	internal	1111.5001	AEJ=III(I)-AEJA (3-3)
				57	internal	1154.4947	gm(r)-AEJ=EJ (3-3)
				58	Internal	1154.4947	EJ=gm(r)-AEJ(3-3)
				59	Internal	1225.5318	gm(r)-AEJ=AEJ (3-3)
				60	Internal	1225.5318	AEJ=gm(r)-AEJ (3-3)
				61	Internal	1299.5686	m(r)-AEJ=m(r)-AEJ (3-3)
				62	Internal	1388.6163	m(r)-AEJ=m(r)-AEJA (3-3)
				63	Internal	1502.6480	m(r)-AEJ=gm(r)-AEJ (3-3)

Table S6- 11. *In silico* fragmentation of the two dimers (gm-AEJ=gm-AEJA (4-3) ang gm-AEJA=gm-AEJ (3-3)) and consolidated list of expected ions

69 70 71

-τer

Internal N-Terminal N-Terminal N-Terminal N-Terminal N-Terminal

N-Terminal

1₂0)

ninal residue

1502.6480 m(r)-AEJ=gm(r)-AEJ (3-3) 1502.6480 gm(r)-AEJ=m(r)-AEJ (3-3) 204.0866 g 481.2028 gm(r) 552.2399 gm(r)-A 681.2825 gm(r)-AE 853.3673 gm(r)-AEJ 1705.7273 gm(r)-AEJ=gm(r)-AEJ (3-3) 1794.7750 gm(r)-AEJ=gm(r)-AEJ (3-3)

gm AEI gm AEI [4-3] gm AEI gm AEI (3-3) 1 20005495 N/A* A 3 30004987 E E 4 173.09207 J J J 5 191.10263 J* N/A A 6 201.08698 AE AE E 7 204.08665 S S S 8 244.12918 JA N/A S 13 302.13466 E N/A J=I 13 320.14523 E N/A I=I 13 345.17686 N/A J=I (3-3) 14 349.16054 m(h/A m(h/A III 15 373.17176 EIA AEIA AEIA 16 391.18234 AEJ EIA JAIIIII 17 343.22454 JA=J(4-3) AEIA IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	lon Nb	m/z		Parts
1 72.0439 A A 2 00.054954 N/A* 3 130.04997 E E 4 173.09207 J J 5 191.10263 J* N/A 6 201.08698 AE AE 7 204.08655 g g 8 244.12918 JA N/A 10 275.12343 m(n) m(r) 11 302.13466 EI EI 12 320.14523 EI N/A 13 345.17686 N/A 14 349.16054 m(r)-A m(r)-A 15 373.17178 EIA EIA 16 391.18234 AEJ EIA 17 434.20889 AEJA N/A 19 462.21945 A-AEJ (4-3) AEIA 21 474.21045 N/A 23 545.26671 N/A 24 52.23991 gm(r)-A gm(r)-A 25 563.26713 JA=EI (4-3) J=AEIA (3-3) 26 603.252055 N/A EI=EIA (3-3) 25 653.2673 JA=EI (4-3) J=AEIA (3-3) 26 603.252			gm-AEJA=gm-AEJ (4-3)	gm-AEJ=gm-AEJA (3-3)
2 0.00-9895 N/A* A** 3 130.04987 F F 4 173.02207 J J 5 191.10263 N/A AE 7 204.08665 8 R 8 244.12918 N/A N/A 9 262.13975 Ar N/A 11 302.13466 H H 12 320.14523 EI N/A 13 343.16054 M(A J=I(3-3) 14 343.16054 M(A AEI 15 373.17178 EIA AEI 16 391.18234 AEI EIA 17 434.22454 J=I(4-3) J=IA(3-3) 18 444.20889 AEIA N/A 19 462.21945 N/A J=EIA(3-3) 24 481.2028 gm(r) gm(r) 25 S3.26713 J=AEI(3-3) J=EIA(3-3) 24 482.2088 gm(r)/AE gm(r)/AE 25 52.39391 gm(r)/AE gm(r)/AE	1	/2.04439	Α	A
3 3.0.0398 / E E 4 1.73.09207 / J J 5 1.91.10263 /* N/A 6 201.08698 AE AE 7 204.08655 g S 8 244.12918 JA N/A 10 275.13243 m(r) m(r) 11 302.13266 EJ EJ 12 320.14523 EJ N/A 13 345.17656 N/A 14 349.16054 m(r)-A m(r)-A 15 373.17178 EJA AEI 16 391.18234 AEI EJA 17 34.22454 JA=[(4-3) J=JA (3-3) 18 444.20839 AEJA M(A 19 462.21995 A-AEJ (4-3) AEJA 20 474.21985 N/A J=EJA (3-3) 21 478.20314 m(r)-AE gm(r)-A gm(r)-A 22 451.2028 gm(r) MA J=EAA (3-3) 22 652.23991 gm(r)-AE gm(r)-A gm(r)-A 23 563.26713 JA=EJ (4-3) J=EAA (3-3) J 24 552.269713 EJA=EI (4-3) gm(r)-AE <	2	90.054954	N/A*	A**
1 1/3 0/20/ J J 1 1/3 0/20/ J N/A 2 01.08689 AE AE 7 204.08665 g g 8 244.12918 JA N/A 9 26.21.3975 As 1(4-3) JA 10 278.12343 m(r) m(r) 11 302.13466 EJ EJ 12 302.14523 EJ N/A 13 345.1758 N/A J=J (3-3) 14 349.16054 m(r)-A m(r)-A 15 373.17178 EJA AEJ 16 391.18234 AEJ EJA 17 434.22454 JA=J (4-3) J=JA (3-3) 18 444.20889 ACJ m(r)-A 19 452.2394 m(r)-AE m(r)-AE 21 478.20314 m(r)-AE m(r)-AE 22 481.2028 gm(r) gm(r) 23 565.2673 JA=EJ (4-3) J=EA (3-3) 24 552.23991 gm(r)-A gm(r)-A 25 663.2673 JA=EJ (4-3) J=AE (3-3) 26 603.26205 N/A I=EI (3-3) 27 63.30425 JA=AEJ (4-3) J=AE (3-3) 28 650.28703 m(r)-AEJ m(r)-AEI 29 662.28849 m(r)-AEJ m(r)-AEI 30 674.29916 N/A EI=EIA (3-3) 31 681.28251 gm(r)-AEJ m(r)-AEI<	3	130.04987	E	E
5 191.10265 // 6 201.08698 AE 7 204.08656 g 8 244.12918 // 10 278.12343 /// 11 302.13466 E 12 30.14523 E // 13 355.17686 N/A // 14 349.16054 m(r)-A m(r)-A 15 373.17178 EIA AEI 16 391.18234 AEJ EIA 17 344.20889 AEIA M/A 18 444.20889 AEIA M/A 19 452.21945 AAEJ AEIA 20 474.21945 N/A J=EIA(3-3) 21 478.20314 m(r)-AE gm(r) 23 552.23991 gm(r)-AE gm(r)-AE 24 552.36975 N/A I=EIA(3-3) 25 563.26713 JAEEIA(3-3) m(r)-AE 24 552.39991 gm(r)-AE gm(r)-AE 25 563.26712 N/A I=EIA(3-3)	4	1/3.09207	J	J
b 2.02.0.0056 g g 7 2.04.0.0056 g g 8 2.44.12918 A N/A 9 262.13975 A=1(4-3) JA 10 27.8.12343 m(r) m(r) 11 302.13466 E) E) 12 320.14528 E N/A 13 349.16054 m(r)-A m(r)-A 15 373.17178 EJA AEJ 16 391.18234 AEJ EIA 17 444.20389 AEJA N/A 18 442.20455 (A-AEJ (4-3) AEJA 19 452.21945 (A-AEJ (4-3) AEJA 14 442.20389 M(r)-AE m(r)-AE 14 42.22454 (A-J) AEJA 12 452.25557 N/A J=EJ(3-3) 14 42.2038 m(r)-AE m(r)-AE 15 552.3671 JA=EJ(4-3) J=AEJA (3-3) 12 634.30425 JA=AEJ (4-3) J=AEJA (3-3) 12 634.20425 JA=AEJ (4-3) m(r)-AEJ 13 651.26251 gm(r)-AE gm(r)-AE 14 91.42.2044 (m(r)-AEIA m(r)-AEI 15	5	191.10263	J [≁]	N/A
2.04.0805 g g 8 2.24.12918 A N/A 9 262.13975 A=1 (A-3) JA 10 278.12343 m(r) m(r) 11 302.13466 EJ EJ 12 320.14525 EJ N/A 13 345.17686 N/A 14 349.16054 m(r)-A m(r)-A 15 373.1717 EIA AEJ 16 391.18234 AEJ EIA 17 434.22454 JA=J (4-3) J=JA (3-3) 18 444.0388 AEJA N/A 19 452.21945 A=AEJ (4-3) AEJA 20 474.21945 N/A 21 478.0214 m(r)-AE m(r)-AE 22 481.2028 gm(r) gm(r) 23 545.25657 N/A 24 522.3991 gm(r)-A gm(r)-A 25 53.26713 JA=EI (4-3) J=EL (3-3) 26 603.26205 N/A 27 634.30425 JA=EI (4-3) J=AEJA (3-3) 28 602.29949 m(r)-AEI 29 668.29849 m(r)-AEI 29 668.29849 m(r)-AEI 29 668.29849 m(r)-AEI 30 674.29916 N/A 31 6612.28251 gm(r)-AE </th <th>6</th> <th>201.08698</th> <th>AE</th> <th>AE</th>	6	201.08698	AE	AE
2 2 2 2 3 2 22.13975 A 10 278.12343 m(r) m(r) 11 302.13466 EI EI 12 320.14525 EI N/A 13 345.17686 N/A J=I (3-3) 14 349.16054 m(r)-A m(r)-A 15 373.17175 EIA AEI 16 391.18234 AEI CA 17 434.22645 J=IA (-3) J=IA (3-3) 18 444.2088 AEIA M/A 19 462.21945 A-AEI (4-3) AEIA 10 474.21945 N/A J=EI (3-3) 14 478.20314 m(r)-AE gan(r) 21 478.20314 m(r)-AE gan(r) 22 481.2028 gm(r) J=AEI (3-3) 23 552.3991 gm(r)-AE gm(r) 24 552.3991 m(r)-AEI gan(2-3) 25 6	/	204.08665	g	g
JB ZB: 1243 m(r) M 10 278: 1243 m(r) m(r) 11 302.13466 E E 12 320.14523 E N/A 13 345.17686 N/A 14 349.16054 m(r)-A m(r)-A 15 373.1717 EIA AEI 16 391.18234 AEI EIA 17 434.22454 IA=1 (4-3) J=IA (3-3) 18 444.0089 AEIA N/A 19 452.21945 A=AEI (4-3) AEIA 20 474.21945 N/A J=EU (3-3) 21 478.20314 m(r)-AE m(r)-AE m(r)-AE 22 545.25657 N/A J=ELG (3-3) 23 545.25657 N/A EI-EU (3-3) 24 525.23991 gm(r)-A gm(r)-AE S 25 603.26205 N/A EI-EU (3-3) 26 603.26205 N/A EI-EU (3-3) 27 643.30425 IA=AEI (4-3) I=AEIA (3-3) 28 668.29849 m(r)-AEI <td< th=""><th>8</th><th>244.12918</th><th></th><th>N/A</th></td<>	8	244.12918		N/A
10 278.12343 m(r) m(r) 11 302.13466 E E 12 320.14523 E N/A 13 345.17686 N/A 14 349.16054 m(r)-A m(r)-A 15 373.17178 E/A AEI 16 391.18234 AE E/A 17 434.22454 JA=J (4-3) J=IA (3-3) 18 444.0088 AE/A N/A 19 452.21945 A=AEJ (4-3) AEIA 20 474.21945 N/A J=EIA (3-3) 21 478.20314 m(r)-AE m(r)-AE 22 481.2028 gm(r) gm(r) 23 542.25657 N/A J=EIA (3-3) 24 550.326713 JA=EJ (4-3) J=EIA (3-3) 25 563.26713 JA=EJ (4-3) J=AEIA (3-3) 26 603.26025 N/A EI=EI (3-3) 27 643.20425 Ja=AEI (4-3) m(r)-AEI 28 650.28793 m(r)-AEI m(r)-AEI 29 668.29849 m(r)-AEI m(r)-AEI 21 723.3561.A=m(r)-AEI m(r)-AEI 23 721.32504 m(r)-AEI m(r)-AEI 24 739.33561.A=m(r)-AEI m(r)-AEI 25 63.36872 m(r)-AEI m(r)-AEI 26 63.398483395 AEI-AEI (4-3) m(r)-AE	9	262.13975	A=J (4-3)	JA
11 302.14523 EJ 12 320.14523 FI 13 345.17686 N/A J-[3-3] 14 349.16054 m(r)-A m(r)-A 15 373.17178 EJA AEI 16 391.18234 AEI EJA 17 434.22454 JA=J (4-3) JEJA (3-3) 18 444.0889 AEIA N/A 20 474.21945 N/A JELI (3-3) 21 478.20314 m(r)-AE m(r)-AE 22 478.20314 m(r)-AE gm(r)-A 23 545.26567 N/A JELI (3-3) 24 552.3991 gm(r)-A gm(r)-A 25 563.26713 JA=EI (4-3) JELI (3-3) 26 650.28733 m(r)-AEI GM 25 650.28733 m(r)-AEI gm(r)-AE 26 650.28733 m(r)-AEI gm(r)-AE 27 634.30425 JA=AEI (4-3) M(A 36 74.29916 N/A EI=AEI (3-3) 371.3525 m(r)-AEI	10	278.12343	m(r)	m(r)
12 345.17666 N/A J=J (3-3) 14 349.16054 m(r)-A m(r)-A 15 373.17178 EIA AEI 16 391.18234 AEI EIA 17 434.22454 JA=J (4-3) J=JA (3-3) 18 444.20883 N/A J=LI (3-3) 18 444.20883 N/A J=LI (3-3) 14 782.0314 m(r)-AE m(r)-AE m(r)-AE 24 782.0314 m(r)-AE m(r)-AE m(r)-AE 25 53.26713 JA=LI (4-3) J=LI (3-3) 26 603.26205 N/A J=AEIA (3-3) 27 634.30425 Ja=AEI (4-3) J=AEIA (3-3) 28 650.28793 m(r)-AEI m(r)-AEI 29 668.29849 m(r)-AEI m(r)-AEI 29 668.29849 m(r)-AEI m(r)-AEI 29 668.29849 m(r)-AEI m(r)-AEI 30 674.29916 N/A EJ=AEI (3-3) 31 681.28251 gm(r)-AEI m(r)-AEI 31 681.28251	11	302.13466	EJ	EJ
13 345.1766 N/A - -3) 14 349.1605 m(r)-A m(r)-A 15 373.17178 EIA AEJ 16 391.18234 AEJ EIA 17 434.22454 JA=1 (-3) 18 444.20889 AEJA N/A 19 462.21945 AEJA N/A 20 474.21945 N/A j=EJ (3-3) 21 478.20314 m(r)-AE m(r)-AE 23 545.26657 N/A j=AEJ (3-3) 24 522.23991 gm(r)-A gm(r)-A 25 632.26205 N/A EJEJA (3-3) 26 632.26205 N/A EJEJA (3-3) 26 650.28793 m(r)-AE m(r)-AE 29 668.29849 m(r)-AE m(r)-AE 29 668.29849 m(r)-AE m(r)-AE 29 668.29849 m(r)-AE m(r)-AE 20 674.29916 N/A EJEAEJ (3-3) 31 61.224521 m(r)-AEIA m(r)-AEIA	12	320 14523	FI	N/A
14 349.1055 m(r)-A 15 373.17178 EIA AEJ 16 391.18234 AEJ EIA 17 434.22454 JAEJ EIA 18 444.20889 AEJ ISIA(3-3) 18 444.20889 AEJA N/A 19 462.21945 A=AEJ (4-3) AEJA 14 474.21945 N/A J=EJ (3-3) 14 782.0314 m(r)-AE m(r)-AE 22 481.2028 gm(r) gm(r) 23 545.26557 N/A J=AEJ (3-3) 24 552.23991 gm(r)-A gm(r)-A 25 563.26713 JA=EJ (4-3) J=AEJA (3-3) 26 602.28793 m(r)-AEI m(r)-AEI 29 668.29849 m(r)-AEI m(r)-AEI 29 668.29849 m(r)-AEI m(r)-AEI 20 672.39916 N/A EI=AEJ (3-3) 31 631.26251 m(r)-AEI M(A 32 692.30973 EIA=EJ (4-3) EI=AEJA (3-3) 37	12	245 17000	LJ N1/A	
14 349.16054 m(r)-A m(r)-A 15 373.17178 EIA AEJ 16 391.18234 AEJ EIA 17 434.22454 JA=1 (4-3) J=JA (3-3) 18 444.20889 AEJA N/A 19 452.21945 A-AEJ (4-3) AEJA 20 474.21945 N/A J=EJ (3-3) 21 478.20314 m(r)-AE m(r)-AE 23 545.26567 N/A J=EJ (3-3) 24 552.23991 gm(r)-A gm(r)-A J=EJ (3-3) 25 563.26713 JA=EJ (4-3) J=AEJ (3-3) J=AEJ (3-3) 26 630.26205 N/A EJ=EJ (3-3) J=AEJ (3-3) 27 634.30425 JA=AEJ (4-3) JAEJ (3-3) m(r)-AEJ 29 650.28793 m(r)-AEJ m(r)-AEJ N/A 20 674.29916 N/A EJ=EJA (3-3) J 31 681.28251 gm(r)-AE gm(r)-AEJ N/A J 33 721.32564 m(r)-AEJ M/A J J 34 739.33561 A=m(r)-AEJ m(r)-AEJ J J 34 739.33561 A=m(r)	12	345.17000	IN/A	J=J (3-3)
15 373.17.17.8 EIA AEI 16 391.18234 AEI EIA 17 434.22454 IA-13 (A-3) IA-13 (A-3) 18 444.20889 AEIA N/A 19 462.21945 A-AEI (A-3) AEIA 20 473.20344 m(r)-AE m(r)-AE 21 478.20314 m(r)-AE m(r)-AE 22 481.2028 m(r) gm(r) 23 545.25657 N/A p-AEI (3-3) 24 552.23991 gm(r)-A gm(r)-A 25 563.26713 JA=EJ (4-3) j-AEI (3-3) 26 603.26205 N/A Ej-ED (3-3) 27 634.30425 JA=AEI (4-3) j-AEI (3-3) 28 650.28793 m(r)-AEI m(r)-AEI 29 668.29849 m(r)-AEI gm(r)-AE 30 674.29916 N/A Ej-ELAI (3-3) 31 681.28251 gm(r)-AEI m(r)-AEI 32 692.30973 EJ-AEI (4-3) Ej-AEI (3-3) 33 721.32504 m(r)-AEI m(r)-AEI 34 739.33551 AAEI AEI AEI 35 745.33627 N/A AEI-AEI (3-3) </th <th>14</th> <th>349.16054</th> <th>m(r)-A</th> <th>m(r)-A</th>	14	349.16054	m(r)-A	m(r)-A
16 391.18234 AEJ EJA 17 434.22854 JA=J (4-3) J=JA (3-3) 18 444.20889 AEJA N/A 19 462.21945 A=AEJ (4-3) AEJA 20 474.2089 A=AEJ (4-3) BEJA 21 478.20314 m(r)-AE m(r)-AE 22 481.2028 gm(r) gm(r) 23 545.25657 N/A J=EJA (3-3) 24 552.23991 gm(r)-A gm(r)-A J=EJA (3-3) 25 653.26713 JA=EJ (4-3) J=AEJA (3-3) J=AEJA (3-3) 26 603.26205 N/A EJ=AEJ (3-3) M(r)-AEJ 29 668.29849 m(r)-AEJ m(r)-AEJ M(r)-AEJ 20 674.29916 N/A EJ=AEJ (3-3) M(r)-AEJ 31 681.288251 gm(r)-AE gm(r)-AE J=AEJA (3-3) M(r)-AEJA M(r)-AEJA 33 721.32504 m(r)-AEJA m(r)-AEJA N/A J=AEJA (3-3) J=AEJA (3-3) <td< th=""><th>15</th><th>373.17178</th><th>EJA</th><th>AEJ</th></td<>	15	373.17178	EJA	AEJ
17 434.22454 JA=J (4-3) J=JA (3-3) 18 444.20889 AEJA N/A 19 462.21945 A=AEJ (4-3) AEJA 20 474.21945 N/A J=EJ (3-3) 21 478.20314 m(r)-AE m(r)-AE m(r)-AE 22 481.2028 gm(r) gm(r) gm(r) 23 545.25657 N/A J=EJA (3-3) 24 552.23991 gm(r)-A gm(r)-A J=EJA (3-3) 25 563.26713 JA=EJ (4-3) J=EJA (3-3) J=EJA (3-3) 26 603.26205 N/A EJ=EJ (3-3) 27 634.30425 JA=AEJ (4-3) J=AEJA (3-3) 28 650.28793 m(r)-AEJ m(r)-AEJ 30 674.29916 N/A EJ=AEJ (3-3) 31 681.28251 gm(r)-AEJ gm(r)-AEJ N/A 33 721.32504 m(r)-AEJA m(r)-AEJA N/A 34 739.33561 A=m(r)-AEJA EJ=AEJA (3-3) J 35 745.33627 N/A AEJ=AEJA (3-3) J 36 763.34684 EJA=AEJ (4-3) EJ=AEJA (3-3) J J 39	16	391.18234	AEJ	ALA
18 444.2088 AEJA N/A 19 462.21945 A=AEJ AEJA 20 474.21945 N/A J=EJ (3-3) 21 478.20314 m(r)-AE m(r)-AE 22 481.2028 m(r) gm(r) 23 545.25657 N/A J=EJA (3-3) 24 552.23991 gm(r)-A gm(r)-A 25 653.26713 JA=EJ (4-3) J=EJA (3-3) 26 603.26205 N/A EJ=LEI (3-3) 27 634.30425 JA=EJ M(r)-AEJ 29 650.28793 m(r)-AEJ m(r)-AEJ 20 674.29916 N/A EJ=AEJ (3-3) 31 681.28251 gm(r)-AE gm(r)-AE 30 674.29916 N/A AEJ=AEJ (3-3) 31 681.28251 gm(r)-AEJ M(r)-AEJA 34 739.33561 A=m(r)-AEJ (4-3) m(r)-AEJ 35 745.33627 N/A AEJ=AEJ (3-3) 36 763.34684 EJ=AEJ (4-3) EJ=AEJ (3-3) 37 822.37272 N/A <th>17</th> <th>121 22151</th> <th>$\Lambda - (A_2)$</th> <th>(-1)(2-2)</th>	17	121 22151	$ \Lambda - (A_2)$	(-1)(2-2)
18 444.20889 AEJA N/A 19 462.21945 A-AEJ (4-3) AEJA 20 474.21945 N/A J=EJ (3-3) 21 478.20314 m(r)-AE m(r)-AE 23 545.25657 N/A J=AEJ (3-3) 24 552.23991 gm(r)-A gm(r)-A 25 563.26713 JA=EJ (4-3) J=ELA (3-3) 26 603.26205 N/A EJ=EL (3-3) 27 634.30425 JA=AEJ (4-3) J=AEJA (3-3) 28 650.28793 m(r)-AEJ m(r)-AEJ 29 668.29849 m(r)-AEJ m(r)-AEJ 20 674.29916 N/A EJ=EJA (3-3) 31 681.28251 gm(r)-AEJ m(r)-AEJ 30 674.29916 N/A EJ=AEJA (3-3) 31 681.28251 gm(r)-AEJ m(r)-AEJA 31 681.28251 gm(r)-AEJ m(r)-AEJA 31 743.33261 A=m(r)-AEJA M/A 32 739.33561 A=m(r)-AEJA M/A 33 72.32504	17	434.22434	JA-J (4-3)	J-JA (2-2)
19 462.1945 N/A J=EJ (3-3) 21 478.20314 m(r)-AE m(r)-AE 22 481.2028 gm(r) gm(r) 23 552.23991 gm(r)-A gm(r)-A 24 552.23991 gm(r)-A gm(r)-A 25 563.26713 J=ELA (3-3) 26 603.26205 N/A EJ=EI (3-3) 27 634.30425 JA=AEJ (4-3) J=AEJA (3-3) 28 650.28793 m(r)-AEJ m(r)-AE 29 668.29849 m(r)-AEJ m(r)-AE 20 674.29916 N/A EJ=AEJ (3-3) 31 681.28251 gm(r)-AE gm(r)-AE J 33 721.32504 m(r)-AEJA N/A 34 739.33561 A=m(r)-AEJ (4-3) m(r)-AEJA N/A 35 745.33627 N/A AEJ=AEJ (3-3) J 36 763.34684 EJA=AEJ (4-3) EJ=AEJA (3-3) J 37 822.37272 N/A M(r)-AEJ=J M(r)-AEJ 38 834.38395 AEJA=AEJ (4-3) EJ=AEJA (3-3) J J J 39 853.3673 gm(r)-AEJ gm(r)-AEJ	18	444.20889	AEJA	N/A
20 474.21945 N/A J=E1 [3-3] 21 478.20314 m(r)-AE m(r)-AE 22 481.2028 gm(r) gm(r) 23 542.2657 N/A J=AEJ [3-3] 24 552.23991 gm(r)-A gm(r)-A 25 553.26713 J=EJ (4-3) J=EJA (3-3) 26 603.2605 N/A EJ=EJ (3-3) 26 603.2605 N/A EJ=EJ (3-3) 27 634.30425 J=AEJ (4-3) J=AEJ (3-3) 28 650.28733 m(r)-AEJ m(r)-AEJ 29 668.29849 m(r)-AEJ M(r)-AEJ 30 674.29916 N/A EJ=AEJ (3-3) 31 681.28251 gm(r)-AE J=AEJ (3-3) 33 721.32504 m(r)-AEJA N/A 35 745.33627 N/A AJ=AEJ (3-3) 36 73.33561 A=m(r)-AEJ (4-3) M(r)-AEJ 37 822.3722 N/A AJ=AEJ (3-3) 37 822.3722 N/A M(r)-AEJ 37 822.37272 N/A	19	462.21945	A=AEJ (4-3)	AEJA
21 478.20314 m(r)-AE m(r)-AE 22 481.2028 gm(r) gm(r) 23 542.25657 N/A J=AEJ (3-3) 24 552.23991 gm(r)-A gm(r)-A 25 563.26713 JAEL (4-3) J=EJA (3-3) 26 603.26205 N/A EI=EJ (3-3) 27 634.30425 JAEAEJ (4-3) J=AEJA (3-3) 28 650.28793 m(r)-AEJ N/A 29 668.29849 m(r)-AE m(r)-AE 29 668.29849 m(r)-AEJ M/A 30 674.29916 N/A EI=EJA (3-3) 31 681.28251 gm(r)-AE m(r)-AEJ 32 692.30973 EJA=EJ (4-3) m(r)-AEJA 34 739.33561 A=m(r)-AEJ M/A 35 745.33627 N/A AEJ=AEJA (3-3) 36 763.34684 EJA=AEJ (4-3) EJAELA (3-3) 37 822.3727 N/A M(r)-AEJA 38 834.38395 AEJA=AEJ (4-3) m(r)-AEJA 40 871.37766	20	474.21945	N/A	J=EJ (3-3)
22 481.2023 m(r) AL m(r) 23 545.25657 N/A J=AEJ (3-3) 24 552.23991 gm(r)-A gm(r)-A 25 563.26015 N/A L=EJ (3-3) 26 603.26205 N/A L=EJ (3-3) 27 634.30425 J=AEJ (4-3) J=AEJA (3-3) 28 650.28793 m(r)-AEJ m(r)-AE 29 668.29849 m(r)-AE m(r)-AE 20 668.29849 m(r)-AE gm(r)-AE 30 674.29916 N/A EJ=AEJ (3-3) 31 681.28251 gm(r)-AE gm(r)-AE 32 692.30973 EJA=EJ (4-3) m(r)-AEJA 33 721.32504 m(r)-AEJA M/A 34 739.33561 A=m(r)-AEJ gm(r)-AE 35 745.33627 N/A AEJ=AEJA (3-3) 36 763.34684 EJA=AEJ (4-3) EJ=AEJA (3-3) 37 822.37272 N/A M(r)-AEJA 40 871.37786 gm(r)-AEJ gm(r)-AE 41 911.4204 m(r)-AEJA m	21	178 20314	m(r)-AF	m(r)-AE
22 481.2028 gm(r) gm(r) 23 545.25657 N/A J=AEJ (3-3) 24 552.23991 gm(r)-A gm(r)-A 25 563.26713 JA=EJ (4-3) J=EJA (3-3) 26 603.26205 N/A EJ=EJ (3-3) 27 634.30425 JA=AEJ (4-3) J=AEJA (3-3) 28 650.28793 m(r)-AEJ m(r)-AEJ 29 668.29849 m(r)-AE gm(r)-AE 29 668.29849 m(r)-AE gm(r)-AE 30 674.29916 N/A EJ=AEJ (3-3) 31 681.28251 gm(r)-AE gm(r)-AE m(r)-AEJA 34 739.33551 A=m(r)-AEJ(4-3) m(r)-AEJA AEJ=AEJ (3-3) 35 745.33627 N/A AEJ=AEJ (3-3) 36 753.34684 EJA=AEJ (4-3) AEJ=AEJA (3-3) 37 822.3772 N/A m(r)-AEJ-A 40 871.37786 gm(r)-AEJ m(r) N/A 41 911.4204 m(r)-AEJA=J (4-3) m(r)-AEJ=J (3-3) 42 924.4044 gm(r)-AEJA=J (4-3) m(r)-AEJ=AEJ (3-3) 44 951.41531 N/A m(r)-AEJ=AEJ (3-3) <	21	478.20314		
22 S42.25657 N/A j=E14 [3-3] 24 552.23991 gm(r)-A gm(r)-A 25 S63.26713 JA=EJ [4-3) J=EIA [3-3] 26 603.26205 N/A EJ=EJ [3-3] 27 634.30425 JA=AEJ [4-3] J=AEJA (3-3) 28 650.28793 m(r)-AEJ m(r)-AEJ 29 668.29849 m(r)-AE gm(r)-AE 30 674.29916 N/A EJ=AEJ (3-3) 31 681.28251 gm(r)-AE gm(r)-AE gm(r)-AE 32 692.30973 EJA=EJ (4-3) EJ=EJA (3-3) m(r)-AEJ 33 721.32504 m(r)-AEJA m(r)-AEJA 34 739.33561 A=m(r)-AEJ (4-3) m(r)-AEJ (3-3) 35 745.33627 N/A AEJ=AEJ (3-3) 36 743.34835 AEJ=AEJ (4-3) EJ=AEJA (3-3) 37 822.37272 N/A m(r)-AEJ (3-3) 38 834.38355 AEJA=AEJ (4-3) AEJ=AEJA (3-3) 39 853.3673 gm(r)-AEJ gm(r)-AEJ 40 871.3776 gm(r)-AEJ m(r)-AEJ 41 911.4204 m(r)-AEJ <th>22</th> <th>481.2028</th> <th>gm(r)</th> <th>gm(r)</th>	22	481.2028	gm(r)	gm(r)
24 552.23991 gm(r)-A gm(r)-A 25 553.26713 JA=L9 (4-3) J=EJA (3-3) 26 603.26205 N/A L]=L3 (3-3) 27 634.30425 JA=AEJ (4-3) m(r)-AEJ 28 650.28793 m(r)-AEJ m(r)-AE 29 668.29849 m(r)-AEJ m(r)-AE 30 674.29916 N/A EJ=AEJ (3-3) 31 681.28251 gm(r)-AE gm(r)-AE 32 692.3073 EJA=EJ (4-3) EJ=EJA (3-3) 33 71.32504 m(r)-AEJ BJ=EJA (3-3) 34 739.33561 A=m(r)-AEJ (4-3) m(r)-AEJA 35 745.34682 LJA=AEJ (4-3) EJ=AEJA (3-3) 36 763.34684 EJA=AEJ (4-3) M(r)-AEJA (3-3) 37 822.37272 N/A M/A M(r)-AEJA (3-3) 38 834.38395 AEJA=AEJ (4-3) M(r)-AEJA (3-3) M(r)-AEJA 39 853.3673 gm(r)-AEJ gm(r)-AEJ M(A 41 911.4204 m(r)-AEJA=J (4-3) m(r)-AEJA (3-3) M(r)-AEJA 42 924.4041 gm(r)-AEJA M(A M(A) M(r)-AEJA 43 942.41498 A=gm(r)-AEJA (4-3) m(r)-AEJA (3-3) M(r)-AEJ	23	545.25657	N/A	J=AEJ (3-3)
25 563.26713 JA-EU (4-3) J=EJA (3-3) 26 603.26205 N/A EJ=EJ (3-3) 27 634.30425 JA-AEL (4-3) J=AEJA (3-3) 28 650.28793 m(r)-AEJ m(r)-AE 29 668.29849 m(r)-AE N/A 30 674.29916 N/A EJ=AEJ (3-3) 31 681.28251 gm(r)-AE gm(r)-AE 32 692.30973 EJA=EJ (4-3) EJ=EJA (3-3) 33 71.32504 m(r)-AEJA N/A 34 739.33561 A=m(r)-AEJ m(r)-AEJ 35 745.336627 N/A ALJ=AEJ (3-3) 36 763.34684 EJA=AEJ (4-3) EJ=AEJA (3-3) 37 822.37272 N/A m(r)-AEJ (3-3) 38 83.43395 AEJA=AEJ (4-3) AEJ=AEJA (3-3) 39 853.3673 gm(r)-AEJ M(r) 40 871.37786 gm(r)-AEJ M(A 41 911.4204 m(r)-AEJA M(A m(r)-AEJA 42 924.40441 gm(r)-AEJA gm(r)-AEJA M(A <th>24</th> <th>552.23991</th> <th>gm(r)-A</th> <th>gm(r)-A</th>	24	552.23991	gm(r)-A	gm(r)-A
26 6032.6205 N/A E1-E1 (3-3) 27 634.30425 JA=AEJ (4-3) J=AEJA (3-3) 28 650.28793 m(r)-AEJ m(r)-AEJ 29 668.29849 m(r)-AEJ N/A 20 668.29849 m(r)-AE m(r)-AE 20 668.29819 m(r)-AE gm(r)-AE 21 681.28251 gm(r)-AE gm(r)-AE 22 692.30973 EJA=EJ (4-3) EJ=EJA (3-3) 33 721.32504 m(r)-AEJA m(r)-AEJA 34 739.33561 A=m(r)-AEJ (4-3) m(r)-AEJA 35 745.33627 N/A AEJ=AEJ (3-3) 36 763.34684 EJA=AEJ (4-3) EJ=AEJA (3-3) 37 822.37272 N/A m(r)-AEJ=J (3-3) 38 834.38395 AEJA=AEJ (4-3) AEJ=AEJA (3-3) 39 853.3673 gm(r)-AEJ gm(r)-AEJ 40 871.3778 gm(r)-AEJ 4(-3) m(r)-AEJ=J (3-3) 41 911.4204 m(r)-AEJA (4-3) gm(r)-AEJA 42 924.40441 gm(r)-AEJA (4-3) gm(r)-AEJA 43 924.41498 A=gm(r)-AEJ (4-3) m(r)-AEJ=J (3-3) 44 951.41531 N/A gm(r)-AEJA 45	25	563 26713	$[\Delta = F] (4-3)$	I=FIΔ (3-3)
26 603.26205 N/A EJEJ (3-3) 27 634.30425 JA=AEJ (4-3) JAEJA (3-3) 28 650.28793 m(r)-AEJ 29 668.29849 m(r)-AE N/A 30 674.29916 N/A EJ=AEJ (3-3) 31 681.28251 gm(r)-AE gm(r)-AE 32 692.30973 EJA=EJ (4-3) EJ=LIA (3-3) 33 721.32504 m(r)-AEJA N/A 34 739.33561 A=m(r)-AEJ (4-3) m(r)-AEJA 35 745.33627 N/A AEJ=AEJ (3-3) 36 763.34684 EJA=AEJ (4-3) EJ=AEJA (3-3) 37 822.37272 N/A m(r)-AEJ=J (3-3) 38 834.38395 EJA=AEJ (4-3) gm(r)-AEJ 39 853.3673 gm(r)-AEJ M/A 40 871.37786 gm(r)-AEJ M/A 41 911.4204 m(r)-AEJA M/A m(r)-AEJ=EJ (3-3) 42 924.40441 gm(r)-AEJA M/A m(r)-AEJ=AEJ (3-3) 43 942.41498 A=gm(r)-AEJ (4-3) m(r)-AEJ=EJ (3-3) </th <th>25</th> <th>505.20715</th> <th>JZ-EJ (+-3)</th> <th>3-E1X (3-3)</th>	25	505.20715	JZ-EJ (+-3)	3-E1X (3-3)
27 634.30425 JA=AEJ (4-3) J=AEJA (3-3) 28 650.28793 m(r)-AEJ m(r)-AEJ 29 668.29849 m(r)-AEJ N/A 30 674.29916 N/A EI=AEJ (3-3) 31 681.28251 gm(r)-AE gm(r)-AE gm(r)-AE 32 692.30973 EJA=EJ (4-3) EJ=EJA (3-3) N/A 33 721.32504 m(r)-AEJA m(r)-AEJA N/A 34 739.33561 A=m(r)-AEJ (4-3) m(r)-AEJA S 35 745.33627 N/A AEJ=AEJ (3-3) AEJ=AEJ (3-3) 36 763.34684 EJA=AEJ (4-3) H=AEJA (3-3) M(3-3) 37 822.37272 N/A m(r)-AEJ a M(A M(A 39 853.3673 gm(r)-AEJ gm(r)-AEJ M(A M(A 40 871.37786 gm(r)-AEJ (4-3) m(r)-AEJ=J (3-3) M(A M(P)-AEJA 41 911.4204 m(r)-AEJA=J (4-3) gm(r)-AEJ (3-3) M(A M(P)-AEJ=J (3-3) 42 924.40441 gm(r)-AEJA (4-3) m(r)-AEJ=J (3-3) M(R)-AEJ=J (3-3) M(R)-AEJ=J (3-3) 44 951.41531 N/A gm(r)-AEJ=AI (3-3) M(R)-AEJ=AI (3-3) M(R)-AE	26	603.26205	N/A	EJ=EJ (3-3)
28 650.28793 m(r)-AEJ m(r)-AEJ 29 668.29849 m(r)-AEJ M/A 30 674.29916 N/A EJ=AEJ (3-3) 31 681.28251 gm(r)-AE gm(r)-AE 32 692.30973 EJA=EJ (4-3) EJ=LAG (3-3) 32 721.32504 m(r)-AEJ m(r)-AEJ 33 721.32504 m(r)-AEJ m(r)-AEJ 34 739.33561 A=m(r)-AEJ (4-3) m(r)-AEJ (3-3) 35 745.33627 N/A ALEJ-AEJ (3-3) 36 763.34684 EJA=AEJ (4-3) EJ=AEJA (3-3) 37 822.37272 N/A M(r)-AEJ (3-3) 38 834.38395 AEJA=AEJ (4-3) gm(r)-AEJ 39 853.3673 gm(r)-AEJ gm(r)-AEJ 39 853.3673 gm(r)-AEJ gm(r)-AEJ 40 911.4204 m(r)-AEJ=J M/A 41 911.4204 m(r)-AEJ=AEJ M(A) 42 924.40441 gm(r)-AEJ gm(r)-AEJ 43 942.41498 A=gm(r)-AEJ M(A) gm(r)-AEJ </th <th>27</th> <th>634.30425</th> <th>JA=AEJ (4-3)</th> <th>J=AEJA (3-3)</th>	27	634.30425	JA=AEJ (4-3)	J=AEJA (3-3)
29 668.29849 m(r)-AEJ N/A 30 674.29916 N/A EJ=AEJ (3-3) 31 681.28251 gm(r)-AE gm(r)-AE 32 692.30973 EJA=EJ (4-3) EJ=EJA (3-3) 33 721.32504 m(r)-AEJA N/A 34 739.33561 A=m(r)-AEJ (4-3) m(r)-AEJA 35 745.33627 N/A AEJ=AEJ (3-3) 36 763.3684 EJA=AEJ (4-3) EJ=AEJA (3-3) 37 822.37272 N/A M(r)-AEJ=J (3-3) 38 834.38395 AEJA=AEJ (4-3) gm(r)-AEJ 39 853.3673 gm(r)-AEJ gm(r)-AEJ 40 871.37786 gm(r)-AEJ gm(r)-AEJ 41 911.4204 m(r)-AEJ=J (4-3) m(r)-AEJ=AEJ (3-3) 42 924.4041 gm(r)-AEJ (4-3) gm(r)-AEJ=AEJ (3-3) 43 942.41498 A=gm(r)-AEJ (4-3) gm(r)-AEJ=AEJ (3-3) 44 951.41531 N/A m(r)-AEJ=AEJ (3-3) 45 1022.4524 N/A gm(r)-AEJ=AEJ (3-3) 46 1025.4521 N/A	28	650.28793	m(r)-AEJ	m(r)-AEJ
Image: second	29	668 29849	m(r)-AFI	N/A
30 6/4.29910 N/A EDFAEJ (3-3) 31 681.28251 gm(r)-AE gm(r)-AE gm(r)-AE 32 692.30973 ELA=EJ (4-3) EJ=EJA (3-3) 33 721.32504 m(r)-AEJA N/A 34 739.33561 A=m(r)-AEJ (4-3) m(r)-AEJA M(r)-AEJA 35 745.33627 N/A AEJ=AEJ (3-3) 36 763.34684 EJA=AEJ (4-3) EJ=AEJA (3-3) 37 822.37272 N/A m(r)-AEJ=J (3-3) 38 83.438395 AEJA=AEJ (4-3) AEJ=AEJA (3-3) 39 853.3673 gm(r)-AEJ gm(r)-AEJ M(A 40 871.37786 gm(r)-AEJ m(r)-AEJ=A N/A 41 911.4204 m(r)-AEJA=J (4-3) m(r)-AEJ=A(3-3) m(r)-AEJ=A 42 924.40441 gm(r)-AEJA m(r)-AEJ=L3 (3-3) m(r)-AEJ=AEJ (3-3) 44 951.41531 N/A m(r)-AEJ=AEJ (3-3) 45 1022.4524 N/A m(r)-AEJ=AEJ (3-3) 46 1025.4521 N/A gm(r)-AEJ=AEJ (3-3) 47 1040.463 m(r)-AEJA=EJ (4-3) gm(r)-AEJ=AEJ (3-3) 48 1111.5001 m(20	C74 2001C	NI/A	
31 681.28251 gm(r)-AE gm(r)-AE 32 692.30973 EJA=EJ (4-3) EJ=EJA (3-3) 33 721.32504 m(r)-AEJA M(r)-AEJA 34 739.33561 A=m(r)-AEJ (4-3) m(r)-AEJA 35 745.33627 N/A AEJ=AEJ (3-3) 36 763.34684 EJA=AEJ (4-3) EJ=AEJA (3-3) 37 822.37272 N/A m(r)-AEJ 38 834.38395 AEJA=AEJ (4-3) AEJ=AEJA (3-3) 39 853.3673 gm(r)-AEJ gm(r)-AEJ 40 871.37786 gm(r)-AEJ m(r)-AEJA (3-3) 41 911.4204 m(r)-AEJA m(r)-AEJA 43 942.41498 A=gm(r)-AEJ (4-3) m(r)-AEJA 44 951.41531 N/A m(r)-AEJA 45 1022.4524 N/A m(r)-AEJ=A(3-3) 46 1025.4521 N/A gm(r)-AEJ=(3-3) 47 1040.463 m(r)-AEJA=EJ (4-3) m(r)-AEJ=AEJA (3-3) 48 1111.5001 m(r)-AEJA=AEJ (4-3) m(r)-AEJ=AEJA (3-3) 49 114.4998 gm(r)-AEJA=EJ (4-3) gm(r)-AEJ=AEJA (3-3) 49 114.4998 gm(r)-AEJA=EJ (4-3) gm(r)-AEJ=AEJA (3-3)	30	674.29916	N/A	EJ=AEJ (3-3)
32 692.30973 EJA=EJ (4-3) EJ=EJA (3-3) 33 721.32204 m(r)-AEJA m(r)-AEJA 34 739.33561 A=m(r)-AEJ (4-3) m(r)-AEJA 35 745.33627 N/A AEJ=AEJ (3-3) 36 763.34684 EJA=AEJ (4-3) EJ=AEJA (3-3) 37 822.37272 N/A m(r)-AEJ=(3-3) 38 834.38395 AEJA=AEJ (4-3) AEJ=AEJA (3-3) 39 853.3673 gm(r)-AEJ gm(r)-AEJA (3-3) 40 871.37786 gm(r)-AEJ m(r)-AEJ=JA (3-3) 41 911.4204 m(r)-AEJA=[4-3) m(r)-AEJ=A (3-3) 42 924.40441 gm(r)-AEJA M(A 43 942.41498 A=gm(r)-AEJ (4-3) gm(r)-AEJ=AEJ (3-3) 44 951.41531 N/A m(r)-AEJ=AEJ (3-3) 45 1022.4524 N/A gm(r)-AEJ=AEJ (3-3) 46 1025.4521 N/A gm(r)-AEJ=AEJ (3-3) 47 1040.463 m(r)-AEJA=EJ (4-3) m(r)-AEJ=AEJA (3-3) 48 1111.5001 m(r)-AEJA=AEJ (4-3) m(r)-AEJ=AEJA (3-3) 49 114.4998 gm(r)-AEJA=EJ (4-3) gm(r)-AEJ=AEJA (3-3) 49 114.4998 gm(r)-AEJA=EJ (4-3)	31	681.28251	gm(r)-AE	gm(r)-AE
33 721.32504 m(r)-AEJA N/A 34 739.33561 A=m(r)-AEJ (4-3) m(r)-AEJA 35 745.33627 N/A AEJ=AEJ (3-3) 36 763.34684 EJA=AEJ (4-3) EJ=AEJA (3-3) 37 822.37272 N/A m(r)-AEJ-I (3-3) 38 834.38395 AEJ=AEJ (4-3) AEJ=AEJA (3-3) 39 853.3673 gm(r)-AEJ gm(r)-AEJ 40 871.37786 gm(r)-AEJ m(r)-AEJ=J (3-3) 41 911.4204 m(r)-AEJA=I (4-3) m(r)-AEJ=IA (3-3) 42 924.40441 gm(r)-AEJ M(A 43 942.41498 A=gm(r)-AEJ (4-3) gm(r)-AEJ=IA (3-3) 44 951.41531 N/A m(r)-AEJ=EJ (3-3) 45 1022.4524 N/A gm(r)-AEJ=AEJ (3-3) 46 1025.4521 N/A gm(r)-AEJ=AEJ (3-3) 47 1040.463 m(r)-AEJ=AEJ (4-3) m(r)-AEJ=AEJ (3-3) 48 1111.5001 m(r)-AEJ=AEJ (4-3) gm(r)-AEJ=AEJ (3-3) 49 114.4998 gm(r)-AEJ=AEJ (4-3) gm(r)-AEJ=AEJ (3-3) <	32	692.30973	EJA=EJ (4-3)	EJ=EJA (3-3)
34 739.33561 A=m(r)-AEJ (4-3) m(r)-AEJA 35 745.33627 N/A AEI=AEJ (3-3) 36 763.34684 EJAAEJ (4-3) EJ=AEJA (3-3) 37 822.37272 N/A m(r)-AEJ=J (3-3) 38 834.38395 AEJA=AEJ (4-3) AEJ=AEJA (3-3) 39 853.3673 gm(r)-AEJ gm(r)-AEJ 40 871.37786 gm(r)-AEJ m(r)-AEJA 41 911.4204 m(r)-AEJA=J (4-3) m(r)-AEJA 42 924.40441 gm(r)-AEJA m(r)-AEJA 43 942.41498 A=gm(r)-AEJ (4-3) m(r)-AEJ=AEJ (3-3) 44 951.41531 N/A m(r)-AEJ=AEJ (3-3) 45 1022.4524 N/A m(r)-AEJ=AEJ (3-3) 46 1025.4521 N/A gm(r)-AEJ=AEJ (3-3) 47 1040.463 m(r)-AEJA=EJ (4-3) m(r)-AEJ=AEJA (3-3) 48 1111.5001 m(r)-AEJA=EJ (4-3) m(r)-AEJ=AEJA (3-3) 49 114.4998 gm(r)-AEJA=J (4-3) gm(r)-AEJA 41 1154.4947 N/A gm(r)-AEJ=AEJA (3-3) 51<	33	721.32504	m(r)-AEJA	N/A
35 745.33627 N/A AEJ=AEJ 35 745.33627 N/A AEJ=AEJ 36 763.34684 EJA=AEJ (4-3) 37 822.37272 N/A m(r)-AEJ=J 38 834.38395 AEJA=AEJ (4-3) 39 853.3673 gm(r)-AEJ gm(r)-AEJ 40 871.37786 gm(r)-AEJ N/A 41 911.4204 m(r)-AEJA=J (4-3) 42 924.40441 gm(r)-AEJ N/A 43 942.41498 A=gm(r)-AEJ M/A 44 951.41531 N/A m(r)-AEJ=EJ (3-3) 45 1022.4524 N/A m(r)-AEJ=AEJ (3-3) 46 1025.4521 N/A gm(r)-AEJ=AEJA (3-3) 47 1040.463 m(r)-AEJA=AEJ (4-3) gm(r)-AEJ=AEJA (3-3) 48 1111.5001 m(r)-AEJA=AEJ (4-3) gm(r)-AEJ=AEJA (3-3) 49 114.4998 gm(r)-AEJA=J (4-3) gm(r)-AEJ=AEJA (3-3) 50 1154.4947 N/A gm(r)-AE	34	739 33561	A - m(r) - A E I (1 - 3)	m(r)-AEIA
35 745.33027 N/A AEJ=AEJ (3-3) 36 763.34684 EJA=AEJ (4-3) EJ=AEJA (3-3) 37 822.37272 N/A m(r)-AEJ=J (3-3) 38 834.38395 AEJA=AEJ (4-3) AEJ=AEJA (3-3) 39 853.3673 gm(r)-AEJ gm(r)-AEJ 40 871.37786 gm(r)-AEJ m(r)-AEJ=A(3-3) 41 911.4204 m(r)-AEJA m(r)-AEJA (3-3) 42 924.40441 gm(r)-AEJA m(r)-AEJA 43 942.41498 A=gm(r)-AEJ (4-3) gm(r)-AEJA 44 951.41531 N/A m(r)-AEJ=A[3-3) 45 1022.4524 N/A gm(r)-AEJ (3-3) 46 1025.4521 N/A gm(r)-AEJ=AEJ (3-3) 47 1040.463 m(r)-AEJA=EJ (4-3) m(r)-AEJ=AEJA (3-3) 48 111.5001 m(r)-AEJA=AEJ (4-3) gm(r)-AEJ=AEJA (3-3) 49 114.4998 gm(r)-AEJA=I (4-3) gm(r)-AEJ=AEJA (3-3) 50 1154.4947 N/A gm(r)-AEJ=AEJA (3-3) 51 1225.5318 N/A gm(r)-AEJ=AEJA (3-3) 52 1243.5424 gm(r)-AEJA=EJ (4-3) gm(r)-AEJ=AEJA (3-3) 53 1299.5686		735.55501		
36 763.34684 EJA=AEJ (4-3) EJ=AEJA (3-3) 37 822.37272 N/A m(r)-AEJ=J (3-3) 38 834.38395 AEJA=AEJ (4-3) AEJ=AEJA (3-3) 39 853.3673 gm(r)-AEJ gm(r)-AEJ 40 871.37786 gm(r)-AEJ m(r)-AEJ 40 871.37786 gm(r)-AEJ m(r)-AEJ 41 911.4204 m(r)-AEJA=J (4-3) m(r)-AEJA (3-3) 42 924.40441 gm(r)-AEJ (4-3) gm(r)-AEJA 43 942.41498 A=gm(r)-AEJ (4-3) gm(r)-AEJA 44 951.41531 N/A m(r)-AEJ=LB (3-3) 45 1022.4524 N/A gm(r)-AEJ=AEJ (3-3) 46 1025.4521 N/A gm(r)-AEJ=JA (3-3) 47 1040.463 m(r)-AEJA=EJ (4-3) m(r)-AEJ=AEJA (3-3) 48 1111.5001 m(r)-AEJA=J (4-3) gm(r)-AEJ=AEJA (3-3) 49 114.4998 gm(r)-AEJ=EJ (4-3) gm(r)-AEJ=AEJA (3-3) 50 1154.4947 N/A gm(r)-AEJ=AEJA (3-3) 51 1225.5318 N/A gm(r)-AEJ=AEJA (3-3)	35	/45.3362/	N/A	AEJ=AEJ (3-3)
37 822.37272 N/A m(r)-AEJ=J (3-3) 38 834.38395 AEJA=AEJ (4-3) AEJ=AEJA (3-3) 39 853.3673 gm(r)-AEJ gm(r)-AEJ 40 871.37786 gm(r)-AEJ N/A 41 911.4204 m(r)-AEJA=J (4-3) m(r)-AEJ=JA (3-3) 42 924.4041 gm(r)-AEJA m(r)-AEJ 43 942.41498 A=gm(r)-AEJ (4-3) gm(r)-AEJ=G (3-3) 44 951.41531 N/A m(r)-AEJ=EJ (3-3) 45 1022.4524 N/A m(r)-AEJ=AEJ (3-3) 46 1025.4521 N/A gm(r)-AEJ=AEJ (3-3) 47 1040.463 m(r)-AEJA=AEJ (4-3) m(r)-AEJ=AEJA (3-3) 48 1111.5001 m(r)-AEJA=AEJ (4-3) gm(r)-AEJ=AEJA (3-3) 49 1114.4998 gm(r)-AEJA=AEJ (4-3) gm(r)-AEJ=AEJA (3-3) 50 1154.4947 N/A gm(r)-AEJ=AEJA (3-3) 51 1225.5318 N/A gm(r)-AEJ=AEJA (3-3) 52 1243.5424 gm(r)-AEJA=AEJ (4-3) gm(r)-AEJ=AEJA (3-3) 53 1299.5686 N/A m(r)-AEJ=m(r)-AEJA (3-3)	36	763.34684	EJA=AEJ (4-3)	EJ=AEJA (3-3)
38 834.38395 AEJA=AEJ (4-3) AEJ=AEJA (3-3) 39 853.3673 gm(r)-AEJ gm(r)-AEJ 40 871.37786 gm(r)-AEJ M(r)-AEJ=A (3-3) 41 911.4204 m(r)-AEJA=J (4-3) m(r)-AEJ=A (3-3) 42 924.4041 gm(r)-AEJA M/A 43 942.41498 Aegm(r)-AEJ (4-3) gm(r)-AEJ=A (3-3) 44 951.41531 N/A m(r)-AEJ=AEJ (3-3) 45 1022.4524 N/A m(r)-AEJ=AEJ (3-3) 46 1025.4521 N/A gm(r)-AEJ=AEJ (3-3) 47 1040.463 m(r)-AEJA=AEJ (4-3) m(r)-AEJ=AEJA (3-3) 48 1111.5001 m(r)-AEJA=AEJ (4-3) gm(r)-AEJ=AEJA (3-3) 49 1114.4998 gm(r)-AEJA=AEJ (4-3) gm(r)-AEJ=AEJA (3-3) 50 1154.4947 N/A gm(r)-AEJ=AEJA (3-3) 51 1225.5318 N/A gm(r)-AEJ=AEJA (3-3) 52 1243.5424 gm(r)-AEJA=AEJ (4-3) gm(r)-AEJ=MI(3-3) 53 1299.5686 N/A m(r)-AEJ=AEJA (3-3) 54 1314.5795 gm(r)-AEJA=AEJ (4-3) <td< th=""><th>37</th><th>822.37272</th><th>N/A</th><th>m(r)-AEJ=J (3-3)</th></td<>	37	822.37272	N/A	m(r)-AEJ=J (3-3)
39 853.3673 gm(r)-AEJ gm(r)-AEJ 40 871.37786 gm(r)-AEJ N/A 41 911.4204 m(r)-AEJA=J (4-3) m(r)-AEJ=JA (3-3) 42 924.40441 gm(r)-AEJA N/A 43 942.41498 A=gm(r)-AEJ (4-3) gm(r)-AEJ=EJ (3-3) 44 951.41531 N/A m(r)-AEJ=EJ (3-3) 45 1022.4524 N/A m(r)-AEJ=EJ (3-3) 46 1025.4521 N/A gm(r)-AEJ=JA (3-3) 47 1040.463 m(r)-AEJA=EJ (4-3) m(r)-AEJ=EJA (3-3) 48 1111.5001 m(r)-AEJA=AEJ (4-3) m(r)-AEJ=AEJA (3-3) 49 114.4998 gm(r)-AEJA=J (4-3) gm(r)-AEJ=AEJA (3-3) 50 1154.4947 N/A gm(r)-AEJ=EJ (3-3) 51 1225.5318 N/A gm(r)-AEJ=EJ (3-3) 52 1243.5424 gm(r)-AEJA=EJ (4-3) gm(r)-AEJ=EJA (3-3) 53 1299.5686 N/A m(r)-AEJ=EJA (3-3) 54 1314.5795 gm(r)-AEJA=AEJ (4-3) gm(r)-AEJ=AEJA (3-3) 55 1388.6163 m(r)-AEJA=AEJ (4-3) gm(r)-AEJ=AEJA (3-3) 56 1502.648 N/A m(r)-AEJ=gm(r)-AEJA (3-3)	38	834 38395	AFIA=AFI (4-3)	AFI=AFIA (3-3)
39 535.3078 gm(r)-AEJ gm(r)-AEJ 40 871.37786 gm(r)-AEJ N/A 41 911.4204 m(r)-AEJA=J (4-3) m(r)-AEJ=JA (3-3) 42 924.40441 gm(r)-AEJA gm(r)-AEJA 43 942.41498 A=gm(r)-AEJ (4-3) gm(r)-AEJA 44 951.41531 N/A m(r)-AEJ=EJ (3-3) 45 1022.4524 N/A gm(r)-AEJ=AEJ (3-3) 46 1025.4521 N/A gm(r)-AEJ=JA (3-3) 47 1040.463 m(r)-AEJA=EJ (4-3) m(r)-AEJ=EJA (3-3) 48 1111.5001 m(r)-AEJA=AEJ (4-3) m(r)-AEJ=AEJA (3-3) 49 114.4998 gm(r)-AEJA=J (4-3) gm(r)-AEJ=AEJ (3-3) 50 1154.4947 N/A gm(r)-AEJ=AEJ (3-3) 51 1225.5318 N/A gm(r)-AEJ=AEJ (3-3) 52 1243.5424 gm(r)-AEJA=EJ (4-3) gm(r)-AEJ=AEJ (3-3) gm(r)-AEJ=AEJ (3-3) 53 1299.5686 N/A m(r)-AEJ=AEJA (3-3) 54 1314.5795 gm(r)-AEJA=AEJ (4-3) gm(r)-AEJ=AEJA (3-3) 55 1388.6163 m(r)-AEJA=m(r)-AEJ (4-3) m(r)-AEJ=am(r)-AEJ (3-3) 56 1502.648 N/A m(r)-AEJ=gm(r	20	052 2672	$r(r) \land E$	am/r) AEI
40 871.37786 gm(r)-AEJ N/A 41 911.4204 m(r)-AEJA=J (4-3) m(r)-AEJ=JA (3-3) 42 924.40441 gm(r)-AEJA M/A 43 942.41498 A=gm(r)-AEJ (4-3) gm(r)-AEJA 44 951.41531 N/A m(r)-AEJ=EJ (3-3) 45 1022.4524 N/A m(r)-AEJ=AEJ (3-3) 46 1025.4521 N/A gm(r)-AEJ=AEJ (3-3) 47 1040.463 m(r)-AEJA=EJ (4-3) m(r)-AEJ=AEJA (3-3) 48 1111.5001 m(r)-AEJA=AEJ (4-3) m(r)-AEJ=AEJA (3-3) 49 114.4998 gm(r)-AEJA=J (4-3) gm(r)-AEJ=AEJA (3-3) 50 1154.4947 N/A gm(r)-AEJ=AEJ (3-3) 51 1225.5318 N/A gm(r)-AEJ=AEJ (3-3) 52 1243.5424 gm(r)-AEJA=EJ (4-3) gm(r)-AEJ=AEJ (3-3) 53 1299.5686 N/A m(r)-AEJ=EJA (3-3) 54 1314.5795 gm(r)-AEJA=AEJ (4-3) gm(r)-AEJ=AEJA (3-3) 55 1388.6163 m(r)-AEJA=gm(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJ (3-3) 56 1502.648 N/A m(r)-AEJ=gm(r)-AEJ (3-3) 57 1591.6956 m(r)-AEJA=gm(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJ (3	39	855.5075	giii(I)-AEJ	giii(i)-AEJ
41911.4204 m(r)-AEJA=J (4-3)m(r)-AEJ=JA (3-3)42924.40441 gm(r)-AEJAm(r)-AEJ=JA (3-3)43942.41498 A=gm(r)-AEJ (4-3)gm(r)-AEJA44951.41531N/Am(r)-AEJ=EJ (3-3)451022.4524N/Am(r)-AEJ=AEJ (3-3)461025.4521N/Agm(r)-AEJ=J (3-3)471040.463 m(r)-AEJA=EJ (4-3)m(r)-AEJ=EJA (3-3)481111.5001 m(r)-AEJA=AEJ (4-3)m(r)-AEJ=AEJA (3-3)491114.4998 gm(r)-AEJA=J (4-3)gm(r)-AEJ=AEJ (3-3)501154.4947N/Agm(r)-AEJ=AEJ (3-3)511225.5318N/Agm(r)-AEJ=AEJ (3-3)521243.5424 gm(r)-AEJA=EJ (4-3)gm(r)-AEJ=EJA (3-3)531299.5686N/Am(r)-AEJ=m(r)-AEJ (3-3)541314.5795 gm(r)-AEJA=AEJ (4-3)gm(r)-AEJ=m(r)-AEJA (3-3)551388.6163 m(r)-AEJA=gm(r)-AEJ (4-3)m(r)-AEJ=gm(r)-AEJA (3-3)561502.648N/Am(r)-AEJ=gm(r)-AEJA (3-3)571591.6956 m(r)-AEJA=gm(r)-AEJ (4-3)m(r)-AEJ=gm(r)-AEJ (3-3)581705.7273N/Agm(r)-AEJ=gm(r)-AEJA (3-3)591794.775 gm(r)-AEJA=gm(r)-AEJ (4-3)gm(r)-AEJ=gm(r)-AEJA (3-3)	40	871.37786	gm(r)-AEJ	N/A
42 924.40441 gm(r)-AEJA gm(r)-AEJA 43 942.41498 A=gm(r)-AEJ (4-3) gm(r)-AEJA 44 951.41531 N/A m(r)-AEJ=EJ (3-3) 45 1022.4524 N/A gm(r)-AEJ=AEJ (3-3) 46 1025.4521 N/A gm(r)-AEJ=AEJ (3-3) 47 1040.463 m(r)-AEJ=AEJ (4-3) m(r)-AEJ=AEJA (3-3) 48 1111.5001 m(r)-AEJA=AEJ (4-3) m(r)-AEJ=AEJA (3-3) 49 1114.4998 gm(r)-AEJA=J (4-3) gm(r)-AEJ=AEJ (3-3) 49 1114.4998 gm(r)-AEJA=J (4-3) gm(r)-AEJ=AEJ (3-3) 50 1154.4947 N/A gm(r)-AEJ=AEJ (3-3) 51 1225.5318 N/A gm(r)-AEJ=AEJ (3-3) 52 1243.5424 gm(r)-AEJA=EJ (4-3) gm(r)-AEJ=AEJ (3-3) 54 1314.5795 gm(r)-AEJA=AEJ (4-3) gm(r)-AEJ=AEJA (3-3) 55 1388.6163 m(r)-AEJ=AEJA (3-3) m(r)-AEJ=m(r)-AEJA (3-3) 56 1502.648 N/A m(r)-AEJ=gm(r)-AEJA (3-3) 57 1591.6956 m(r)-AEJ=gm(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJA (3-3) 58	41	911.4204	m(r)-AEJA=J (4-3)	m(r)-AEJ=JA (3-3)
43 942.41498 A=gm(r)-AEJ (4-3) gm(r)-AEJA 44 951.41531 N/A m(r)-AEJ=EJ (3-3) 45 1022.4524 N/A m(r)-AEJ=AEJ (3-3) 46 1025.4521 N/A gm(r)-AEJ=J (3-3) 47 1040.463 m(r)-AEJA=EJ (4-3) m(r)-AEJ=EJA (3-3) 48 1111.5001 m(r)-AEJA=AEJ (4-3) m(r)-AEJ=AEJA (3-3) 49 114.4998 gm(r)-AEJA=J (4-3) gm(r)-AEJ=AEJ (3-3) 50 1154.4947 N/A gm(r)-AEJ=EJ (3-3) 51 1225.5318 N/A gm(r)-AEJ=EJ (3-3) 52 1243.5424 gm(r)-AEJA=EJ (4-3) gm(r)-AEJ=EJA (3-3) 53 1299.5686 N/A m(r)-AEJ=m(r)-AEJ (3-3) 54 1314.5795 gm(r)-AEJA=EJ (4-3) gm(r)-AEJ=m(r)-AEJ (3-3) 55 1388.6163 m(r)-AEJA=(4-3) gm(r)-AEJ=m(r)-AEJA (3-3) 56 1502.648 N/A m(r)-AEJ=gm(r)-AEJA (3-3) 57 1591.6956 m(r)-AEJA=gm(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJA (3-3) 58 1705.7273 N/A gm(r)-AEJ=gm(r)-AEJ (3-3) 59 1794.775 gm(r)-AEJA=gm(r)-AEJ (4-3) gm(r)-AEJ=gm(r)-AEJ (3-3)	42	924.40441	gm(r)-AEJA	N/A
44 951.41531 N/A m(r)-AEJ=EJ (3-3) 45 1022.4524 N/A m(r)-AEJ=AEJ (3-3) 46 1025.4521 N/A gm(r)-AEJ=Id (3-3) 47 1040.463 m(r)-AEJA=EJ (4-3) m(r)-AEJ=EJA (3-3) 48 1111.5001 m(r)-AEJA=AEJ (4-3) m(r)-AEJ=AEJA (3-3) 49 114.4998 gm(r)-AEJA=J (4-3) gm(r)-AEJ=AEJA (3-3) 50 1154.4947 N/A gm(r)-AEJ=AEJ (3-3) 51 1225.5318 N/A gm(r)-AEJ=AEJ (3-3) 52 1243.5424 gm(r)-AEJA=EJ (4-3) gm(r)-AEJ=HA (3-3) 53 1299.5686 N/A m(r)-AEJ=m(r)-AEJ (3-3) 54 1314.5795 gm(r)-AEJA=AEJ (4-3) gm(r)-AEJ=AEJA (3-3) 55 1388.6163 m(r)-AEJA=m(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJA (3-3) 56 1502.648 N/A m(r)-AEJ=gm(r)-AEJA (3-3) 57 1591.6956 m(r)-AEJA=gm(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJA (3-3) 58 1705.7273 N/A gm(r)-AEJ=gm(r)-AEJA (3-3) 59 1794.775 gm(r)-AEJA=gm(r)-AEJ (4-3) gm(r)-AEJ=gm(r)-AEJA (3-3)	43	942 41498	$\Delta = gm(r) - \Delta F I (4-3)$	gm(r)-AFIA
44 951.41531 N/A m(r)-AEJ=EJ (3-3) 45 1022.4524 N/A m(r)-AEJ=AEJ (3-3) 46 1025.4521 N/A gm(r)-AEJ=J (3-3) 47 1040.463 m(r)-AEJA=EJ (4-3) m(r)-AEJ=EJA (3-3) 48 1111.5001 m(r)-AEJA=AEJ (4-3) m(r)-AEJ=AEJA (3-3) 49 1114.4998 gm(r)-AEJA=J (4-3) gm(r)-AEJ=JA (3-3) 50 1154.4947 N/A gm(r)-AEJ=LJ (3-3) 51 1225.5318 N/A gm(r)-AEJ=AEJ (3-3) 52 1243.5424 gm(r)-AEJA=EJ (4-3) gm(r)-AEJ=m(r)-AEJ (3-3) 53 1299.5686 N/A m(r)-AEJ=m(r)-AEJ (3-3) 54 1314.5795 gm(r)-AEJA=AEJ (4-3) gm(r)-AEJ=m(r)-AEJ (3-3) 55 1388.6163 m(r)-AEJA=gm(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJ (3-3) 56 1502.648 N/A m(r)-AEJ=gm(r)-AEJ (3-3) 57 1591.6956 m(r)-AEJA=gm(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJ (3-3) 58 1705.7273 N/A gm(r)-AEJ=gm(r)-AEJ (3-3) 59 1794.775 gm(r)-AEJA=gm(r)-AEJ (4-3) gm(r)-AEJ=gm(r)-AEJ (3-3)	45	054 44524		
45 1022.4524 N/A m(r)-AEJ=AEJ (3-3) 46 1025.4521 N/A gm(r)-AEJ=J (3-3) 47 1040.463 m(r)-AEJA=EJ (4-3) m(r)-AEJ=EJA (3-3) 48 1111.5001 m(r)-AEJA=AEJ (4-3) m(r)-AEJ=AEJA (3-3) 49 1114.4998 gm(r)-AEJA=J (4-3) gm(r)-AEJ=JA (3-3) 50 1154.4947 N/A gm(r)-AEJ=EJ (3-3) 51 1225.5318 N/A gm(r)-AEJ=EJ (3-3) 52 1243.5424 gm(r)-AEJA=EJ (4-3) gm(r)-AEJ=EJA (3-3) 53 1299.5686 N/A m(r)-AEJ=m(r)-AEJ (3-3) 54 1314.5795 gm(r)-AEJA=AEJ (4-3) gm(r)-AEJ=m(r)-AEJ (3-3) 55 1388.6163 m(r)-AEJA=m(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJ (3-3) 56 1502.648 N/A m(r)-AEJ=gm(r)-AEJ (3-3) 57 1591.6956 m(r)-AEJA=gm(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJ (3-3) 58 1705.7273 N/A gm(r)-AEJ=gm(r)-AEJ (3-3) 59 1794.775 gm(r)-AEJA=gm(r)-AEJ (4-3) gm(r)-AEJ=gm(r)-AEJ (3-3)	44	951.41531	N/A	III(I)-AEJ=EJ (3-3)
46 1025.4521 N/A gm(r)-AEJ=J (3-3) 47 1040.463 m(r)-AEJA=EJ (4-3) m(r)-AEJ=EJA (3-3) 48 1111.5001 m(r)-AEJA=AEJ (4-3) m(r)-AEJ=AEJA (3-3) 49 1114.4998 gm(r)-AEJA=J (4-3) gm(r)-AEJ=JA (3-3) 50 1154.4947 N/A gm(r)-AEJ=EJ (3-3) 51 1225.5318 N/A gm(r)-AEJ=AEJ (3-3) 52 1243.5424 gm(r)-AEJA=EJ (4-3) gm(r)-AEJ=EJA (3-3) 53 1299.5686 N/A m(r)-AEJ=m(r)-AEJ (3-3) 54 1314.5795 gm(r)-AEJA=AEJ (4-3) gm(r)-AEJ=AEJA (3-3) 55 1388.6163 m(r)-AEJA=m(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJA (3-3) 56 1502.648 N/A m(r)-AEJ=gm(r)-AEJ (3-3) 57 1591.6956 m(r)-AEJA=gm(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJ (3-3) 58 1705.7273 N/A gm(r)-AEJ=gm(r)-AEJ (3-3) 59 1794.775 gm(r)-AEJA=gm(r)-AEJ (4-3) gm(r)-AEJ=gm(r)-AEJ (3-3)	45	1022.4524	N/A	m(r)-AEJ=AEJ (3-3)
471040.463 m(r)-AEJA=EJ (4-3)m(r)-AEJ=EJA (3-3)481111.5001 m(r)-AEJA=AEJ (4-3)m(r)-AEJ=AEJA (3-3)491114.4998 gm(r)-AEJA=J (4-3)gm(r)-AEJ=JA (3-3)501154.4947N/Agm(r)-AEJ=EJ (3-3)511225.5318N/Agm(r)-AEJ=AEJ (3-3)521243.5424 gm(r)-AEJA=EJ (4-3)gm(r)-AEJ=m(r)-AEJ (3-3)531299.5686N/Am(r)-AEJ=m(r)-AEJ (3-3)541314.5795 gm(r)-AEJA=AEJ (4-3)gm(r)-AEJ=m(r)-AEJ (3-3)551388.6163 m(r)-AEJA=m(r)-AEJ (4-3)m(r)-AEJ=gm(r)-AEJA (3-3)561502.648N/Am(r)-AEJ=gm(r)-AEJ (3-3)571591.6956 m(r)-AEJA=gm(r)-AEJ (4-3)m(r)-AEJ=gm(r)-AEJ (3-3)581705.7273N/Agm(r)-AEJ=gm(r)-AEJ (3-3)591794.775 gm(r)-AEJA=gm(r)-AEJ (4-3)gm(r)-AEJ=gm(r)-AEJ (3-3)	46	1025.4521	N/A	gm(r)-AEJ=J (3-3)
48 1111.5001 m(r)-AEJA=AEJ (4-3) m(r)-AEJ=AEJA (3-3) 49 1114.4998 gm(r)-AEJA=J (4-3) gm(r)-AEJ=JA (3-3) 50 1154.4947 N/A gm(r)-AEJ=EJ (3-3) 51 1225.5318 N/A gm(r)-AEJ=AEJ (3-3) 52 1243.5424 gm(r)-AEJA=EJ (4-3) gm(r)-AEJ=EJA (3-3) 53 1299.5686 N/A m(r)-AEJ=m(r)-AEJ (3-3) 54 1314.5795 gm(r)-AEJA=AEJ (4-3) gm(r)-AEJ=m(r)-AEJ (3-3) 55 1388.6163 m(r)-AEJA=m(r)-AEJ (4-3) gm(r)-AEJ=gm(r)-AEJ (3-3) 56 1502.648 N/A m(r)-AEJ=gm(r)-AEJ (3-3) 57 1591.6956 m(r)-AEJA=gm(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJ (3-3) 58 1705.7273 N/A gm(r)-AEJ=gm(r)-AEJ (3-3) 59 1794.775 gm(r)-AEJA=gm(r)-AEJ (4-3) gm(r)-AEJ=gm(r)-AEJ (3-3)	47	1040.463	m(r)-AEJA=EJ (4-3)	m(r)-AEJ=EJA (3-3)
48 1111.5001 m(r)-AEJA=AEJ (4-3) m(r)-AEJ=AEJA (3-3) 49 1114.4998 gm(r)-AEJA=J (4-3) gm(r)-AEJ=JA (3-3) 50 1154.4947 N/A gm(r)-AEJ=EJ (3-3) 51 1225.5318 N/A gm(r)-AEJ=AEJ (3-3) 52 1243.5424 gm(r)-AEJA=EJ (4-3) gm(r)-AEJ=EJA (3-3) 53 1299.5686 N/A m(r)-AEJ=m(r)-AEJ (3-3) 54 1314.5795 gm(r)-AEJA=AEJ (4-3) gm(r)-AEJ=m(r)-AEJA (3-3) 55 1388.6163 m(r)-AEJA=m(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJA (3-3) 56 1502.648 N/A m(r)-AEJ=gm(r)-AEJ (3-3) 57 1591.6956 m(r)-AEJA=gm(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJ (3-3) 58 1705.7273 N/A gm(r)-AEJ=gm(r)-AEJ (3-3) 59 1794.775 gm(r)-AEJA=gm(r)-AEJ (4-3) gm(r)-AEJ=gm(r)-AEJA (3-3)	40	1111 5001	$m(r) \wedge F(A = A F(A = 2))$	$m(r) \wedge E = \Delta E \Delta (2, 2)$
49 1114.4998 gm(r)-AEJA=J (4-3) gm(r)-AEJ=JA (3-3) 50 1154.4947 N/A gm(r)-AEJ=EJ (3-3) 51 1225.5318 N/A gm(r)-AEJ=AEJ (3-3) 52 1243.5424 gm(r)-AEJA=EJ (4-3) gm(r)-AEJ=EJA (3-3) 53 1299.5686 N/A m(r)-AEJ=m(r)-AEJ (3-3) 54 1314.5795 gm(r)-AEJA=AEJ (4-3) gm(r)-AEJ=m(r)-AEJ (3-3) 55 1388.6163 m(r)-AEJA=m(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJA (3-3) 56 1502.648 N/A m(r)-AEJ=gm(r)-AEJ (3-3) 57 1591.6956 m(r)-AEJA=gm(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJ (3-3) 58 1705.7273 N/A gm(r)-AEJ=gm(r)-AEJ (3-3) 59 1794.775 gm(r)-AEJA=gm(r)-AEJ (4-3) gm(r)-AEJ=gm(r)-AEJ (3-3)	40	1111.5001		
50 1154.4947 N/A gm(r)-AEJ=EJ (3-3) 51 1225.5318 N/A gm(r)-AEJ=AEJ (3-3) 52 1243.5424 gm(r)-AEJA=EJ (4-3) gm(r)-AEJ=EJA (3-3) 53 1299.5686 N/A m(r)-AEJ=m(r)-AEJ (3-3) 54 1314.5795 gm(r)-AEJA=AEJ (4-3) gm(r)-AEJ=AEJA (3-3) 55 1388.6163 m(r)-AEJA=m(r)-AEJ (4-3) m(r)-AEJ=m(r)-AEJA (3-3) 56 1502.648 N/A m(r)-AEJ=gm(r)-AEJ (3-3) 57 1591.6956 m(r)-AEJA=gm(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJ (3-3) 58 1705.7273 N/A gm(r)-AEJ=gm(r)-AEJ (3-3) 59 1794.775 gm(r)-AEJA=gm(r)-AEJ (4-3) gm(r)-AEJ=gm(r)-AEJ (3-3)	49	1114.4998	gm(r)-AEJA=J (4-3)	gm(r)-AEJ=JA (3-3)
51 1225.5318 N/A gm(r)-AEJ=AEJ (3-3) 52 1243.5424 gm(r)-AEJA=EJ (4-3) gm(r)-AEJ=EJA (3-3) 53 1299.5686 N/A m(r)-AEJ=m(r)-AEJ (3-3) 54 1314.5795 gm(r)-AEJA=AEJ (4-3) gm(r)-AEJ=AEJA (3-3) 55 1388.6163 m(r)-AEJA=m(r)-AEJ (4-3) m(r)-AEJ=m(r)-AEJA (3-3) 56 1502.648 N/A m(r)-AEJ=gm(r)-AEJ (3-3) 57 1591.6956 m(r)-AEJA=gm(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJ (3-3) 58 1705.7273 N/A gm(r)-AEJ=gm(r)-AEJ (3-3) 59 1794.775 gm(r)-AEJA=gm(r)-AEJ (4-3) gm(r)-AEJ=gm(r)-AEJA (3-3)	50	1154.4947	N/A	gm(r)-AEJ=EJ (3-3)
52 1243.5424 gm(r)-AEJA=EJ (4-3) gm(r)-AEJ=EJA (3-3) 53 1299.5686 N/A m(r)-AEJ=m(r)-AEJ (3-3) 54 1314.5795 gm(r)-AEJA=AEJ (4-3) gm(r)-AEJ=AEJA (3-3) 55 1388.6163 m(r)-AEJA=m(r)-AEJ (4-3) m(r)-AEJ=m(r)-AEJA (3-3) 56 1502.648 N/A m(r)-AEJ=gm(r)-AEJ (3-3) 57 1591.6956 m(r)-AEJA=gm(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJ (3-3) 58 1705.7273 N/A gm(r)-AEJ=gm(r)-AEJ (3-3) 59 1794.775 gm(r)-AEJA=gm(r)-AEJ (4-3) gm(r)-AEJ=gm(r)-AEJA (3-3)	51	1225.5318	N/A	gm(r)-AEJ=AEJ (3-3)
52 129.544 Bill(1/AEJ-EJA (5-3)) 53 1299.5686 N/A m(r)-AEJ=m(r)-AEJ (3-3) 54 1314.5795 gm(r)-AEJA=AEJ (4-3) gm(r)-AEJ=AEJA (3-3) 55 1388.6163 m(r)-AEJ(4-3) m(r)-AEJ=m(r)-AEJA (3-3) 56 1502.648 N/A m(r)-AEJ=gm(r)-AEJ (3-3) 57 1591.6956 m(r)-AEJ=gm(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJ (3-3) 58 1705.7273 N/A gm(r)-AEJ=gm(r)-AEJ (3-3) 59 1794.775 gm(r)-AEJ=gm(r)-AEJ (4-3) gm(r)-AEJ=gm(r)-AEJA (3-3)	E.2	1242 5124	$gm(r)-\Delta F \Delta = F (1-3)$	$gm(r)-\Delta FI-FI\Delta (3-3)$
53 1299.5686 N/A m(r)-AEJ=m(r)-AEJ (3-3) 54 1314.5795 gm(r)-AEJA=AEJ (4-3) gm(r)-AEJ=AEJA (3-3) 55 1388.6163 m(r)-AEJA=m(r)-AEJ (4-3) m(r)-AEJ=m(r)-AEJA (3-3) 56 1502.648 N/A m(r)-AEJ=gm(r)-AEJ (3-3) 57 1591.6956 m(r)-AEJA=gm(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJ (3-3) 58 1705.7273 N/A gm(r)-AEJ=gm(r)-AEJ (3-3) 59 1794.775 gm(r)-AEJA=gm(r)-AEJ (4-3) gm(r)-AEJ=gm(r)-AEJA (3-3)	52	1200 5000	5 · · · (·) - ALJA-LJ (+-J)	
54 1314.5795 gm(r)-AEJA=AEJ (4-3) gm(r)-AEJ=AEJA (3-3) 55 1388.6163 m(r)-AEJA=m(r)-AEJ (4-3) m(r)-AEJ=m(r)-AEJA (3-3) 56 1502.648 N/A m(r)-AEJ=gm(r)-AEJ (3-3) 57 1591.6956 m(r)-AEJA=gm(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJA (3-3) 58 1705.7273 N/A gm(r)-AEJ=gm(r)-AEJ (3-3) 59 1794.775 gm(r)-AEJA=gm(r)-AEJ (4-3) gm(r)-AEJ=gm(r)-AEJA (3-3)	53	1299.5686	N/A	m(r)-AEJ=m(r)-AEJ (3-3)
55 1388.6163 m(r)-AEJA=m(r)-AEJ (4-3) m(r)-AEJ=m(r)-AEJA (3-3) 56 1502.648 N/A m(r)-AEJ=gm(r)-AEJ (3-3) 57 1591.6956 m(r)-AEJA=gm(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJA (3-3) 58 1705.7273 N/A gm(r)-AEJ=gm(r)-AEJ (3-3) 59 1794.775 gm(r)-AEJA=gm(r)-AEJ (4-3) gm(r)-AEJ=gm(r)-AEJA (3-3)	54	1314.5795	gm(r)-AEJA=AEJ (4-3)	gm(r)-AEJ=AEJA (3-3)
56 1502.648 N/A m(r)-AEJ=gm(r)-AEJ (3-3) 57 1591.6956 m(r)-AEJA=gm(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJA (3-3) 58 1705.7273 N/A gm(r)-AEJ=gm(r)-AEJ (3-3) 59 1794.775 gm(r)-AEJA=gm(r)-AEJ (4-3) gm(r)-AEJ=gm(r)-AEJA (3-3)	55	1388.6163	m(r)-AEJA=m(r)-AEJ (4-3)	m(r)-AEJ=m(r)-AEJA (3-3)
57 1591.6956 m(r)-AEJa=gm(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJ (3-3) 58 1705.7273 N/A gm(r)-AEJ =gm(r)-AEJ (3-3) 59 1794.775 gm(r)-AEJ =gm(r)-AEJ (3-3)	56	1502.649	N/A	$m(r)-\Delta EI=\sigma m(r)-\Delta EI(3-3)$
57 1591.0956 m(r)-AEJA=gm(r)-AEJ (4-3) m(r)-AEJ=gm(r)-AEJA (3-3) 58 1705.7273 N/A gm(r)-AEJ=gm(r)-AEJ (3-3) 59 1794.775 gm(r)-AEJA=gm(r)-AEJ (4-3) gm(r)-AEJ=gm(r)-AEJA (3-3)		1501 0050		$m(r) \land \Box = gm(r) \land \Box = g(r)$
58 1705.7273 N/A gm(r)-AEJ=gm(r)-AEJ (3-3) 59 1794.775 gm(r)-AEJA=gm(r)-AEJ (4-3) gm(r)-AEJ=gm(r)-AEJA (3-3)	5/	1231.0320	m(r)-AEJA=gm(r)-AEJ (4-3)	m(r)-AEJ=gm(r)-AEJA (3-3)
59 1794.775 gm(r)-AEJA=gm(r)-AEJ (4-3) gm(r)-AEJ=gm(r)-AEJA (3-3)	58	1705.7273	N/A	gm(r)-AEJ=gm(r)-AEJ (3-3)
	59	1794.775	gm(r)-AEJA=gm(r)-AEJ (4-3)	gm(r)-AEJ=gm(r)-AEJA (3-3)

* N/A, Not applicable (does not exist) ** C-terminal residue (+H₂O)

Amongst the 47 trimers identified (Table S6-10), 26 matched only a single structure, Given the complexity of each MS/MS analysis and the prior validation of dimers, we chose to simply assign ambiguous trimer matches to the structure built from the most abundant dimer linked to the most abundant donor. For example, structures like gm-AEJA=gm-AEJA=gm-AEJA (4-3, 4-3) were chosen over structures like gm-AEJA-gm-AEJ (4-3, 3-3). The final list of 121 muropeptides including monomers, dimers, porins, and their modified counterparts is described in Table S6-12.

Structure	Crosslink	RT (min)	TheoMw	Intensity	Abundance	Acceptor	Donor
1 gm-AEJA=gm-AEJA 2	3-4	17.52	1864.8048	9.28E+08	12.41%		gm-AEJAX
2 gm-AEJA=gm-AEJ 2	3-3	15.90	1793.7676	8.75E+08	11.71%	giii-ALJA	gm-AEJX
3 gm-AEJG=gm-AEJ 2	3-3	27.88	1779.7520	6.16E+08	8.23%	gm-AEJG	gm-AEJX
4 gm-AEJ=gm-AEJ 2	3-3	14.01	1722.7305	4.88E+08	6.52%	gm-AEJ	gm-AEJX
5 gm-AEJG=gm-AEJA 2	3-4	14.82	1850.7891	2.96E+08	3.96%	gm-AEJG	gm-AEJAX
6 gm-AEJF=gm-AEJ 2	3-3	23.53	1869.7989	9.12E+07	1.22%	gm-AEJF	gm-AEJX
7 gm-AEJN=gm-AEJ 2	3-3	12.96	1836.7734	7.00E+07	0.94%	gm-AEJN	gm-AEJX
gm-AEJAG=gm-AEJA 2	3_1	15.05	1021 8263	5 05F±07	0.80%	gm-AEJAG	gm-AEJAX
gm-AEJQ=gm-AEJA 2	5-4	15.05	1921.8205	J.JJL107	0.80%	gm-AEJQ	gm-AEJAX
9 gm-AEJS=gm-AEJ 2	3-3	13.21	1809.7625	4.50E+07	0.60%	gm-AEJS	gm-AEJX
10 gm-AEJAI=gm-AEJA 2	3-4	24.15	1977.8888	3.42E+07	0.46%	gm-AEJAI	gm-AEJAX
11 gm-AEJAI=gm-AEJ 2	3-3	22.72	1906.8518	3.31E+07	0.44%	gm-AEJAI	gm-AEJX
12 gm-AEJAF=gm-AEJ 2	3-3	24.36	1940.8360	3.19E+07	0.43%	gm-AEJAF	gm-AEJX
13 gm-AEJN=gm-AEJA 2	3-4	10.54	1907.8105	2.19E+07	0.29%	gm-AEJN	gm-AEJAX
14 gm-AEJS=gm-AEJA 2	3-4	14.08	1880.7996	1.86E+07	0.25%	gm-AEJS	gm-AEJAX
15 gm-AEJY=gm-AEJ 2	3-3	19.61	1885.7938	1.79E+07	0.24%	gm-AEJY	gm-AEJX
16 gm-AEJAD=gm-AEJA 2	3-4	16.08	1979.8317	1.60E+07	0.21%	gm-AEJAD	gm-AEJAX
17 gm-AEJK=gm-AEJ 2	3-3	13.07	1850.8254	1.54E+07	0.21%	gm-AEJK	gm-AEJX
18 gm-AEJM=gm-AEJ 2	3-3	18.83	1853.7710	1.41E+07	0.19%	gm-AEJM	gm-AEJX
19 gm-AEJAA=gm-AEJA 2	3-4	16.33	1935.8419	1.34E+07	0.18%	gm-AEJAA	gm-AEJAX
20 gm-AEJI=gm-AEJ 2	3-3	22.38	1835.8145	1.31E+07	0.18%	gm-AEJI	gm-AEJX
21 gm-AEJH=gm-AEJ 2	3-3	12.63	1859.7894	1.27E+07	0.17%	gm-AEJH	gm-AEJX
22 gm-AEJH=gm-AEJA 2	3-4	13.27	1930.8265	8.00E+06	0.11%	gm-AEJH	gm-AEJAX
23 gm-AEJT=gm-AEJ 2	3-3	13.72	1823.7782	7.76E+06	0.10%	gm-AEJT	gm-AEJX
24 gm-AEJY=gm-AEJA 2	3-4	16.77	1956.8309	6.87E+06	0.09%	gm-AEJY	gm-AEJAX
25 gm-AEJAW=gm-AEJ 2	3-3	26.49	1979.8469	6.11E+06	0.08%	gm-AEJAW	gm-AEJX
26 gm-AEJAM=gm-AEJA 2	3-4	20.43	1995.8452	5.82E+06	0.08%	gm-AEJAM	gm-AEJAX
27 gm-AEJAM=gm-AEJ 2	3-3	19.41	1924.8081	5.68E+06	0.08%	gm-AEJAM	gm-AEJX
28 gm-AEJK=gm-AEJA 2	3-4	10.28	1921.8625	5.35E+06	0.07%	gm-AEJK	gm-AEJAX
29 gm-AEJD=gm-AEJA 2	3-4	14.49	1908.7946	5.32E+06	0.07%	gm-AEJD	gm-AEJAX
30 gm-AEJAW=gm-AEJA 2	3-4	27.00	2050.8840	5.24E+06	0.07%	gm-AEJAW	gm-AEJAX
31 gm-AEJV=gm-AEJ 2	3-3	19.14	1821.7989	4.19E+06	0.06%	gm-AEJV	gm-AEJX
32 gm-AEJF=gm-AEJA 2	3-4	25.26	1940.8360	3.89E+06	0.05%	gm-AEJF	gm-AEJAX
33 gm-AEJD=gm-AEJ 2	3-3	13.69	1837.7574	3.85E+06	0.05%	gm-AEJD	gm-AEJX
34 gm-AEJT=gm-AEJA 2	3-4	14.67	1894.8153	3.71E+06	0.05%	gm-AEJT	gm-AEJAX
35 gm-AEJAR=gm-AEJ 2	3-3	11.09	1949.8687	3.46E+06	0.05%	gm-AEJAR	gm-AEJX
36 gm-AEJAF=gm-AEJA 2	3-4	25.89	2011.8731	2.95E+06	0.04%	gm-AEJAF	gm-AEJAX
37 gm-AEJW=gm-AEJ 2	3-3	21.70	1908.8098	2.86E+06	0.04%	gm-AEJW	gm-AEJX
38 gm-AEJAH=gm-AEJA 2	3-4	11.02	2001.8636	2.26E+06	0.03%	gm-AEJAH	gm-AEJAX
39 gm-AEJV=gm-AEJA 2	3-4	19.89	1892.8360	2.08E+06	0.03%	gm-AEJV	gm-AEJAX
40 gm-AEJI=gm-AEJA 2	3-4	23.45	1906.8518	2.02E+06	0.03%	gm-AEJI	gm-AEJAX
41 gm-AEJM=gm-AEJA 2	3-4	19.76	1924.8081	1.53E+06	0.02%	gm-AEJM	gm-AEJAX
42 gm-AEJAE=gm-AEJA 2	3-4	15.81	1993.8473	1.14E+06	0.02%	gm-AEJAE	gm-AEJAX
43 gm-AEJ=gm-AEJA 2	3-4	14.30	1793.7676	1.09E+06	0.01%	gm-AEJ	gm-AEJAX
44 gm-AEJAK=gm-AEJA 2	3-4	13.82	1992.8997	1.01E+06	0.01%	gm-AEJAK	gm-AEJAX
45 gm-AEJAQ=gm-AEJA 2	3-4	16.36	1992.8633	9.37E+05	0.01%	gm-AEJAQ	gm-AEJAX
46 gm-AEJAA=gm-AEJ 2	3-3	17.60	1864.8048	6.93E+05	0.01%	gm-AEJAA	gm-AEJX

Table S6- 12. List of validated dimers

	Structure	Theo mass	Aver	age	% mono-	Cumulated	Average		Intensity		
QVI	Structure	(Da)	Intensity	%	di-, trimers	abundance	RT (min)	TY1	TY2	TY3	
1	gm-AEJA 1	941.4078	1.07E+09	14.361%	47.90%	47.90%	9.50 ± 0.03	9.1E+08	1.1E+09	1.2E+09	
2	gm-AEJG 1	927.3922	2.43E+08	3.249%	10.84%	58.74%	6.85 ± 0.04	2.5E+08	2.3E+08	2.5E+08	
3	gm-AEJ 1	870.3707	1.64E+08	2.199%	7.33%	66.07%	5.48 ± 0.03	1.7E+08	1.3E+08	1.9E+08	
4	gm-AEJAI 1	1054.4919	1.54E+08	2.054%	6.85%	72.92%	21.40 ± 0.00	1.6E+08	1.5E+08	1.5E+08	
5	gm-AEJF 1	1017.4391	1.48E+08	1.976%	6.59%	79.51%	21.62 ± 0.00	1.4E+08	1.5E+08	1.5E+08	
6	gm-AEJAA 1	1012.4450	6.55E+07	0.875%	2.92%	82.43%	10.86 ± 0.02	5.0E+07	7.3E+07	7.4E+07	
7	gm-AEJAF 1	1088.4762	5.00E+07	0.669%	2.23%	84.67%	15.38 ± 0.00	6.0E+07	5.2E+07	3.9E+07	
8	gm-AEJ[AG/Q] 1	998.4292	4.61E+07	0.616%	2.06%	86.72%	8.80 ± 0.05	4.3E+07	4.3E+07	5.3E+07	
9	gm-AEJK 1	998.4656	4.60E+07	0.615%	2.05%	88.77%	7.71 ± 0.06	6.2E+07	4.0E+07	3.6E+07	
10	gm-AEJY 1	1033.4340	3.38E+07	0.452%	1.51%	90.28%	16.84 ± 0.01	3.3E+07	3.6E+07	3.2E+07	
11	gm-AEJAM 1	1072.4483	2.87E+07	0.384%	1.28%	91.56%	16.56 ± 0.00	2.8E+07	3.1E+07	2.7E+07	
12	gm-AEJN 1	984.4136	2.70E+07	0.361%	1.20%	92.77%	5.03 ± 0.02	2.5E+07	2.4E+07	3.1E+07	
13	gm-AEJAD 1	1056.4348	2.52E+07	0.337%	1.13%	93.89%	9.61 ± 0.04	3.7E+07	2.0E+07	1.9E+07	
14	gm-AEJM 1	1001.4112	2.03E+07	0.272%	0.91%	94.80%	15.59 ± 0.01	2.1E+07	2.3E+07	1.7E+07	
15	gm-AEJI 1	983.4547	1.81E+07	0.242%	0.81%	95.61%	20.26 ± 0.01	2.0E+07	1.8E+07	1.6E+07	
16	gm-AE 1	698.2859	1.73E+07	0.232%	0.77%	96.38%	8.74 ± 0.05	2.3E+07	1.1E+07	1.8E+07	
17	gm-AEJS 1	957.4027	1.58E+07	0.211%	0.70%	97.09%	5.79 ± 0.13	1.7E+07	1.0E+07	2.0E+07	
18	gm-AEJAW 1	1127.4871	1.47E+07	0.197%	0.66%	97.74%	25.19 ± 0.00	1.4E+07	1.6E+07	1.4E+07	
19	gm-AEJH 1	1007.4296	9.64E+06	0.129%	0.43%	98.17%	5.93 ± 0.01	1.1E+07	8.7E+06	8.9E+06	
20	gm-AEJW 1	1056.4500	6.41E+06	0.086%	0.29%	98.46%	23.89 ± 0.00	6.5E+06	7.0E+06	5.8E+06	
21	gm-AEJV 1	969.4391	6.29E+06	0.084%	0.28%	98.74%	14.56 ± 0.01	6.1E+06	6.5E+06	6.3E+06	
22	gm-AEJAQ 1	1069.4664	5.40E+06	0.072%	0.24%	98.98%	11.03 ± 0.03	5.5E+06	5.0E+06	5.7E+06	
23	gm-AEJAH 1	1078.4667	5.02E+06	0.067%	0.22%	99.20%	7.25 ± 0.05	5.6E+06	5.0E+06	4.5E+06	
24	gm-AEJT 1	971.4184	4.51E+06	0.060%	0.20%	99.40%	7.63 ± 0.05	4.5E+06	4.7E+06	4.4E+06	
25	gm-AEJAE 1	1070.4504	4.51E+06	0.060%	0.20%	99.61%	10.67 ± 0.04	3.8E+06	3.7E+06	6.0E+06	
26	gm-AEJAR 1	1097.5089	4.11E+06	0.055%	0.18%	99.79%	9.07 ± 0.06	3.7E+06	4.1E+06	4.5E+06	
27	gm-AEJAK 1	1069.5028	2.53E+06	0.034%	0.11%	99.90%	7.30 ± 0.06	3.6E+06	2.1E+06	1.9E+06	
28	gm-AEJD 1	985.3976	2.21E+06	0.030%	0.10%	100.00%	7.03 ± 0.19	2.8E+06	7.6E+05	3.1E+06	29.98%
29	gm-AEJA=gm-AEJA 2	1864.8048	9.28E+08	12.411%	24.39%	24.39%	15.70 ± 0.01	9.5E+08	9.2E+08	9.2E+08	
30	gm-AEJA=gm-AEJ 2	1793.7676	8.75E+08	11.707%	23.01%	47.39%	14.91 ± 0.02	8.6E+08	8.9E+08	8.7E+08	
31	gm-AEJG=gm-AEJ 2	1779.7520	6.16E+08	8.234%	16.18%	63.57%	13.63 ± 0.01	6.3E+08	6.3E+08	5.9E+08	
32	gm-AEJ=gm-AEJ 2	1722.7305	4.88E+08	6.524%	12.82%	76.39%	13.93 ± 0.01	5.1E+08	4.9E+08	4.6E+08	
33	gm-AEJG=gm-AEJA 2	1850.7891	2.96E+08	3.964%	7.79%	84.18%	14.37 ± 0.02	2.9E+08	3.0E+08	3.0E+08	

Table S6-13. List of unmodified muropeptides in TY1, TY2 and TY3 samples

Nh Structure		Theo mass	Avera	ige	% mono-	Cumulated	Average		Intensity	
ND	Structure	(Da)	Intensity	%	di-, trimers	abundance	RT (min)	TY1	TY2	TY3
34	gm-AEJF=gm-AEJ 2	1869.7989	9.12E+07	1.220%	2.40%	86.58%	23.89 ± 0.07	1.1E+08	8.5E+07	7.5E+07
35	gm-AEJN=gm-AEJ 2	1836.7734	7.00E+07	0.936%	1.84%	88.42%	12.95 ± 0.01	7.2E+07	7.0E+07	6.7E+07
36	gm-AEJ[AG/Q]=gm-AEJA 2	1921.8262	5.95E+07	0.796%	1.56%	89.99%	15.02 ± 0.02	5.8E+07	6.2E+07	5.9E+07
37	gm-AEJS=gm-AEJ 2	1809.7625	4.50E+07	0.602%	1.18%	91.17%	13.31 ± 0.01	4.6E+07	4.9E+07	4.0E+07
38	gm-AEJAI=gm-AEJA 2	1977.8888	3.42E+07	0.457%	0.90%	92.07%	24.15 ± 0.00	3.2E+07	3.3E+07	3.7E+07
39	gm-AEJAI=gm-AEJ 2	1906.8517	3.31E+07	0.442%	0.87%	92.94%	22.72 ± 0.01	3.4E+07	3.3E+07	3.3E+07
40	gm-AEJAF=gm-AEJ 2	1940.8360	3.19E+07	0.426%	0.84%	93.77%	24.69 ± 0.01	3.2E+07	3.2E+07	3.1E+07
41	gm-AEJN=gm-AEJA 2	1907.8105	2.19E+07	0.292%	0.57%	94.35%	13.64 ± 0.01	2.4E+07	2.2E+07	1.9E+07
42	gm-AEJS=gm-AEJA 2	1880.7996	1.86E+07	0.249%	0.49%	94.84%	14.06 ± 0.02	1.8E+07	2.0E+07	1.8E+07
43	gm-AEJY=gm-AEJ 2	1885.7938	1.79E+07	0.240%	0.47%	95.31%	19.59 ± 0.01	2.0E+07	1.7E+07	1.7E+07
44	gm-AEJAD=gm-AEJA 2	1979.8317	1.60E+07	0.214%	0.42%	95.73%	16.77 ± 0.01	2.3E+07	1.6E+07	8.7E+06
45	gm-AEJK=gm-AEJ 2	1850.8254	1.54E+07	0.206%	0.40%	96.13%	12.57 ± 0.01	2.0E+07	1.5E+07	1.2E+07
46	gm-AEJM=gm-AEJ 2	1853.7710	1.41E+07	0.188%	0.37%	96.50%	19.04 ± 0.01	1.6E+07	1.5E+07	1.2E+07
47	gm-AEJAA=gm-AEJA 2	1935.8419	1.34E+07	0.179%	0.35%	96.86%	16.27 ± 0.05	1.2E+07	1.4E+07	1.4E+07
48	gm-AEJI=gm-AEJ 2	1835.8145	1.31E+07	0.175%	0.34%	97.20%	22.78 ± 0.01	1.3E+07	1.4E+07	1.2E+07
49	gm-AEJH=gm-AEJ 2	1859.7894	1.27E+07	0.170%	0.33%	97.53%	12.62 ± 0.01	1.3E+07	1.3E+07	1.2E+07
50	gm-AEJH=gm-AEJA 2	1930.8265	8.00E+06	0.107%	0.21%	97.74%	13.27 ± 0.01	8.5E+06	8.0E+06	7.5E+06
51	gm-AEJT=gm-AEJ 2	1823.7782	7.76E+06	0.104%	0.20%	97.95%	13.93 ± 0.01	8.4E+06	7.9E+06	7.0E+06
52	gm-AEJY=gm-AEJA 2	1956.8309	6.87E+06	0.092%	0.18%	98.13%	20.40 ± 0.01	7.2E+06	6.9E+06	6.5E+06
53	gm-AEJAW=gm-AEJ 2	1979.8469	6.11E+06	0.082%	0.16%	98.29%	26.49 ± 0.01	6.4E+06	6.3E+06	5.6E+06
54	gm-AEJAM=gm-AEJA 2	1995.8452	5.82E+06	0.078%	0.15%	98.44%	20.43 ± 0.01	5.8E+06	6.2E+06	5.5E+06
55	gm-AEJAM=gm-AEJ 2	1924.8081	5.68E+06	0.076%	0.15%	98.59%	19.39 ± 0.02	6.6E+06	5.8E+06	4.6E+06
56	gm-AEJK=gm-AEJA 2	1921.8626	5.35E+06	0.072%	0.14%	98.73%	13.22 ± 0.01	5.7E+06	5.6E+06	4.8E+06
57	gm-AEJD=gm-AEJA 2	1908.7946	5.32E+06	0.071%	0.14%	98.87%	14.49 ± 0.02	7.4E+06	4.8E+06	3.8E+06
58	gm-AEJAW=gm-AEJA 2	2050.8840	5.24E+06	0.070%	0.14%	99.01%	27.00 ± 0.01	6.2E+06	5.6E+06	3.9E+06
59	gm-AEJV=gm-AEJ 2	1821.7989	4.19E+06	0.056%	0.11%	99.12%	19.13 ± 0.01	4.7E+06	4.3E+06	3.6E+06
60	gm-AEJF=gm-AEJA 2	1940.8360	3.89E+06	0.052%	0.10%	99.22%	25.27 ± 0.00	4.0E+06	3.9E+06	3.8E+06
61	gm-AEJD=gm-AEJ 2	1837.7574	3.85E+06	0.051%	0.10%	99.32%	13.68 ± 0.01	3.9E+06	4.1E+06	3.5E+06
62	gm-AEJT=gm-AEJA 2	1894.8153	3.71E+06	0.050%	0.10%	99.42%	14.67 ± 0.02	3.6E+06	3.7E+06	3.8E+06
63	gm-AEJAR=gm-AEJ 2	1949.8687	3.46E+06	0.046%	0.09%	99.51%	13.95 ± 0.06	3.9E+06	3.5E+06	3.1E+06
64	gm-AEJAF=gm-AEJA 2	2011.8731	2.95E+06	0.039%	0.08%	99.59%	25.90 ± 0.01	3.1E+06	2.7E+06	3.0E+06
65	gm-AEJW=gm-AEJ 2	1908.8098	2.86E+06	0.038%	0.08%	99.66%	25.36 ± 0.01	3.2E+06	3.0E+06	2.4E+06
66	gm-AEJAH=gm-AEJA 2	2001.8636	2.26E+06	0.030%	0.06%	99.72%	13.94 ± 0.01	2.3E+06	2.3E+06	2.2E+06
67	gm-AEJV=gm-AEJA 2	1892.8360	2.08E+06	0.028%	0.05%	99.78%	19.89 ± 0.01	2.0E+06	2.9E+06	1.4E+06
68	gm-AEJI=gm-AEJA 2	1906.8517	2.02E+06	0.027%	0.05%	99.83%	23.45 ± 0.00	2.1E+06	2.1E+06	1.9E+06

Nik Christenne		Theo mass	Avera	age	% mono-	Cumulated	Average		Intensity		
ND	Structure	(Da)	Intensity	%	di trimers	abundance	RT (min)	TY1	TY2	TY3	
69	gm-AEJM=gm-AEJA 2	1924.8081	1.53E+06	0.020%	0.04%	99.87%	19.74 ± 0.01	1.7E+06	1.8E+06	1.1E+06	
70	gm-AEJAE=gm-AEJA 2	1993.8473	1.14E+06	0.015%	0.03%	99.90%	15.81 ± 0.01	1.0E+06	1.2E+06	1.2E+06	
71	gm-AEJ=gm-AEJA 2	1793.7676	1.09E+06	0.015%	0.03%	99.93%	14.29 ± 0.01	9.8E+05	1.1E+06	1.1E+06	
72	gm-AEJAK=gm-AEJA 2	1992.8997	1.01E+06	0.014%	0.03%	99.96%	13.82 ± 0.01	1.1E+06	1.1E+06	9.1E+05	
73	gm-AEJAQ=gm-AEJA 2	1992.8633	9.37E+05	0.013%	0.02%	99.98%	16.36 ± 0.01	8.7E+05	9.6E+05	9.8E+05	
74	gm-AEJAA=gm-AEJ 2	1864.8048	6.93E+05	0.009%	0.02%	100.00%	17.60 ± 0.01	7.2E+05	4.9E+05	9.0E+05	50.89%
75	gm-AEJA=gm-AEJA=gm-AEJA 3	2788.2017	2.61E+08	3.493%	18.26%	18.26%	18.62 ± 0.01	2.7E+08	2.6E+08	2.6E+08	
76	gm-AEJA=gm-AEJ=gm-AEJA 3	2717.1646	2.59E+08	3.461%	18.09%	36.35%	18.10 ± 0.01	2.7E+08	2.6E+08	2.5E+08	
77	gm-AEJA=gm-AEJ=gm-AEJ 3	2646.1274	2.27E+08	3.042%	15.90%	52.25%	17.61 ± 0.01	2.2E+08	2.4E+08	2.2E+08	
78	gm-AEJG=gm-AEJ=gm-AEJ 3	2632.1118	1.73E+08	2.311%	12.08%	64.32%	16.45 ± 0.01	1.7E+08	1.7E+08	1.8E+08	
79	gm-AEJ=gm-AEJ=gm-AEJ 3	2575.0903	1.34E+08	1.787%	9.34%	73.66%	16.74 ± 0.01	1.4E+08	1.4E+08	1.3E+08	
80	gm-AEJG=gm-AEJ=gm-AEJA 3	2703.1489	1.07E+08	1.429%	7.47%	81.13%	16.97 ± 0.01	1.0E+08	1.1E+08	1.1E+08	
81	gm-AEJG=gm-AEJA=gm-AEJA 3	2774.1860	7.71E+07	1.031%	5.39%	86.52%	17.48 ± 0.01	7.7E+07	7.5E+07	7.9E+07	
82	gm-AEJF=gm-AEJ=gm-AEJ 3	2722.1587	2.89E+07	0.387%	2.02%	88.54%	25.04 ± 0.00	3.2E+07	2.9E+07	2.6E+07	
83	gm-AEJ[AG/Q]=gm-AEJA=gm-AEJA 3	2845.2231	1.92E+07	0.257%	1.34%	89.88%	18.00 ± 0.01	1.8E+07	2.0E+07	2.0E+07	
84	gm-AEJN=gm-AEJ=gm-AEJ 3	2689.1332	1.85E+07	0.247%	1.29%	91.17%	15.84 ± 0.01	1.8E+07	1.9E+07	1.8E+07	
85	gm-AEJS=gm-AEJ=gm-AEJ 3	2662.1223	1.29E+07	0.173%	0.90%	92.08%	16.15 ± 0.01	1.3E+07	1.4E+07	1.2E+07	
86	gm-AEJF=gm-AEJ=gm-AEJA 3	2793.1958	1.10E+07	0.148%	0.77%	92.85%	25.55 ± 0.01	1.1E+07	1.1E+07	1.1E+07	
87	gm-AEJN=gm-AEJ=gm-AEJA 3	2760.1703	1.03E+07	0.138%	0.72%	93.57%	16.34 ± 0.01	1.0E+07	1.0E+07	1.0E+07	
88	gm-AEJN=gm-AEJA=gm-AEJA 3	2831.2074	7.61E+06	0.102%	0.53%	94.10%	16.86 ± 0.01	7.4E+06	8.1E+06	7.3E+06	
89	gm-AEJF=gm-AEJA=gm-AEJA 3	2864.2329	7.16E+06	0.096%	0.50%	94.60%	26.00 ± 0.01	7.1E+06	7.6E+06	6.7E+06	
90	gm-AEJS=gm-AEJ=gm-AEJA 3	2733.1594	6.25E+06	0.084%	0.44%	95.04%	16.66 ± 0.00	6.5E+06	6.5E+06	5.7E+06	
91	gm-AEJAI=gm-AEJ=gm-AEJ 3	2759.2115	5.65E+06	0.076%	0.40%	95.43%	24.60 ± 0.01	5.1E+06	5.6E+06	6.2E+06	
92	gm-AEJH=gm-AEJ=gm-AEJ 3	2712.1492	5.07E+06	0.068%	0.35%	95.79%	17.02 ± 1.06	5.3E+06	4.8E+06	5.0E+06	
93	gm-AEJAI=gm-AEJA=gm-AEJA 3	2901.2857	5.06E+06	0.068%	0.35%	96.14%	25.57 ± 0.01	4.8E+06	5.0E+06	5.4E+06	
94	gm-AEJS=gm-AEJA=gm-AEJA 3	2804.1965	4.96E+06	0.066%	0.35%	96.49%	17.19 ± 0.01	5.1E+06	5.1E+06	4.7E+06	
95	gm-AEJAI=gm-AEJ=gm-AEJA 3	2830.2486	4.63E+06	0.062%	0.32%	96.81%	25.05 ± 0.01	4.9E+06	4.5E+06	4.5E+06	
96	gm-AEJY=gm-AEJ=gm-AEJ 3	2738.1536	4.36E+06	0.058%	0.30%	97.12%	21.34 ± 0.01	4.4E+06	4.8E+06	3.9E+06	
97	gm-AEJM=gm-AEJ=gm-AEJ 3	2706.1308	4.04E+06	0.054%	0.28%	97.40%	21.39 ± 0.70	4.2E+06	4.3E+06	3.6E+06	
98	gm-AEJK=gm-AEJ=gm-AEJ 3	2703.1852	3.43E+06	0.046%	0.24%	97.64%	15.44 ± 0.02	4.4E+06	3.3E+06	2.6E+06	
99	gm-AEJAA=gm-AEJA=gm-AEJA 3	2859.2388	3.39E+06	0.045%	0.24%	97.88%	18.99 ± 0.02	3.3E+06	3.2E+06	3.6E+06	
##	gm-AEJI=gm-AEJ=gm-AEJ 3	2688.1743	3.21E+06	0.043%	0.22%	98.10%	23.98 ± 0.01	3.5E+06	3.4E+06	2.8E+06	
##	gm-AEJAD=gm-AEJ=gm-AEJA 3	2832.1915	2.83E+06	0.038%	0.20%	98.30%	18.86 ± 0.01	3.9E+06	3.1E+06	1.5E+06	
##	gm-AEJM=gm-AEJ=gm-AEJA 3	2777.1679	2.13E+06	0.028%	0.15%	98.45%	21.52 ± 0.00	2.8E+06	2.4E+06	1.3E+06	
##	gm-AEJT=gm-AEJ=gm-AEJ 3	2676.1380	2.12E+06	0.028%	0.15%	98.60%	16.72 ± 0.01	2.2E+06	2.0E+06	2.2E+06	
##	gm-AEJH=gm-AEJ=gm-AEJA 3	2783.1863	1.89F+00	0.025%	0.13%	98.73%	16.03 ± 0.01	2.0E+06	1.8E+06	1.9E+06	

Nb	Structure	Theo mass Average		% mono-	Cumulated	Average	Intensity				
		(Da)	Intensity	%	di-, trimers	abundance	RT (min)	TY1	TY2	TY3	
##	gm-AEJAM=gm-AEJ=gm-AEJA 3	2848.2050	1.77E+06	0.024%	0.12%	98.85%	22.00 ± 0.00	2.0E+06	1.9E+06	1.4E+06	
##	gm-AEJAW=gm-AEJ=gm-AEJ 3	2832.2067	1.74E+06	0.023%	0.12%	98.97%	27.07 ± 0.01	1.6E+06	1.7E+06	1.9E+06	
##	gm-AEJY=gm-AEJ=gm-AEJA 3	2809.1907	1.70E+06	0.023%	0.12%	99.09%	21.83 ± 0.00	1.5E+06	1.8E+06	1.7E+06	
##	gm-AEJY=gm-AEJA=gm-AEJA 3	2880.2278	1.68E+06	0.023%	0.12%	99.21%	22.54 ± 0.00	1.5E+06	1.8E+06	1.8E+06	
##	gm-AEJAW=gm-AEJA=gm-AEJA 3	2974.2809	1.48E+06	0.020%	0.10%	99.31%	24.47 ± 4.82	3.3E+06	6.0E+05	6.0E+05	
##	gm-AEJH=gm-AEJA=gm-AEJA 3	2854.2234	1.38E+06	0.018%	0.10%	99.41%	16.49 ± 0.03	2.0E+06	1.1E+06	9.9E+05	
##	gm-AEJV=gm-AEJ=gm-AEJ 3	2674.1587	1.19E+06	0.016%	0.08%	99.49%	21.35 ± 0.01	1.3E+06	1.2E+06	1.1E+06	
##	gm-AEJAM=gm-AEJA=gm-AEJA 3	2919.2421	1.02E+06	0.014%	0.07%	99.56%	22.46 ± 0.01	1.2E+06	1.1E+06	7.7E+05	
##	gm-AEJK=gm-AEJ=gm-AEJA 3	2774.2224	9.90E+05	0.013%	0.07%	99.63%	15.94 ± 0.01	1.1E+06	8.8E+05	1.0E+06	
##	gm-AEJD=gm-AEJ=gm-AEJ 3	2690.1172	9.18E+05	0.012%	0.06%	99.70%	17.22 ± 0.01	6.5E+05	6.7E+05	1.4E+06	
##	gm-AEJW=gm-AEJ=gm-AEJ 3	2761.1696	8.96E+05	0.012%	0.06%	99.76%	26.15 ± 0.01	1.1E+06	9.1E+05	7.1E+05	
##	gm-AEJW=gm-AEJA=gm-AEJA 3	2903.2438	8.42E+05	0.011%	0.06%	99.82%	27.44 ± 0.02	8.3E+05	6.0E+05	1.1E+06	
##	gm-AEJT=gm-AEJ=gm-AEJA 3	2747.1751	7.13E+05	0.010%	0.05%	99.87%	17.21 ± 0.01	7.3E+05	7.4E+05	6.6E+05	
##	gm-AEJAD=gm-AEJA=gm-AEJA 3	2903.2286	6.36E+05	0.009%	0.04%	99.91%	19.31 ± 0.02	9.1E+05	6.6E+05	3.4E+05	
##	gm-AEJV=gm-AEJ=gm-AEJA 3	2745.1958	4.90E+05	0.007%	0.03%	99.95%	21.71 ± 0.00	4.3E+05	5.5E+05	4.9E+05	
##	gm-AEJV=gm-AEJA=gm-AEJA 3	2816.2329	4.31E+05	0.006%	0.03%	99.98%	22.10 ± 0.01	4.4E+05	4.2E+05	4.4E+05	
##	gm-AEJAD=gm-AEJ=gm-AEJ 3	2761.1544	3.08E+05	0.004%	0.02%	100.00%	17.83 ± 0.01	4.3E+05	2.5E+05	2.4E+05	19.13%
			7.48E+09	100.000%							

6.3.7 Step 5: Final quantification of muropeptides and comparison of growth conditions

Growing *R. leguminosarum* in minimal media (MM) as opposed to rich TY media leads to changes in the muropeptide profile, suggesting that PG remodelling occurs under these conditions. We sought to apply the strategy described (and summarized in Fig. 6-5) to compare the PG structure of *R. leguminosarum* grown in rich and minimal media.

To perform a final quantification, we combined the monomers from DB_2 with DB_4's porin muropeptides from RopA1,2,3, RopB, and pRL90069 to generate the database DB_5 (Table S6-13). This database was then used to perform a "one off" search using PGFinder's new bulk processing feature. All three TY and MM datasets were searched with a low mass tolerance (5 ppm) and anhydroMurNAc modifications and 3-3 / 4-3 multimers enabled. Individual search outputs were consolidated and manually checked wherever retention times had a standard deviation of more than 0.5 min. Dimer and trimer ambiguities were resolved using the strategy described in Step 4.

The final list of muropeptides contains 255 structures found across all three biological replicates of either condition: 65 monomers, 97 dimers, and 93 trimers. 111 muropeptides were exclusively found in the TY datasets, and 25 were exclusively found in MM. Comparing the two conditions reveals subtle differences in PG remodelling (Table 6-3 and Fig.6-8). Growth in MM was associated with a significantly lower crosslinking index (28.5% \pm 0.4% vs 31.3% \pm 0.5%; P=0.002) and a significant increase in glycan chain length (18.4 vs 21.1 residues; P=0.011). Interestingly, we found a moderate but significant increase of 3-3 cross-links in the MM samples ($64.7\% \pm 0.3\%$ vs $62.6\% \pm$ 0.4%; P=0.003). Whilst 3-3 cross-linking increased for dimers and trimers, L,Dtranspeptidase-mediated exchange activity (which leads to non-canonical residues in the fourth position) drastically dropped in the MM samples: non-canonical AEJX peptide stems represented only $2.18\% \pm 0.4\%$ of all muropeptides as compared to 26.6% ± 0.5% in the TY samples (P>0.001). A significant decrease in the proportion non-canonical AEJAX peptide stems was also found in the MM samples (16.6% ± 0.3% vs 5.7% ± 0.4%; P<0.001). Finally, a significant increase in the proportion of PG-bound porin peptides was observed in the MM datasets (5.2% \pm 0.9% vs 9.1% \pm 2.0%; P=0.037).

	ТҮ		MM		Unpaired t-test	
	Average	SD	Average	SD	P value (significance)	
Monomers (inc. porins)	30.72% ±	1.06%	35.57% ±	1.17%		
Dimers (inc. porins)	49.65% ±	0.85%	42.94% ±	0.31%		
Trimers (inc. porins)	19.63% ±	0.22%	21.49% ±	0.90%		
3-3	62.56% ±	0.45%	64.68% ±	0.31%	0.003 (**)	
4-3	37.44% ±	0.45%	35.32% ±	0.30%	0.002 (**)	
Crosslinking index	31.30% ±	0.49%	28.53% ±	0.43%	0.002 (**)	
AEJX monomers (excl. AEJA)	4.86% ±	0.21%	1.56% ±	0.34%		
AEJX dimers (excl. AEJA)	16.01% ±	0.50%	0.50% ±	0.06%		
AEJX trimers (excl. AEJA)	5.70% ±	0.05%	0.12% ±	0.01%		
All AEJX	26.58% ±	0.52%	2.18% ±	0.41%	<0.001 (***)	
AEJAX monomers (excl. AEJAA)	4.01% ±	0.09%	3.20% ±	0.17%		
AEJAX dimers (excl. AEJAA)	5.28% ±	0.11%	1.78% ±	0.27%		
AEJAX trimers (excl. AEJAA)	7.31% ±	0.26%	0.67% ±	0.14%		
All AEJAX	16.59% ±	0.29%	5.65% ±	0.41%	<0.001 (***)	
Anh monomers (inc. porins)	2.18% ±	0.08%	2.26% ±	0.15%		
Anh dimers (inc. porins)	5.95% ±	0.17%	4.98% ±	0.16%		
Anh trimers (inc. porins)	0.89% ±	0.02%	0.07% ±	0.02%		
All anhydro	9.02% ±	0.14%	7.77% ±	0.32%		
Chain length	18.35 ±	0.23	20.96 ±	1.03	0.011 (*)	
Porin monomers	3.83% ±	0.63%	5.28% ±	1.26%		
Porin dimers	1.17% ±	0.23%	3.08% ±	0.60%		
Porin trimers	0.23% ±	0.04%	0.75% ±	0.17%		
All porins	5.23% ±	0.87%	9.12% ±	2.01%	0.037 (*)	

Table 6-3. Comparative muropeptide analysis of PG extracted from cells grown in TY or MM.



Figure 6-8. *R. leguminosarum* grown in minimal media (MM) have different PG composition compared to those grown in rich media (TY).

The Sankey driagram shows total PG composition broken down first by oligomerisation state, then by stem peptides. Branch size is proportional to percentage, and only peptides stems are represented. A, L- or D-alanine; E, γ-D-glutamic acid; J, meso-diaminopimelic acid; X, any residue except Alanine.

6.4 Discussion

The numerous functionality improvements to PGFinder's existing MS tool, as well as the new, PGLang-enabled mass calculation and MS/MS fragment prediction features, were key to our PG analysis strategy. By making these improved tools accessible through an easy-to-use web-interface and laying out our approach in step-by-step tutorial, we hope to encourage others to adopt our rigorous and reproducible LC-MS/MS pipeline. Though PG structural analysis remains challenging we feel that the improvements made to PGFinder throughout this work are a significant step towards the eventual elimination of labour-intensive and error-prone manual analysis.

Since PGFinder was first described, two other tools dedicated to the LC-MS analysis of PG have been published (Hsu et al., 2023; Kwan et al., 2024). Every existing tool, PGFinder included, comes with its own trade-offs and rank differently when it comes to flexibility, completeness, and ease-of-use. HAMA, for example, is one of the more complete tools, covering the whole LC-MS/MS pipeline, but is written in MATLAB and lacks a GitHub repository. This makes it impossible to adapt or extend without a MATLAB licence and difficult to contribute those improvements back to HAMA. Additionally, HAMA does not currently build 3-3 cross-linked dimers, because they are often confused with 4-3 dimers due to the similarity of their fragmentation spectra. In contrast, PGFinder is written entirely using open-source programming languages, follows software best-practices, and is hosted on GitHub where anyone can easily contribute improvements back to the project. Though PGFinder does not yet automate MS/MS analysis, our signature ion approach makes it possible to know for certain if 3-3 or 4-3 cross-links are present, a critical distinction when it comes to assessing things like antibiotic resistance or L,D-transpeptidase activity. Another

powerful tool for the in silico fragmentation of muropeptides, PGN MS2, suffers from an incomplete description of its Python dependencies, making the installation process challenging. Additionally, by operating entirely at the atomic level, the fragment generation process is slow, and fragments can only be described in SMILES, making it difficult to tell at a glance which fragment came from which part of the muropeptide. Finally, though its fragment prediction is currently more complete than either HAMA or PGFinder's, it is limited to this task only; users will need to use another tool for the actual MS and MS/MS analysis (Kwan et al., 2024). PGFinder, on the other hand, requires no installation whatsoever, and its residue-graph abstraction makes generating fragments orders of magnitude faster than PGN MS2 whilst giving them each useful PGLang-like names. Additionally, MS analysis can be done within PGFinder itself, only requiring additional software for data deconvolution / feature extraction. The next steps for PGFinder are clear: more of the LC-MS/MS analysis pipeline can be covered by further automating tasks like cross-replicate consolidation, summary statistic generation, and MS/MS analysis/disambiguation. These changes, along with incorporating data deconvolution into PGFinder directly (eliminating the need for MaxQuant or Byos), would bring PGFinder closer to being a true, one-click muropeptide analysis tool.

We believe that the wider adoption of PGLang could help address the inconsistency of muropeptide descriptions throughout the literature. This inconsistency can make it difficult to understand the composition of many muropeptides; for example, the monomer GlcNAc-MurNAc-Ala-Glu-mDAP-Gly (gm-AEJG in PGLang) has been described in many ways: GM-Tripeptide + Gly (Coullon et al., 2020), (NAG)(NAM)-AemG (Kwan et al., 2024), AEmG (Anderson et al., 2019), Tri-Gly (Peltier et al., 2011), DS-TP-Gly (Popham et al., 1999), M3G (Desmarais et al., 2015), B-M-I(-A-E-H-G) (Kwan et al., 2024), or even as numbers originally defined in other publications (de Jonge et al., 1996). The description of dimers, trimers, and modifications are likewise inconsistent. By building on the existing intuition of those in the field, PGLang aims to remain intuitive whilst striking the right balance between concision and unambiguity. The automated monoisotopic mass calculation of PGLang structures will also help to address a surprising inconsistency in masses reported by the literature. For example, the theoretical monoisotopic mass of the major reduced monomer in E. coli (gm-AEJA, 941.407702 Da) is reported variably as: 941.4099 in (Kühner et al., 2014) (Δ ppm=2.3), 941.4030 in (Anderson et al., 2019) (Δ ppm=5.0), 941.41 in (Hernández et al., 2022) (Δ ppm=2.4) or 941.4064 in (Bui et al., 2009) (Δ ppm=1.4). Additionally, the PGLang to SMILES translator can be used to get stereoisomer-resolved structures that make obtaining a chemical formula, chemical drawing, or protein-ligand docking trivial (using a tool like Boltz-1; Wohlwend et al., 2024). Note that whilst PGN MS2 does output SMILES structures for the muropeptides it fragments (Kwan et al., 2024), these structures do not contain stereoisomer information and are therefore unsuitable for docking into stereospecific enzymes like L,D- and D,D-transpeptidases. By designing PGLang to be easy for humans to read and by including a number of useful tools for its translation and manipulation, we hope that it can become a standard nomenclature capable of improving consistency throughout the field.

Our proof-of-concept study describing PGFinder v0.02 (Patel et al., 2021) was largely limited to a description of the software. The step-by-step strategy laid out in this paper

allows any user with a basic understanding of PG structure to perform comprehensive structural analyses. Using *R. leguminosarum* as a model system, we identified 265 muropeptides, which represents (by far) the most comprehensive PG analysis to date and the first PG characterization of this organism. This work provides a solid foundation for exploring cell envelope remodelling occurring throughout the rhizobial life cycle: from a free-living soil-dwelling bacterium to a terminally differentiated bacteroid that can fix atmospheric nitrogen. The remodelling of PG has been associated with morphogenetic changes during growth and exposure to various stressors, but how specific enzymes contribute to this adaptation remains poorly understood. The level of detail of our analysis will allow us to more easily investigate the roles played by these PG remodelling enzymes in the future.

Another aspect of PG analysis that requires highly sensitive tools is the description of covalently anchored proteins. In R. leguminosarum, we demonstrated that a large proportion of muropeptides contain N-terminal residues from β -barrel proteins. This covalent anchoring of β -barrel proteins is known to tether the outer membrane in the closely related genera Coxiella and Brucella and play a role in maintaining cell envelope integrity (Godessart et al., 2021; Sandoz et al., 2021a). The increase in the proportion of PG fragments with β -barrel "scars" in MM may therefore be indicative of an increase in envelope stress. Further studying the dynamics of this process and establishing if distinct β -barrel proteins are preferentially anchored under different conditions, as has been shown in Coxiella burnetii (Sandoz et al., 2021a), would provide a valuable insight into rhizobial adaptation and symbiosis. An increase in 3-3 cross-linking by L,D-transpeptidases has also previously been implicated in stress resistance and cell envelope homeostasis (Morè et al., 2019). In the case of Rhizobium, growth in MM has a significant (albeit subtle) impact on the abundance of 3-3 cross-links, but a dramatic impact on the abundance of gm-AEJX muropeptides. Looking at these specific L,D-transpeptidation products will be useful for better understanding the role that individual L,D-transpeptidases play in PG remodelling and how they contribute to cellular fitness. Finally, we demonstrated that our PG analysis strategy can uncover unexpected muropeptides like those containing unusual amino acids in the fifth position of their peptide stem (gm-AEJAX in Table 3, 16.6% of total muropeptides). The biological significance of these unusual stems and the enzymes responsible remain unknown, but warrant investigation, as unusual residues in the fifth position are likely to impact PBP-mediated PG polymerization.

Overall, the granularity of the PG analysis described by this work makes it possible to monitor minor changes in PG structure and composition like never before and transforms how we study the bacterial cell wall and the role it plays in helping species like *R. leguminosarum* thrive in a highly dynamic environment.

6.5 Experimental Procedures

6.5.1 Bacterial strains and growth conditions.

R. leguminosarum bv. *viciae* strain 3841 (Wheatley et al., 2017) was grown at 28 °C in TY (5 g/L Tryptone + 3 g/L Yeast Extract + 1.3 g/L CaCl₂.6H₂O) broth or agar (15 g/L). The recipe for minimal medium is described in (Wheatley et al., 2017). Liquid cultures were grown in 2 L flasks under agitation (180 rpm).

6.5.2 Peptidoglycan extraction and muropeptide preparation.

Cells corresponding to 500 mL of culture were spun down and resuspended in 20 mL of boiling Milli-Q water prior to the addition of SDS 5% (w/v) final. After 30 min at 100 °C, peptidoglycan was recovered by centrifugation (2 h at 125,000 x *g*, room temperature) and washed three times in Milli-Q water. Samples were treated with trypsin (100 μ g/mL) for 4 h at 37 °C in 50 mM Tris-HCI (pH 7.5). Trypsin was heat-inactivated (10 min at 65 °C) and removed by washes in Milli-Q water. The material was freeze-dried and resuspended at a concentration of 10 mg/mL.

6.5.3 LC-MS/MS analysis

2 mg of purified peptidoglycan was digested for 16 h using 250 units of mutanolysin (Sigma-Aldrich) in 20 mM phosphate buffer (pH 5.5) in a final volume of 200 µL. Following heat inactivation (5 min at 100 °C), the soluble disaccharide peptides were mixed with an equal volume of 250 mM borate buffer (pH 9.25) and reduced via the addition of 25 µL of a sodium borohydride solution at 25 mg/mL. After 20 min at room temperature, the pH was adjusted to 5.0 using phosphoric acid. An Ultimate 3000 Ultra High-Performance Chromatography (UHPLC; Dionex / Thermo Fisher Scientific) system coupled with a high-resolution Q Exactive Focus mass spectrometer (Thermo Fisher Scientific) was used for LC-MS/MS analysis. Muropeptides were separated using a C18 analytical column (Hypersil Gold aQ, 1.9 µm particles, 150 x 2.1 mm; Thermo Fisher Scientific), at a temperature of 50 °C. Muropeptide elution was performed by applying a mixture of solvent A (water, 0.1% (v/v) formic acid) and solvent B (acetonitrile, 0.1% (v/v) formic acid). Following a 10 µL sample injection, MS/MS spectra were recorded over a 40 min gradient: 0-12.5% B for 25 min; 12.5-20% B for 5 min; held at 20% B for 5 min, followed by column re-equilibration for 10 min under the initial conditions. The Q Exactive Focus was operated under electrospray ionization (H-ESI II) in positive mode. Full scan (m/z 150-2250) used resolution 70,000 (FWHM) at m/z 200, with an automatic gain control (AGC) target of 1x10⁶ ions and an automated maximum ion injection time (IT). MS/MS spectra were recorded in "Top 3" datadependent mode using the following parameters: resolution 17,500; AGC 1x10⁵ ions, maximum IT 50 ms, NCE 25%, and a dynamic exclusion time of 5 seconds.

6.5.4 Determination of glycan chain length and crosslinking index

Cross-linking index and glycan chain length were calculated based on the formulae described previously (Glauner, 1988). The cross-linking was calculated as:

$$\frac{1}{2}$$
(% of all dimers) + $\frac{2}{3}$ (% of all trimers)

No glycosidically-linked multimers were identified, so all dimers and trimers included in this calculation were peptide cross-linked.

Glycan chain length was inferred from the abundance of anhydroMurNAc groups, which are found at the ends of glycan chains:

100

(% of anhydro monomers) + $\frac{1}{2}$ (% of anhydro dimers) + $\frac{1}{3}$ (% of anhydro trimers)

Because no di-anhydro muropeptides were included in the search process, they have also been excluded from the formula above.

6.5.5 Byos® searches

Unbiased searches were performed using Byonic[™] v5.2.5. For monomer searches, a FASTA file containing each peptide stem was used and glycan moieties (gm, 480.1955 Da) were added as N-terminal modifications. For PG-anchored proteins, searches were performed against the entire *R. leguminosarum* proteome. Modified peptides with a mass of 852.3600 Da (gm-AEJ) permitted once per peptide on any residue within the peptide were searched using non-specific cleavage parameters. Precursor mass tolerance was set at 8 ppm and fragment mass tolerance was set to 20 ppm for HCD fragmentation. Spectra corresponding to peptides containing an N-terminal disaccharide-tripeptide were examined manually.

6.6 Concluding remarks

This chapter presents the initial version of a submitted manuscript detailing the development of novel software tools for peptidoglycan structural analysis (peptidoglycomics) and a workflow showing the application of these tools to *Rhizobium leguminosarum* peptidoglycan composition analysis. Our research utilized LC-MS/MS data to analyse *R. leguminosarum* peptidoglycan grown under two distinct conditions (minimal and rich media), revealing an unprecedented complexity of over 250 muropeptides.

This chapter demonstrates how these new tools (PGLang, Mass Calculator and Fragment Generator), combined with an optimized workflow, facilitate the detailed characterisation of peptidoglycan composition, including modifications, distinct crosslinking patterns, and outer membrane protein tethering, confirming significant structural changes associated with growth conditions. To achieve the analysis, a key innovation was PGLang, a novel language for describing peptidoglycan structures, integrated to PGFinder (Patel et al., 2021), the open-source software I have been testing and the workflow I have been improving throughout my thesis.

Within the context of this thesis, this chapter presents the most updated workflow for analysing peptidoglycan (PG) composition. The addition of two new modules: the Mass Calculator and the Fragment Predictor, significantly streamlined the process of analysing peptidoglycan. As discussed in Chapter 3, a major concern associated with manual database generation is the risk of mass calculation errors. The Mass Calculator module addresses this issue by normalizing the masses of all residues and automatizing the calculation of the databases.

The Fragment Predictor module represents another significant advance. As explained in Chapter 5, the initial implementation allowed the generation of non-modified monomers and dimers using a command-line script. Unlike the previous version, this new version introduces a graphical user interface and expands capabilities to predict modified muropeptides and 3-1 crosslinks (Chapter 5), This improvement significantly simplifies and reduces the time required for MS/MS analysis to confirm the type of crosslinking and modifications.

Further improvements related to data handling include automatic consolidation of single datasets, optimization of search output and the ability to submit multiple datasets simultaneously. These improvements greatly reduced the time required for data analysis. The single dataset search consolidation feature reduces the effort of consolidating the results to combining results from multiple searches into a single output. To support this feature, the output format was optimized. Previously, information such as ion count, charge order and XIC coordinates were displayed. These columns have been replaced with Δ ppm and Abundance columns. This new format simplifies analysis and improves efficiency by eliminating the need of manual Δ ppm calculation.

Rhizobium leguminosarum exhibits complex interactions and adaptations as a symbiotic bacterium with plants and their response to environmental changes. *Rhizobium* is a useful model for studying microbial cell wall structure and dynamics, as well as for understanding how mutualistic relationships can be affected by

environmental factors (Aliashkevich et al., 2021; Cava et al., 2011; Jun et al., 2020; Torrens and Cava, 2024). Therefore, it is crucial to develop tools and mechanisms to be able to describe any minor changes that occur during the process of symbiosis.

Overall, PGFinder v1.2.1, with its novel language, new modules and proposed workflow, may become a valuable tool for peptidoglycomics research. The user-friendly interface and the stepwise methodology presented in this thesis have allowed the identification of a large number of PG fragments, making PGFinder a powerful tool for studying peptidoglycan structure. While manual inspection of fragmentation spectra remains necessary, further improvements towards complete automation are certainly possible, but this version represents a significant improvement compared to earlier versions described in my thesis.

Chapter 7

7. General Discussion

7.1 Contributions to Peptidoglycomics Research: Enhancing Mass Spectrometry Data Analysis

This thesis shows my contributions to peptidoglycomics research, focusing on mass spectrometry data analysis using PGFinder. The aim of the project was to facilitate a more comprehensive and reproducible analysis of LC-MS/MS data with user-friendly software tools. I contributed to develop a model strategy for PG analysis. In doing so, I identified areas of improvement and did beta-testing of newly added PGFinder features.

Conventional peptidoglycan (PG) composition analysis is complex and timeconsuming. The strategy followed since the late 80's typically involves purifying PG sacculi from bacterial cells, enzymatically digesting them to generate muropeptides, and separating the muropeptides by reverse-phase high-performance liquid chromatography (HPLC) (Glauner, 1988). The Identification of disaccharide-peptides then relies on the analysis of UV chromatograms using chemometric methods (Kumar et al., 2017) or mass spectrometry (MS) (Kühner et al., 2014; Porfírio et al., 2019). Chemometric methods use statistics to compare the UV chromatograms. However, these methods are somewhat "coarse" since they assume that each peak correspond to a single muropeptide and require correction of certain artifacts (e.g., irrelevant segments, offset, and retention time drifts) (Kumar et al., 2017). They also require the use of the proprietary software MATLAB. A major flaw is also that they rely on the prior identification of muropeptides in the samples used as a reference. This method is therefore not providing any evidence about the structure of muropeptides identified. The combination of these limitations may explain why this method has not been used except in a couple of seminal articles published by the lab that described it (Hsu et al., 2023; Kumar et al., 2017). The alternative to Chemometrics relies on the analysis of MS data. The mass-to-charge ratio (m/z) of ions is used to calculate monoisotopic masses of molecules and infer their structure, which can be confirmed by MS/MS. Ion intensity can be used to quantify the abundance of muropeptides (Patel et al., 2021). Nuclear Magnetic Resonance (NMR) can also be used for structural determination and characterization of peptidoglycan by determining the number of amide groups in muropeptide peptide stems. Two-dimensional NMR techniques, such as COSY (correlated spectroscopy), TOCSY (total correlated spectroscopy), and ROESY (rotating frame nuclear Overhauser effect spectroscopy), are used to further analyze muropeptide structure by identifying the connectivity of amino acid residues through amide bonds. NMR is useful for confirming the presence of anhydro groups in anhydromuramic acid or specific crosslinks (Atrih et al., 1998; Espaillat et al., 2016). Although solution state NMR has been successfully used to study the structure of individual muropeptides but extracting pure PG fragments for structural characterization via solution-state NMR is challenging (Hernández and Cava, 2021). Solid state NMR, on the other hand, allows for the study of whole cells and isolated cell walls, though it requires highly specialized equipment and expert interpretation due

to its lower spectral resolution and sensitivity compared to solution state NMR (Kim et al., 2015b).

The complexity associated with the methods for PG structural analysis has limited the number and breadth of PG composition studies. However, recent articles have described strategies for the analysis of LC-MS data using proprietary or open-access software (Anderson et al., 2019; Bern et al., 2017; Hsu et al., 2023; Kwan et al., 2024; Patel et al., 2021). My work has contributed to develop and test new tools using various bacteria with distinct PG compositions.

7.2 PGFinder improvements over the course of this project

Many improvements of PGFinder have been made by Brooks Rady in the lab (with some support of the Research Software Engineer team) over the course of my PhD to facilitate the PG analyses I carried out. They involved (i) an increased automation of the matching output processing, (ii) a faster, user-friendly interface, (iii) a databases builder and (iv) a module to predict ion fragmentation. The evolution of PGFinder has been driven by the issues met during the analysis of *C. difficile, G. oxydans* and *R. leguminosarum* PG.

7.2.1 Moving to more user-friendly platform

The initial versions of PGFinder (v.0.0.2 and v.0.1.1a) required manual database construction and server-side operation in a Jupyter Notebook environment (Patel et al., 2021). Using version 0.1.1a, we attempted to optimize the quality of our fragmentation data (MS/MS) collection (Chapter 3). We compared the MS output using different LC-MS/MS conditions and the issues encountered with the Jupyter notebook were a real bottleneck, with the server being sometimes down or too busy to run searches. The time investment required by PGFinder analysis, compared to Byonic[™] analysis, motivated us to transition to a client-side operation to reduce loading times. On a few occasions, the PGFinder script had to be changed to keep it compatible with the Jupyter Notebook. From PGFinder v1.0.3 onward, versions use a web interface that is run locally on the user computer, improving response time and user experience. The newer versions also feature an improved user interface with informative tags for more intuitive webpage usage. The addition of drag and drop boxes was a great improvement.

7.2.2 Making database construction, standard and reliable

Another major limitation we faced dealt with the calculation the monoisotopic masses of muropeptides to build databases. For large databases (over 150 structures, like those used in Chapters 3 and 6), manual calculation became tedious and error prone. PGFinder v1.2.1 featured a Mass Calculator module to generate databases automatically from a text file. The Mass Calculator module reduces the risk of misidentifying muropeptides due to mass calculation errors, which are common for larger and more complex databases. Although the calculation of monoisotopic masses is a trivial process, our analysis of the literature revealed a lot of mistakes so this change should help to fix this issue.

7.2.3 Using accuracy of mass identification (Δ ppm)

A common challenge we faced was resolving the instances where multiple muropeptides have very similar or identical masses (mass coincidence). Mass coincidences can result in inaccurate muropeptide assignments and therefore lead to the underestimation or overestimation of certain structures. This becomes problematic for low abundance muropeptides or those with unexpected compositions. Although the Mass Calculator can reduce the risk of mass coincidence due to a more accurate calculation of monoisotopic masses, some tight mass coincidences cannot be resolved without MS/MS data. MS/MS structure validation is time consuming. When the project began, neither the Mass Calculator nor the Fragment Predictor module were available. Additionally, since no MS/MS setup performed better than Top5 (Chapter3), we needed to identify trends and characteristic patterns to effectively distinguish mass coincidences and efficiently focus our efforts. Our first observation (Chapter 3) was that the structure with the lowest Δ ppm is more likely to be the correct. During Chapters 4 and 5, we manually calculated Δ ppm for every matched structure. PGFinder v1.2.1 calculates and displays Δ ppm in the search results for each possible match and picks by default the structure with the lowest Δ ppm.

7.2.4 Fragment predictor module for MS/MS analyses

Resolving mass coincidences requires the fragmentation of ions by tandem mass spectrometry (MS/MS) to determine the sequence of muropeptides (monomers or multimers). In Chapter 3, we used the Byonic[™] module from the Byos® suite to avoid a manual analysis of fragmentation data. The automated annotation of MS/MS spectra by Byonic[™] was used to create monomer databases validated by MS/MS to reduce the original search space to a minimum. However, Byonic[™] cannot analyse the fragmentation of multimers due to their unusual crosslinking (they contain two Ntermini). As an alternative to the proprietary software Byos®, we decided to develop a PGFinder module for MS/MS analysis. This required the creation of PGLang, a formal language for describing muropeptide structures, introduced in PGFinder v1.2.1. This represented a significant progress, enabling in silico generation of fragmentation products for any muropeptide (modified or not, branched or not and multimers with 4-3, 3-3, or 1-3 crosslinks). The fragment predictor enabled the interpretation of MS/MS spectra to determine the structure of R. leguminosarum PG (Chapter 6). While manual inspection of fragmentation spectra remains necessary for complex muropeptides, this module represents a substantial improvement. Being part of our open-access software, it allows users with no access to the Byos® platform to perform MS/MS analysis of muropeptides.

7.2.5 Analysis of 1-3 crosslinks

The characterisation of 1-3 crosslinks in *G. oxydans* (Chapter 5) using PGFinder v1.0.3 required a manual calculation of monoisotopic masses to create a bespoke database including specific monomers and dimers. The addition of 1-3 crosslinked dimer prediction and calculation capabilities in PGFinder v1.1.0 helped 1-3 crosslink identification. However, the initial fragment predictor was limited to predicting 3-3 and 4-3 dimers for unmodified muropeptides. In consequence, during 1-3 crosslink validation, manual calculation of the mass of MurNAc-L-Ala=mDAP (m-A=J) and

comparison with available product ions in the fragmentation data were necessary to identify signature ions in potential 1-3 crosslinked dimers.

7.2.6 Batch processing of datasets

R. leguminosarum characterization (Chapter 6) presented a new challenge, requiring parallel analysis of six datasets (three corresponding to rich media (TY) and three to minimal media (MM). The transition of PGFinder to client-side operation enabled batch submission of multiple datasets (PGFinder v1.2.1), significantly reducing processing time compared to single-dataset submission required by the Jupyter Notebook interface. Furthermore, automated consolidation of findings from individual searches improved readability and analysis efficiency for multiple datasets. Finally, although this feature has not been used for this work, I have contributed to design the automated consolidation across replicates in PGFInder v1.4.0, which will represent a significant gain of time and will increase the reproducibility of analyses.

7.2.7 Drawing PG fragments and predicting their structure using SMILES

The newest version of PGFinder (v1.4.0) generates Simplified Molecular Input Line Entry System (SMILES) strings from PG structure. This represents a major step forward to draw PG molecules and predict their 3D structures. Predicted SMILES consider the chirality of PG fragments and therefore open the possibility to dock PG ligand in protein structures either determined experimentally or predicted with AlphaFold. The recent improvement in AlphaFold molecular modelling will pave the way to understand the structure/function of enzymes involved in PG metabolism. For example, it will provide key information to explain the specific role of Ldts.

7.2.8 PGFinder analysis caveats

Analyzing PG structure using PGFinder is a powerful approach. However, to fully interpret the results, users should be aware of its inherent limitations and caveats, understanding what the data represents and what aspects might be overlooked.

As previously explained, the method involves purifying and digesting the sacculi from a population of cells. Consequently, the analysis provides an average PG composition across the entire population but lacks information about the structural arrangement within the intact PG network. Standard digestion protocols and LC separation methods are designed to analyse disaccharide-peptides (muropeptides). Outer membrane proteins or lipoproteins can be covalently attached to PG. Although they can in theory be identified using PGFinder, this is a challenging task that requires preliminary proteomic analysis to identify candidates potentially linked to PG. Once these have been identified, PGFinder can easily be used to search for specific "scars" of these proteins, *i.e.* disacharride-peptides containing amino acids from covalently linked proteins. This approach has been validated with both R. leguminosarum and B. abortus (Alamán-Zárate et al., 2025; Sandoz et al., 2021b). Finally, PGFinder is not suitable to explore the structure of other polymers such as WTAs or polysaccharides linked to PG. At best PGFinder can identify extra phosphate groups resulting from the removal of such polymers during the purification of PG (Atrih and Foster, 1999; Heydenreich et al., 2025).

Population level limitations

Standard PG analysis is performed on PG isolated from a population of cells, often at various stages of the cell cycle. As a result, the data represent an average composition across the population and cell surfaces. PGFinder analysis does not provide information about structural differences in architecture that might exist in specific locations within a single cell, such as the cell poles or the division septum. While comparative analyses between different growth phases are possible, heterogeneity within a single mixed population cannot be captured unless physical separation of a specific state is performed beforehand.

Database dependency and manual inspection

PGFinder relies on matching observed masses against a database of known or predicted muropeptide structures. Any muropeptide structure not included in the database will be missed. For example, in the R. leguminosarum analysis, during the last search, on average 19.41% of the total intensity captured in the datasets was identified (as monomers, dimers, trimers, or modifications). The remaining unidentified intensity corresponds to "dark matter", representing structures not present in the database during searches or molecules unrelated to PG. While PGFinder aims to reduce bias and improve detection of low-abundance species compared to other analytical methods, its effectiveness still depends on the design of the search database. Unusual modifications, or novel structures absent from the database need to be identified using manual inspection. This strategy should allow to identify low abundance muropeptides, especially those with uncoded amino acids (e.g. D-lactate, Lanthionine, ornithine, D-canavanine) (Deghorain et al., 2007; Vasstrand, 1981). Manual verification remains crucial to filter out potential misidentifications due to mass coincidences or resolve ambiguous structures. Unusual or uninterpretable mass data or fragmentation patterns may be overlooked if their structure cannot be confidently assigned or validated. Distinguishing structural isomers (3-3 and 4-3 crosslinks) requires manual verification, and determining the exact location of modifications or resolving mass coincidences in multimers can be challenging, particularly when MS/MS spectra lack sufficient peptide fragment ions.

Quantification and analytical constraints

Muropeptides are quantified based on ion intensity in the mass spectrometry data. While this intensity is useful for comparing relative compositional changes, it does not necessarily represent absolute molar abundance. However, our comprehensive strategy in *R. leguminosarum* identified a large number or muropeptides (>250), suggesting that much of the complexity is captured within the muropeptide pool. Still, this refers only to the material that was effectively digested, separated, and detected by LC-MS/MS. Large unresolved fragments or the intact network are not analysed, and determining their precise composition and structural arrangement within the native cell wall requires complementary techniques. While PGFinder can identify low abundance muropeptides, material absent from the database or yielding uninterpretable spectra will be missed. Ultimately, users should recognize that PGFinder analysis provides a snapshot of the digested fragments from a population average, rather than a complete picture of the intact, dynamic, and potentially spatially heterogeneous PG network within individual cells.

7.3 How can we further improve PGFinder?

Although PGFinder enables users with no command skills to analyse complex LC-MS/MS data, several aspects of our pipeline need to be improved.

7.3.1 Database building

Whilst calculation of monoisotopic masses is now automatic, the list of muropeptides must be created by the user and this process is somewhat time-consuming. In the future, it would be good to offer the possibility to build a list of monomers automatically, asking the user to provide basic properties of the PG they study: what are the amino acids in position 2 and 3P Are these residues substituted by a lateral chainP If so, which residuesP Do you want to include peptide stems resulting from exchange activityP If so, 1-3 (thereby including gm-dipeptides)P 3-3 (thereby including gm-tetrapeptides)P or 4-3 (thereby including gm-pentapeptides)P Based on this information, we could make the mass calculator build a bespoke list of structure with their monoisotopic masses. This could be useful when studying the peptidoglycan with an unknown composition, allowing for "try and error" searches until the suitable strategy is identified.

7.3.2 Database building

A major limitation in PGFinder is that it is exclusively able to generate multimers for PG containing mDAP in position 3. The formation of 4-3, 3-3 and 1-3 multimers currently follows a fixed structural assembly, limiting its ability to predict all the diversity of multimers. Currently, the software exclusively assigns gm-AEJA, gm-AEJ or gm-A as donor stems to the previously identified monomers. Consequently, multimers from most Gram-positive peptidoglycans that contain lysine and a lateral chain cannot be predicted. This limitation also applies to peptidoglycans containing amidated mDAP (*Bacillus subtilis*) ornithine (*Leptospira interrogans*) or lanthionine (*Fusobacterium nucleatum*) at position 3. This issue is currently being fixed in the laboratory as a matter of urgency (Tia Duh, personal communication) and should be available before summer 2025. The idea is to identify the most abundant monomers and use this information to define the composition of the donor stem that needs to be used to generate 4-3, 3-3 or 1-3 crosslinks.

7.3.3 Integrating MS/MS analysis to the search output

PGFinder v1.2.1 has significantly enhanced the user-friendliness and analytical capabilities of PGFinder v0.0.2. However, while the Fragment Predictor module is a valuable tool, the inspection of fragmentation spectra remains a manual operation. Further development is needed to fully integrate automated MS/MS data analysis within PGFinder to streamlining the validation process. As a first step, we could add a column to the search output to indicate if MS/MS data is available. Chapter 3 revealed that the MS/MS coverage is not always satisfactory. Identifying putative PG masses with no MS/MS data could allow the user to design a complementary experiment to carry out targeted fragmentation on the corresponding ions. Ideally, PGFinder could also generate a list of ions to fragment. Practically, this means that the user could run a sample, analyse the data as soon as it is available and run it again to generate missing MS/MS data. Targeted fragmentation requires a retention time, target m/z value and ion charge for optimal results. Being able to carry out this operation shortly after the

data has been collected would ensure optimal fragmentation. We could even make PGFinder build a list of ions with missing MS/MS data with all the parameters required (m/z, charge, RT ions). This would guarantee the completeness of fragmentation.

The most challenging task will be to include MS/MS characterization to PGFinder in a similar way to Byonic[™]. A first task will be to enable PGFinder to "read" deconvoluted data files which contain information for MS and MS/MS analysis. Now, MS analysis is carried out using either .FTRS or .txt files generated by Byos® or MaxQuant, respectively. These formats do not contain any MS/MS data. A first step will be to enable PGFinder to read mzML files, an open-source and vendor-neutral file format. Next, we will need to create a script that compares the monoisotopic masses of predicted fragments with those in observed and establish a scoring system. This scoring system will need to predict the likelihood of fragmentation and in the case of mass coincidence consider signature ions to discriminate between structures. Adding MS/MS capability to PGFinder is currently work in progress in the lab (Brooks Rady). This feature should be available in the coming months.

7.3.4 Novel strategies to Identify PG "dark matter"

Using E. coli PG as a proof of concept revealed that approximately 50-60% of the masses identified can be assigned to PG structures (Patel et al., 2021). There are therefore hundreds of remaining masses that could potentially correspond to PG fragments. Manual inspection of MS spectra from unassigned masses occasionally reveals signature ions (e.g., [M+H]⁺ at 204.09, corresponding to GlcNAc residues), suggesting that certain PG properties remain to be identified. While we previously claimed that PGFinder enables an "unbiased" peptidoglycan analysis by facilitating a comprehensive search using databases containing all possible peptide stem compositions, it could be argued that, although this represents a significant improvement over previous studies, which often fail to describe the search space used, it does not a truly unbiased search. A rigorous unbiased search should be made with a random library of PG structures, built with a minimal set of rules defining the connectivity between disaccharides and amino acids. Such an approach is theoretically feasible but not entirely smart, as PG structure and composition are generally extremely conserved. The so-called "dark matter" muropeptides constitute a relatively small proportion of the total muropeptide population; to date, no PG samples have been analysed containing major, unidentified species. To identify the structure of putative PG fragments with signature ions in their MS/MS spectra, we could generate libraries of offset masses to known structures. This would be a first step to identify modifications found on several muropeptides. The formal identification of such unusual mass increments would still require additional analyses such as NMR. Some examples of uncommon amino acids (e.g. D-canavanine, lanthionine or alaninol) can be helpful (Aliashkevich et al., 2021; Coullon et al., 2020; Fredriksen et al., 1991).

7.4 How did PGFinder and this work contribute to our understanding of PG biology?

The granularity of PG analysis is critical for identify PG properties present at low abundance. Because PG is an essential molecule, its biosynthesis pathway is conserved and unless bacteria are grown in very particular conditions (like in the presence of antibiotics), it is expected that the PG composition will not be subject to a lot of variations. It is therefore not surprising that the activity of enzymes involved in PG synthesis (in particular PG hydrolases) are inferred from the *in vitro* activity of recombinant proteins rather than PG isolated from mutants. The use of PGFinder has led to interesting discoveries that relied on the unbiased and systematic search of muropeptides in deconvoluted data. These discoveries have also relied on the high sensitivity of the LC-MS/MS approach. A few examples are given below.

7.4.1 The high sensitivity of LC-MS/MS can identify low abundance features

The identification of minor PG modifications requires a very sensitive method. For example, a low amount of deacetylated monomers (0.1%) was detected in *E. coli* (Patel et al., 2021). The biological relevance of PG acetylation remains to be explored but this suggests that such modifications could play a role in particular physiological conditions and are therefore worth being monitored in future studies. Previous work on *C. burnetii* PG has also revealed the existence of "proteins scars" on muropeptides (Sandoz et al., 2021a). The analysis of *R. leguminosarum* PG also revealed minor species resulting from the anchoring of one β -barrel protein (pRL90069) representing less than 0.05% of monomers. Another example is the identification of so-called "denuded glycan strands" identified as di- tri- or tetrasaccharides (Anderson et al., 2020; Patel et al., 2021). In *E. coli* this denuded glycan strands could be deacetylated by SddA modulating a switch to activate septal peptidoglycan splitting in the early stages of cell division (Hernandez-Rocamora et al., 2025).

7.4.2 Using high-resolution LC-MS to revisit the specific activity of enzymes involved in PG synthesis

Our work on C. difficile Ldt_{Cd1}, Ldt_{Cd2}, Ldt_{Cd3}, has revealed that the high sensitivity of LC-MS is a great asset to understand the functional specialisation of enzymes. Our work revealed that a previous MS analysis was relatively crude and failed to discover the activity of some of these enzymes such as the capacity of Ldt_{Cd1} to carry out exchange or transpeptidation reactions (Galley et al., 2024; Sütterlin et al., 2018). The quantification of transpeptidation products in vitro generated by C. difficile Ldts allowed us to determine the preferential activity of each enzyme in vitro. With an automated analysis to perform cross-replicate consolidation, we could easily envisage kinetic analyses requiring low amounts of purified material to further explore the structure/function properties of enzymes. Although this would require some effort, we could also envisage to use LC-MS and PGFinder to explore the respective contribution of individual enzymes such as Ldts to the composition of PG. For example, in E. coli, 3 Ldts have been associated with protein anchoring (LdtA, B and C) and 2 (LdtD and E) have been associated with PG crosslinking. The respective contribution of LdtA, B, C to protein anchoring is unknown. The automation of LC-MS data analysis paves the way for systematic analyses to re-investigate the contribution of individual PG metabolic enzymes.

7.4.3 C. difficile, 3-3 crosslinks and antimicrobial resistance

The first study exploring the PG structure of C. difficile reported a contribution of two Ldts to the PG structure of this organism (Peltier et al., 2011). Using the same strains as those described in the publication and biological triplicates, we could not confirm the conclusions of this work (S. Mesnage and N. Galey, unpublished), suggesting that a comparative PG analysis based on a single sample can be misleading. Our work on a different model strain (R20291) also confirmed the none of the canonical Ldts encoded by C. difficile have a major impact on transpeptidation when cells are grown in rich media. The discovery of the Ldts harbouring a VanW domain (Bollinger et al., 2024) and the creation of isogenic mutants with multiple deletions revealed that the presence of 3-3 crosslinks in C. difficile is essential. This result is surprising and contrasts with E. coli and M. smegmatis, in which all Idt genes can be inactivated simultaneously (Francis et al., 2023; Sanders and Pavelka, 2013). The catalytic activities of C. difficile VanW Ldts awaits more detailed analyses to determine their preferential role in PG remodelling. Another question that remains open is their contribution to antibiotic resistance. While vanW is present in the loci encoding vancomycin resistance in Gram positive bacteria (Stogios and Savchenko, 2020), the contribution of VanW proteins to resistance has not been demonstrated. The essential nature of L,D-transpeptidation in C. difficile means that this organism produces enzymes that maintain a low abundance of pentapeptide stems in the peptidoglycan (c.a. 0.15% of all muropeptides).

7.4.4 Functional diversity of Ldts

Ldts can perform several enzymatic reactions (crosslinking, exchange as well as the cleavage of monomers or dimer stems). This work revealed that some Ldts distantly related to canonical Ldts (with a YkuD domain) like G. oxydans Ldt_{Go2} enzyme can also make 1-3 crosslinks. By analogy with the catalytic activity of Ldts forming 3-3 crosslinks, we proposed that Ldt_{Go2} could use a disaccharide-dipeptide as a donor substrate. Based on the work published by our competitors, it seems unlikely that this hypothesis is correct. Based on their in vitro assays using recombinant enzyme from Burkholderia cenocepacia Ldt_{Bc} and purified PG sacculi, they showed that Ldts catalysing 1-3 bonds engaging non-terminal amino acids in the crosslinking process (Espaillat et al., 2024). Another interesting property is the ability of Ldt_{Bc} to exchange both L- and D- amino acids (Espaillat et al., 2024). These remarkable properties illustrate how Ldts have evolved to perform a wide range of enzymatic reactions to modify or polymerise bacterial PGs. The contribution of Ldts to bacterial physiology seems to be associated with resistance to stress and survival (Gupta et al., 2010; Lavollay et al., 2008; Magnet et al., 2007; Morè et al., 2019). It has been proposed that 1-3 crosslinks contribute to cell envelope stability in high acetic acid concentrations which are common encountered by Actetobacteraceae. In G. oxydans, non-canonical Ldts might also facilitate evasion of the Drosophila immune system. Insect Peptidoglycan Recognition Protein LE (PGRP-LE) has been shown to selectively bind to mDAP-containing PG (Tindwa et al., 2013). This recognition activates immune pathways and autophagy (Tindwa et al., 2013). By cleaving mDAP from its peptide stems in the crosslinks, G. oxydans could potentially evade immune recognition and the production of antimicrobial peptides. Furthermore, increasing the levels L.D.-transpeptidation (1-3 crosslinks) can inhibit the activity of LTs. This inhibition also influences the release of anhydromuropeptides which are finally signalling molecules of PGRP-LE (Alvarez et al.,

2024). Finally, the role of the Ldt_{Go2} homologue Ldt_{Go1} awaits further analysis. Gene duplication is a common evolutionary phenomenon, and the evolutionary pressures maintaining duplicated genes, such as neofunctionalization, sub functionalization, gene dosage amplification, or backup compensation, are often complex and not fully understood (Kuzmin et al., 2022).

7.4.5 Exchange reaction

A recurring theme throughout this work is to explore PG remodelling during bacterial growth and division. We focused on enzymes forming 1-3 (G. oxydans), 3-3 (C. difficile) and 4-3 (R. leguminosarum) crosslinks. Beside the polymerisation activities of D,D- and L,D-transpeptidases, these enzymes also have cleavage and exchange activities; G. oxydans L,D-transpeptidase generates disaccharide-dipeptides, C. difficile L,D transpeptidases generates disaccharide-tetrapeptides. Interestingly, this work revealed that R. leguminosarum PG contains many pentapeptide stems with noncanonical amino acids at their C-terminus. This could be either associated with a nonspecific incorporation of amino acids in the D-Ala-D-Ala peptide by the Ddl ligase or an exchange reaction. The latter is more likely to explain the presence of unusual amino acids in position 5. Exchange activity has been described in studies using fluorescent D-amino acids as probes for *in situ* PG labelling (Kuru et al., 2019; Taguchi et al., 2019). Some penicillin-binding proteins (PBPs) catalyze D-amino acid exchange in vitro (Lupoli et al., 2014; Taguchi et al., 2019). PBP transpeptidases catalyze D-amino acid exchange by forming a covalent complex with the donor peptide, replacing terminal D-Ala (Cochrane and Lohans, 2020). Certain low molecular weight PBPs, such as S. aureus PBP4 (SaPBP4), E. faecalis EfPBPX and Streptococcus gordonii SgPBPX, can perform fifth residue D-amino acid exchange (Welsh et al., 2017) and catalyze cyclic muropeptide formation in vitro (Maya-Martinez et al., 2018).

The presence of many non-canonical residues in position 5 has not been described in the PG composition of bacteria, most certainly because it has not been searched. We found these in several Gram-negative organisms such as *Flavobacterium johnsoniae*, *Fusobacterium nucleatum* or *R. leguminosarum*. The ability of PBPs to exchange amino acids in position 5 awaits further analysis. *R. leguminosarum* preliminary work found more than ten PBPs (Olson et al., 2023), but none are homologous to *Sa*PBP4. How the presence of amino acids different from D-Ala affect PG polymerisation is also an important question to address.

8. References

- Abdek-Akher M, Hamilton JK and Smith F (1951) The Reduction of Sugars with Sodium Borohydride 1. *Journal of the American Chemical Society* 73(10): 4691–4692. DOI: 10.1021/ja01154a061.
- Alamán-Zárate MG, Rady BJ, Evans CA, et al. (2024) Unusual 1-3 peptidoglycan crosslinks in *Acetobacteraceae* are made by L,D-transpeptidases with a catalytic domain distantly related to YkuD domains. *The Journal of biological chemistry* 300(1): 105494. DOI: 10.1016/j.jbc.2023.105494.
- Alamán-Zárate MG, Rady BJ, Ledermann R, et al. (2025) A software tool and strategy for peptidoglycomics, the high-resolution analysis of bacterial peptidoglycans via LC-MS/MS. *Communications chemistry* 8(1): 91. DOI: 10.1038/s42004-025-01490-6.
- Albers S-V and Meyer BH (2011) The archaeal cell envelope. *Nature reviews. Microbiology* 9(6): 414–426. DOI: 10.1038/nrmicro2576.
- Alderwick LJ, Harrison J, Lloyd GS, et al. (2015) The Mycobacterial Cell Wall--Peptidoglycan and Arabinogalactan. *Cold Spring Harbor perspectives in medicine* 5(8): a021113. DOI: 10.1101/cshperspect.a021113.
- Aliashkevich A and Cava F (2021) LD-transpeptidases: The great unknown among the peptidoglycan cross-linkers. *The FEBS journal*. Epub ahead of print 10 June 2021. DOI: 10.1111/febs.16066.
- Aliashkevich A, Guest T, Alvarez L, et al. (2024) LD-transpeptidation is crucial for fitness and polar growth in *Agrobacterium tumefaciens*. *PLoS genetics* 20(10): e1011449. DOI: 10.1371/journal.pgen.1011449.
- Aliashkevich A, Howell M, Brown PJB, et al. (2021) d-canavanine affects peptidoglycan structure, morphogenesis and fitness in Rhizobiales. *Environmental microbiology* 23(10): 5823–5836. DOI: 10.1111/1462-2920.15513.
- Allen NE and Hobbs JN (1995) Induction of vancomycin resistance in *Enterococcus faecium* by non-glycopeptide antibiotics. *FEMS microbiology letters* 132(1-2): 107–114. DOI: 10.1111/j.1574-6968.1995.tb07819.x.
- Alvarez L, Hernandez SB, Pedro MA de, et al. (2016) Ultra-Sensitive, High-Resolution Liquid Chromatography Methods for the High-Throughput Quantitative Analysis of Bacterial Cell Wall Chemistry and Structure. *Methods in molecular biology* (*Clifton, N.J.*) 1440: 11–27. DOI: 10.1007/978-1-4939-3676-2_2.
- Alvarez L, Hernandez SB, Torrens G, et al. (2024) Control of bacterial cell wall autolysins by peptidoglycan crosslinking mode. *Nature communications* 15(1): 7937. DOI: 10.1038/s41467-024-52325-2.
- Anderson EM, Greenwood NA, Brewer D, et al. (2020) Semi-Quantitative Analysis of Peptidoglycan by Liquid Chromatography Mass Spectrometry and Bioinformatics.

Journal of visualized experiments JoVE. Epub ahead of print 13 October 2020. DOI: 10.3791/61799.

- Anderson EM, Shaji Saji N, Anderson AC, et al. (2022) Pseudomonas aeruginosa
 Alters Peptidoglycan Composition under Nutrient Conditions Resembling Cystic
 Fibrosis Lung Infections. mSystems 7(3): e00156 22. DOI:
 10.1128/msystems.00156-22.
- Anderson EM, Sychantha D, Brewer D, et al. (2019) Peptidoglycomics reveals compositional changes in peptidoglycan between biofilm- and planktonic-derived *Pseudomonas aeruginosa*. *The Journal of biological chemistry* 295(2): 504–516.
 DOI: 10.1074/jbc.RA119.010505.
- Araki Y and Ito E (1989) Linkage units in cell walls of gram-positive bacteria. *Critical reviews in microbiology* 17(2): 121–135. DOI: 10.3109/10408418909105745.
- Armstrong T, Fenn SJ and Hardie KR (2021) JMM Profile: Carbapenems: a broadspectrum antibiotic: This article is part of the JMM Profiles collection. *Journal of Medical Microbiology* 70(12). DOI: 10.1099/jmm.0.001462.
- Atrih A, Bacher G, Allmaier G, et al. (1999) Analysis of peptidoglycan structure from vegetative cells of *Bacillus subtilis* 168 and role of PBP 5 in peptidoglycan maturation. *Journal of bacteriology* 181(13): 3956–3966. DOI: 10.1128/jb.181.13.3956-3966.1999.
- Atrih A and Foster SJ (1999) The role of peptidoglycan structure and structural dynamics during endospore dormancy and germination. *Antonie van Leeuwenhoek* 75(4): 299–307. DOI: 10.1023/a:1001800507443.
- Atrih A, Zöllner P, Allmaier G, et al. (1996) Structural analysis of *Bacillus subtilis* 168 endospore peptidoglycan and its role during differentiation. *Journal of bacteriology* 178(21): 6173–6183. DOI: 10.1128/jb.178.21.6173-6183.1996.
- Atrih A, Zöllner P, Allmaier G, et al. (1998) Peptidoglycan structural dynamics during germination of *Bacillus subtilis* 168 endospores. *Journal of bacteriology* 180(17): 4603–4612. DOI: 10.1128/jb.180.17.4603-4612.1998.
- Azuma I, Thomas DW, Adam A, et al. (1970) Occurrence of N-glycolylmuramic acid in bacterial cell walls. *Biochimica et Biophysica Acta (BBA) - General Subjects* 208(3): 444–451. DOI: 10.1016/0304-4165(70)90217-5.
- Bahadur R, Chodisetti PK and Reddy M (2021) Cleavage of Braun's lipoprotein Lpp from the bacterial peptidoglycan by a paralog of l,d-transpeptidases, LdtF. *Proceedings of the National Academy of Sciences of the United States of America* 118(19). DOI: 10.1073/pnas.2101989118.
- Baloh M, Nerber HN and Sorg JA (2022) Imaging *Clostridioides difficile* Spore Germination and Germination Proteins. *Journal of bacteriology* 204(7): e0021022. DOI: 10.1128/jb.00210-22.
- Baranowski C, Welsh MA, Sham L-T, et al. (2018a) Maturing *Mycobacterium smegmatis* peptidoglycan requires non-canonical crosslinks to maintain shape. *eLife* 7. DOI: 10.7554/eLife.37516.

- Baranowski C, Welsh MA, Sham L-T, et al. (2018b) Maturing *Mycobacterium smegmatis* peptidoglycan requires non-canonical crosslinks to maintain shape. *eLife* 7: e37516. DOI: 10.7554/eLife.37516.
- Barreteau H, Kovac A, Boniface A, et al. (2008) Cytoplasmic steps of peptidoglycan biosynthesis. *FEMS microbiology reviews* 32(2): 168–207. DOI: 10.1111/j.1574-6976.2008.00104.x.
- Bateman A and Rawlings ND (2003) The CHAP domain: a large family of amidases including GSP amidase and peptidoglycan hydrolases. *Trends in biochemical sciences* 28(5): 234–237. DOI: 10.1016/S0968-0004(03)00061-6.
- Beaud Benyahia B, Taib N, Beloin C, et al. (2025) Terrabacteria: redefining bacterial envelope diversity, biogenesis and evolution. *Nature Reviews Microbiology* 23(1): 41–56. DOI: 10.1038/s41579-024-01088-0.
- Beeby M, Gumbart JC, Roux B, et al. (2013) Architecture and assembly of the Grampositive cell wall. *Molecular microbiology* 88(4): 664–672. DOI: 10.1111/mmi.12203.
- Bera A, Herbert S, Jakob A, et al. (2005) Why are pathogenic staphylococci so lysozyme resistant? The peptidoglycan O-acetyltransferase OatA is the major determinant for lysozyme resistance of *Staphylococcus aureus*. *Molecular microbiology* 55(3): 778–787. DOI: 10.1111/j.1365-2958.2004.04446.x.
- Bern M, Beniston R and Mesnage S (2017) Towards an automated analysis of bacterial peptidoglycan structure. *Analytical and Bioanalytical Chemistry* 409(2): 551–560. DOI: 10.1007/s00216-016-9857-5.
- Bernal-Cabas M, Ayala JA and Raivio TL (2015) The Cpx envelope stress response modifies peptidoglycan cross-linking via the L,D-transpeptidase LdtD and the novel protein YgaU. *Journal of bacteriology* 197(3): 603–614. DOI: 10.1128/jb.02449-14.
- Bernard E, Rolain T, Courtin P, et al. (2011) Characterization of O-acetylation of Nacetylglucosamine: a novel structural variation of bacterial peptidoglycan. *The Journal of biological chemistry* 286(27): 23950–23958. DOI: 10.1074/jbc.M111.241414.
- Bernhardt TG and Boer PAJ de (2003) The *Escherichia coli* amidase AmiC is a periplasmic septal ring component exported via the twin-arginine transport pathway. *Molecular microbiology* 48(5): 1171–1182. DOI: 10.1046/j.1365-2958.2003.03511.x.
- Beveridge TJ (1978) The response of cell walls of *Bacillus subtilis* to metals and to electron-microscopic stains. *Canadian Journal of Microbiology* 24(2): 89–104. DOI: 10.1139/m78-018.
- Beveridge TJ (1999) Structures of Gram-Negative Cell Walls and Their Derived Membrane Vesicles. *Journal of bacteriology* 181(16): 4725–4733. DOI: 10.1128/JB.181.16.4725-4733.1999.

- Billot-Klein D, Shlaes D, Bryant D, et al. (1996) Peptidoglycan structure of *Enterococcus faecium* expressing vancomycin resistance of the VanB type. *Biochemical Journal* 313 (Pt 3)(Pt 3): 711–715. DOI: 10.1042/bj3130711.
- Blackman SA, Smith TJ and Foster SJ (1998) The role of autolysins during vegetative growth of *Bacillus subtilis* 168. *Microbiology (Reading, England)* 144 (Pt 1): 73–82. DOI: 10.1099/00221287-144-1-73.
- Bollinger KW, Müh U, Ocius KL, et al. (2024) Identification of a family of peptidoglycan transpeptidases reveals that *Clostridioides difficile* requires noncanonical cross-links for viability. *Proceedings of the National Academy of Sciences of the United States of America* 121(34): e2408540121. DOI: 10.1073/pnas.2408540121.
- Boneca IG, Dussurget O, Cabanes D, et al. (2007) A critical role for peptidoglycan N-deacetylation in *Listeria* evasion from the host innate immune system.
 Proceedings of the National Academy of Sciences of the United States of America 104(3): 997–1002. DOI: 10.1073/pnas.0609672104.
- Boneca IG, Xu N, Gage DA, et al. (1997) Structural characterization of an abnormally cross-linked muropeptide dimer that is accumulated in the peptidoglycan of methicillin- and cefotaxime-resistant mutants of *Staphylococcus aureus*. *The Journal of biological chemistry* 272(46): 29053–29059. DOI: 10.1074/jbc.272.46.29053.
- Böth D, Steiner EM, Stadler D, et al. (2013) Structure of LdtMt2, an L,D-transpeptidase from *Mycobacterium tuberculosis*. *Acta crystallographica*. *Section D, Biological crystallography* 69(Pt 3): 432–441. DOI: 10.1107/S0907444912049268.
- Bouhss A, Trunkfield AE, Bugg TDH, et al. (2008) The biosynthesis of peptidoglycan lipid-linked intermediates. *FEMS microbiology reviews* 32(2): 208–233. DOI: 10.1111/j.1574-6976.2007.00089.x.
- Braun V (1975) Covalent lipoprotein from the outer membrane of *Escherichia coli*. *Biochimica et Biophysica Acta (BBA) Reviews on Biomembranes* 415(3): 335–377.
 DOI: 10.1016/0304-4157(75)90013-1.
- Brennan PJ and Nikaido H (1995) The envelope of mycobacteria. *Annual review of biochemistry* 64: 29–63. DOI: 10.1146/annurev.bi.64.070195.000333.
- Brogan AP and Rudner DZ (2023) Regulation of peptidoglycan hydrolases: localization, abundance, and activity. *Current opinion in microbiology* 72: 102279. DOI: 10.1016/j.mib.2023.102279.
- Brown AR, Gordon RA, Hyland SN, et al. (2020) Chemical Biology Tools for Examining the Bacterial Cell Wall. *Cell chemical biology* 27(8): 1052–1062. DOI: 10.1016/j.chembiol.2020.07.024.
- Brown S, Santa Maria JP and Walker S (2013) Wall teichoic acids of gram-positive bacteria. *Annual review of microbiology* 67: 313–336. DOI: 10.1146/annurev-micro-092412-155620.
- Buddle JE and Fagan RP (2023) Pathogenicity and virulence of *Clostridioides difficile*. *Virulence* 14(1): 2150452. DOI: 10.1080/21505594.2022.2150452.
- Bugg TD (1999) 3.10 Bacterial Peptidoglycan Biosynthesis and its Inhibition. In: Barton D, Nakanishi K and Meth-Cohn O (eds) *Comprehensive Natural Products Chemistry:* Oxford: Pergamon, pp. 241–294.
- Bui NK, Gray J, Schwarz H, et al. (2009) The peptidoglycan sacculus of *Myxococcus xanthus* has unusual structural features and is degraded during glycerol-induced myxospore development. *Journal of bacteriology* 191(2): 494–505. DOI: 10.1128/JB.00608-08.
- Camacho C, Coulouris G, Avagyan V, et al. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10(1). DOI: 10.1186/1471-2105-10-421.
- Caparrós M, Pisabarro AG and Pedro MA de (1992) Effect of D-amino acids on structure and synthesis of peptidoglycan in *Escherichia coli*. *Journal of bacteriology* 174(17): 5549–5559.
- Carballido-López R, Formstone A, Li Y, et al. (2006) Actin homolog MreBH governs cell morphogenesis by localization of the cell wall hydrolase LytE. *Developmental cell* 11(3): 399–409. DOI: 10.1016/j.devcel.2006.07.017.
- Cava F, Pedro MA de, Lam H, et al. (2011) Distinct pathways for modification of the bacterial cell wall by non-canonical D-amino acids. *The EMBO journal* 30(16): 3442–3453. DOI: 10.1038/emboj.2011.246.
- Chancock I (2002) Bacterial Cell WallsAn Overview. In: *Molecular Medical Microbiology:* Elsevier, pp. 33–46.
- Chang JD, Wallace AG, Foster EE, et al. (2018) Peptidoglycan Compositional Analysis of *Enterococcus faecalis* Biofilm by Stable Isotope Labeling by Amino Acids in a Bacterial Culture. *Biochemistry* 57(7): 1274–1283. DOI: 10.1021/acs.biochem.7b01207.
- Chapot-Chartier M-P and Kulakauskas S (2014) Cell wall structure and function in lactic acid bacteria. *Microbial cell factories* 13 Suppl 1(Suppl 1): S9. DOI: 10.1186/1475-2859-13-S1-S9.
- Cho H, Uehara T and Bernhardt TG (2014) Beta-lactam antibiotics induce a lethal malfunctioning of the bacterial cell wall synthesis machinery. *Cell* 159(6): 1300–1311. DOI: 10.1016/j.cell.2014.11.017.
- Chodisetti PK and Reddy M (2019) Peptidoglycan hydrolase of an unusual cross-link cleavage specificity contributes to bacterial cell wall synthesis. *Proceedings of the National Academy of Sciences of the United States of America* 116(16): 7825–7830. DOI: 10.1073/pnas.1816893116.
- Christie G and Setlow P (2020) *Bacillus* spore germination: Knowns, unknowns and what we need to learn. *Cellular signalling* 74: 109729. DOI: 10.1016/j.cellsig.2020.109729.
- Chu H and Mazmanian SK (2013) Innate immune recognition of the microbiota promotes host-microbial symbiosis. *Nature immunology* 14(7): 668–675. DOI: 10.1038/ni.2635.

- Claret C, Salmon JM, Romieu C, et al. (1994) Physiology of *Gluconobacter oxydans* during dihydroxyacetone production from glycerol. *Applied Microbiology and Biotechnology* 41(3): 359–365. DOI: 10.1007/BF00221232.
- Clarke AJ (1993) Extent of peptidoglycan O acetylation in the tribe Proteeae. *Journal of bacteriology* 175(14): 4550–4553. DOI: 10.1128/jb.175.14.4550-4553.1993.
- Cochrane SA and Lohans CT (2020) Breaking down the cell wall: Strategies for antibiotic discovery targeting bacterial transpeptidases. *European journal of medicinal chemistry* 194: 112262. DOI: 10.1016/j.ejmech.2020.112262.
- Cohen SA and Schure MR (2008) Multidimensional liquid chromatography: Theory and applications in industrial chemistry and the life sciences. Hoboken N.J.: Wiley-Interscience.
- Cordillot M, Dubée V, Triboulet S, et al. (2013) In vitro cross-linking of *Mycobacterium tuberculosis* peptidoglycan by L,D-transpeptidases and inactivation of these enzymes by carbapenems. *Antimicrobial agents and chemotherapy* 57(12): 5940– 5945. DOI: 10.1128/AAC.01663-13.
- Corrigan RM, Abbott JC, Burhenne H, et al. (2011) c-di-AMP is a new second messenger in *Staphylococcus aureus* with a role in controlling cell size and envelope stress. *PLoS pathogens* 7(9): e1002217. DOI: 10.1371/journal.ppat.1002217.
- Coullon H, Rifflet A, Wheeler R, et al. (2018) N-Deacetylases required for muramic- δ lactam production are involved in *Clostridium difficile* sporulation, germination, and heat resistance. *The Journal of biological chemistry* 293(47): 18040–18054. DOI: 10.1074/jbc.RA118.004273.
- Coullon H, Rifflet A, Wheeler R, et al. (2020) Peptidoglycan analysis reveals that synergistic deacetylase activity in vegetative *Clostridium difficile* impacts the host response. *The Journal of biological chemistry* 295(49): 16785–16796. DOI: 10.1074/jbc.RA119.012442.
- Da Silva GAR, Oliveira SSdS, Lima SF, et al. (2022) The industrial versatility of *Gluconobacter oxydans:* current applications and future perspectives. *World journal of microbiology & biotechnology* 38(8): 134. DOI: 10.1007/s11274-022-03310-8.
- Daitch AK and Goley ED (2020) Uncovering Unappreciated Activities and Niche Functions of Bacterial Cell Wall Enzymes. *Current biology CB* 30(19): R1170-R1175. DOI: 10.1016/j.cub.2020.07.004.
- Dajkovic A, Tesson B, Chauhan S, et al. (2017) Hydrolysis of peptidoglycan is modulated by amidation of meso-diaminopimelic acid and Mg²⁺ in *Bacillus subtilis*. *Molecular microbiology* 104(6): 972–988. DOI: 10.1111/mmi.13673.
- Das D, Hervé M, Elsliger M-A, et al. (2013) Structure and function of a novel LDcarboxypeptidase a involved in peptidoglycan recycling. *Journal of bacteriology* 195(24): 5555–5566. DOI: 10.1128/JB.00900-13.
- de Jonge BL, Chang YS, Xu N, et al. (1996) Effect of exogenous glycine on peptidoglycan composition and resistance in a methicillin-resistant

Staphylococcus aureus strain. Antimicrobial agents and chemotherapy 40(6): 1498–1503.

- Deghorain M, Goffin P, Fontaine L, et al. (2007) Selectivity for D-lactate incorporation into the peptidoglycan precursors of Lactobacillus plantarum: role of Aad, a VanXlike D-alanyl-D-alanine dipeptidase. *Journal of bacteriology* 189(11): 4332–4337. DOI: 10.1128/jb.01829-06.
- Delhaye A, Collet J-F and Laloux G (2019) A Fly on the Wall: How Stress Response Systems Can Sense and Respond to Damage to Peptidoglycan. *Frontiers in cellular and infection microbiology* 9: 380. DOI: 10.3389/fcimb.2019.00380.
- Dembek M, Barquist L, Boinett CJ, et al. (2015) High-throughput analysis of gene essentiality and sporulation in *Clostridium difficile*. *mBio* 6(2): e02383. DOI: 10.1128/mBio.02383-14.
- Deng LL, Alexander AA, Lei S, et al. (2010) The Cell Wall Teichuronic Acid Synthetase (TUAS) Is an Enzyme Complex Located in the Cytoplasmic Membrane of *Micrococcus luteus*. *Biochemistry research international* 2010: 395758. DOI: 10.1155/2010/395758.
- Deppenmeier U, Hoffmeister M and Prust C (2002) Biochemistry and biotechnological applications of *Gluconobacter* strains. *Applied Microbiology and Biotechnology* 60(3): 233–242. DOI: 10.1007/s00253-002-1114-5.
- Desmarais SM, Pedro MA de, Cava F, et al. (2013) Peptidoglycan at its peaks: How chromatographic analyses can reveal bacterial cell wall structure and assembly. *Molecular microbiology* 89(1): 1–13. DOI: 10.1111/mmi.12266.
- Desmarais SM, Tropini C, Miguel A, et al. (2015) High-throughput, Highly Sensitive Analyses of Bacterial Morphogenesis Using Ultra Performance Liquid Chromatography^{*}. *The Journal of biological chemistry* 290(52): 31090–31100. DOI: 10.1074/jbc.M115.661660.
- Desvaux M, Candela T and Serror P (2018a) Surfaceome and Proteosurfaceome in Parietal Monoderm Bacteria: Focus on Protein Cell-Surface Display. *Frontiers in microbiology* 9: 100. DOI: 10.3389/fmicb.2018.00100.
- Desvaux M, Candela T and Serror P (2018b) Surfaceome and Proteosurfaceome in Parietal Monoderm Bacteria: Focus on Protein Cell-Surface Display. *Frontiers in microbiology* 9: 100. DOI: 10.3389/fmicb.2018.00100.
- Dmitriev B, Toukach F and Ehlers S (2005) Towards a comprehensive view of the bacterial cell wall. *Trends in microbiology* 13(12): 569–574. DOI: 10.1016/j.tim.2005.10.001.
- Dmitriev BA, Toukach FV, Schaper K-J, et al. (2003) Tertiary Structure of Bacterial Murein: the Scaffold Model. *Journal of bacteriology* 185(11): 3458–3468. DOI: 10.1128/JB.185.11.3458-3468.2003.
- Donahue EH, Dawson LF, Valiente E, et al. (2014) Clostridium difficile has a single sortase, SrtB, that can be inhibited by small-molecule inhibitors. *BMC microbiology* 14: 219. DOI: 10.1186/s12866-014-0219-1.

- Dörr T, Moynihan PJ and Mayer C (2019) Editorial: Bacterial Cell Wall Structure and Dynamics. *Frontiers in microbiology* 10: 2051. DOI: 10.3389/fmicb.2019.02051.
- Dramsi S, Magnet S, Davison S, et al. (2008) Covalent attachment of proteins to peptidoglycan. *FEMS microbiology reviews* 32(2): 307–320. DOI: 10.1111/j.1574-6976.2008.00102.x.
- Driks A and Eichenberger P (2016) The Spore Coat. *Microbiology Spectrum* 4(2). DOI: 10.1128/microbiolspec.tbs-0023-2016.
- Dubée V, Arthur M, Fief H, et al. (2012) Kinetic analysis of Enterococcus faecium L,Dtranspeptidase inactivation by carbapenems. *Antimicrobial agents and chemotherapy* 56(6): 3409–3412. DOI: 10.1128/AAC.06398-11.
- Ealand CS, Machowski EE and Kana BD (2018) β -lactam resistance: The role of low molecular weight penicillin binding proteins, β -lactamases and Id-transpeptidases in bacteria associated with respiratory tract infections. *IUBMB life* 70(9): 855–868. DOI: 10.1002/iub.1761.
- Egan AJF, Biboy J, van't Veer I, et al. (2015) Activities and regulation of peptidoglycan synthases. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 370(1679). DOI: 10.1098/rstb.2015.0031.
- Egan AJF, Errington J and Vollmer W (2020) Regulation of peptidoglycan synthesis and remodelling. *Nature reviews. Microbiology* 18(8): 446–460. DOI: 10.1038/s41579-020-0366-3.
- Egan AJF, Jean NL, Koumoutsi A, et al. (2014) Outer-membrane lipoprotein LpoB spans the periplasm to stimulate the peptidoglycan synthase PBP1B. *Proceedings of the National Academy of Sciences of the United States of America* 111(22): 8197–8202. DOI: 10.1073/pnas.1400376111.
- Ehrt C, Krause B, Schmidt R, et al. (2020) SMARTS.plus A Toolbox for Chemical Pattern Design. *Molecular informatics* 39(12): e2000216. DOI: 10.1002/minf.202000216.
- Eliuk S and Makarov A (2015) Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annual review of analytical chemistry (Palo Alto, Calif.)* 8: 61–80. DOI: 10.1146/annurev-anchem-071114-040325.
- Espaillat A, Alvarez L, Torrens G, et al. (2024) A distinctive family of L,Dtranspeptidases catalyzing L-Ala-mDAP crosslinks in Alpha- and Betaproteobacteria. *Nature communications* 15(1): 1343. DOI: 10.1038/s41467-024-45620-5.
- Espaillat A, Forsmo O, El Biari K, et al. (2016) Chemometric Analysis of Bacterial Peptidoglycan Reveals Atypical Modifications That Empower the Cell Wall against Predatory Enzymes and Fly Innate Immunity. *Journal of the American Chemical Society* 138(29): 9193–9204. DOI: 10.1021/jacs.6b04430.
- Fibriansah G, Gliubich FI and Thunnissen A-MWH (2012) On the mechanism of peptidoglycan binding and cleavage by the endo-specific lytic transglycosylase

MItE from Escherichia coli. *Biochemistry* 51(45): 9164–9177. DOI: 10.1021/bi300900t.

- Filipe SR and Tomasz A (2000) Inhibition of the expression of penicillin resistance in Streptococcus pneumoniae by inactivation of cell wall muropeptide branching genes. Proceedings of the National Academy of Sciences of the United States of America 97(9): 4891–4896. DOI: 10.1073/pnas.080067697.
- Foster SJ (1992) Analysis of the autolysins of *Bacillus subtilis* 168 during vegetative growth and differentiation by using renaturing polyacrylamide gel electrophoresis. *Journal of bacteriology* 174(2): 464–470.
- Francis MB and Sorg JA (2016) Detecting Cortex Fragments During Bacterial Spore Germination. *Journal of visualized experiments JoVE*. Epub ahead of print 25 June 2016. DOI: 10.3791/54146.
- Francis ZK, Sanders AN, Crick DC, et al. (2023) LdtC Is a Key I,d-Transpeptidase for Peptidoglycan Assembly in *Mycobacterium smegmatis*. *Journal of bacteriology* 205(1): e0042422. DOI: 10.1128/jb.00424-22.
- Fredriksen A, Vasstrand EN and Jensen HB (1991) Peptidoglycan precursor from *Fusobacterium nucleatum* contains lanthionine. *Journal of bacteriology* 173(2):
 900–902. DOI: 10.1128/jb.173.2.900-902.1991.
- Frère JM (1977) Mechanism of action of beta-lactam antibiotics at the molecular level. Biochemical pharmacology 26(23): 2203–2210. DOI: 10.1016/0006-2952(77)90280-5.
- Fricke PM, Lürkens M, Hünnefeld M, et al. (2021) Highly tunable TetR-dependent target gene expression in the acetic acid bacterium *Gluconobacter oxydans*. *Applied microbiology and biotechnology* 105(18): 6835–6852. DOI: 10.1007/s00253-021-11473-x.
- Frirdich E and Gaynor EC (2013) Peptidoglycan hydrolases, bacterial shape, and pathogenesis. *Current opinion in microbiology* 16(6): 767–778. DOI: 10.1016/j.mib.2013.09.005.
- Fukushima T, Afkham A, Kurosawa S-I, et al. (2006) A new D,L-endopeptidase gene product, YojL (renamed CwlS), plays a role in cell separation with LytE and LytF in *Bacillus subtilis. Journal of bacteriology* 188(15): 5541–5550. DOI: 10.1128/jb.00188-06.
- Fukushima T, Yao Y, Kitajima T, et al. (2007) Characterization of new L,Dendopeptidase gene product CwlK (previous YcdD) that hydrolyzes peptidoglycan in *Bacillus subtilis*. *Molecular genetics and genomics MGG* 278(4): 371–383. DOI: 10.1007/s00438-007-0255-8.
- Galley NF, Greetham D, Alamán-Zárate MG, et al. (2024) *Clostridioides difficile* canonical L,D-transpeptidases catalyze a novel type of peptidoglycan cross-links and are not required for beta-lactam resistance. *The Journal of biological chemistry* 300(1): 105529. DOI: 10.1016/j.jbc.2023.105529.

- Gamba P, Hamoen LW and Daniel RA (2016) Cooperative Recruitment of FtsW to the Division Site of *Bacillus subtilis*. *Frontiers in microbiology* 7: 1808. DOI: 10.3389/fmicb.2016.01808.
- Gan L, Chen S and Jensen GJ (2008) Molecular organization of Gram-negative peptidoglycan. *Proceedings of the National Academy of Sciences* 105(48): 18953–18957. DOI: 10.1073/pnas.0808035105.
- Garde S, Chodisetti PK, Reddy M, et al. (2021) Peptidoglycan: Structure, Synthesis, and Regulation. *EcoSal Plus* 9(2). DOI: 10.1128/ecosalplus.ESP-0010-2020.
- Garner EC (2021) Toward a Mechanistic Understanding of Bacterial Rod Shape Formation and Regulation. *Annual review of cell and developmental biology* 37: 1– 21. DOI: 10.1146/annurev-cellbio-010521-010834.
- Ghosh AS, Chowdhury C and Nelson DE (2008) Physiological functions of D-alanine carboxypeptidases in *Escherichia coli*. *Trends in microbiology* 16(7): 309–317. DOI: 10.1016/j.tim.2008.04.006.
- Ghuysen JM (1968) Use of bacteriolytic enzymes in determination of wall structure and their role in cell metabolism. *Bacteriological reviews* 32(4 Pt 2): 425–464.
- Giesbrecht P, Kersten T, Maidhof H, et al. (1998) *Staphylococcal* cell wall: morphogenesis and fatal variations in the presence of penicillin. *Microbiology and molecular biology reviews: MMBR* 62(4): 1371–1414. DOI: 10.1128/MMBR.62.4.1371-1414.1998.
- Gilmore MC and Cava F (2025) Bacterial peptidoglycan recycling. *Trends in microbiology* 33(3): 340–353. DOI: 10.1016/j.tim.2024.11.004.
- Gilmore MC, Yadav AK, Espaillat A, et al. (2024) A peptidoglycan N-deacetylase specific for anhydroMurNAc chain termini in *Agrobacterium tumefaciens*. *The Journal of biological chemistry* 300(2): 105611. DOI: 10.1016/j.jbc.2023.105611.
- Gilmore ME, Bandyopadhyay D, Dean AM, et al. (2004) Production of muramic deltalactam in *Bacillus subtilis* spore peptidoglycan. *Journal of bacteriology* 186(1): 80– 89. DOI: 10.1128/jb.186.1.80-89.2004.
- Glauner B (1988) Separation and quantification of muropeptides with highperformance liquid chromatography. *Analytical biochemistry* 172(2): 451–464. DOI: 10.1016/0003-2697(88)90468-X.
- Glauner B, Höltje JV and Schwarz U (1988) The composition of the murein of *Escherichia coli*. *The Journal of biological chemistry* 263(21): 10088–10095.
- Godessart P, Lannoy A, Dieu M, et al. (2021) β -Barrels covalently link peptidoglycan and the outer membrane in the α -proteobacterium *Brucella abortus*. *Nature microbiology* 6(1): 27–33. DOI: 10.1038/s41564-020-00799-3.
- Gomes RJ, Borges MdF, Rosa MdF, et al. (2018) Acetic Acid Bacteria in the Food Industry: Systematics, Characteristics and Applications. *Food technology and biotechnology* 56(2): 139–151. DOI: 10.17113/ftb.56.02.18.5593.

- Gondré B, Flouret B and van Heijenoort J (1973) Release of D-alanyl-D-alanine from the precursor of the cell wall peptidoglycan by a peptidase of *Escherichia coli* K 12. *Biochimie* 55(6): 685–691. DOI: 10.1016/S0300-9084(73)80022-7.
- Gonzalez-Santana A and Diaz Heijtz R (2020) Bacterial Peptidoglycans from Microbiota in Neurodevelopment and Behavior. *Trends in molecular medicine* 26(8): 729–743. DOI: 10.1016/j.molmed.2020.05.003.
- Grasekamp KP, Beaud Benyahia B, Taib N, et al. (2023) The MIa system of diderm Firmicute *Veillonella parvula* reveals an ancestral transenvelope bridge for phospholipid trafficking. *Nature communications* 14(1): 7642. DOI: 10.1038/s41467-023-43411-y.
- Gupta R, Lavollay M, Mainardi J-L, et al. (2010) The *Mycobacterium tuberculosis* protein LdtMt2 is a nonclassical transpeptidase required for virulence and resistance to amoxicillin. *Nature medicine* 16(4): 466–469. DOI: 10.1038/nm.2120.
- Gupta RS (1998) Life's third domain (Archaea): an established fact or an endangered paradigm? *Theoretical population biology* 54(2): 91–104. DOI: 10.1006/tpbi.1998.1376.
- Hahn FE and Ciak J (1957) Penicillin-Induced Lysis of *Escherichia coli*. *Science* 125(3238): 119–120. DOI: 10.1126/science.125.3238.119.
- Handwerger S, Pucci MJ, Volk KJ, et al. (1994) Vancomycin-resistant *Leuconostoc mesenteroides* and *Lactobacillus casei* synthesize cytoplasmic peptidoglycan precursors that terminate in lactate. *Journal of bacteriology* 176(1): 260–264. DOI: 10.1128/jb.176.1.260-264.1994.
- Harper CE and Hernandez CJ (2020) Cell biomechanics and mechanobiology in bacteria: Challenges and opportunities. *APL bioengineering* 4(2): 21501. DOI: 10.1063/1.5135585.
- Harz H, Burgdorf K and Höltje JV (1990) Isolation and separation of the glycan strands from murein of Escherichia coli by reversed-phase high-performance liquid chromatography. *Analytical biochemistry* 190(1): 120–128. DOI: 10.1016/0003-2697(90)90144-x.
- Hashimi A and Tocheva El (2024) Cell envelope diversity and evolution across the bacterial tree of life. *Nature microbiology* 9(10): 2475–2487. DOI: 10.1038/s41564-024-01812-9.
- Hayhurst EJ, Kailas L, Hobbs JK, et al. (2008) Cell wall peptidoglycan architecture in Bacillus subtilis. Proceedings of the National Academy of Sciences of the United States of America 105(38): 14603–14608. DOI: 10.1073/pnas.0804138105.
- Healy C, Gouzy A and Ehrt S (2020) Peptidoglycan Hydrolases RipA and Ami1 Are Critical for Replication and Persistence of *Mycobacterium tuberculosis* in the Host. *mBio* 11(2). DOI: 10.1128/mBio.03315-19.
- Heidrich C, Templin MF, Ursinus A, et al. (2001) Involvement of N-acetylmuramyl-Lalanine amidases in cell separation and antibiotic-induced autolysis of *Escherichia coli*. *Molecular microbiology* 41(1): 167–178. DOI: 10.1046/j.1365-2958.2001.02499.x.

- Herbert S, Bera A, Nerz C, et al. (2007) Molecular basis of resistance to muramidase and cationic antimicrobial peptide activity of lysozyme in staphylococci. *PLoS pathogens* 3(7): e102. DOI: 10.1371/journal.ppat.0030102.
- Herlihey FA, Moynihan PJ and Clarke AJ (2014) The essential protein for bacterial flagella formation FlgJ functions as a β -N-acetylglucosaminidase. *The Journal of biological chemistry* 289(45): 31029–31042. DOI: 10.1074/jbc.M114.603944.
- Hernández SB, Castanheira S, Pucciarelli MG, et al. (2022) Peptidoglycan editing in non-proliferating intracellular *Salmonella* as source of interference with immune signaling. *PLoS pathogens* 18(1): e1010241. DOI: 10.1371/journal.ppat.1010241.
- Hernández SB and Cava F (2021) New approaches and techniques for bacterial cell wall analysis. *Current opinion in microbiology* 60: 88–95. DOI: 10.1016/j.mib.2021.01.010.
- Hernandez-Rocamora V, Martorana AM, Belloso A, et al. (2025) A novel peptidoglycan deacetylase modulates daughter cell separation in *E. coli. bioRxiv the preprint server for biology*. Epub ahead of print 19 February 2025. DOI: 10.1101/2025.02.18.638797.
- Hett EC and Rubin EJ (2008) Bacterial growth and cell division: a mycobacterial perspective. *Microbiology and molecular biology reviews MMBR* 72(1): 126-56. DOI: 10.1128/MMBR.00028-07.
- Heydenreich R, Nacita J, Lin C-W, et al. (2025) *Revisiting bacterial spore germination in the presence of peptidoglycan fragments (Accepted for publishing)*.
- Hirota Y, Suzuki H, Nishimura Y, et al. (1977) On the process of cellular division in Escherichia coli: A mutant of E. coli lacking a murein-lipoprotein. Proceedings of the National Academy of Sciences of the United States of America 74(4): 1417– 1420. DOI: 10.1073/pnas.74.4.1417.
- Ho CS, Wong CTH, Aung TT, et al. (2025) Antimicrobial resistance: a concise update. *The Lancet. Microbe* 6(1): 100947. DOI: 10.1016/j.lanmic.2024.07.010.
- Höltje JV (1998) Growth of the stress-bearing and shape-maintaining murein sacculus of *Escherichia coli*. *Microbiology and molecular biology reviews: MMBR* 62(1): 181–203. DOI: 10.1128/MMBR.62.1.181-203.1998.
- Höltje JV, Mirelman D, Sharon N, et al. (1975) Novel type of murein transglycosylase in *Escherichia coli. Journal of bacteriology* 124(3): 1067–1076. DOI: 10.1128/jb.124.3.1067-1076.1975.
- Hommel RK (2014) *Gluconobacter*. In: *Encyclopedia of Food Microbiology:* Elsevier, pp. 99–105.
- Hsu Y-C, Su P-R, Huang L-J, et al. (2023) High-throughput Automated Muropeptide Analysis (HAMA) Reveals Peptidoglycan Composition of Gut Microbial Cell Walls. *eLife*. DOI: 10.7554/eLife.88491.2.
- Huang Z, Wang J, Xu X, et al. (2019) Antibody neutralization of microbiota-derived circulating peptidoglycan dampens inflammation and ameliorates autoimmunity. *Nature microbiology* 4(5): 766–773. DOI: 10.1038/s41564-019-0381-1.

- Humphrey BA and Vincent JM (1965) The effect of calcium nutrition on the production of diffusible antigens by *Rhizobium trifolii*. *Journal of general microbiology* 41(1): 109–118. DOI: 10.1099/00221287-41-1-109.
- Hwang H, Paracini N, Parks JM, et al. (2018) Distribution of mechanical stress in the *Escherichia coli* cell envelope. *Biochimica et biophysica acta. Biomembranes* 1860(12): 2566–2575. DOI: 10.1016/j.bbamem.2018.09.020.
- Jacobs C, Frère JM and Normark S (1997) Cytosolic intermediates for cell wall biosynthesis and degradation control inducible beta-lactam resistance in gramnegative bacteria. *Cell* 88(6): 823–832. DOI: 10.1016/s0092-8674(00)81928-5.
- Jeon W-J and Cho H (2022) A Cell Wall Hydrolase MepH Is Negatively Regulated by Proteolysis Involving Prc and NIpI in Escherichia coli. *Frontiers in microbiology* 13: 878049. DOI: 10.3389/fmicb.2022.878049.
- Jeske O, Schüler M, Schumann P, et al. (2015) *Planctomycetes* do possess a peptidoglycan cell wall. *Nature communications* 6: 7116. DOI: 10.1038/ncomms8116.
- Johanesen PA, Mackin KE, Hutton ML, et al. (2015) Disruption of the Gut Microbiome: *Clostridium difficile* Infection and the Threat of Antibiotic Resistance. *Genes* 6(4): 1347–1360. DOI: 10.3390/genes6041347.
- Johannsen L (1993) Biological properties of bacterial peptidoglycan. *APMIS* 101(1-6): 337–344. DOI: 10.1111/j.1699-0463.1993.tb00119.x.
- Johnson JW, Fisher JF and Mobashery S (2013) Bacterial cell-wall recycling. *Annals of the New York Academy of Sciences* 1277(1): 54–75. DOI: 10.1111/j.1749-6632.2012.06813.x.
- Jonge BL de, Sidow T, Chang YS, et al. (1993) Altered muropeptide composition in *Staphylococcus aureus* strains with an inactivated femA locus. *Journal of bacteriology* 175(9): 2779–2782. DOI: 10.1128/jb.175.9.2779-2782.1993.
- Jumper J, Evans R, Pritzel A, et al. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596(7873): 583–589. DOI: 10.1038/s41586-021-03819-2.
- Jun D, Idem U and Dahms TES (2020) Altered Envelope Structure and Nanomechanical Properties of a C-Terminal Protease A-Deficient *Rhizobium leguminosarum*. *Microorganisms* 8(9). DOI: 10.3390/microorganisms8091421.
- Jupyter P, Bussonnier M, Forde J, et al. (2018) Binder 2.0 Reproducible, interactive, sharable environments for science at scale. In: *Proceedings of the 17th Python in Science Conference:* Austin, Texas, July 9 - July 15 2018, pp. 113–120. SciPy.
- Kampff Z, van Sinderen D and Mahony J (2023) Cell wall polysaccharides of streptococci: A genetic and structural perspective. *Biotechnology advances* 69: 108279. DOI: 10.1016/j.biotechadv.2023.108279.
- Kandler O and König H (1978) Chemical composition of the peptidoglycan-free cell walls of methanogenic bacteria. *Archives of microbiology* 118(2): 141–152. DOI: 10.1007/BF00415722.
- Kathayat D, Huang Y, Denis J, et al. (2025) LD-transpeptidase-mediated cell envelope remodeling enables developmental transitions and survival in *Coxiella burnetii* and

Legionella pneumophila. Journal of bacteriology: e00247 - 24. DOI: 10.1128/jb.00247-24.

- Kaus GM, Snyder LF, Müh U, et al. (2020) Lysozyme Resistance in *Clostridioides difficile* Is Dependent on Two Peptidoglycan Deacetylases. *Journal of bacteriology* 202(22). DOI: 10.1128/jb.00421-20.
- Kawai Y and Errington J (2023) Dissecting the roles of peptidoglycan synthetic and autolytic activities in the walled to L-form bacterial transition. *Frontiers in microbiology* 14: 1204979. DOI: 10.3389/fmicb.2023.1204979.
- Keck W, van Leeuwen AM, Huber M, et al. (1990) Cloning and characterization of mepA, the structural gene of the penicillin-insensitive murein endopeptidase from *Escherichia coli*. *Molecular microbiology* 4(2): 209–219. DOI: 10.1111/j.1365-2958.1990.tb00588.x.
- Kieser KJ and Rubin EJ (2014) How sisters grow apart: mycobacterial growth and division. *Nature Reviews Microbiology* 12(8): 550–562. DOI: 10.1038/nrmicro3299.
- Kim C, Lee M, Birhanu BT, et al. (2023) Synthesis of Muramyl- δ -Lactam in Spore Peptidoglycan of *Clostridioides difficile*. *Chembiochem a European journal of chemical biology* 24(11): e202300282. DOI: 10.1002/cbic.202300282.
- Kim HS, Im HN, An DR, et al. (2015a) The Cell Shape-determining Csd6 Protein from *Helicobacter pylori* Constitutes a New Family of L,D-Carboxypeptidase. *The Journal of biological chemistry* 290(41): 25103–25117. DOI: 10.1074/jbc.M115.658781.
- Kim SJ, Chang J and Singh M (2015b) Peptidoglycan architecture of Gram-positive bacteria by solid-state NMR. *Biochimica et biophysica acta* 1848(1 Pt B): 350–362. DOI: 10.1016/j.bbamem.2014.05.031.
- Kim SJ, Singh M, Preobrazhenskaya M, et al. (2013) *Staphylococcus aureus* Peptidoglycan Stem Packing by Rotational-Echo Double Resonance NMR Spectroscopy. *Biochemistry* 52(21): 3651–3659. DOI: 10.1021/bi4005039.
- Kirk JA, Banerji O and Fagan RP (2017) Characteristics of the *Clostridium difficile* cell envelope and its importance in therapeutics. *Microbial biotechnology* 10(1): 76– 90. DOI: 10.1111/1751-7915.12372.
- Koch AL (1998) The three-for-one model for gram-negative wall growth: a problem and a possible solution. *FEMS microbiology letters* 162(1): 127–134. DOI: 10.1111/j.1574-6968.1998.tb12989.x.
- Kosolapova AO, Belousov MV, Sulatskaya AI, et al. (2019) Two Novel Amyloid Proteins, RopA and RopB, from the Root Nodule Bacterium *Rhizobium leguminosarum*. *Biomolecules* 9(11): 694. DOI: 10.3390/biom9110694.
- Kosslak RM and Bohlool BB (1984) Suppression of nodule development of one side of a split-root system of soybeans caused by prior inoculation of the other side. *Plant physiology* 75(1): 125–130. DOI: 10.1104/pp.75.1.125.
- Kruijff B de, van Dam V and Breukink E (2008) Lipid II: A central component in bacterial cell wall synthesis and a target for antibiotics. *Prostaglandins*,

leukotrienes, and essential fatty acids 79(3-5): 117–121. DOI: 10.1016/j.plefa.2008.09.020.

- Kühner D, Stahl M, Demircioglu DD, et al. (2014) From cells to muropeptide structures in 24 h: peptidoglycan mapping by UPLC-MS. *Scientific reports* 4: 7494. DOI: 10.1038/srep07494.
- Kumar K and Cava F (2019) Chromatographic analysis of peptidoglycan samples with the aid of a chemometric technique: introducing a novel analytical procedure to classify bacterial cell wall collection. *Analytical Methods* 11(12): 1671–1679. DOI: 10.1039/c8ay02501k.
- Kumar K, Espaillat A and Cava F (2017) PG-Metrics: A chemometric-based approach for classifying bacterial peptidoglycan data sets and uncovering their subjacent chemical variability. *PloS one* 12(10): e0186197. DOI: 10.1371/JOURNAL.PONE.0186197.
- Kuru E, Radkov A, Meng X, et al. (2019) Mechanisms of Incorporation for D-Amino Acid Probes That Target Peptidoglycan Biosynthesis. *ACS chemical biology* 14(12): 2745–2756. DOI: 10.1021/acschembio.9b00664.
- Kuzmin E, Taylor JS and Boone C (2022) Retention of duplicated genes in evolution. *Trends in Genetics* 38(1): 59–72. DOI: 10.1016/j.tig.2021.06.016.
- Kwan JMC, Liang Y, Ng EWL, et al. (2024) In silico MS/MS prediction for peptidoglycan profiling uncovers novel anti-inflammatory peptidoglycan fragments of the gut microbiota. *Chemical science* 15(5): 1846–1859. DOI: 10.1039/d3sc05819k.
- Lanéelle M-A, Tropis M and Daffé M (2013) Current knowledge on mycolic acids in *Corynebacterium glutamicum* and their relevance for biotechnological processes. *Applied Microbiology and Biotechnology* 97(23): 9923–9930. DOI: 10.1007/s00253-013-5265-3.
- Larson TR and Yother J (2017) *Streptococcus pneumoniae* capsular polysaccharide is linked to peptidoglycan via a direct glycosidic bond to β-D-N-acetylglucosamine. *Proceedings of the National Academy of Sciences of the United States of America* 114(22): 5695–5700. DOI: 10.1073/pnas.1620431114.
- Lavollay M, Arthur M, Fourgeaud M, et al. (2008) The peptidoglycan of stationaryphase *Mycobacterium tuberculosis* predominantly contains cross-links generated by L,D-transpeptidation. *Journal of bacteriology* 190(12): 4360–4366. DOI: 10.1128/JB.00239-08.
- Lawson PA, Citron DM, Tyrrell KL, et al. (2016) Reclassification of *Clostridium difficile* as *Clostridioides difficile* (Hall and O'Toole 1935) Prévot 1938. *Anaerobe* 40: 95–99. DOI: 10.1016/j.anaerobe.2016.06.008.
- Lecoq L, Dubée V, Triboulet S, et al. (2013) Structure of Enterococcus faeciuml,dtranspeptidase acylated by ertapenem provides insight into the inactivation mechanism. *ACS chemical biology* 8(6): 1140–1146. DOI: 10.1021/cb4001603.
- Lederer E, Adam A, Ciorbaru R, et al. (1975) Cell walls of *Mycobacteria* and related organisms; chemistry and immunostimulant properties. *Molecular and cellular biochemistry* 7(2): 87–104. DOI: 10.1007/BF01792076.

- Lee C-R, Son JE, Park SH, et al. (2024) Lytic transglycosylase repertoire diversity enables daughter cell separation and antibiotic resistance in *Escherichia coli* under acidic stress. *bioRxiv the preprint server for biology*. DOI: 10.1101/2024.02.16.580723.
- Leisico F, V Vieira D, Figueiredo TA, et al. (2018) First insights of peptidoglycan amidation in Gram-positive bacteria - the high-resolution crystal structure of *Staphylococcus aureus* glutamine amidotransferase GatD. *Scientific reports* 8(1): 5313. DOI: 10.1038/s41598-018-22986-3.
- Léonard RR, Sauvage E, Lupo V, et al. (2022) Was the Last Bacterial Common Ancestor a Monoderm after All? *Genes* 13(2): 376. DOI: 10.3390/genes13020376.
- Lessa FC, Gould CV and McDonald LC (2012) Current status of *Clostridium difficile* infection epidemiology. *Clinical infectious diseases an official publication of the Infectious Diseases Society of America* 55 Suppl 2(Suppl 2): S65-70. DOI: 10.1093/cid/cis319.
- Li Z, Jiang B, Zhang X, et al. (2020) The role of bacterial cell envelope structures in acid stress resistance in *E. coli*. *Applied microbiology and biotechnology* 104(7): 2911–2921. DOI: 10.1007/s00253-020-10453-x.
- Lithgow T, Stubenrauch CJ and Stumpf MPH (2023) Surveying membrane landscapes: a new look at the bacterial cell surface. *Nature reviews. Microbiology* 21(8): 502–518. DOI: 10.1038/s41579-023-00862-w.
- Liu L, Chen Y, Yu S, et al. (2021) Simultaneous transformation of five vectors in *Gluconobacter oxydans*. *Plasmid* 117: 102588. DOI: 10.1016/j.plasmid.2021.102588.
- Lovering AL, Safadi SS and Strynadka NCJ (2012) Structural perspective of peptidoglycan biosynthesis and assembly. *Annual review of biochemistry* 81: 451–478. DOI: 10.1146/annurev-biochem-061809-112742.
- Lupoli TJ, Lebar MD, Markovski M, et al. (2014) Lipoprotein activators stimulate Escherichia coli penicillin-binding proteins by different mechanisms. Journal of the American Chemical Society 136(1): 52–55. DOI: 10.1021/ja410813j.
- Ma Y, Li B, Zhang X, et al. (2022) Production of Gluconic Acid and Its Derivatives by Microbial Fermentation: Process Improvement Based on Integrated Routes. *Frontiers in bioengineering and biotechnology* 10: 864787. DOI: 10.3389/fbioe.2022.864787.
- Machowski EE, Senzani S, Ealand C, et al. (2014) Comparative genomics for mycobacterial peptidoglycan remodelling enzymes reveals extensive genetic multiplicity. *BMC microbiology* 14: 75. DOI: 10.1186/1471-2180-14-75.
- Mada PK and Alam MU (2025) *Clostridioides difficile* infection. Available at: https://www.ncbi.nlm.nih.gov/books/NBK431054/ (accessed 28 March 2025).
- Magnet S, Bellais S, Dubost L, et al. (2007) Identification of the L,D-transpeptidases responsible for attachment of the Braun lipoprotein to *Escherichia coli* peptidoglycan. *Journal of bacteriology* 189(10): 3927–3931. DOI: 10.1074/jbc.M610911200.

- Mahapatra S, Crick DC, McNeil MR, et al. (2008) Unique structural features of the peptidoglycan of *Mycobacterium leprae*. *Journal of bacteriology* 190(2): 655–661. DOI: 10.1128/JB.00982-07.
- Mainardi JL, Legrand R, Arthur M, et al. (2000) Novel mechanism of beta-lactam resistance due to bypass of DD-transpeptidation in *Enterococcus faecium*. *The Journal of biological chemistry* 275(22): 16490–16496. DOI: 10.1074/jbc.M909877199.
- Mainardi J-L, Fourgeaud M, Hugonnet J-E, et al. (2005a) A novel peptidoglycan crosslinking enzyme for a beta-lactam-resistant transpeptidation pathway. *The Journal of biological chemistry* 280(46): 38146–38152. DOI: 10.1074/jbc.M507384200.
- Mainardi J-L, Fourgeaud M, Hugonnet J-E, et al. (2005b) A novel peptidoglycan crosslinking enzyme for a beta-lactam-resistant transpeptidation pathway. *The Journal of biological chemistry* 280(46): 38146–38152. DOI: 10.1074/jbc.M507384200.
- Makarov A (2000) Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Analytical chemistry* 72(6): 1156–1162. DOI: 10.1021/ac991131p.
- Maróti G and Kondorosi E (2014) Nitrogen-fixing *Rhizobium*-legume symbiosis: are polyploidy and host peptide-governed symbiont differentiation general principles of endosymbiosis? *Frontiers in microbiology* 5: 326. DOI: 10.3389/fmicb.2014.00326.
- Mathelié-Guinlet M, Asmar AT, Collet J-F, et al. (2020) Bacterial Cell Mechanics Beyond Peptidoglycan. *Trends in microbiology* 28(9): 706–708. DOI: 10.1016/j.tim.2020.04.013.
- Maya-Martinez R, Alexander JAN, Otten CF, et al. (2018) Recognition of Peptidoglycan Fragments by the Transpeptidase PBP4 From *Staphylococcus aureus*. *Frontiers in microbiology* 9: 3223. DOI: 10.3389/fmicb.2018.03223.
- Meeske AJ, Sham L-T, Kimsey H, et al. (2015) MurJ and a novel lipid II flippase are required for cell wall biogenesis in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences of the United States of America* 112(20): 6437–6442. DOI: 10.1073/pnas.1504967112.
- Merfort M, Herrmann U, Ha S-W, et al. (2006) Modification of the membrane-bound glucose oxidation system in *Gluconobacter oxydans* significantly increases gluconate and 5-keto-D-gluconic acid accumulation. *Biotechnology journal* 1(5): 556–563. DOI: 10.1002/biot.200600032.
- Meroueh SO, Bencze KZ, Hesek D, et al. (2006) Three-dimensional structure of the bacterial cell wall peptidoglycan. *Proceedings of the National Academy of Sciences* 103(12): 4404–4409. DOI: 10.1073/pnas.0510182103.
- Miller SI and Salama NR (2018) The gram-negative bacterial periplasm: Size matters. *PLoS biology* 16(1): e2004935. DOI: 10.1371/journal.pbio.2004935.

- Mistou M-Y, Sutcliffe IC and van Sorge NM (2016) Bacterial glycobiology: rhamnosecontaining cell wall polysaccharides in Gram-positive bacteria. *FEMS microbiology reviews* 40(4): 464–479. DOI: 10.1093/femsre/fuw006.
- Mistry J, Chuguransky S, Williams L, et al. (2021) Pfam: The protein families database in 2021. *Nucleic acids research* 49(D1): D412-D419. DOI: 10.1093/nar/gkaa913.
- Miyachiro MM, Contreras-Martel C and Dessen A (2019) Penicillin-Binding Proteins (PBPs) and Bacterial Cell Wall Elongation Complexes. *Sub-cellular biochemistry* 93: 273–289. DOI: 10.1007/978-3-030-28151-9_8.
- Morè N, Martorana AM, Biboy J, et al. (2019) Peptidoglycan Remodeling Enables *Escherichia coli* To Survive Severe Outer Membrane Assembly Defect. *mBio* 10(1). DOI: 10.1128/mBio.02729-18.
- Mukhopadhyay B (2024) A reminder-peptidoglycan cell walls indeed occur in the archaeal domain, specifically in the members of *Methanobacteria* and *Methanopyri* classes. *Frontiers in microbiology* 15: 1329047. DOI: 10.3389/fmicb.2024.1329047.
- Münch D, Roemer T, Lee SH, et al. (2012) Identification and in vitro analysis of the GatD/MurT enzyme-complex catalyzing lipid II amidation in *Staphylococcus aureus*. *PLoS pathogens* 8(1): e1002509. DOI: 10.1371/journal.ppat.1002509.
- Muynck C de, Pereira CSS, Naessens M, et al. (2007) The genus *Gluconobacter oxydans*: comprehensive overview of biochemistry and biotechnological applications. *Critical reviews in biotechnology* 27(3): 147–171. DOI: 10.1080/07388550701503584.
- Naamala J, Jaiswal SK and Dakora FD (2016) Antibiotics Resistance in *Rhizobium*: Type, Process, Mechanism and Benefit for Agriculture. *Current microbiology* 72(6): 804–816. DOI: 10.1007/s00284-016-1005-0.
- Nakano MM and Hulett FM (1997) Adaptation of *Bacillus subtilis* to oxygen limitation. *FEMS microbiology letters* 157(1): 1–7. DOI: 10.1111/j.1574-6968.1997.tb12744.x.
- Navarre WW and Schneewind O (1999) Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiology and molecular biology reviews: MMBR* 63(1): 174–229. DOI: 10.1128/MMBR.63.1.174-229.1999.
- Nicolai X, Liang Y, Ruaudel F, et al. (2024) Inner membrane protein OutB is covalently attached to peptidoglycan in the γ -proteobacterium Dickeya dadantii.
- Nicoud Q, Barrière Q, Busset N, et al. (2021) *Sinorhizobium meliloti* Functions Required for Resistance to Antimicrobial NCR Peptides and Bacteroid Differentiation. *mBio* 12(4): e0089521. DOI: 10.1128/mBio.00895-21.
- Nyholm SV and McFall-Ngai MJ (2021) A lasting symbiosis: how the Hawaiian bobtail squid finds and keeps its bioluminescent bacterial partner. *Nature Reviews Microbiology* 19(10): 666–679. DOI: 10.1038/s41579-021-00567-y.

- Oldroyd GED and Downie JA (2008) Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annual review of plant biology* 59: 519–546. DOI: 10.1146/annurev.arplant.59.032607.092839.
- Olson RD, Assaf R, Brettin T, et al. (2023) Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. *Nucleic acids research* 51(D1): D678-D689. DOI: 10.1093/nar/gkac1003.
- Orsini Delgado ML, Gamelas Magalhaes J, Morra R, et al. (2024) Muropeptides and muropeptide transporters impact on host immune response. *Gut Microbes* 16(1): 2418412. DOI: 10.1080/19490976.2024.2418412.
- Otten C, Brilli M, Vollmer W, et al. (2018) Peptidoglycan in obligate intracellular bacteria. *Molecular microbiology* 107(2): 142–163. DOI: 10.1111/mmi.13880.
- Paiva AMO, Courtin P, Charpentier G, et al. (2024) The L,D-transpeptidation pathway is inhibited by antibiotics of the β -lactam class in Clostridioides difficile.
- Pan H and Wang D (2017) Nodule cysteine-rich peptides maintain a working balance during nitrogen-fixing symbiosis. *Nature plants* 3(5): 17048. DOI: 10.1038/nplants.2017.48.
- Park JT and Strominger JL (1957) Mode of action of penicillin. *Science* 125(3238): 99–101. DOI: 10.1126/science.125.3238.99.
- Pasquina-Lemonche L, Burns J, Turner RD, et al. (2020) The architecture of the Grampositive bacterial cell wall. *Nature* 582(7811): 294–297. DOI: 10.1038/s41586-020-2236-6.
- Patel AV (2021) Development of an automated, high-resolution analysis pipeline for bacterial peptidoglycan structural analyes. PhD thesis, University of Sheffield. Faculty of Sciences.
- Patel AV, Turner RD, Rifflet A, et al. (2021) PGFinder, a novel analysis pipeline for the consistent, reproducible, and high-resolution structural analysis of bacterial peptidoglycans. *eLife* 10. DOI: 10.7554/eLife.70597.
- Patti GJ, Chen J, Schaefer J, et al. (2008a) Characterization of structural variations in the peptidoglycan of vancomycin-susceptible *Enterococcus faecium:* understanding glycopeptide-antibiotic binding sites using mass spectrometry. *Journal of the American Society for Mass Spectrometry* 19(10): 1467–1475. DOI: 10.1016/j.jasms.2008.06.020.
- Patti GJ, Kim SJ and Schaefer J (2008b) Characterization of the peptidoglycan of vancomycin-susceptible *Enterococcus faecium*. *Biochemistry* 47(32): 8378–8385. DOI: 10.1021/bi8008032.
- Pazos M and Peters K (2019) Peptidoglycan. *Sub-cellular biochemistry* 92: 127–168. DOI: 10.1007/978-3-030-18768-2_5.
- Peabody MA, Laird MR, Vlasschaert C, et al. (2016) PSORTdb: Expanding the bacteria and archaea protein subcellular localization database to better reflect diversity in cell envelope structures. *Nucleic acids research* 44(D1): D663-8. DOI: 10.1093/nar/gkv1271.

Pedro MA de and Cava F (2015) Structural constraints and dynamics of bacterial cell wall architecture. *Frontiers in microbiology* 6: 449. DOI: 10.3389/fmicb.2015.00449.

- Peltier J, Courtin P, El Meouche I, et al. (2011) *Clostridium difficile* has an original peptidoglycan structure with a high level of N-acetylglucosamine deacetylation and mainly 3-3 cross-links. *The Journal of biological chemistry* 286(33): 29053–29062. DOI: 10.1074/jbc.M111.259150.
- Perry AM, Ton-That H, Mazmanian SK, et al. (2002) Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. III. Lipid II is an in vivo peptidoglycan substrate for sortase-catalyzed surface protein anchoring. *Journal of Biological Chemistry* 277(18): 16241–16248. DOI: 10.1074/jbc.M109194200.
- Pfanzagl B, Zenker A, Pittenauer E, et al. (1996) Primary structure of cyanelle peptidoglycan of *Cyanophora paradoxa*: a prokaryotic cell wall as part of an organelle envelope. *Journal of bacteriology* 178(2): 332–339. DOI: 10.1128/jb.178.2.332-339.1996.
- Poole K (2004) Resistance to beta-lactam antibiotics. *Cellular and molecular life sciences CMLS* 61(17): 2200–2223. DOI: 10.1007/s00018-004-4060-9.
- Popham DL and Bernhards CB (2015) Spore Peptidoglycan. *Microbiology Spectrum* 3(6). DOI: 10.1128/microbiolspec.tbs-0005-2012.
- Popham DL, Helin J, Costello CE, et al. (1996a) Analysis of the peptidoglycan structure of *Bacillus subtilis* endospores. *Journal of bacteriology* 178(22): 6451–6458.
- Popham DL, Helin J, Costello CE, et al. (1996b) Muramic lactam in peptidoglycan of Bacillus subtilis spores is required for spore outgrowth but not for spore dehydration or heat resistance. Proceedings of the National Academy of Sciences 93(26): 15405–15410. DOI: 10.1073/pnas.93.26.15405.
- Popham DL, Meador-Parton J, Costello CE, et al. (1999) Spore Peptidoglycan Structure in a cwID dacB Double Mutant of *Bacillus subtilis*. *Journal of bacteriology* 181(19): 6205–6209.
- Porfírio S, Carlson RW and Azadi P (2019) Elucidating Peptidoglycan Structure: An Analytical Toolset. *Trends in microbiology* 27(7): 607–622. DOI: 10.1016/j.tim.2019.01.009.
- Prell J and Poole P (2006) Metabolic changes of rhizobia in legume nodules. *Trends in microbiology* 14(4): 161–168. DOI: 10.1016/j.tim.2006.02.005.
- Pronk JT, Levering PR, Olijve W, et al. (1989) Role of NADP-dependent and quinoprotein glucose dehydrogenases in gluconic acid production by *Gluconobacter oxydans. Enzyme and Microbial Technology* 11(3): 160–164. DOI: 10.1016/0141-0229(89)90075-6.
- Qiao Y, Lebar MD, Schirner K, et al. (2014) Detection of Lipid-Linked Peptidoglycan Precursors by Exploiting an Unexpected Transpeptidase Reaction. *Journal of the American Chemical Society* 136(42): 14678–14681. DOI: 10.1021/ja508147s.
- Rady BJ and Mesnage S (2024) PGFinder, an Open-Source Software for Peptidoglycomics: The Structural Analysis of Bacterial Peptidoglycan by LC-MS.

Methods in molecular biology (Clifton, N.J.) 2836: 111–132. DOI: 10.1007/978-1-0716-4007-4 8.

- Raymond JB, Mahapatra S, Crick DC, et al. (2005) Identification of the namH gene, encoding the hydroxylase responsible for the N-glycolylation of the mycobacterial peptidoglycan. *The Journal of biological chemistry* 280(1): 326–333. DOI: 10.1074/jbc.M411006200.
- Rico-Pérez G, Pezza A, Pucciarelli MG, et al. (2016) A novel peptidoglycan D,Lendopeptidase induced by *Salmonella* inside eukaryotic cells contributes to virulence. *Molecular microbiology* 99(3): 546–556. DOI: 10.1111/mmi.13248.
- Rimal B, Senzani S, Ealand C, et al. (2022) Peptidoglycan compositional analysis of *Mycobacterium smegmatis* using high-resolution LC–MS. *Scientific reports* 12(1): 11061. DOI: 10.1038/s41598-022-15324-1.
- Ripoll M, Lerma-Escalera JA, Morones-Ramírez JR, et al. (2023) New perspectives into *Gluconobacter*-catalysed biotransformations. *Biotechnology advances* 65: 108127. DOI: 10.1016/j.biotechadv.2023.108127.
- Rodrigues-Oliveira T, Belmok A, Vasconcellos D, et al. (2017) Archaeal S-Layers: Overview and Current State of the Art. *Frontiers in microbiology* 8: 2597. DOI: 10.3389/fmicb.2017.02597.
- Rogers HJ and Perkins HR (1968) Cell walls and membranes. London: Spon.
- Rohde M, Fischetti VA, Novick RP, et al. (2019) The Gram-Positive Bacterial Cell Wall. *Microbiology Spectrum* 7(3): 7.3.10. DOI: 10.1128/microbiolspec.GPP3-0044-2018.
- Rohs PDA and Bernhardt TG (2021) Growth and Division of the Peptidoglycan Matrix. *Annual review of microbiology* 75: 315–336. DOI: 10.1146/annurev-micro-020518-120056.
- Romaniuk JAH and Cegelski L (2015) Bacterial cell wall composition and the influence of antibiotics by cell-wall and whole-cell NMR. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 370(1679). DOI: 10.1098/rstb.2015.0024.
- Roo AC de and Regenbogen SE (2020) *Clostridium difficile* Infection: An Epidemiology Update. *Clinics in colon and rectal surgery* 33(2): 49–57. DOI: 10.1055/s-0040-1701229.
- Sacco E, Hugonnet J-E, Josseaume N, et al. (2010) Activation of the L,Dtranspeptidation peptidoglycan cross-linking pathway by a metallo-D,Dcarboxypeptidase in *Enterococcus faecium*. *Molecular microbiology* 75(4): 874– 885. DOI: 10.1111/j.1365-2958.2009.07014.x.
- Sanders AN and Pavelka MS (2013) Phenotypic analysis of *Eschericia coli* mutants lacking L,D-transpeptidases. *Microbiology (Reading, England)* 159(Pt 9): 1842– 1852. DOI: 10.1021/bi00714a008.
- Sandoz KM, Moore RA, Beare PA, et al. (2021a) β-Barrel proteins tether the outer membrane in many Gram-negative bacteria. *Nature microbiology* 6(1): 19–26. DOI: 10.1038/s41564-020-00798-4.

- Sandoz KM, Moore RA, Beare PA, et al. (2021b) β-Barrel proteins tether the outer membrane in many Gram-negative bacteria. *Nature microbiology* 6(1): 19–26. DOI: 10.1038/s41564-020-00798-4.
- Sauvage E, Kerff F, Terrak M, et al. (2008) The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. *FEMS microbiology reviews* 32(2): 234–258. DOI: 10.1111/j.1574-6976.2008.00105.x.
- Schaub RE and Dillard JP (2017) Digestion of Peptidoglycan and Analysis of Soluble Fragments. *Bio-protocol* 7(15). DOI: 10.21769/BioProtoc.2438.
- Scheffers D-J and Pinho MG (2005) Bacterial cell wall synthesis: new insights from localization studies. *Microbiology and molecular biology reviews MMBR* 69(4): 585–607. DOI: 10.1128/MMBR.69.4.585-607.2005.
- Schindler M, Mirelman D and Schwarz U (1976) Quantitative determination of Nacetylglucosamine residues at the non-reducing ends of peptidoglycan chains by enzymic attachment of 14C-D-galactose. *European journal of biochemistry* 71(1): 131–134. DOI: 10.1111/j.1432-1033.1976.tb11098.x.
- Schirner K, Marles-Wright J, Lewis RJ, et al. (2009) Distinct and essential morphogenic functions for wall- and lipo-teichoic acids in *Bacillus subtilis*. *The EMBO journal* 28(7): 830–842. DOI: 10.1038/emboj.2009.25.
- Schleifer KH and Kandler O (1972) Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriological reviews* 36(4): 407–477. DOI: 10.1128/br.36.4.407-477.1972.
- Schmitz AM, Pian B, Medin S, et al. (2021) Generation of a *Gluconobacter oxydans* knockout collection for improved extraction of rare earth elements. *Nature Communications* 12(1). DOI: 10.1038/s41467-021-27047-4.
- Schulte CCM, Ramachandran VK, Papachristodoulou A, et al. (2022) Genome-Scale Metabolic Modelling of Lifestyle Changes in *Rhizobium leguminosarum*. *mSystems* 7(1): e0097521. DOI: 10.1128/msystems.00975-21.
- Sekiya H, Tamai E, Kawasaki J, et al. (2021) Structural and biochemical characterizations of the novel autolysin Acd24020 from *Clostridioides difficile* and its full-function catalytic domain as a lytic enzyme. *Molecular microbiology* 115(4): 684–698. DOI: 10.1111/mmi.14636.
- Seligman SJ (1987) Cell division in staphylococci: a clue to the three-dimensional structure of peptidoglycan. *The Journal of Infectious Diseases* 155(3): 423–432. DOI: 10.1093/infdis/155.3.423.
- Seligman SJ and Pincus MR (1987) A model for the three-dimensional structure of peptidoglycan in staphylococci. *Journal of Theoretical Biology* 124(3): 275–292. DOI: 10.1016/s0022-5193(87)80116-9.
- Seltmann G and Holst O (2002) *The Bacterial Cell Wall*. Berlin, Heidelberg: Springer Berlin Heidelberg.

- Shaku M, Ealand C, Matlhabe O, et al. (2020) Peptidoglycan biosynthesis and remodeling revisited. *Advances in applied microbiology* 112: 67–103. DOI: 10.1016/bs.aambs.2020.04.001.
- Sham L-T, Butler EK, Lebar MD, et al. (2014) Bacterial cell wall. MurJ is the flippase of lipid-linked precursors for peptidoglycan biogenesis. *Science (New York, N.Y.)* 345(6193): 220–222. DOI: 10.1126/science.1254522.
- Shang G, Liu X, Lu D, et al. (2012) Structural insight into how *Pseudomonas aeruginosa* peptidoglycanhydrolase Tse1 and its immunity protein Tsi1 function. *The Biochemical journal* 448(2): 201–211. DOI: 10.1042/BJ20120668.
- Sheng B, Ni J, Gao C, et al. (2014) Draft Genome Sequence of the *Gluconobacter oxydans* Strain DSM 2003, an Important Biocatalyst for Industrial Use. *Genome Announcements* 2(2): e00417 - 14. DOI: 10.1128/genomeA.00417-14.

Sibinelli-Sousa S, Hespanhol JT and Bayer-Santos E (2021) Targeting the Achilles' Heel of Bacteria: Different Mechanisms To Break Down the Peptidoglycan Cell Wall during Bacterial Warfare. *Journal of bacteriology* 203(7). DOI: 10.1128/JB.00478-20.

Siegel SD, Liu J and Ton-That H (2016) Biogenesis of the Gram-positive bacterial cell envelope. *Current opinion in microbiology* 34: 31–37. DOI: 10.1016/j.mib.2016.07.015.

Silhavy TJ, Kahne D and Walker S (2010a) The bacterial cell envelope. *Cold Spring Harbor perspectives in biology* 2(5): a000414. DOI: 10.1101/cshperspect.a000414.

Silhavy TJ, Kahne D and Walker S (2010b) The bacterial cell envelope. *Cold Spring Harbor perspectives in biology* 2(5): a000414. DOI: 10.1101/cshperspect.a000414.

- Singh SK, Parveen S, SaiSree L, et al. (2015) Regulated proteolysis of a cross-linkspecific peptidoglycan hydrolase contributes to bacterial morphogenesis. *Proceedings of the National Academy of Sciences of the United States of America* 112(35): 10956–10961. DOI: 10.1073/pnas.1507760112.
- Sjodt M, Rohs PDA, Gilman MSA, et al. (2020) Structural coordination of polymerization and crosslinking by a SEDS-bPBP peptidoglycan synthase complex. *Nature microbiology* 5(6): 813–820. DOI: 10.1038/s41564-020-0687-z.

Sleytr UB, Schuster B, Egelseer E-M, et al. (2014) S-layers: principles and applications. *FEMS microbiology reviews* 38(5): 823–864. DOI: 10.1111/1574-6976.12063.

- Sonnenfeld EM, Beveridge TJ, Koch AL, et al. (1985) Asymmetric distribution of charge on the cell wall of *Bacillus subtilis*. *Journal of bacteriology* 163(3): 1167–1171. DOI: 10.1128/jb.163.3.1167-1171.1985.
- Steen H and Mann M (2004) The ABC's (and XYZ's) of peptide sequencing. *Nature reviews. Molecular cell biology* 5(9): 699–711. DOI: 10.1038/nrm1468.
- Stock JB, Rauch B and Roseman S (1977) Periplasmic space in Salmonella typhimurium and Escherichia coli. The Journal of biological chemistry 252(21): 7850–7861.
- Stogios PJ and Savchenko A (2020) Molecular mechanisms of vancomycin resistance. *Protein science a publication of the Protein Society* 29(3): 654–669. DOI: 10.1002/pro.3819.

- Straume D, Piechowiak KW, Kjos M, et al. (2021) Class A PBPs: It is time to rethink traditional paradigms. *Molecular microbiology* 116(1): 41–52. DOI: 10.1111/mmi.14714.
- Straume D, Piechowiak KW, Olsen S, et al. (2020) Class A PBPs have a distinct and unique role in the construction of the pneumococcal cell wall. *Proceedings of the National Academy of Sciences of the United States of America* 117(11): 6129–6138. DOI: 10.1073/pnas.1917820117.
- Strupat K, Scheibner O and Bromirski M (2016) High-Resolution, Accurate-Mass Orbitrap Mass Spectrometry - Definitions, Opportunities, and Advantages: Technical Note 64287.
- Sun J, Rutherford ST, Silhavy TJ, et al. (2022) Physical properties of the bacterial outer membrane. *Nature reviews. Microbiology* 20(4): 236–248. DOI: 10.1038/s41579-021-00638-0.
- Sütterlin L, Edoo Z, Hugonnet J-E, et al. (2018) Peptidoglycan Cross-Linking Activity of I,d-Transpeptidases from *Clostridium difficile* and Inactivation of These Enzymes by β -Lactams. *Antimicrobial agents and chemotherapy* 62(1). DOI: 10.1128/AAC.01607-17.
- Sychantha D, Brott AS, Jones CS, et al. (2018) Mechanistic Pathways for Peptidoglycan O-Acetylation and De-O-Acetylation. *Frontiers in microbiology* 9: 2332. DOI: 10.3389/fmicb.2018.02332.
- Szwedziak P, Wang Q, Bharat TAM, et al. (2014) Architecture of the ring formed by the tubulin homologue FtsZ in bacterial cell division. *eLife* 3: e04601. DOI: 10.7554/eLife.04601.
- Taguchi A, Kahne D and Walker S (2019) Chemical tools to characterize peptidoglycan synthases. *Current opinion in chemical biology* 53: 44–50. DOI: 10.1016/j.cbpa.2019.07.009.
- Tamaki S, Sato T and Matsuhashi M (1971) Role of lipopolysaccharides in antibiotic resistance and bacteriophage adsorption of *Escherichia coli* K-12. *Journal of bacteriology* 105(3): 968–975. DOI: 10.1128/jb.105.3.968-975.1971.
- Templin MF, Ursinus A and Höltje JV (1999) A defect in cell wall recycling triggers autolysis during the stationary growth phase of *Escherichia coli*. *The EMBO journal* 18(15): 4108–4117. DOI: 10.1093/emboj/18.15.4108.
- Thermo fisher Inc (2013) Thermo Xcalibur: Qual Browser. Thermo fisher Inc.
- Tinajero-Trejo M, Aindow M, Pasquina-Lemonche L, et al. (2025) Control of morphogenesis during the Staphylococcus aureus cell cycle. *Science advances* 11(15): eadr5011. DOI: 10.1126/sciadv.adr5011.
- Tindwa H, Patnaik BB, Kim DH, et al. (2013) Cloning, characterization and effect of TmPGRP-LE gene silencing on survival of *Tenebrio molitor* against *Listeria monocytogenes* infection. *International journal of molecular sciences* 14(11): 22462–22482. DOI: 10.3390/ijms141122462.
- Tipper DJ (1970) Structure and function of peptidoglycans. *International Journal of Systematic Bacteriology* 20(4): 361–377. DOI: 10.1099/00207713-20-4-361.

- Tipper DJ and Strominger JL (1965) Mechanism of action of penicillins: a proposal based on their structural similarity to acyl-D-alanyl-D-alanine. *Proceedings of the National Academy of Sciences of the United States of America* 54(4): 1133–1141. DOI: 10.1073/pnas.54.4.1133.
- Tobin MJ, Cho SY, Profy W, et al. (2023) Reconstituting Spore Cortex Peptidoglycan Biosynthesis Reveals a Deacetylase That Catalyzes Transamidation. *Biochemistry* 62(8): 1342–1346. DOI: 10.1021/acs.biochem.3c00100.
- Tocheva El, Ortega DR and Jensen GJ (2016) Sporulation, bacterial cell envelopes and the origin of life. *Nature Reviews Microbiology* 14(8): 535–542. DOI: 10.1038/nrmicro.2016.85.
- Torrens G and Cava F (2024) Mechanisms conferring bacterial cell wall variability and adaptivity. *Biochemical Society transactions* 52(5): 1981–1993. DOI: 10.1042/BST20230027.
- Toth M, Stewart NK, Smith CA, et al. (2022) The I,d-Transpeptidase LdtAb from *Acinetobacter baumannii* Is Poorly Inhibited by Carbapenems and Has a Unique Structural Architecture. *ACS infectious diseases* 8(9): 1948–1961. DOI: 10.1021/acsinfecdis.2c00321.
- Touhami A, Jericho MH and Beveridge TJ (2004) Atomic force microscopy of cell growth and division in *Staphylococcus aureus*. *Journal of bacteriology* 186(11): 3286–3295. DOI: 10.1128/JB.186.11.3286-3295.2004.
- Trachtenberg S (2005) Mollicutes. *Current biology CB* 15(13): R483-4. DOI: 10.1016/j.cub.2005.06.049.
- Triboulet S, Dubée V, Lecoq L, et al. (2013) Kinetic features of L,D-transpeptidase inactivation critical for β -lactam antibacterial activity. *PloS one* 8(7): e67831. DOI: 10.1371/journal.pone.0067831.
- Turello LA, Consul A, Yip C, et al. (2025) Differential gene expression analysis shows that cephalosporin resistance is intrinsic to *Clostridioides difficile* strain 630. *The Journal of antibiotics* 78(2): 113–125. DOI: 10.1038/s41429-024-00795-3.
- Turner RD, Hurd AF, Cadby A, et al. (2013) Cell wall elongation mode in Gram-negative bacteria is determined by peptidoglycan architecture. *Nature communications* 4(1): 1496. DOI: 10.1038/ncomms2503.
- Turner RD, Mesnage S, Hobbs JK, et al. (2018) Molecular imaging of glycan chains couples cell-wall polysaccharide architecture to bacterial cell morphology. *Nature communications* 9(1): 1263. DOI: 10.1038/s41467-018-03551-y.
- Turner RD, Ratcliffe EC, Wheeler R, et al. (2010) Peptidoglycan architecture can specify division planes in *Staphylococcus aureus*. *Nature communications* 1: 26. DOI: 10.1038/ncomms1025.
- Turner RD, Vollmer W and Foster SJ (2014) Different walls for rods and balls: the diversity of peptidoglycan. *Molecular microbiology* 91(5): 862–874. DOI: 10.1111/mmi.12513.

- Uehara T and Park JT (2007) An anhydro-N-acetylmuramyl-L-alanine amidase with broad specificity tethered to the outer membrane of *Escherichia coli*. *Journal of bacteriology* 189(15): 5634–5641. DOI: 10.1128/JB.00446-07.
- van de Velde W, Zehirov G, Szatmari A, et al. (2010) Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science (New York, N.Y.)* 327(5969): 1122–1126. DOI: 10.1126/science.1184057.
- van der Aart LT, Spijksma GK, Harms A, et al. (2018) High-Resolution Analysis of the Peptidoglycan Composition in *Streptomyces coelicolor*. *Journal of bacteriology* 200(20). DOI: 10.1128/JB.00290-18.
- van Heijenoort J (2011) Peptidoglycan hydrolases of *Escherichia coli*. *Microbiology and molecular biology reviews MMBR* 75(4): 636–663. DOI: 10.1128/MMBR.00022-11.
- van Heijenoort J and Gutmann L (2000) Correlation between the structure of the bacterial peptidoglycan monomer unit, the specificity of transpeptidation, and susceptibility to beta-lactams. *Proceedings of the National Academy of Sciences of the United States of America* 97(10): 5028–5030. DOI: 10.1073/pnas.97.10.5028.
- van Kempen M, Kim SS, Tumescheit C, et al. (2024) Fast and accurate protein structure search with Foldseek. *Nature biotechnology* 42(2): 243–246. DOI: 10.1038/s41587-023-01773-0.
- van Wolferen M, Pulschen AA, Baum B, et al. (2022) The cell biology of archaea. Nature microbiology 7(11): 1744–1755. DOI: 10.1038/s41564-022-01215-8.
- Vasstrand EN (1981) Lysozyme digestion and chemical characterization of the peptidoglycan of Fusobacterium nucleatum Fev 1. *Infection and immunity* 33(1): 75–82. DOI: 10.1128/iai.33.1.75-82.1981.
- Vergalli J, Bodrenko IV, Masi M, et al. (2020) Porins and small-molecule translocation across the outer membrane of Gram-negative bacteria. *Nature reviews. Microbiology* 18(3): 164–176. DOI: 10.1038/s41579-019-0294-2.
- Vermassen A, Leroy S, Talon R, et al. (2019) Cell Wall Hydrolases in Bacteria: Insight on the Diversity of Cell Wall Amidases, Glycosidases and Peptidases Toward Peptidoglycan. *Frontiers in microbiology* 10: 331. DOI: 10.3389/fmicb.2019.00331.
- Voedts H, Dorchêne D, Lodge A, et al. (2021) Role of endopeptidases in peptidoglycan synthesis mediated by alternative cross-linking enzymes in *Escherichia coli*. *The EMBO journal* 40(19): e108126. DOI: 10.15252/embj.2021108126.
- Vollmer W (2008) Structural variation in the glycan strands of bacterial peptidoglycan. *FEMS microbiology reviews* 32(2): 287–306. DOI: 10.1111/j.1574-6976.2007.00088.x.
- Vollmer W, Blanot D and Pedro MA de (2008a) Peptidoglycan structure and architecture. *FEMS microbiology reviews* 32(2): 149–167. DOI: 10.1111/j.1574-6976.2007.00094.x.

- Vollmer W and Höltje J-V (2004) The architecture of the murein (peptidoglycan) in gram-negative bacteria: vertical scaffold or horizontal layer(s)? *Journal of bacteriology* 186(18): 5978–5987. DOI: 10.1128/JB.186.18.5978-5987.2004.
- Vollmer W, Joris B, Charlier P, et al. (2008b) Bacterial peptidoglycan (murein) hydrolases. *FEMS microbiology reviews* 32(2): 259–286. DOI: 10.1111/j.1574-6976.2007.00099.x.
- Vollmer W, Massidda O and Tomasz A (2019) The Cell Wall of *Streptococcus* pneumoniae. Microbiology Spectrum 7(3). DOI: 10.1128/microbiolspec.GPP3-0018-2018.
- Vollmer W and Tomasz A (2000) The pgdA gene encodes for a peptidoglycan Nacetylglucosamine deacetylase in *Streptococcus pneumoniae*. *The Journal of biological chemistry* 275(27): 20496–20501. DOI: 10.1074/jbc.M910189199.
- Vonberg R-P, Kuijper EJ, Wilcox MH, et al. (2008) Infection control measures to limit the spread of *Clostridium difficile*. *Clinical microbiology and infection the official publication of the European Society of Clinical Microbiology and Infectious Diseases* 14 Suppl 5: 2–20. DOI: 10.1111/j.1469-0691.2008.01992.x.
- Walter A and Mayer C (2019) Peptidoglycan Structure, Biosynthesis, and Dynamics During Bacterial Growth. In: Cohen E and Merzendorfer H (eds) Extracellular Sugar-Based Biopolymers Matrices: Cham: Springer International Publishing, pp. 237–299.
- Wang G, Olczak A, Forsberg LS, et al. (2009) Oxidative stress-induced peptidoglycan deacetylase in *Helicobacter pylori*. *The Journal of biological chemistry* 284(11): 6790–6800. DOI: 10.1074/jbc.M808071200.
- Wang J, Alvarez L, Bulgheresi S, et al. (2021) PBP4 Is Likely Involved in Cell Division of the Longitudinally Dividing Bacterium Candidatus *Thiosymbion Oneisti*. *Antibiotics (Basel, Switzerland)* 10(3). DOI: 10.3390/antibiotics10030274.
- Wang J, Chitsaz F, Derbyshire MK, et al. (2023) The conserved domain database in 2023. *Nucleic acids research* 51(D1): D384-D388. DOI: 10.1093/nar/gkac1096.
- Wang T-SA, Manning SA, Walker S, et al. (2008) Isolated peptidoglycan glycosyltransferases from different organisms produce different glycan chain lengths. *Journal of the American Chemical Society* 130(43): 14068–14069. DOI: 10.1021/ja806016y.
- Ward JB (1973) The chain length of the glycans in bacterial cell walls. *The Biochemical journal* 133(2): 395–398. DOI: 10.1042/bj1330395.
- Ward JB and Curtis CA (1982) The biosynthesis and linkage of teichuronic acid to peptidoglycan in *Bacillus licheniformis*. *European journal of biochemistry* 122(1): 125–132. DOI: 10.1111/j.1432-1033.1982.tb05857.x.
- Watanabe T, Izaki K and Takahashi H (1982) New polyenic antibiotics active against gram-positive and -negative bacteria. I. Isolation and purification of antibiotics produced by *Gluconobacter* sp. W-315. *The Journal of antibiotics* 35(9): 1141–1147. DOI: 10.7164/antibiotics.35.1141.

- Waz NT, Oliveira S, Girardello R, et al. (2022) Influence of the Polysaccharide Capsule on the Bactericidal Activity of Indolicidin on *Streptococcus pneumoniae*. *Frontiers in microbiology* 13: 898815. DOI: 10.3389/fmicb.2022.898815.
- Weadge JT and Clarke AJ (2006) Identification and characterization of Oacetylpeptidoglycan esterase: a novel enzyme discovered in *Neisseria gonorrhoeae*. *Biochemistry* 45(3): 839–851. DOI: 10.1021/bi051679s.
- Weadge JT, Pfeffer JM and Clarke AJ (2005) Identification of a new family of enzymes with potential O-acetylpeptidoglycan esterase activity in both Gram-positive and Gram-negative bacteria. *BMC microbiology* 5: 49. DOI: 10.1186/1471-2180-5-49.
- Webby MN, Oluwole AO, Pedebos C, et al. (2022) Lipids mediate supramolecular outer membrane protein assembly in bacteria. *Science advances* 8(44): eadc9566. DOI: 10.1126/sciadv.adc9566.
- Weidel W, Frank H and Martin HH (1960) The rigid layer of the cell wall of *Escherichia coli* strain B. *Journal of general microbiology* 22: 158–166. DOI: 10.1099/00221287-22-1-158.
- Weidel W and Pelzer H (1964) Bagshaped Macromolecules A New Outlook on Bacterial Cell Walls. Advances in enzymology and related subjects of biochemistry 26: 193–232. DOI: 10.1002/9780470122716.ch5.
- Weidenmaier C and Peschel A (2008) Teichoic acids and related cell-wall glycopolymers in Gram-positive physiology and host interactions. *Nature reviews. Microbiology* 6(4): 276–287. DOI: 10.1038/nrmicro1861.
- Welsh MA, Taguchi A, Schaefer K, et al. (2017) Identification of a Functionally Unique Family of Penicillin-Binding Proteins. *Journal of the American Chemical Society* 139(49): 17727–17730. DOI: 10.1021/jacs.7b10170.
- Wheatley RM, Ford BL, Li L, et al. (2020) Lifestyle adaptations of *Rhizobium* from rhizosphere to symbiosis. *Proceedings of the National Academy of Sciences of the United States of America* 117(38): 23823–23834. DOI: 10.1073/pnas.2009094117.
- Wheatley RM, Ramachandran VK, Geddes BA, et al. (2017) Role of O2 in the Growth of *Rhizobium leguminosarum* bv. viciae 3841 on Glucose and Succinate. *Journal of bacteriology* 199(1). DOI: 10.1128/JB.00572-16.
- Wheeler R, Chevalier G, Eberl G, et al. (2014) The biology of bacterial peptidoglycans and their impact on host immunity and physiology. *Cellular Microbiology* 16(7): 1014–1023. DOI: 10.1111/cmi.12304.
- Wheeler R, Mesnage S, Boneca IG, et al. (2011) Super-resolution microscopy reveals cell wall dynamics and peptidoglycan architecture in ovococcal bacteria. *Molecular microbiology* 82(5): 1096–1109. DOI: 10.1111/j.1365-2958.2011.07871.x.
- Wheeler R, Turner RD, Bailey RG, et al. (2015) Bacterial Cell Enlargement Requires Control of Cell Wall Stiffness Mediated by Peptidoglycan Hydrolases. *mBio* 6(4): e00660. DOI: 10.1128/mbio.00660-15.

- Wickramage I, Spigaglia P and Sun X (2021) Mechanisms of antibiotic resistance of *Clostridioides difficile. The Journal of antimicrobial chemotherapy* 76(12): 3077– 3090. DOI: 10.1093/jac/dkab231.
- Wietzerbin J, Das BC, Petit JF, et al. (1974) Occurrence of D-alanyl-(D)-mesodiaminopimelic acid and meso-diaminopimelyl-meso-diaminopimelic acid interpeptide linkages in the peptidoglycan of *Mycobacteria*. *Biochemistry* 13(17): 3471–3476. DOI: 10.1021/bi00714a008.
- Wingen-Heimann SM, Davies K, Viprey VF, et al. (2023) *Clostridioides difficile* infection (CDI): A pan-European multi-center cost and resource utilization study, results from the Combatting Bacterial Resistance in Europe CDI (COMBACTE-CDI). *Clinical Microbiology and Infection* 29(5): 651.e1 - 651.e8. DOI: 10.1016/j.cmi.2022.12.019.
- Winkle M, Hernández-Rocamora VM, Pullela K, et al. (2021) DpaA Detaches Braun's Lipoprotein from Peptidoglycan. *mBio* 12(3). DOI: 10.1128/mBio.00836-21.
- Wohlwend J, Corso G, Passaro S, et al. (2024) Boltz-1 Democratizing Biomolecular Interaction Modeling. *bioRxiv the preprint server for biology*. Epub ahead of print 27 December 2024. DOI: 10.1101/2024.11.19.624167.
- Wright J and Heckels JE (1975) The teichuronic acid of cell walls of *Bacillus subtilis* W23 grown in a chemostat under phosphate limitation. *The Biochemical journal* 147(1): 187–189. DOI: 10.1042/bj1470187.
- Wu X, Paskaleva EE, Mehta KK, et al. (2016) Wall Teichoic Acids Are Involved in the Medium-Induced Loss of Function of the Autolysin CD11 against *Clostridium difficile*. *Scientific reports* 6: 35616. DOI: 10.1038/srep35616.
- Wydau-Dematteis S, El Meouche I, Courtin P, et al. (2018) Cwp19 Is a Novel Lytic Transglycosylase Involved in Stationary-Phase Autolysis Resulting in Toxin Release in *Clostridium difficile. mBio* 9(3). DOI: 10.1128/mBio.00648-18.
- Yadav AK, Espaillat A and Cava F (2018) Bacterial Strategies to Preserve Cell Wall Integrity Against Environmental Threats. *Frontiers in microbiology* 9: 2064. DOI: 10.3389/fmicb.2018.02064.
- Yamaguchi H, Furuhata K, Fukushima T, et al. (2004) Characterization of a new *Bacillus subtilis* peptidoglycan hydrolase gene, yvcE (named cwlO), and the enzymatic properties of its encoded protein. *Journal of bioscience and bioengineering* 98(3): 174–181. DOI: 10.1016/S1389-1723(04)00262-2.
- Yang DC, Tan K, Joachimiak A, et al. (2012) A conformational switch controls cell wallremodelling enzymes required for bacterial cell division. *Molecular microbiology* 85(4): 768–781. DOI: 10.1111/j.1365-2958.2012.08138.x.
- Yao X, Jericho M, Pink D, et al. (1999) Thickness and elasticity of gram-negative murein sacculi measured by atomic force microscopy. *Journal of bacteriology* 181(22): 6865–6875. DOI: 10.1128/JB.181.22.6865-6875.1999.
- Young JPW, Moeskjær S, Afonin A, et al. (2021) Defining the *Rhizobium leguminosarum* Species Complex. *Genes* 12(1): 111. DOI: 10.3390/genes12010111.

- Zapun A, Contreras-Martel C and Vernet T (2008a) Penicillin-binding proteins and beta-lactam resistance. *FEMS microbiology reviews* 32(2): 361–385. DOI: 10.1111/j.1574-6976.2007.00095.x.
- Zapun A, Vernet T and Pinho MG (2008b) The different shapes of cocci. *FEMS microbiology reviews* 32(2): 345–360. DOI: 10.1111/j.1574-6976.2007.00098.x.
- Zhang G, Meredith TC and Kahne D (2013) On the essentiality of lipopolysaccharide to Gram-negative bacteria. *Current opinion in microbiology* 16(6): 779–785. DOI: 10.1016/j.mib.2013.09.007.
- Zhang S, Gu S, Rycroft P, et al. (2022) Scaffolding Protein GspB/OutB Facilitates Assembly of the *Dickeya dadantii* Type 2 Secretion System by Anchoring the Outer Membrane Secretin Pore to the Inner Membrane and to the Peptidoglycan Cell Wall. *mBio* 13(3): e0025322. DOI: 10.1128/mbio.00253-22.
- Zuber B, Chami M, Houssin C, et al. (2008) Direct visualization of the outer membrane of mycobacteria and corynebacteria in their native state. *Journal of bacteriology* 190(16): 5672–5680. DOI: 10.1128/JB.01919-07.

Appendix 1

This appendix represents my contribution to paper produced at the during my thesis study.

Galley, N. F., Greetham, D., **Alamán-Zárate, M. G.**, Williamson, M. P., Evans, C. A., Spittal, W. D., Buddle, J. E., Freeman, J., Davis, G. L., Dickman, M. J., Wilcox, M. H., Lovering, A. L., Fagan, R. P., & Mesnage, S. (2024). *Clostridioides difficile* canonical L,Dtranspeptidases catalyze a novel type of peptidoglycan cross-links and are not required for β-lactam resistance. The Journal of Biological Chemistry, 300(1), 105529. https://doi.org/10.1016/j.jbc.2023.105529



Clostridioides difficile canonical L,D-transpeptidases catalyze a novel type of peptidoglycan cross-links and are not required for beta-lactam resistance

Received for publication, October 13, 2023, and in revised form, November 19, 2023 Published, Papers in Press, December 1, 2023, https://doi.org/10.1016/j.jbc.2023.105529

Nicola F. Galley¹[®], Darren Greetham¹[®], Marcel G. Alamán-Zárate¹[®], Mike P. Williamson¹[®], Caroline A. Evans²[®], William D. Spittal^{3,4}[®], Jessica E. Buddle¹[®], Jane Freeman^{3,4}[®], Georgina L. Davis^{3,4}, Mark J. Dickman²[®], Mark H. Wilcox^{3,4}[®], Andrew L. Lovering⁵[®], Robert P. Fagan¹[®], and Stéphane Mesnage^{1,*}[®]

From the ¹School of Biosciences, and ²Department of Chemical and Biological Engineering, University of Sheffield, Sheffield, UK; ³Department of Microbiology, Leeds Teaching Hospitals NHS Trust, Leeds Institute of Medical Research, University of Leeds, Leeds, UK; ⁴Healthcare Associated Infections Research Group, Leeds Institute of Medical Research University of Leeds, Leeds, UK; ⁵School of Biosciences, University of Birmingham, Birmingham, UK

Reviewed by members of the JBC Editorial Board. Edited by Chris Whitfield

Clostridioides difficile is the leading cause of antibioticassociated diarrhea worldwide with significant morbidity and mortality. This organism is naturally resistant to several betalactam antibiotics that inhibit the polymerization of peptidoglycan, an essential component of the bacteria cell envelope. Previous work has revealed that C. difficile peptidoglycan has an unusual composition. It mostly contains 3-3 cross-links, catalyzed by enzymes called L,D-transpeptidases (Ldts) that are poorly inhibited by beta-lactams. It was therefore hypothesized that peptidoglycan polymerization by these enzymes could underpin antibiotic resistance. Here, we investigated the catalytic activity of the three canonical Ldts encoded by C. difficile (Ldt_{Cd1}, Ldt_{Cd2}, and Ldt_{Cd3}) in vitro and explored their contribution to growth and antibiotic resistance. We show that two of these enzymes catalyze the formation of novel types of peptidoglycan cross-links using meso-diaminopimelic acid both as a donor and an acceptor, also observed in peptidoglycan sacculi. We demonstrate that the simultaneous deletion of these three genes only has a minor impact on both peptidoglycan structure and resistance to beta-lactams. This unexpected result therefore implies that the formation of 3-3 peptidoglycan cross-links in C. difficile is catalyzed by as yet unidentified noncanonical Ldt enzymes.

Clostridioides difficile is a spore-forming Gram-positive obligate anaerobe that can cause hospital-associated diarrhea worldwide, representing increasing healthcare resource and economic burden (1). Although *C. difficile* has been recognized as a major cause of healthcare-associated infections since the 1970s, the more recent increase in morbidity and mortality is linked to the emergence of virulent epidemic strains including those belonging to ribotype 027 (2). *C. difficile* infections are underpinned by the natural resistance of this organism to several antibiotics including broad-spectrum beta-lactams such as cephalosporins. The dysbiosis caused by an antibiotic

* For correspondence: Stéphane Mesnage, s.mesnage@sheffield.ac.uk.

treatment creates an environment conducive to the germination of *C. difficile* spores and the production of virulence factors including toxins and several surface proteins (3).

The resistance of C. difficile to beta-lactams is poorly understood. These antibiotics covalently bind to D,D-transpeptidases (also known as penicillin-binding proteins) and irreversibly inhibit the enzymatic activity of these enzymes (4). In most bacteria, inhibition of D,Dtranspeptidation disrupts the polymerization of peptidoglycan, the major and essential component of the bacterial cell wall and prevents bacterial growth (5). The peptidoglycan of C. difficile has an unusual composition. It is mostly polymerized by a class of enzymes called L,D transpeptidases (Ldts) (6). Unlike D,D-transpeptidases, which form bonds between the amino acids in positions 3 and 4 of peptidoglycan peptide stems (4-3 cross-links), Ldts form bonds between two amino acids in positions 3 (3-3 cross-links). The activity of Ldts involves a catalytic mechanism distinct from the mechanism of D,D-transpeptidases, and Ldts are not inhibited by beta-lactams, with the exception of penems and carbapenems (7). The C. difficile genome encodes three enzymes called Ldt_{Cd1}, Ldt_{Cd2}, and Ldt_{Cd3}, which contain a canonical Ldt domain (YkuD). The contribution of these three enzymes to the peptidoglycan structure was investigated in strain 630 (6). Despite attempts to generate a triple knockout strain, only genes encoding Ldt_{Cd1} and Ldt_{Cd2} could be inactivated simultaneously (6) so it was suggested that 3-3 cross-links were required for viability. Analysis of the peptidoglycan structure in the double mutant strain revealed a limited decrease of 3-3 cross-links (6). The mutant remained able to perform 3-3 cross-links in the presence of ampicillin, suggesting that C. difficile Ldts were insensitive to this antibiotic (6). In vitro experiments revealed that these enzymes display distinct enzymatic activities and inhibition by beta-lactams (8). All enzymes were reported to have L,D carboxypeptidase activity, but L,D-transpeptidation and exchange of the amino acid in position 4 could only be detected for Ldt_{Cd2} and Ldt_{Cd3} . Interestingly, Ldt_{Cd3} could not be

SASBMB

J. Biol. Chem. (2024) 300(1) 105529 1

© 2023 THE AUTHORS. Published by Elsevier Inc on behalf of American Society for Biochemistry and Molecular Biology. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

acylated by any of the beta-lactams tested. The acylation efficacy of Ldt_{Cd1} and Ldt_{Cd2} by penicillin and cephalosporin antibiotics was much lower than the acylation by carbapenems, and the hydrolysis of these antibiotics was more efficient. It was therefore concluded that Ldt_{Cd} activity could only be inhibited by carbapenems (8).

Outstanding questions remain on the individual role of Ldt_{Cd} enzymes in peptidoglycan polymerization, the essentiality of the L,D-transpeptidation pathway in C. difficile and its contribution to antibiotic resistance. In this work, we further investigate the enzymatic activity of C. difficile Ldts, both in vitro and during vegetative growth. We show that Ldt_{Cd2} and Ldt_{Cd3} display novel enzymatic activities and that the genes encoding the three canonical Ldts can be deleted simultaneously. High-resolution structure of the wildtype and triple mutant peptidoglycan only revealed a minor impact on muropeptide composition, and no change in resistance to betalactams could be detected in the mutant strain. This work therefore provides new insights into the catalytic activities of Ldts and implies that the existence of another unidentified type of enzyme(s) that does not contain a canonical YkuD domain is able to catalyze the formation of 3-3 cross-links in C. difficile.

Results

In vitro assays with recombinant Ldt_{Cd1} , Ldt_{Cd2} , and Ldt_{Cd3} reveal distinct activities and a novel type of L,Dtranspeptidation

We sought to investigate the activities of the three Ldt_{Cd} enzymes to identify their specific roles in peptidoglycan remodeling. The recombinant enzymes were purified (Fig. S1) to test their enzymatic activities using four types of purified substrates: (i) a disaccharide-tetrapeptide alone (GlcNAc-Mur-NAc-L-Ala-D-isoGlu-meso-DAP-D-Ala; gm-AEJA, where "J" represents meso-diaminopimelic acid [DAP]) to test L,Dcarboxypeptidase and L,D-transpeptidase activities (Fig. 1A); (ii) the same disaccharide-tetrapeptide (gm-AEJA) in the presence of D-methionine to test fourth amino acid exchange (Fig. 1B); (iii) a 4-3 cross-linked dimer ((GlcNAc-MurNAc-L-Ala-D-isoGlu-meso-DAP-D-Ala)2; gm-AEJA=gm-AEJA, where "=" represents a peptidoglycan cross-link) (Fig. 1C); and (iv) a 3-3 cross-linked dimer (GlcN-MurNAc-L-Ala-D-isoGlu-meso-DAP-D-Ala-GlcN-MurNAc-L-Ala-D-isoGlu-meso-DAP (g(-Ac)m-AEJA=g(-Ac)m-AEJA) to test endopeptidase activities ((Fig. 1D). The monomer and 4-3 cross-linked dimer were purified from the *Escherichia coli* $\Delta 6ldt$ strain and therefore contain fully acetylated sugars (9). The 3-3 cross-linked dimer was purified from C. difficile and therefore contained deacetylated GlcNAc (GlcN).

 Ldt_{Cd} recombinant enzymes were active against all substrates tested and revealed distinct preferential activities. Ldt_{Cd1} displayed a low carboxypeptidase and transpeptidase activity and only converted half of the substrate during the exchange reaction. No endopeptidase activity was detected with any of the dimers.

 Ldt_{Cd2} had a preferential carboxypeptidase activity on the gm-AEJA substrate that was mostly converted into a

disaccharide-tripeptide (gm-AEJ) and transformed all the monomer into gm-AEJM. A very weak carboxypeptidase activity was detected with the 4-3 dimer while all the 3-3 dimers were completely cleaved, releasing disaccharide-tripeptides (g(-Ac)m-AEJ) as the most abundant products. Ldt_{Cd2} was only active on 3-3 cross-linked dimers. Interestingly, several multimers matching the expected mass for dimers, trimers, and tetramers lacking a molecule of water were detected (labeled as a, b, c, a*, b* and d*; Fig. 1*E*). These were further analyzed by NMR.

Ldt_{Cd3} had the highest L,D-transpeptidase activity of all enzymes, and the 3-3 cross-linked dimer was the most abundant product generated from the gm-AEJA substrate. This enzyme also displayed some carboxypeptidase activity, using the monomer or both dimers and converted all the monomer into gm-AEJM. Surprisingly, the carboxypeptidase activity of Ldt_{Cd3} was higher on the 4-3 dimer than on the 3-3 dimer. Ldt_{Cd3} also produced two transpeptidation products matching the mass of a 3-3 dimer lacking a molecule of water (peak b, also detected with Ldt_{Cd2}) and the mass of a 4-3 dimer lacking a molecule of water (peak c). The structures of all expected and previously described muropeptides produced by Ldt_{Cd1}, Ldt_{Cd2}, and Ldt_{Cd3} are described in Figure 1*E*.

Tandem mass spectrometry and NMR analyses of Ldt_{cd2} and Ldt_{cd3} transpeptidation products reveal a novel type of peptidoglycan cross-links

The muropeptide contained in peak a (Fig. 1) was analyzed by tandem mass spectrometry (MS/MS). The Fragmentation spectrum confirmed the inferred structure for a dimer with doubly cross-linked *meso*-DAP residues used both as an acceptor and as a donor group (Fig. 2*A*; see ions labeled). Several signature ions were found, including a doubly crosslinked DAP–DAP fragment.

Peptidoglycan fragments in peaks 1, a, and b (Fig. 1*E*) were purified and further analyzed by NMR. One-dimensional (1D) NMR spectra of the peptidoglycan fragments demonstrated a high purity for each. The monomer gm-AEJ (peak 1) used as a control (Fig. 2*B*, bottom panel) had all the amide signals expected, namely, two sugar *N*-acetyl signals, and one signal each for the Ala, DAP, and isoGlu (iE) residues. There is only one signal for DAP because one amine forms an amide with iE, while the other is a free amino group and therefore exchanges too fast with water to be visible. All the other signals are as expected, including the presence of two *N*-acetyl methyl singlets from the two sugars and two methyl doublets from Ala and the lactyl group on MurNAc.

The dimer corresponding to peak a has a remarkably simple NMR spectrum (Fig. 2*B*, middle panel). The amide region contains only six amide doublets, two *N*-acetyl methyl singlets, and two methyl doublets, as seen in the monomer (Fig. S2). This simplicity very strongly suggests a symmetrical dimer, and the similarity of the amide chemical shifts between the monomer and dimer implies a similar covalent structure.

There are two amide signals from the diaminopimelate, visible in the total correlation spectroscopy (TOCSY)



Novel peptidoglycan remodeling activities in C. difficile

Figure 1. HPLC-MS chromatograms of Ldt in vitro assays. Recombinant enzymes were incubated in the presence of a disaccharide-tetrapeptide substrate to test carboxypeptidase and transpeptidase activity and exchange (*A* and *B*, respectively). Endopeptidase activity was tested using either a 4-3 crosslinked or a 3-3 cross-linked dimer (*C* and *D*, respectively). The inferred structures from LC-MS analysis (*E*) as well as expected structures (*F*) are described. All traces correspond to Total lon Chromatograms (LC-MS data) corresponding to average intensity values from three independent experiments. The difference between observed and theoretical masses (Δppm) was calculated ((Theoretical mass – Observed mass)/Theoretical mass*10E6).

spectrum because they belong to the same spin system (Fig. 2*C*), indicating that both amines in the diaminopimelate take part in amide bonds. Chemical shift assignments for the dimer are listed in Table S1. In the nuclear Overhauser enhancement spectroscopy (NOESY) spectrum, there are the expected sequential NOEs present between NH_i and protons in residue (*i*-1), as indicated in Figure 2*D*. Crucially, these include NOEs between DAP NH^a and iE CγH₂, and the other "sequential" NOE of DAP NH^b to DAP CaH (Fig. 2, *D* and *E*).

Similarly, the spectrum of trimer b (Fig. 2*B*, top panel) is also very similar. The chemical shifts remain very similar to the monomer and dimer, and again there is only one set of signals, indicating a symmetrical trimer. The NMR spectra are thus fully consistent with the structures described in Figure 3, and the simplicity of the spectrum means that no unsymmetrical structure is possible. Based on our NMR data, we conclude that the muropeptides in peaks a, b, c, and d all correspond to multiply cross-linked structures (Figs. 2 and 3).



Figure 2. MS/MS and NMR analysis of unusual muropeptides a and b. *A*, MS/MS analysis of the ion corresponding to peak a (m/z = 1705.7211). Nine fragment ions corresponding to peptides with doubly bonded m-DAP residues are indicated. *B*, 1D NMR spectra of peptidoglycan fragments identified during *in vitro* assays (see Fig. 1): peak 1 (gm-AEJ) was used as a control; peak a corresponds to a dimer (gm-AEJ=gm-AEJ) with an unusual cross-link; peak b corresponds to a trimer (gm-AEJ=gm-AEJ) with an unusual cross-link The identity of each amide proton is indicated on the spectra. *C*, part of the COCSY spectrum of the muropeptide in peak a, showing connectivities between amide protons and side chains. The signals linked by *red lines* are the connectivities for the DAP^a and DAP^b amides, showing that they connect to identical side-chain frequencies and are therefore part of the same spin system. *D*, TOCSY (*red*) and NOESY (*black*) spectra of the dimer. Significant peaks are marked. *E*, structures of (*left*) dimer (with NOEs indicated) and (*right*) trimer. Only the central part of the trimer is shown, with arrows indicating where the AEJ chains are attached. A, alanine; DAP^a, *meso*-diaminopimelic acid backbone (directly bonded to the isoglutamate); DAP^b, *meso*-diaminopimelic acid side chain; GlcNAc, *N*-acetylglucosamine; iE, isoglutamate; MurNAc, *N*-acetylglure

TraDIS data mining and gene deletion reveal that the three canonical C. difficile Ldts are nonessential

Previous attempts to build a mutant with deletions in all three genes encoding the canonical Ldts were unsuccessful (6), suggesting either that one of them is essential or that 3-3 cross-linking is essential. We took advantage of transposondirected insertion site sequencing (TraDIS) data previously published in a study that identified essential genes in



Novel peptidoglycan remodeling activities in C. difficile

Figure 3. Unusual peptidoglycan cross-links catalyzed by Ldt_{cd2} and Ldt_{cd3}. Based on NMR data, the structure of muropeptides corresponding to peaks a, b, c, and d are shown next to the structure of the canonical 3-3 or 4-3 dimers, 3-3 trimer, and 3-3 tetramer. Dimer c contains both types of cross-links (3-3 and 4-3), resulting from D,D- and L,D-transpeptidation. Muropeptides a*, b*, d* display the same cross-links as a, b, d but contain GlcN instead of GlcNAc.

C. difficile R20291 (10). Using the number of transposon insertions in each *ldt*_{Cd} gene as a proxy to determine essentiality, we concluded that none of these genes was essential (Fig. 4), leaving the possibility that the combined deletion of ldt_{Cdl} , ldt_{Cd2} , and ldt_{Cd3} could be non viable. To test this hypothesis, we sought to generate a series of in-frame deletions in ldt_{Cdl} , ldt_{Cd2} , and ldt_{Cd3} . All genes were deleted individually or simultaneously. All the combinations of deletions, including the triple deletion mutant, could be obtained, showing that these genes are not required for viability. Since this result was unexpected, we performed whole genome sequencing on the mutants and confirmed the deletion of the *ldt* genes in the strains sequenced. Single nucleotide polymorphism analysis identified a unique single mutation in the triple deletion mutant (File S1 and Fig. S3). The mutation (T>TA) occurred at position 581480 in the intergenic mutation between CD0482 and glsA, two genes with no known link to peptidoglycan polymerization, encoding a putative phosphoribulokinase/ uridine kinase and a putative glutaminase, respectively. This result therefore suggests that the deletion of the ldt_{Cd} genes does not lead to genetic mutations likely to compensate for the lack of L,D-transpeptidase activity.

High-resolution structure of the wildtype and triple mutant peptidoglycans

Peptidoglycan was extracted from vegetative cells in stationary phase, and soluble fragments released after mutanolysin digestion were analyzed by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). Surprisingly, the chromatograms of the WT and triple mutant were virtually identical, indicating a very minor contribution of the three Ldts to peptidoglycan structure (Fig. 5).

To investigate subtle differences associated with the simultaneous deletion of the three ldt_{Cd} genes, we performed a highresolution analysis of the LC-MS/MS datasets using the Byos (Protein Metrics by Dotmatics) and PGFinder software (11). A bespoke search strategy was designed (Fig. S4). A first search was performed to identify the monomer search space using the Byonic module from the proprietary software Byos. Thirty-four disaccharide-peptides with a fragmentation showing more than half of the expected *b* and *y* ions were identified (Fig. S5). These included 24 deacetylated monomers containing di-, tri-, tetra-, and pentapeptide stems (g(-Ac)m-AE, g(-Ac)m-AEJ, g(-Ac)m-AEJX, and g(-Ac)m-AEJAX) and 10 fully acetylated monomers (gm-AE, gm-AEJ, gm-AEJX, and gm-AEJAX), where X can be any amino acid. A database called DB_0 made of these 34 monomers was used to perform a PGFinder search (step 2 in Fig. S4) to identify the most abundant monomers. Thirteen disaccharide-peptides accounting for more than 98% of the monomers identified were selected to create a second database (DB_1) containing dimers resulting from 3-3 and 4-3 crosslinking. A third PGFinder search (step 3 in Fig. S4) was performed to identify the most abundant dimers and generate the next database containing monomers, dimers, and trimers (DB_2). The next search with PGFinder and DB_2 (step 4 in Fig. S4) identified the most abundant trimers. A final database called DB_3 was created using all the information from sequential searches; it contained all MS/MS-checked monomers, 26 dimers, 16 trimers, and all AnhydroMurNAc derivatives of the 10 most abundant mono-, di-, and trimers, as well as 4 unusually cross-linked structures identified during in vitro assays.



Figure 4. TraDIS analysis of the ldtCd loci. The number of transposition events in ldtCd1 (A), ldtCd2 (B), and ldtCd3 (C) are shown as histograms depicting the localization (x-axis) and the frequency (y-axis) of transposon insertion sites (in *red* are antisense insertions; in *blue* are sense insertions).

The result of the PGFinder search using DB_3 and biological replicates from the WT and triple mutant is described in Table 1. The search strategy described here combining both LC-MS and LC-MS/MS analysis allowed us to identify 97 muropeptides, which is an unprecedentedly detailed analysis. The comparison between the two strains revealed a remarkable similarity between the two peptidoglycan compositions (Table 2). No significant difference was found when comparing the proportion of monomers, dimers, trimers, or glycan chain length. Crosslinking index as well as the proportion of 3-3 cross-links was also similar. The only difference found was a significant decrease in the exchange reaction (23.4 ± 0.7% in the WT and 16.9 ± 1.4% in the triple mutant). Overall, our analysis therefore demonstrated that the three canonical L,D-transpeptidases Ldt_{Cd1}, Ldt_{Cd2}, and Ldt_{Cd3} only

contribute marginally to remodel the peptidoglycan of *C. difficile* vegetative cells.

Comparative phenomics of the parental R20291 strain and its isogenic Δ 3ldt derivative

A comprehensive set of experiments were carried out to compare the phenotype of the R20291 strain and the triple ldt_{Cd} mutant. As expected, based on the results from peptidoglycan analysis, no significant differences were observed between the two strains in cell size (Fig. S6), sporulation (Fig. S7), or toxin release (Fig. S8). Minimum inhibitory concentrations (MICs) for several beta-lactams were also tested for all the mutants generated in this study and did not reveal any difference in the resistance against any of these antibiotics (Table S2).

Discussion

Recombinant L,D-transpeptidases represent a class of enzymes that are amenable to study in vitro since they can use soluble peptidoglycan fragments as a substrate. In vitro assays with distinct peptidoglycan fragments purified from intact sacculi were used to explore the catalytic activities of the three C. difficile Ldts. Our comparative analysis based on an LC-MS/MS assay with several substrates provided information about the preferential activity of each Ldt. Previous studies reported that Ldt_{Cd1} only displayed carboxypeptidase activity (8). Our data confirmed this result and also revealed that it can also perform transpeptidation and exchange reactions, even though this enzyme was poorly active on all substrates tested. Ldt_{Cd2} was able to perform all reactions but preferentially acted as a carboxypeptidase. Remarkably, Ldt_{Cd2} was the only enzyme with endopeptidase activity, exclusively using 3-3 dimers as a substrate. Unlike other endopeptidases that cleave 3-3 cross-links (MepA, MepM, and MepK), Ldt_{Cd2} has a strict substrate specificity for 3-3 cross-links since no activity against 4-3 cross-links could be detected (12, 13) (Table S3). Ldt_{Cd2} activity therefore appears to be unique since it is the first enzyme described that is only active on 3-3 cross-links.

 Ldt_{Cd3} displayed the highest transpeptidase/exchange activity and a relatively weak carboxypeptidase and endopeptidase activity, preferentially against 3-3 cross-linked dimers. Unlike Ldt_{Cd2} , Ldt_{Cd3} could cleave 4-3 dimers with low efficiency.

Besides the exhaustive description of expected Ldt_{Cd} activities, our *in vitro* assays also revealed that Ldt_{Cd2} and Ldt_{Cd3} can generate a novel type of peptidoglycan cross-links. These result from double transpeptidation reactions that use meso-DAP both as a donor and an acceptor group. Interestingly, we identified double cross-linked dimers containing either two 3-3 cross-links or a mixture of 4-3 and 3-3 cross-links. In hindsight, this result is not entirely surprising since the catalytic reaction leading to this type of bond is the same as the reaction leading to the formation of "normal" 3-3 bonds. Double cross-links can be detected in the peptidoglycan from *C. difficile* as well as in the peptidoglycan from Gram-negative

Novel peptidoglycan remodeling activities in C. difficile



Figure 5. HPLC-MS chromatogram of C. difficile reduced disaccharide-peptides. Each total ion chromatogram corresponds to a biological replicate of strain R20291 (WT) (A) and its isogenic derivative with in-frame deletions in genes *ldt_{Cd1}*, *ldt_{Cd2}*, and *ldt_{Cd3}* (Δ3*ldt*) (B).

organisms producing Ldts (*E. coli, Rhizobium leguminosarum,* and *Brucella abortus*; S. Mesnage, unpublished). Interestingly, double 4-3 cross-links resulting from D,D-transpeptidation have also been described in *Staphylococcus aureus* (14). The physiological role of these peptidoglycan cross-links remains unknown and awaits further studies.

Based on the impact of L,D-transpeptidation on antibiotic resistance in Enterococcus faecium, it is tempting to assume that L,D-transpeptidation in C. difficile could underpin betalactam resistance. This remains an open question since we were unable to generate a mutant devoid of 3-3 cross-links. The mutant harboring deletions in the genes encoding the three canonical Ldts still contained 78% of 3-3 cross-links, indicating that this organism encodes (an)other enzyme(s) that does not contain a YkuD domain but is (are) able to make 3-3 cross-links. Our findings are surprising and somewhat contrasting with a previous study, where the combined deletion of ldt_{Cd1} and ldt_{Cd2} led to a decrease in the cross-linking index (18.2% as compared with 33.8% for the WT strain). This discrepancy is difficult to explain but could be attributed either to the different strain analyzed (C. difficile 630 versus R20291) or to the different strategies followed for peptidoglycan analysis. The work by Peltier et al. involved offline analysis of individual fractions from a single replicate and a quantification of muropeptides based on UV whilst our analysis involved LC-MS analysis (online) of biological triplicates and a quantification using ion intensity. Based on the remarkably similar peptidoglycan structure of the WT and mutant strain described here (Tables 1 and 2), we are confident that the three canonical Ldt_{Cd} enzymes have a minor contribution to the formation of 3-3 cross-links and only contribute significantly to incorporate noncanonical amino acids in the position 4 of peptide stems.

The interaction of recombinant Ldts with beta-lactams has been extensively studied in vitro, and the data reported support the idea that these enzymes play a role in resistance to these antibiotics. Ldts are acylated by beta-lactams, but enzyme inactivation only occurs in the presence of carbapenems and penems. Other beta-lactams such as cephems (cephalosporin) are poor inhibitors since acylation is slow and the thioester bond formed in the enzyme-antibiotic adduct is prone to hydrolysis (15). This has been shown for model organisms including E. faecium (15), Mycobacterium tuberculosis (16), and C. difficile Ldts (8) and is true for most Ldts studied to date despite some exceptions for M. tuberculosis Ldt_{Mt5} (16), Acinetobacter baumanii Ldt_{Ab} (17), and C. difficile Ldt_{Cd3} (8), which are not inhibited by carbapenems. The contribution of Ldts to betalactam resistance has been documented in E. faecium (18), M. tuberculosis (19), Mycobacterium smegmatis (20), and A. baumanii (17). In C. difficile, the inactivation of two of the three Ldts did not lead to a change in beta-lactam resistance (6). Our data revealed that the inactivation of all canonical C. difficile ldts has no impact on beta-lactam resistance, as expected, based on the results from peptidoglycan analysis.

The identification of alternative (noncanonical) Ldt(s) encoded by *C. difficile* is therefore required to investigate (i) whether L,D-transpeptidation is essential in this organism and (ii) whether this mode of peptidoglycan polymerization underpins beta-lactam resistance.

Experimental procedures

Bacterial strains, plasmids, oligonucleotides, and growth conditions

Bacterial strains, plasmids, and oligonucleotides are described in Table S4. C. difficile R20291 (ribotype 027) and

SASBMB

Table 1

Muropeptide analysis of	C. difficile WT	and triple Idtcd1,	Idtcd2, Idtcd3	mutant (Δ3 <i>ldt</i>)

	I dt activity/	ince (%) ^c	%) ^c		
Muropeptide ^a	crosslink ^b	WT	$\Delta 3 l d t$	RT (min)	Δppm
gm(-Ac)-AEIA 1	N/A	21.594% ± 1.050%	28.192% ± 4.671%	7.90 ± 0.02	2.71
gm(-Ac)-AEJ 1	Carboxypeptidase	$9.002\% \pm 1.348\%$	$8.235\% \pm 1.026\%$	5.39 ± 0.03	2.85
gm(-Ac)-AEJG 1	Exchange	$9.814\% \pm 0.907\%$	$4.696\%~\pm~1.177\%$	6.16 ± 0.02	2.55
gm(-Ac)-AEJF 1	Exchange	$1.820\% \pm 0.806\%$	$2.083\% \pm 0.886\%$	20.72 ± 0.03	2.14
gm(-Ac)-AE 1	N/A	$3.174\% \pm 0.185\%$	$2.297\% \pm 0.520\%$	7.74 ± 0.05	2.31
gm(Ac) A FIV[1]	N/A Exchange	$0.512\% \pm 0.061\%$ $0.944\% \pm 0.111\%$	$0.716\% \pm 0.211\%$ $0.570\% \pm 0.113\%$	8.84 ± 0.02 13.83 ± 0.02	1.50
gm(-Ac)-AEII	Exchange	$0.944\% \pm 0.111\%$ $0.294\% \pm 0.076\%$	$0.389\% \pm 0.113\%$	13.83 ± 0.02 18.20 ± 0.04	0.66
gm(-Ac)-AEJY 1	Exchange	0.172%	0.328% 0.072%	15.82 ± 0.04	0.99
gm(-Ac)-AEJAA 1	N/A	$0.244\% \pm 0.007\%$	$0.311\% \pm 0.053\%$	9.63 ± 0.08	1.68
gm-AEJA (Anh) 1	N/A	$0.168\% \pm 0.010\%$	$0.270\% \pm 0.085\%$	7.97 ± 0.11	2.24
gm(-Ac)-AEJS 1	Exchange	$0.302\% \pm 0.036\%$	$0.234\% \pm 0.039\%$	5.52 ± 0.02	2.04
gm-AEJ[1	N/A	$0.197\% \pm 0.025\%$	$0.207\% \pm 0.130\%$	9.41 ± 4.13	1.89
gm(-Ac)-AE O 1 $gm(-Ac)-AE AG 1$	Exchange	$0.324\% \pm 0.086\%$ $0.107\% \pm 0.011\%$	$0.184\% \pm 0.055\%$ $0.095\% \pm 0.022\%$	4.55 ± 0.06 8 41 + 1 39	2.08
gm(-Ac)-AEIA (Anh) 1	N/A	$0.033\% \pm 0.002\%$	$0.084\% \pm 0.010\%$	11.35 ± 0.02	0.31
gm-AEJG 1	Exchange	$0.115\% \pm 0.026\%$	$0.071\% \pm 0.010\%$	7.74 ± 0.75	1.08
gm-AE 1	N/A	$0.081\% \pm 0.016\%$	$0.062\% \pm 0.019\%$	8.65 ± 0.03	0.41
gm(-Ac)-AEJAT 1	N/A	$0.177\% \pm 0.021\%$	$0.096\% \pm 0.022\%$	6.63 ± 0.04	0.61
gm(-Ac)-AEJAF 1	N/A	$0.042\% \pm 0.015\%$	$0.055\% \pm 0.018\%$	24.36 ± 0.02	1.30
gm-AEJI I gm(Ag) AEINII	Exchange	$0.015\% \pm 0.009\%$	$0.042\% \pm 0.028\%$	16.08 ± 4.43	0.88
gm(-Ac)-AE H 1	Exchange	$0.235\% \pm 0.005\%$ $0.104\% \pm 0.080\%$	0.002% ± 0.004%	5.05 ± 0.05 5.19 + 0.20	2.01
gm(-Ac)-AEIE[1	Exchange	$0.062\% \pm 0.030\%$	$0.041\% \pm 0.003\%$	7.51 ± 0.03	1.37
gm-AEIF 1	Exchange	$0.019\% \pm 0.013\%$	$0.020\% \pm 0.014\%$	21.95 ± 0.01	1.35
gm-AEJK 1, gm(-Ac)-AEJAV 1 (mixture)	Exchange	$0.000\% \pm 0.000\%$	$0.000\% \pm 0.000\%$	6.69 ± 3.18	1.88
gm(-Ac)-AEJAI 1	N/A	$0.021\% \pm 0.003\%$	$0.023\%~\pm~0.003\%$	19.36 ± 0.02	0.76
gm(-Ac)-AEJAK 1	N/A	$0.019\% \pm 0.001\%$	$0.024\% \pm 0.006\%$	7.22 ± 0.12	1.79
gm(-Ac)-AE)D[1	N/A	$0.125\% \pm 0.027\%$	$0.024\% \pm 0.004\%$	6.27 ± 0.04	2.03
gm(Ac) A FIT 1	Exchange	$0.014\% \pm 0.007\%$ $0.025\% \pm 0.002\%$	$ND = 0.020\% \pm 0.002\%$	6.40 ± 0.02 6.40 ± 0.14	1.09
gm(-Ac)-AE (Anh) 1	N/A	$0.029\% \pm 0.002\%$	$0.015\% \pm 0.002\%$	12.00 ± 0.01	1.58
gm-AEJAA 1	N/A	$0.010\% \pm 0.001\%$	$0.010\% \pm 0.000\%$	10.79 ± 1.54	1.10
gm(-Ac)-AEJ (Anh) 1	N/A	$0.025\% \pm 0.005\%$	$0.013\% \pm 0.003\%$	8.56 ± 0.01	0.58
gm(-Ac)-AEJAY 1	N/A	0.005%	$0.005\% \pm 0.004\%$	17.18 ± 0.03	0.22
gm(-Ac)-AEJG (Anh) 1	Exchange	$0.015\% \pm 0.004\%$	$0.006\% \pm 0.003\%$	9.31 ± 0.10	0.46
gm(-Ac)-AEJAA (Anh) 1	Exchange	ND	$0.001\% \pm 0.000\%$	9.49 ± 0.04	1.74
$gm_A FIA F[1]$	N/A	$0.001\% \pm 0.001\%$	0.000% $0.001\% \pm 0.001\%$	24.42 ± 0.03 24.07 ± 0.01	1.59
gm(-Ac)-AEIA=gm(-Ac)-AEI/2	3-3	$16.323\% \pm 0.074\%$	$19.316\% \pm 1.031\%$	13.09 ± 0.08	1.31
gm(-Ac)-AEJA=gm(-Ac)-AEJA 2	3-4	6.806% ± 1.727%	8.963% ± 1.708%	13.93 ± 0.11	1.20
gm(-Ac)-AEJ=gm(-Ac)-AEJ 2	3-3	$5.903\% \pm 1.156\%$	$5.446\% \pm 0.636\%$	11.99 ± 0.03	1.57
gm(-Ac)-AEJG=gm(-Ac)-AEJ 2	3-3	5.741% ± 0.730%	$2.688\% \pm 0.881\%$	11.54 ± 0.04	1.39
gm-AEJA=gm(-Ac)-AEJ/2	3-3	$1.238\% \pm 0.141\%$	$1.677\% \pm 0.398\%$	13.80 ± 0.03	1.77
gm(-Ac)-AE)F=gm(-Ac)-AE)/2 $gm_AEIA = gm(-Ac)-AEIA/2$	3-3	$0.682\% \pm 0.213\%$ $0.560\% \pm 0.052\%$	$0.719\% \pm 0.283\%$ $0.713\% \pm 0.151\%$	22.56 ± 0.01 14.78 ± 0.08	1.85
gm(-Ac)-AEIG=gm(-Ac)-AEIA 2	3-4	$1.112\% \pm 0.202\%$	$0.467\% \pm 0.131\%$	12.40 ± 0.00	0.68
gm-AEJA=gm(-Ac)-AEJ (Anh) 2	3-3	$0.244\% \pm 0.036\%$	$0.374\% \pm 0.131\%$	19.22 ± 0.10	0.83
gm(-Ac)-AEJV=gm(-Ac)-AEJ 2	3-3	$0.308\% \pm 0.047\%$	$0.205\% \pm 0.057\%$	17.77 ± 0.02	0.57
gm(-Ac)-AEJA=gm(-Ac)-AEJ (Anh) 2	3-3	$0.337\% \pm 0.040\%$	$0.244\% \pm 0.023\%$	15.68 ± 0.32	1.47
gm(-Ac)-AEJI=gm(-Ac)-AEJ 2	3-3	$0.138\% \pm 0.055\%$	$0.176\% \pm 0.073\%$	20.98 ± 0.02	0.88
gm-AE = gm(-Ac)-AE 2 gm(-Ac)-AE A = gm(-Ac)-AE A = AE A = b 2	3-3	$0.250\% \pm 0.011\%$ $0.112\% \pm 0.007\%$	$0.250\% \pm 0.092\%$ $0.176\% \pm 0.011\%$	12.69 ± 0.03 16.00 ± 0.02	1.26
gm(-Ac)-AE/A=gm(-Ac)-AE/A(Am)/2 gm(-Ac)-AF/E=gm(-Ac)-AF/A/2	3-4	$0.113\% \pm 0.007\%$ $0.161\% \pm 0.061\%$	$0.170\% \pm 0.011\%$ $0.130\% \pm 0.043\%$	10.99 ± 0.02 23.14 + 0.02	1 43
gm(-Ac)-AEIY=gm(-Ac)-AEII2	3-3	$0.063\% \pm 0.084\%$	$0.107\% \pm 0.034\%$	18.74 ± 1.18	0.37
gm(-Ac)-AEJS=gm(-Ac)-AEJ 2	3-3	$0.207\% \pm 0.005\%$	$0.133\% \pm 0.014\%$	11.41 ± 0.14	0.63
gm-AEJA=gm(-Ac)-AEJA (Anh) 2	3-4	$0.072\% \pm 0.019\%$	$0.125\% \pm 0.012\%$	14.02 ± 0.30	9.40
gm(-Ac)-AEJK=gm(-Ac)-AEJ 2	3-3	0.154% ± 0.033%	$0.080\% \pm 0.022\%$	10.20 ± 0.02	0.27
gm(-Ac)-AEJI=gm(-Ac)-AEJA 2	3-4	$0.037\% \pm 0.013\%$	$0.042\% \pm 0.016\%$	21.78 ± 0.02	0.42
gm(-Ac)-AEJV = gm(-Ac)-AEJA/2	3-4	$0.06\% \pm 0.006\%$	$0.038\% \pm 0.011\%$	18.60 ± 0.02 12.00 ± 0.02	0.35
gm(-Ac)-AEJS=gm(-Ac)-AEJA/2 gm(-Ac)-AEJA/2	3-4	$0.033\% \pm 0.006\%$ $0.013\% \pm 0.006\%$	$0.046\% \pm 0.008\%$ $0.027\% \pm 0.002\%$	12.09 ± 0.02 13.96 ± 0.51	0.48
gm(-Ac)-AEI=gm(-Ac)-AEI (Anh) 2	3-3	$0.010\% \pm 0.000\%$ $0.110\% \pm 0.021\%$	0.011%	14.84 ± 0.22	0.30
gm(-Ac)-AEJN=gm(-Ac)-AEJ 2	3-3	0.094% ± 0.025%	$0.028\% \pm 0.005\%$	10.88 ± 0.02	0.64
gm(-Ac)-AEJK=gm(-Ac)-AEJA 2	3-4	$0.044\% \pm 0.008\%$	$0.027\%~\pm~0.002\%$	10.97 ± 0.02	0.56
gm(-Ac)-AEJY=gm(-Ac)-AEJA 2	3-4	$0.013\% \pm 0.000\%$	$0.021\% \pm 0.005\%$	18.92 ± 0.02	0.31
gm(-Ac)-AEJA=gm(-Ac)-AEJ (-H2O) 2	3-3*	$0.138\% \pm 0.000\%$	$0.016\% \pm 0.005\%$	14.67 ± 0.54	0.55
gm(-Ac)-AEJN=gm(-Ac)-AEJA/2 gm(-Ac)-AEJG=gm(-Ac)-AEJA/(Apb)/2	3-4	$0.021\% \pm 0.006\%$	$0.009\% \pm 0.001\%$	11.62 ± 0.02 15.38 ± 0.00	1.08
gm(-Ac)-AE (F=gm(-Ac)-AE) (Anh)/2	3-3	0.004%	0.000% $0.014\% \pm 0.014\%$	20.96 ± 4.59	5.28
gm(-Ac)-AEJG=gm(-Ac)-AEJ (Anh) 2	3-3	$0.126\% \pm 0.017\%$	0.001%	15.56 ± 2.77	0.79
gm(-Ac)-AEJI=gm(-Ac)-AEJ (Anh) 2	3-3	$0.002\% \pm 0.001\%$	$0.001\% \pm 0.000\%$	23.32 ± 0.03	1.04
gm(-Ac)-AEJV=gm(-Ac)-AEJ (Anh) 2	3-3	$0.015\% \pm 0.011\%$	0.001%	19.18 ± 3.13	2.32
gm(-Ac)-AEJA=gm(-Ac)-AEJ=gm(-Ac)-AEJ 3	3-3	$3.428\% \pm 0.114\%$	$3.041\% \pm 0.233\%$	15.39 ± 0.04	1.42
gm(-Ac)-AEJA=gm(-Ac)-AEJA=gm(-Ac)-AEJ 3	3-3, 3-4	$1.823\% \pm 0.087\%$	$2.273\% \pm 0.327\%$	15.98 ± 0.02	1.09
gm-AEJA=gm(-Ac)-AEJA=gm(-Ac)-AEJ[3	3-3, 3-4	$0.344\% \pm 0.047\%$ 0.505% ± 0.052%	$0.446\% \pm 0.133\%$ $0.449\% \pm 0.135\%$	16.76 ± 0.02 16.19 ± 0.02	0.27
gm(-Ac)-AEIA=gm(-Ac)-AEIA=gm(-Ac)-AEIA 3	3-4	$0.348\% \pm 0.052\%$	$0.468\% \pm 0.155\%$	16.55 ± 0.02	0.40
a contraction of the second se	· ·			0.00	

8 J. Biol. Chem. (2024) 300(1) 105529

SASBMB

Table 1—Continued

	I dt activity/	Abundance (%) ^c			Averaged
Muropeptide ^a	crosslink ^b	WT	$\Delta 3ldt$	RT (min)	Δppm
gm-AEJA=gm(-Ac)-AEJ=gm(-Ac)-AEJ (Anh) 3	3-3	0.259% ± 0.026%	0.342% ± 0.073%	20.34 ± 0.03	1.20
gm(-Ac)-AEJ=gm(-Ac)-AEJ=gm(-Ac)-AEJ 3	3-3	$0.638\% \pm 0.165\%$	0.395% ± 0.046%	14.46 ± 0.02	0.48
gm(-Ac)-AEJG=gm(-Ac)-AEJ=gm(-Ac)-AEJ 3	3-3	0.565% ± 0.090%	$0.194\% \pm 0.068\%$	14.13 ± 0.02	0.23
gm-AEJA=gm(-Ac)-AEJA=gm(-Ac)-AEJA 3	3-4	0.095% ± 0.015%	$0.142\% \pm 0.032\%$	17.30 ± 0.02	0.31
gm(-Ac)-AEJG=gm(-Ac)-AEJA=gm(-Ac)-AEJ 3	3-3, 3-4	0.253% ± 0.052%	0.115% ± 0.038%	14.69 ± 0.01	0.31
gm-AEJA=gm(-Ac)-AEJA=gm(-Ac)-AEJ (Anh) 3	3-3, 3-4	0.097% ± 0.008%	0.136% ± 0.017%	20.18 ± 1.58	2.32
gm(-Ac)-AEJF=gm(-Ac)-AEJ=gm(-Ac)-AEJ 3	3-3	$0.061\% \pm 0.018\%$	$0.041\% \pm 0.019\%$	23.03 ± 0.01	1.05
gm(-Ac)-AEJA=gm(-Ac)-AEJ=gm(-Ac)-AEJ (Anh) 3	3-3	$0.174\% \pm 0.019\%$	$0.049\% \pm 0.008\%$	17.55 ± 0.16	0.30
gm(-Ac)-AEJA=gm(-Ac)-AEJA=gm(-Ac)-AEJ (Anh) 3	3-3, 3-4	$0.062\% \pm 0.007\%$	$0.051\% \pm 0.005\%$	18.40 ± 0.09	0.40
gm(-Ac)-AEJF=gm(-Ac)-AEJ=gm(-Ac)-AEJA 3	3-3, 3-4	0.036% ± 0.013%	$0.026\% \pm 0.010\%$	23.39 ± 0.02	0.40
gm(-Ac)-AEJG=gm(-Ac)-AEJA=gm(-Ac)-AEJA 3	3-4	$0.047\% \pm 0.012\%$	$0.021\% \pm 0.008\%$	15.23 ± 0.01	0.59
gm(-Ac)-AEJG=gm(-Ac)-AEJ=gm(-Ac)-AEJ (Anh) 3	3-3	$0.047\% \pm 0.007\%$	ND	16.24 ± 0.02	0.22
gm(-Ac)-AEJ=gm(-Ac)-AEJ=gm(-Ac)-AEJ (-H2O) 3	3-3*	0.056% ± 0.015%	ND	16.98 ± 0.03	1.18
gm(-Ac)-AEJA=gm(-Ac)-AEJA=gm(-Ac)-AEJA (Anh) 3	3-4	$0.013\% \pm 0.001\%$	0.017% ± 0.002%	19.02 ± 0.04	0.40
gm(-Ac)-AEJG=gm(-Ac)-AEJ=gm(-Ac)-AEJA (Anh) 3	3-3, 3-4	$0.011\% \pm 0.001\%$	ND	16.82 ± 0.12	0.25
gm(-Ac)-AEJA=gm(-Ac)-AEJ=gm(-Ac)-AEJ (-H2O) 3	3-3	0.002%	ND	16.83 ± 0.00	2.64
gm-AEJA=gm(-Ac)-AEJA=gm(-Ac)-AEJA (Anh) 3	3-4	$0.004\% \pm 0.006\%$	$0.016\% \pm 0.007\%$	17.35 ± 1.74	4.89
gm(-Ac)-AEJ=gm(-Ac)-AEJ=gm(-Ac)-AEJ (Anh) 3	3-3	0.083% ± 0.006%	$0.005\% \pm 0.003\%$	16.68 ± 0.28	1.56
gm(-Ac)-AEJF=gm(-Ac)-AEJ=gm(-Ac)-AEJ (Anh) 3	3-3	$0.002\% \pm 0.000\%$	$0.005\% \pm 0.005\%$	23.51 ± 3.39	3.09
gm(-Ac)-AEJ=gm(-Ac)-AEJ=gm(-Ac)-AEJ (Anh) 3	3-3	$0.083\% \pm 0.006\%$	$0.005\% \pm 0.003\%$	16.68 ± 0.28	1.56
gm(-Ac)-AEJF=gm(-Ac)-AEJ=gm(-Ac)-AEJ (Anh) 3	3-3	$0.002\% \pm 0.000\%$	$0.005\% \pm 0.005\%$	23.51 ± 3.39	3.09

Abbreviation: RT, retention time.

⁴ When multiple dimer structures are possible, the most likely structure is proposed based on the abundance of acceptor stems.

^b N/A, not applicable; 3-3* crosslinks correspond to L,D-transpeptidation products doubly crosslinked.

^e ND, not detected; no standard deviation is provided when muropeptides were only identified in a single replicate.

^d Absolute mass difference between observed and theoretical mass in parts per million (ppm).

isogenic derivatives were grown on BHI agar plates or in TY broth. During selection of mutants, strains were grown on *C. difficile* minimal medium (21) supplemented with 5-fluorocytosine (50 μ g/ml) when required. Cultures were incubated at 37 °C in an anaerobic cabinet under an atmosphere containing 80% nitrogen, 10% hydrogen, and 10% carbon dioxide. *E. coli* was grown on Luria Bertani (LB) agar plates or in LB broth at 37 °C. When needed, thiamphenicol was added (30 μ g/ml).

Construction of C. difficile deletion mutants

C. difficile mutant strains were constructed by homologous recombination. Briefly, 1.2 kb upstream and downstream of the region to be deleted was synthesized as a single DNA fragment (Genewiz) and cloned between BamHI and SacI sites in pJAK112, yielding pNG007 (ldt_{Cd1} deletion), pNG008 (ldt_{Cd2} deletion), and pNG009 (ldt_{Cd3} deletion). Plasmids were introduced into *C. difficile* strain R20291 by conjugation (22), and allelic exchange was carried out as described (10).

Determination of minimum inhibitory concentrations

MICs were determined according to an agar dilution method using Wilkins Chalgren agar and as recommended by the Clinical and Laboratory Standards Institute guidelines. *C. difficile* isolates were cultured on fresh blood agar plates,

Table 2

Summary of WT and 3Δldt PG properties

Muropeptides/properties	WT	$\Delta 3 l dt$	
Monomers	49.93% ± 2.82%	49.96% ± 3.13%	
Dimers	$41.12\% \pm 2.15\%$	41.94% ± 2.19%	
Trimers	$8.95\% \pm 1.00\%$	8.09% ± 1.04%	
Cross-linking index	$23.51\% \pm 1.40\%$	23.64% ± 1.44%	
Glycan chain length	96.6 ± 1.8	87.3 ± 18.7	
gm-AEJX	23.39% ± 0.70%	$16.86\% \pm 1.41\%$	
3-3 cross-links	81.33% ± 2.54%	78.35% ± 2.68%	

prior to inoculation of single colonies into prereduced Schaedler Anaerobic Broths and anaerobic culture for 24 h. Cultures were diluted in prereduced phosphate-buffered saline to achieve a 1 McFarland standard equivalent, and 10^5 colony-forming units were spotted on Wilkins Chalgren agar containing doubling antibiotic dilutions and non-antibiotic-containing controls. Agar plates containing amoxicillin clavulanate were prepared with a fixed concentration of 2 mg/l clavulanate, and those containing piperacillin tazobactam were prepared with a fixed concentration of 4 mg/l tazobactam, as recommended by European Committee on Antimicrobial Susceptibility Testing guidelines. Agar plates were incubated anaerobically for 48 h before reading. The MIC was defined as the lowest concentration of antibiotic completely preventing growth, significantly reducing it to a haze or one to three discrete colonies.

Chromosomal DNA extraction, sequencing, and genome analysis

Genomic DNA was purified using phenol-chloroform extraction as described (10), and whole genome sequencing was performed by MicrobesNG using their standard Illumina service. Sequence analysis was performed using a custom script, as described (23). In brief, reads were aligned to the *C. difficile* R20291 reference (accession number: FN545816) using BWA-mem (v0.7.17) and sorted using SAMtools (v1.43) (24). PCR duplicates were removed *via* Picard (v2.25.2) (http://broadinstitute.github.io/picard/). SAMtools (v1.43) mpileup was used to generate the mpileup prerequisite for Varscan. Varscan (v2.4.3-1) (25) was then used to call variants using parameters previously described (23), and snpEff (v5.0) (26) was used to annotate variants. Variants that co-occurred in the WT were removed to generate a list of mutations unique to mutant strains. Mutations were visualized on the genome
Novel peptidoglycan remodeling activities in C. difficile

using a previously published custom script in RStudio (v4.1.0) using the Plotrix package (27).

TraDIS analysis

The construction of the transposon library, the sequencing of insertion sites, and the mapping to their corresponding reference sequences were described (10). Visualization of insertion sites was done using the Artemis genome browser (28).

Peptidoglycan extraction

C. difficile strains were grown overnight in 10 ml of TY broth from a single colony. The starter cultures were used to inoculate 100 ml TY medium (1/100 dilution). After 48 h at 37 °C, cells were spun, supernatant was discarded, and cell pellet was snap frozen in liquid nitrogen; the cell pellet was then resuspended in 20 ml of boiling MilliQ water (MQ) before the addition of 5 ml warm 20% (w/v) SDS (4% SDS final concentration). After 30 min at 100 °C, the cells were allowed to cool down to room temperature. Insoluble cell walls were pelleted at 45,000g for 20 min and washed 5 times using warm MQ water. Proteins covalently bound to peptidoglycan were removed by pronase treatment (final concentration of 2 mg/ml, 4 h at 60 °C). Protease-treated cell walls were washed 6 times with 30 ml of MQ water before covalently bound polymers were removed by incubation in 1 M HCl for 5 h at 37 °C. Insoluble pure peptidoglycan was washed 6 times with MQ water, snap frozen in liquid nitrogen, freeze-dried and resuspended at a final concentration of 10 mg/ml.

Ldt_{cd} production and purification

The plasmids for protein production were designed as described (8). Ldt_{Cd1} and Ldt_{Cd3} were expressed as full-length His-tagged proteins. Ldt_{Cd2} could not be produced as a stable full-length protein, so the catalytic domain was purified. Recombinant Ldt_{Cd} were produced in E. coli BL21(DE3) grown in LB broth. One-liter cultures were inoculated at an OD_{600nm} of 0.05, and protein expression was induced with 1 mM isopropyl ß-D-1-thiogalactopyranoside when the cultures reached an OD_{600nm} of 0.7. They were then cooled down to 20 °C and incubated for 16 h at this temperature. Cells were harvested, resuspended in a buffer containing 50 mM Tris-HCl (pH8.0) + 500 mM NaCl, and mechanically broken using a French press (2 passages at 1250 psi). Cell debris were removed by centrifuging the crude cell extract at 45,000g for 30 min at 4 °C. The entire soluble fraction was loaded on a 5-ml HiTrap column equilibrated in buffer A at a flow rate of 5 ml/min. Elution was performed using a 10 column volume gradient to 250 mM imidazole in buffer A. Fractions containing the Ldt_{Cd} proteins were pooled, concentrated to 2 mg/ml and further purified by gel filtration chromatography using a Hiload 16/600 superdex 75 column equilibrated in 50 mM Tris-HCl pH 8.5 + 250 mM NaCl). Ldt_{Cd} proteins were concentrated on an Amicon centrifugal filter to a final concentration of 2 mg/ml and stored at -80 °C until further use.

Purification of substrates for in vitro assays

Peptidoglycan fragments used as substrates were purified from *E. coli* or *C. difficile* sacculi digested with mutanolysin and reduced with sodium borohydride. Digestion products were separated by reversed-phase HPLC using a Hypersil column (4.6 mm × 250 mm, 5 μ m particle size) using a water– acetonitrile–0.1% (v/v) formic acid gradient. Fractions containing the muropeptides of interest were freeze-dried and quantified by NMR using trimethylsilyl propionate as a standard.

In vitro Ldt assays

Each *in vitro* assay was carried out in triplicate, and average chromatograms are shown in Figure 1. Each reaction was carried out in a phosphate saline buffer (pH 8.0) in a final volume of 50 μ l and contained 100 μ M substrate and 10 μ M enzyme. For exchange reactions, D-methionine was added at a concentration of 1 mM. Reactions were incubated at 37 °C for 4 h.

Preparation of soluble muropeptides for peptidoglycan structural analysis

Purified peptidoglycan (1 mg) was digested overnight in 50 mM phosphate buffer (pH 5.5) supplemented with 200 U of mutanolysin (Sigma) in a final volume of 125 μ l. Following heat inactivation of mutanolysin (5 min at 100 °C), soluble disaccharide peptides were mixed with an equal volume of 250 mM borate buffer (pH 9.25) and reduced with 0.2% (w/v) sodium borohydride. After 20 min at room temperature, the pH was adjusted to 4.5 to 5.5 using phosphoric acid.

Ultrahigh-Performance chromatography coupled to tandem mass spectrometry

An Ultimate 3000 UHPLC (Dionex/Thermo Fisher Scientific) system coupled with a high-resolution Q Exactive Focus mass spectrometer (Thermo Fisher Scientific) was used for LC-MS analysis. Muropeptides were separated using a C18 analytical column (Hypersil Gold aQ, 1.9-µm particles, 150 mm × 2.1 mm; Thermo Fisher Scientific) at a temperature of 50 °C for peptidoglycan analysis or on a smaller C18 column for in vitro assays (Hypersil Gold aQ, 1.9-µm particles, 50 mm × 2.1 mm; Thermo Fisher Scientific). For peptidoglycan analysis, muropeptide elution was performed at 0.25 ml/min by applying a mixture of solvent A (water, 0.1% [v/v] formic acid) and solvent B (acetonitrile, 0.1% [v/v] formic acid). Liquid chromatography conditions were 0 to 12.5% B for 25 min increasing to 20% B for 10 min. After 5 min at 95%, the column was re-equilibrated for 10 min with 100% buffer A. For in vitro assays, a flow rate of 0.4 ml/min was used. PG fragments were eluted with a 5-min gradient to 15% B followed by 2 min at 95% B. The column was re-equilibrated for 6 min with 100% buffer A.

The Orbitrap Exploris 240 was operated under electrospray ionization (H-ESI high flow)-positive mode, full scan (m/z 150–2250) at resolution 120,000 (full width at half maximum) at m/z 200, with normalized AGC Target 100%, and automated

Novel peptidoglycan remodeling activities in C. difficile

maximum ion injection time. Data-dependent MS/MS were acquired on a "Top 5" data-dependent mode using the following parameters: resolution 30,000; AGC 100%, automated injection time, with normalized collision energy 25%.

Nuclear magnetic resonance

Purified peptidoglycan fragments were dissolved in 90% $H_2O/10\%$ D_2O . They were analyzed by NMR at 298 K on a Bruker DRX-600 equipped with a cryoprobe. TOCSY spectra were acquired using a 60-ms spin-lock with a field strength of 10 kHz. NOESY spectra used a 200-ms mixing time. All data were analyzed using Topspin 4.0.5.

Analysis of PG structure

LC-MS datasets were deconvoluted with the Byos software v3.11 (Protein Metrics). Sequential searches were carried out with PGFinder v1.0.3, with default settings (10 ppm tolerance, 0.5 min cleanup window) following the strategy described in Fig. S4. Data from individual matching output was consolidated as previously described to calculate average intensities, retention times, observed monoisotopic masses, and ppm differences. The output from individual searches and consolidated data are described in File S2). Cross-linking index and glycan chain length were determined as previously (29). The cross-linking index is defined as 0.5 * (% of dimers) + 0.66 * (% of trimers); glycan chain length was inferred from the abundance of anhydroMurNAc groups, which are found at the end of glycan chains. It is defined as 1/(% of AnhydroMurNAc monomers + 0.5 * (% of AnhydroMurNAc dimers) + 0.33 * (% of AnhydroMurNAc trimers).

Flow cytometry

Cells corresponding to biological replicates were grown overnight, diluted 1:100 into fresh broth (OD₆₀₀ ~ 0.02), and grown to mid-exponential phase (OD₆₀₀ ~ 0.5). Bacteria were diluted 1:100 in filtered phosphate-buffered saline and analyzed by flow cytometry using Millipore Guava Easy Cyte system. Light scatter data were obtained with logarithmic amplifiers for 2500 events. Forward scattered and side-scattered light values were compared using Student *t* test with Welch's correction using GraphPad Prism.

Sporulation

Sporulation efficiency was assessed as described (30). Briefly, stationary phase cultures of *C. difficile* were incubated anaerobically for 5 days and the total and heat-resistant (spore) colony-forming units (65 $^{\circ}$ C for 30 min) were determined every 24 h. Strains were assayed in technical triplicate and the data presented as the mean and standard deviation.

Toxin release assays

Toxin production was detected in whole cell lysates or concentrated culture supernatants by Western blot. For both fractions, material corresponding to the equivalent of 20 ml of culture at $OD_{600nm} = 1$ was loaded onto a 6% SDS PAGE,

SASBMB

transferred on a polyvinylidene fluoride membrane, and probed with a mouse monoclonal antibody (MA1-7413, Thermo Fisher) against toxin B at a 1/1000 dilution. A secondary rabbit anti-mouse antibody coupled to horseradish peroxidase (#31450, Thermo Fisher) was used at a 1/10,000 dilution. Blots were revealed by chemiluminescence using a BioRad chemidoc system.

Data availability

LC-MS/MS datasets have been deposited in the GLYCO-POST repository (GPST000371). NMR assignments have been deposited in the Biological Magnetic Resonance Data Bank (52169). Sequencing data were deposited in the NCBI Sequence Read Archive (SRA) under Bioproject ID PRJNA1026070.

Supporting information—This article contains supporting information (Figs. S1–S8, Tables S1–S4 and Files S1 and S2).

Acknowledgments—The authors would like to thank Nick W. Turner (University of Sheffield) for insightful suggestions to draw unusual peptidoglycan fragments.

Author contributions—S. M. conceptualization; N. F. G. and S. M. data curation; S. M. formal analysis; S. M., M. H. W., R. P. F., and A. L. L. funding acquisition; N. F. G., D. G., M. G. A.-Z., M. P. W., C. A. E., W. D. S., J. E. B., and J. F. investigation; N. F. G., C. A. E., M. J. D., and R. P. F. methodology; S. M. and G. L. D. project administration; S. M., R. P. F., and G. L. D. supervision; N. F. G. and S. M. validation; M. G. A.-Z., M. P. W., and S. M. visualization; S.M. writing – original draft; N. F. G., D. G., M. G. A.-Z., M. P. W., C. A. E., J. F., G. L. D., A. L. L., R. P. F., and S. M. writing – review and editing.

Funding and additional information—This work was supported by a Medical Research Council grant to S. M., R. F. P., M. H. W., and A. L. L. (MR/S009272/1). C. A. E. and M. J. D. received support from the Biotechnology and Biological Science Research Council (BB/ M012166/1). M. G. A.-Z. is funded by a Mexican government PhD scholarship (CONAHCYT, 2021-00007-01EXTF-00221); J. E. B. is funded by a studentship from the MRC Discovery Medicine North (DiMeN) Doctoral Training Partnership (MR/R015902/1).

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: DAP, diaminopimelic acid; GlcN, glucosamine; LC-MS/MS, liquid chromatography coupled to mass spectrometry; Ldt, L,D-transpeptidase; MQ, MilliQ water; MS/MS, tandem mass spectrometry; MurNAc, *N*-acetylmuramic acid; TraDIS, transposon-directed insertion site sequencing; MIC, minimum inhibitory concentration.

References

 Wingen-Heimann, S. M., Davies, K., Viprey, V. F., Davis, G., Wilcox, M. H., Vehreschild, M., et al. (2023) *Clostridioides difficile* infection (CDI): a pan-European multi-center cost and resource utilization study, results from the Combatting Bacterial Resistance in Europe CDI (COMBACTE-CDI). *Clin. Microbiol. Infect.* 29, 651.e1–651.e8

J. Biol. Chem. (2024) 300(1) 105529 11

Novel peptidoglycan remodeling activities in C. difficile

- Lessa, F. C., Gould, C. V., and McDonald, L. C. (2012) Current status of *Clostridium difficile* infection epidemiology. *Clin. Infect. Dis.* 55, S65–S70
- Buddle, J. E., and Fagan, R. P. (2023) Pathogenicity and virulence of Clostridioides difficile. Virulence 14, 2150452
- Frere, J. M. (1977) Mechanism of action of beta-lactam antibiotics at the molecular level. *Biochem. Pharmacol.* 26, 2203–2210
- Cho, H., Uehara, T., and Bernhardt, T. G. (2014) Beta-lactam antibiotics induce a lethal malfunctioning of the bacterial cell wall synthesis machinery. *Cell* 159, 1300–1311
- Peltier, J., Courtin, P., El Meouche, I., Lemee, L., Chapot-Chartier, M. P., and Pons, J. L. (2011) *Clostridium difficile* has an original peptidoglycan structure with a high level of N-acetylglucosamine deacetylation and mainly 3-3 cross-links. *J. Biol. Chem.* 286, 29053–29062
- Aliashkevich, A., and Cava, F. (2021) L,D-transpeptidases: the great unknown among the peptidoglycan cross-linkers. *FEBS J.* 289, 4718–4730
- Sutterlin, L., Edoo, Z., Hugonnet, J. E., Mainardi, J. L., and Arthur, M. (2018) Peptidoglycan cross-linking activity of L₄D-Transpeptidases from *Clostridium difficile* and inactivation of these enzymes by beta-lactams. *Antimicrob. Agents Chemother.* 62, e01607-17
- More, N., Martorana, A. M., Biboy, J., Otten, C., Winkle, M., Serrano, C. K. G., et al. (2019) Peptidoglycan remodeling enables *Escherichia* coli to survive severe outer membrane assembly defect. mBio 10, e02729-18
- Dembek, M., Barquist, L., Boinett, C. J., Cain, A. K., Mayho, M., Lawley, T. D., et al. (2015) High-throughput analysis of gene essentiality and sporulation in *Clostridium difficile. mBio* 6, e02383
- Patel, A. V., Turner, R. D., Rifflet, A., Acosta-Martin, A. E., Nichols, A., Awad, M. M., *et al.* (2021) PGFinder, a novel analysis pipeline for the consistent, reproducible and high-resolution structural analysis of bacterial peptidoglycans. *eLife* 10, e70597
- Voedts, H., Dorchene, D., Lodge, A., Vollmer, W., Arthur, M., and Hugonnet, J. E. (2021) Role of endopeptidases in peptidoglycan synthesis mediated by alternative cross-linking enzymes in Escherichia coli. *EMBO J.* 40, e108126
- Chodisetti, P. K., and Reddy, M. (2019) Peptidoglycan hydrolase of an unusual cross-link cleavage specificity contributes to bacterial cell wall synthesis. *Proc. Natl. Acad. Sci. U. S. A.* 116, 7825–7830
- Boneca, I. G., Xu, N., Gage, D. A., de Jonge, B. L., and Tomasz, A. (1997) Structural characterization of an abnormally cross-linked muropeptide dimer that is accumulated in the peptidoglycan of methicillin- and cefotaxime-resistant mutants of *Staphylococcus aureus*. J. Biol. Chem. 272, 29053–29059
- Triboulet, S., Dubee, V., Lecoq, L., Bougault, C., Mainardi, J. L., Rice, L. B., et al. (2013) Kinetic features of L,D-transpeptidase inactivation critical for beta-lactam antibacterial activity. *PLoS One* 8, e67831
- Cordillot, M., Dubee, V., Triboulet, S., Dubost, L., Marie, A., Hugonnet, J. E., et al. (2013) In vitro cross-linking of Mycobacterium tuberculosis peptidoglycan by L,D-transpeptidases and inactivation of these enzymes by carbapenems. Antimicrob. Agents Chemother. 57, 5940–5945

- Toth, M., Stewart, N. K., Smith, C. A., Lee, M., and Vakulenko, S. B. (2022) The l,d-transpeptidase ldt(Ab) from *Acinetobacter baumannii* is poorly inhibited by carbapenems and has a unique structural architecture. *ACS Infect. Dis.* 8, 1948–1961
- Mainardi, J. L., Legrand, R., Arthur, M., Schoot, B., van Heijenoort, J., and Gutmann, L. (2000) Novel mechanism of beta-lactam resistance due to bypass of DD-transpeptidation in *Enterococcus faecium. J. Biol. Chem.* 275, 16490–16496
- 19. Gupta, R., Lavollay, M., Mainardi, J. L., Arthur, M., Bishai, W. R., and Lamichhane, G. (2010) The *Mycobacterium tuberculosis* protein LdtMt2 is a nonclassical transpeptidase required for virulence and resistance to amoxicillin. *Nat. Med.* 16, 466–469
- 20. Baranowski, C., Welsh, M. A., Sham, L. T., Eskandarian, H. A., Lim, H. C., Kieser, K. J., et al. (2018) Maturing Mycobacterium smegmatis peptidoglycan requires non-canonical crosslinks to maintain shape. eLife 7, e37516
- Karasawa, T., Ikoma, S., Yamakawa, K., and Nakamura, S. (1995) A defined growth medium for *Clostridium difficile*. *Microbiology* (*Reading*) 141, 371–375
- Kirk, J. A., and Fagan, R. P. (2016) Heat shock increases conjugation efficiency in *Clostridium difficile. Anaerobe* 42, 1–5
- [preprint] Buddle, J. E., Wright, R. C. T., Turner, C. E., Chaudhuri, R. R., Brockhurst, M. A., and Fagan, R. P. (2023) Multiple evolutionary pathways lead to vancomycin resistance in *Clostridioides difficile. bioRxiv.* https://doi.org/10.1101/2023.1109.1115.557922
- Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760
- Koboldt, D. C., Chen, K., Wylie, T., Larson, D. E., McLellan, M. D., Mardis, E. R., *et al.* (2009) VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* 25, 2283–2285
- 26. Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6, 80-92
- Harrison, E., Guymer, D., Spiers, A. J., Paterson, S., and Brockhurst, M. A. (2015) Parallel compensatory evolution stabilizes plasmids across the parasitism-mutualism continuum. *Curr. Biol.* 25, 2034–2039
- Carver, T., Harris, S. R., Berriman, M., Parkhill, J., and McQuillan, J. A. (2012) Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28, 464–469
- Glauner, B. (1988) Separation and quantification of muropeptides with high-performance liquid chromatography. *Anal. Biochem.* 172, 451–464
- 30. Kirk, J. A., Gebhart, D., Buckley, A. M., Lok, S., Scholl, D., Douce, G. R., et al. (2017) New class of precision antimicrobials redefines role of *Clostridium difficile* S-layer in virulence and viability. *Sci. Transl. Med.* 9, eaah6813

Appendix 2

This appendix represents my first co-authorship produced during my thesis study.

Alamán-Zárate, M. G^{*}., Rady, B. J.^{*}, Evans, C. A., Pian, B., Greetham, D., Marecos-Ortiz, S., Dickman, M. J., Lidbury, I. D. E. A., Lovering, A. L., Barstow, B. M., & Mesnage, S. (2024). Unusual 1-3 peptidoglycan cross-links in *Acetobacteraceae* are made by L,D-transpeptidases with a catalytic domain distantly related to YkuD domains. The Journal of Biological Chemistry, 300(1), 105494. <u>https://doi.org/10.1016/j.jbc.2023.105494</u> *First co-authors



Unusual 1-3 peptidoglycan cross-links in *Acetobacteraceae* are made by L,D-transpeptidases with a catalytic domain distantly related to YkuD domains

Received for publication, November 2, 2023, and in revised form, November 20, 2023 Published, Papers in Press, November 23, 2023, https://doi.org/10.1016/j.jbc.2023.105494

Marcel G. Alamán-Zárate^{1,‡}, Brooks J. Rady^{1,‡}, Caroline A. Evans², Brooke Pian³, Darren Greetham¹, Sabrina Marecos-Ortiz³, Mark J. Dickman², Ian D. E. A. Lidbury¹, Andrew L. Lovering⁴, Buz M. Barstow³, and Stéphane Mesnage^{1,*}

From the ¹Molecular Microbiology, Biochemistry to Disease, School of Biosciences, and ²Department of Chemical and Biological Engineering, ChELSI Institute, University of Sheffield, Sheffield, UK; ³Department of Biological and Environmental Engineering, Cornell University, Ithaca, USA; ⁴School of Biosciences, University of Birmingham, Birmingham, UK

Reviewed by members of the JBC Editorial Board. Edited by Chris Whitfield

Peptidoglycan is an essential component of the bacterial cell envelope that contains glycan chains substituted by short peptide stems. Peptide stems are polymerized by D,Dtranspeptidases, which make bonds between the amino acid in position four of a donor stem and the third residue of an acceptor stem (4-3 cross-links). Some bacterial peptidoglycans also contain 3-3 cross-links that are formed by another class of enzymes called L,D-transpeptidases which contain a YkuD catalytic domain. In this work, we investigate the formation of unusual bacterial 1-3 peptidoglycan cross-links. We describe a version of the PGFinder software that can identify 1-3 crosslinks and report the high-resolution peptidoglycan structure of Gluconobacter oxydans (a model organism within the Acetobacteraceae family). We reveal that G. oxydans peptidoglycan contains peptide stems made of a single alanine as well as several dipeptide stems with unusual amino acids at their Cterminus. Using a bioinformatics approach, we identified a G. oxydans mutant from a transposon library with a drastic reduction in 1-3 cross-links. Through complementation experiments in G. oxydans and recombinant protein production in a heterologous host, we identify an L,D-transpeptidase enzyme with a domain distantly related to the YkuD domain responsible for these non-canonical reactions. This work revisits the enzymatic capabilities of L,D-transpeptidases, a versatile family of enzymes that play a key role in bacterial peptidoglycan remodelling.

Peptidoglycan is an essential component of the bacterial cell envelope that confers cell shape and resistance to a high internal osmotic pressure (1). This bag-shaped macromolecule surrounding the cytoplasmic membrane is made of disaccharide-peptides as building blocks. Their polymerization forms glycan chains alternating *N*-acetylglucosamine and *N*acetylmuramic acid (MurNAc) residues, substituted by short pentapeptide stems containing L- and D-amino acids (2).

[‡] These authors contributed equally to this work.

* For correspondence: Stéphane Mesnage, s.mesnage@sheffield.ac.uk.

Depending on the bacterial species considered, the composition of peptidoglycan building blocks can vary (2), but in most bacteria (including *Escherichia coli*), pentapeptide stems are made of the sequence L-Ala-isoD-Glu-*meso*-DAP-D-Ala-D-Ala, (where DAP is diaminopimelic acid).

The polymerization of peptidoglycan has been extensively studied since the late 1950s, when it was discovered that this process is inhibited by penicillin, a beta-lactam antibiotic widely used to combat infections (3, 4). The ubiquitous enzymes that polymerize peptidoglycan, D,D-transpeptidases, are also called Penicillin Binding Proteins (PBPs). They recognize the C-terminal D-Ala-D-Ala extremity of a donor peptide stem, form an acyl-enzyme intermediate with the amino acid in position 4, and link this residue to the side-chain amino group of the amino acid in position three of an acceptor stem (4-3 cross-link). Beta-lactams are structural analogs of the D-Ala-D-Ala stems and can be used as suicide substrates (5), leading to growth arrest and cell death (6). Alternative 3-3 peptidoglycan cross-links were originally described in Mycobacteria (7). These types of bonds are prevalent in the peptidoglycan of important pathogens such as Mycobacterium tuberculosis (8), Mycobacterium leprae (9) and Clostridium difficile (10). In Enterococcus faecium, resistance to betalactams and glycopeptides can emerge when 4-3 cross-links are replaced by 3-3 cross-links. The complete bypass of the D,D-transpeptidation pathway in E. faecium led to the identification of the enzyme catalysing the formation of 3-3 bonds (11) which is an L,D-transpeptidase. Instead of recognizing the D-Ala-D-Ala extremity of the pentapeptide donor stem, L,Dtranspeptidases use a tetrapeptide stem as a substrate. These enzymes can perform several activities depending on the substrate they use as an acceptor. They can act as a carboxypeptidase (cleaving the fourth residue of the donor stem) (12), as a transpeptidase (forming 3-3 cross-linked muropeptides or covalently anchoring proteins to peptidoglycan) (13, 14), or as an endopeptidase (cleaving 3-3 cross-links or the link between peptidoglycan and covalently attached proteins) (15-17). Finally, L,D-transpeptidases can also exchange the fourth amino acid of a peptide stem for another amino acid (18, 19).

SASBMB

J. Biol. Chem. (2024) 300(1) 105494 **1**

^{© 2023} THE AUTHORS. Published by Elsevier Inc on behalf of American Society for Biochemistry and Molecular Biology. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

The peptidoglycan structural changes catalyzed by L,Dtranspeptidases (called remodeling) play an important role in cell shape (12), resistance to abiotic stress (20), pathogenesis, and host immunity (21). All L,D-transpeptidases described to date contain a YkuD (Pfam: PF03734)) domain.

A recent study described the existence of peptidoglycan 1-3 cross-links in Acetobacteraceae and proposed that this unusual type of cross-link could play a role in the survival of these organisms in the context of their interaction with the fly immune system and during competition with other organisms (22). In this work, we describe a version of the PGFinder software that can automate the analysis of peptidoglycans with 1-3 cross-links. Using this tool, we determine the highresolution structure of Gluconobacter oxydans peptidoglycan and reveal that it contains a high proportion of previously undescribed disaccharide-dipeptides with non-canonical amino acids at their C-terminus. Using a transposon mutant and its complemented derivative, as well as heterologous expression experiments, we demonstrate that G. oxydans 1-3 cross-links are formed by an enzyme with a domain distantly related to the YkuD domain of canonical L,D-transpeptidases. Collectively, our data show that L,D-transpeptidases have evolved to carry out enzymatic reactions using either tetrapeptide or dipeptide stems as donors.

Results

Building a software tool for the structural analysis of 1-3 cross-linked peptidoglycans

Prior to this study, the PGFinder software (v1.0.3; https:// mesnage-org.github.io/pgfinder/) could only generate dimers and trimers crosslinked via 3-3 and 4-3 bonds (23). To perform the structural analysis of *G. oxydans* peptidoglycan, we modified PGFinder and its graphical user interface to enable the creation of dynamic databases containing dimers and trimers with 1-3 cross-links. This upgrade (v1.1.0) was tested using datasets from *G. oxydans*.

High-resolution analysis of G. oxydans B58 peptidoglycan

Peptidoglycan was purified from G. oxydans B58 cells harvested during both exponential (Fig. 1A) and stationary phase (Fig. 1B). As expected, the muropeptide profiles revealed changes indicative of major peptidoglycan remodeling during the stationary phase. We used a combination of automated tools previously described to determine the high-resolution structure of G. oxydans peptidoglycan (18, 23). A two-step custom search strategy was followed (Fig. S1). We first used the proprietary Byonic software to identify monomers based on tandem mass spectrometry data. The search space contained mono-, di-, tri-, tetra- and pentapeptides containing Alanine in position 1, glutamic acid (E) or glutamine (Q) in position 2, meso-DAP (J) or amidated meso-DAP (Z) in position 3, any possible amino acids (X) in position 4, and pentapeptides containing AX dipeptides at their C-terminus (Table S1).

Seventeen monomers showing more than half of the expected b and y ions in their fragmentation spectra were

2 J. Biol. Chem. (2024) 300(1) 105494



Figure 1. HPLC-MS chromatogram of *G. oxydans* reduced disaccharidepeptides. Strain B58 was grown in YPM media to exponential (A) or stationary phase (B). The numbers refer to the muropeptide structures described in Table 1.

identified by Byonic (Fig. S2). Interestingly, these included several muropeptides with a dipeptide stem other than AE that were not previously identified and a lack of tetrapeptide stems with unusual amino acids formed by canonical L,Dtranspeptidases. In addition to amidated meso-DAP residues, we also found the presence of deacetylated N-acetylglucosamine residues (glucosamine) which were not previously reported. The disaccharide-peptides corresponding to these validated monomers were combined to create the database called DB0 Go (Table S2). We next performed a PGFinder search, enabling the identification of dimers and trimers with 4-3, 3-3, and 1-3 cross-links as well as modified disaccharides (deacetylated and containing MurNAc residues). A total of 61 masses matching the monoisotopic mass of theoretical structures were identified (Table 1), revealing a far more complex structure than previously reported. Peptidoglycan from cells harvested during the exponential and stationary phases showed that cross-linking index is higher in the stationary phase (19.9% versus 15.9%), partly due to an increased proportion of 1-3 cross-links in the stationary phase (16.6% versus 5.8% in exponential phase). The higher proportion of 1-3 cross-links was concomitant with the higher proportion of

Table 1 G oxydans pentidoalycan composition

		W	7T				Theoretical	
No.	Muropeptide ^a	Expo	Stat	Ldt _{Go1}	Ldt _{Go2}	RT (min) ^b	Mass (Da)	Δppm
1	Gm-AEJ _{NH2}	36.519%	41.440%	48.001%	34.926%	5.32 ± 0.05	869.3867	4.01
2	gm-AE	4.239%	10.518%	9.628%	6.887%	6.79 ± 0.03	698.2859	2.77
3	gm-AEJA	22.572%	3.270%	4.019%	28.131%	6.80 ± 0.01	941.4078	3.38
4	gm-AE (Anh)	0.880%	1.158%	0.656%	0.505%	11.06 ± 0.00	678.2597	1.52
5	gm-AEJ _{NH2} (Anh)	0.522%	0.682%	0.472%	0.161%	8.61 ± 0.01	849.3605	2.29
6	gm-AEJ _{NH} 2A	0.254%	0.290%	0.343%	0.012%	6.37 ± 0.27	940.4238	1.39
7	gm-AEJ	0.711%	0.356%	0.193%	0.014%	5.66 ± 0.04	870.3707	2.14
8	gm-AEJAG	0.494%	0.064%	0.130%	0.247%	6.49 ± 0.02	998.4293	1.46
9	gm-AF	0.072%	0.171%	0.096%	ND	14.54 ± 0.01	716.3117	0.89
10	gm-AEJAA	0.353%	0.074%	0.094%	0.067%	8.08 ± 0.39	1012.4450	2.97
11	gm-A	0.335%	1.385%	0.534%	0.024%	6.26 ± 0.03	569.2433	1.8
12	gm-AEJA (Ann)	0.425%	0.093%	0.079%	0.131%	10.42 ± 0.01	921.3816	1.45
15	gm-ALJAK	0.114%	0.021%	0.042%	0.05770 ND	5.80 ± 0.05 10.72 ± 0.01	722 2066	1.07
14	gm-AI	0.032%	0.100%	0.042%	ND	10.72 ± 0.01 12.89 ± 0.00	692 3274	0.52
16	g_{m-AI} (Aph)	ND	0.070%	ND	ND	12.89 ± 0.00	662 3012	0.01
17	gm-AFIAR	0.121%	0.029%	0.037%	0.062%	642 ± 0.03	1097 5089	1.57
18	gm-AFIAH	0.121%	0.035%	0.035%	0.050%	5.99 ± 0.03	1078 4667	1.37
19	gm-AEJAN	0.022%	0.069%	0.033%	0.038%	10.37 ± 4.33	1055.4508	1.72
20	gm-AI (-Ac)	0.055%	ND	ND	ND	6.29	640.3169	0.24
21	gm-AEI _{NH2} (-Ac)	0.036%	0.015%	0.023%	0.022%	5.05 ± 0.06	827.3762	1.15
22	gm-AO	0.057%	0.026%	0.022%	0.047%	5.77 ± 0.04	697.3019	0.97
23	gm-AEIAE	0.029%	0.019%	0.018%	0.031%	7.29 ± 0.01	1070.4504	0.50
24	gm-AEJ (Anh)	0.011%	ND	0.017%	ND	9.03 ± 0.00	850.3445	1.15
25	gm-AEJA (-Ac)	0.021%	0.002%	0.007%	0.027%	6.43 ± 0.02	899.3973	1.11
26	gm-AEJ _{NH2} A (Anh)	ND	ND	0.006%	ND	9.93	920.3976	1.17
27	gm-AE (-Ac)	ND	0.030%	0.003%	ND	5.62 ± 0.01	656.2754	0.14
28	gm-AF (-Ac)	ND	ND	0.003%	ND	16.28	674.3012	1.24
29	gm-AEJ _{NH2} =gm-AEJA	13.782%	15.265%	18.938%	15.714%	8.43 ± 0.00	1792.7836	3.35
30	gm-AEJ _{NH2} =gm-A	2.592%	13.391%	7.317%	0.037%	8.60 ± 0.00	1420.6194	1.75
31	gm-AEJ _{NH2} =gm-AEJA (Anh)	3.741%	4.098%	4.220%	4.109%	10.87 ± 0.00	1772.7574	2.38
32	gm-AEJA=gm-AEJA	6.802%	2.822%	2.318%	6.976%	9.59 ± 0.00	1864.8047	2.51
33	gm-AEJ _{NH2} =gm-A (Anh)	0.243%	0.794%	0.609%	ND	11.43 ± 0.01	1400.5932	0.85
34	gm-AEJA=gm-A	2.520%	2.028%	0.512%	0.011%	10.06 ± 0.00	1492.6405	1.02
35	gm-AEJA=gm-AEJA (Anh)	0.729%	0.388%	0.199%	0.601%	12.01 ± 0.00	1844.7785	1.37
36	gm-AEJ=gm-AEJA	0.186%	0.136%	0.12/%	ND	9.03 ± 0.05	1/93./6/6	0.68
3/	gm-AEJA=gm-A (Ann)	0.317%	0.197%	0.074%	ND 0.0789/	12.87 ± 0.02	14/2.0143	0.41
20	gm-AEJAG=gm-AEJA	0.134%	0.035%	0.049%	0.078%	9.18 ± 0.00	1921.8262	1.54
40	gm-AFIAG-gm-A	0.029%	0.023%	0.034%	0.013%	10.40 ± 0.02	1549 6617	3.24
40	gm-AEJ-AG=gm-A	0.008%	0.030%	0.022%	0.013%	938 ± 0.00	1421 6034	0.94
42	$gm_A EIA - gm_A EIA (Anb)$	ND	0.014%	ND	ND	9.38 ± 0.00	1844 7786	0.09
43	gm-AFIAA=gm-AFIA	0.034%	0.009%	0.016%	0.023%	9.84 ± 0.01	1935 8419	0.67
44	gm-AEI _{NH2} A=gm-AEIA	0.007%	0.008%	0.012%	ND	9.32 ± 0.04	1863.8207	0.70
45	gm-AEIAG=gm-AEIA (Anh)	0.032%	0.014%	0.010%	0.016%	11.53 ± 0.01	1901.8000	1.14
46	gm-AEJAR=gm-AEJA	0.017%	ND	0.006%	0.007%	9.00 ± 0.00	2020.9058	0.32
47	gm-AEJAN=gm-AEJA	ND	ND	0.005%	0.004%	8.76 ± 0.01	1978.8477	0.81
48	gm-AEJ _{NH2} A=gm-AEJA (Anh)	ND	ND	0.004%	ND	11.56	1843.7945	0.37
49	gm-AEJ _{NH2} =gm-AEJA (-Ac)	0.005%	ND	0.003%	0.004%	8.18 ± 0.03	1750.7731	2.03
50	gm-AEJAH=gm-AEJA	0.013%	ND	ND	0.006%	8.64 ± 0.00	2001.8636	0.74
51	gm-AEJ _{NH2} A=gm-A		ND	ND	0.005%	16.44	1491.6565	0.55
52	gm-AEJAA=gm-AEJA (Anh)	0.006%	ND	0.002%	0.004%	12.12 ± 0.01	1915.8157	0.63
53	gm-AEJAG=gm-A (Anh)	0.005%	ND	0.002%	ND	9.71 ± 2.72	1529.6355	1.74
54	gm-AEJAR=gm-AEJA (Anh)	0.004%	ND	ND	ND	11.24	2000.8796	0.61
55	gm-AEJ _{NH2} =gm-AEJA=gm-AEJA	0.370%	0.384%	0.551%	0.554%	9.97 ± 0.00	2716.1805	0.96
56	gm-AEJ _{NH2} =gm-AEJA=gm-AEJA (Anh)	0.180%	0.196%	0.210%	0.229%	11.84 ± 0.00	2696.1543	0.78
57	gm-AEJA=gm-AEJA=gm-AEJA	0.099%	0.045%	0.119%	0.136%	10.84 ± 0.00	2788.2016	1.89
58	gm-AEJA=gm-AEJA=gm-AEJA (Anh)	0.032%	0.015%	0.024%	0.033%	12.67 ± 0.03	2768.1754	0.28
59	gm-AEJ=gm-AEJA=gm-AEJA	ND	ND	0.019%	ND	10.59 ± 0.00	2717.1645	1.18
60	gm-AEJ=gm-AEJA=gm-AEJA (Anh)	ND	ND	0.005%	ND	12.51	2697.1383	0.52
61	gm-AEJAG=gm-AEJA=gm-AEJA	0.005%	ND	0.004%	0.003%	10.49 ± 0.01	2845.2231	0.53

^{*a*} g, GlcNAc; m, MurNAc; A, Alanine; E, isoglutamic acid; J, *meso*-diaminopimelic acid; J_{NH2}, amidated meso-diaminopimelic acid. ^{*b*} Standard deviations in bold are determined from two values only.

disaccharide-dipeptide structures detected in the stationary phase (13.5% versus 5.7%). Very little variation was observed in the glycan chain length between the exponential and stationary phase (determined using the proportion of AnhydroMurNAc residues), with the average length being equal to 24 and 22 disaccharides, respectively.

mediate with a disaccharide-alanine. Identification of the L,D-transpeptidase catalyzing the formation of 1-3 cross-links in G. oxydans

By analogy with the transpeptidation reaction leading to the formation of 3-3 bonds (Fig. 2A), we hypothesized that the formation of 1-3 bonds uses muropeptides with a dipeptide

Interestingly, G. oxydans peptidoglycan does not contain any tetrapeptide stems with unusual amino acids at their Cterminus, which are characteristic of canonical L,D-

stem as donor substrates (Fig. 2B). According to this hypoth-

esis, the enzyme is predicted to form an acyl enzyme inter-



Figure 2. Schematic representation of the L,D-transpeptidation reactions leading to the formation of 3-3 and 1-3 cross-links. The enzymatic reactions carried out by L,D-transpeptidases in organisms with 3-3 cross-links are described (*A*). By analogy with these L,D transpeptidation reactions, we propose a model that leads to distinct reactions in *G. oxydans* (*B*). We hypothesize that an unidentified endopeptidase generates disaccharide dipeptides. These muropeptides are used as substrates to form an acyl-enzyme intermediate. Depending on the acceptor group, the reaction can lead to a carboxypeptidase reaction or a transpeptidation reaction that generates either a dimer or a disaccharide-dipeptide. DS, disaccharide (GlcNAc-MurNAc); DS-Tetra, disaccharide-tetrapeptide; DS-Tri, disaccharide-tripeptide; GlcNAc, *N*-acetylglucosamine; LDT, L,D-transpeptidase; MurNAc, *N*-acetylmuramic acid; X, any D amino acid.

transpeptidase enzymatic activity (Fig. 2*A*). We therefore hypothesized that in *G. oxydans*, L,D-transpeptidases could perform a similar enzymatic reaction using disaccharidedipeptides as substrates instead (Fig. 2*B*). We searched the *G. oxydans* genome to identify genes encoding homologs of the L,D-transpeptidases and found two putative L,D-transpeptidases (labeled GOX1074, 337 residues and GOX2269, 171 residues in *G. oxydans* 621H) related to the YkuD catalytic domain (Pfam: PF03734). We hypothesized that one or both enzymes could catalyze the formation of 1-3 cross-links and renamed these putative L,D-transpeptidases Ldt_{Go1} and Ldt_{Go2}. To test this, we took advantage of the *G. oxydans* B58 Sudoku library previously described (24) and analyzed the peptidoglycan structure of the two transposon mutants with an insertion in each of the ldt_{Go} genes by LC-MS. Comparison of the total ion chromatography (TIC) profiles indicated the presence of a peak corresponding to the major 1-3 dimer (gm-AEJ_{NH2}=gm-A) in the wild type (Fig. 3*A*) and ldt_{Go1} mutant (Fig. 3*B*), whilst no equivalent peak was detected in the ldt_{Go2} mutant (Fig. 3*C*). Analysis of the extracted ion chromatograms for all molecules eluted between 7.5 and 11.5 min revealed a drastic reduction of 1-3 cross-links in the ldt_{Go2} peptidoglycan sample (0.035% *versus* 13.4% in the WT). The ldt_{Go2} mutation was also associated with a reduction of



Figure 3. LC-MS detection of 1-3 2-1 cross-links in *G. oxydans* **peptidoglycan.** TIC of *G. oxydans* B58 (A) and mutants with a transposon insertion in the ldt_{Go1} (B) and ldt_{Go2} genes (C) are shown on the *left-hand side*. Extracted ions corresponding to the muropeptides eluted between 7.5 and 11.5 min are shown on the *right-hand side*. The major dimer with a 1-3 cross-link (shown with an *arrow* on the TIC and on the *top* MS spectrum) is associated with two major protonated ions: a singly charged ion with an *m/z* at 1421.32 and a doubly charged ion with an *m/z* at 711.32. None of these ions were detected in the peptidoglycan from the ldt_{Go2} mutant, demonstrating that this gene is essential for the formation of 1-3 cross-links.

disaccharide-dipeptides and 1-3 crosslinked dimers as compared to the parental strain and the ldt_{Go1} mutant (Table 1). Collectively, our LC-MS data showed that Ldt_{Go2} plays a major role in the unusual L,D-transpeptidation reactions in *G. oxydans*, including the formation of 1-3 cross-links.

Complementation and heterologous expression experiments show that the Ldt_{Go2} enzyme is sufficient to catalyse peptidoglycan 1-3 cross-links

To verify that the drastic reduction of 1-3 cross-links was associated with the disruption of ldt_{Go2} and not a secondary mutation, we built a complementation strain expressing Ldt_{Go2} under the anhydrotetracycline-inducible promoter (25). The production of Ldt_{Go2} in the ldt_{Go2} transposon mutant background clearly restored the presence of a peak corresponding to the major 1-3 cross-linked dimers (Fig. 4).

We further confirmed the enzymatic activity of Ldt_{Go2} by producing the full-length protein in *E. coli*. Since *E. coli* peptidoglycan contains disaccharide-dipeptides (gm-AE) that represent the proposed substrate for the 1-3 transpeptidation reaction, we anticipated that recombinant Ldt_{Go2} could generate the expected products found in *G. oxydans* in this heterologous host. Peptidoglycan was purified from *E. coli* transformed with either the empty pET expression vector or a recombinant derivative expressing ldt_{Go2} , digested with mutanolysin, and analyzed by reverse-phase HPLC (Fig. 5*A*, top and bottom trace, respectively). A simple search strategy was followed to identify and quantify muropeptides resulting from unusual L,D-transpeptidation reactions (Fig. S3).

First, we identified monomers based on MS/MS data using the Byonic module from Byos. We searched for all possible disaccharide-peptides containing one to five amino acids (A, AX, AEJ, AEJX, and AEJAX, where J is meso-diaminopimelic acid and X any amino acid) adding sugar deacetylation previously identified in E. coli (23) as a potential glycan modification (Table S3). Twenty-eight monomers validated by MS/ MS analysis were selected to create a database called DB0_Ec (Table S4). This monomer database was then run through PGFinder to identify, compare, and quantify muropeptides in the E. coli expression strain and its derivative expressing Ldt_{Go2}. To focus on 1-3 L,D transpeptidation products, we only enabled the search for 1-3 dimers that contain the gm-A moiety. Interestingly, the PGFinder search revealed a very low amount of 1-3 transpeptidation products in E. coli (gm-A, 0.44% and gm-AEJA=gm-A, 0.056%) (Fig. 5B). A striking



Figure 4. Complementation of the Idt_{Go2} transposon insertion restores 1-3 cross-links. Wild-type train B58 (A), the IdtGo2 insertion mutant (B) and the complemented mutant (C) were grown in YPM media. The expression of Idt_{Go2} in the complemented strain was induced at $OD_{GO0mm} = 0.5$ with 100 ng/ml anhydrotetracycline. The peaks corresponding to the major species are labelled. Numbers refer to the muropeptides described in Table 1.

increase in gm-A (9.0%), gm-AX (4.3%) monomers and dimers resulting from 1-3 cross-linking (9.5%) was detected in the peptidoglycan of the strain expressing Ldt_{Go2}. demonstrating that this enzyme is an L,D-transpeptidase that can perform all the reactions described in Figure 2 (carboxypeptidation, exchange, and 1-3 transpeptidation).

Ldt_{Go2} is characterized by an atypical YkuD-like catalytic domain that can be found in distant families of bacteria with 1-3 peptidoglycan cross-links

To place *G. oxydans*'s L,D-transpeptidases in a broader evolutionary context, homologues were extracted from genomes of numerous alphaproteobacterial species (Table S5), including those previously shown to contain 1-3 peptidoglycan cross-links (22). Four other organisms with characterized L,Dtranspeptidases (*E. coli, C. difficile, M. tuberculosis* and *E. faecium*) were added. Phylogenetic reconstruction of all putative alphaproteobacterial L,D-transpeptidases revealed

6 J. Biol. Chem. (2024) 300(1) 105494

that both Ldt_{Go1} and Ldt_{Go2} homologues form distinct clades representing previously uncharacterised transpeptidase subfamilies (Fig. 6A).

When annotated using InterProScan and its default significance thresholds, Ldt_{Go1} homologs are shown to contain a canonical YkuD (Pfam: PF03734) domain, but the more distantly related Ldt_{Go2} homologs typically lack this annotation. Instead, most are annotated with a YkuD-like (CDD: cd16913) domain, and others contain no domain annotations at all (Fig. 6*B*). Although Ldt_{Go2} is annotated with a canonical YkuD domain, it should be noted that the E-value for this annotation is high (2.0e-1), indicative of a significant divergence from the canonical YkuD domain.

Finally, to better understand the distribution of this unusual L,D-transpeptidase subfamily beyond the Alphaproteobacteria, the catalytic domain of Ldt_{Go2} was searched against the entirety of the NCBI RefSeq Select database (26). Out of the 307 hits returned from unique bacterial species, roughly 37% could be attributed to Acetobacteraceae like G. oxydans, but an even greater percentage of hits (45%) came from the Burkholderiaceae (Fig. 6C). Though the Acetobacteraceae and Burkholderiaceae encompass the majority of Ldt_{Go2} homologs, others are found sprinkled throughout the broader Burkholderiales and even beyond the Pseudomonadota, with homologs in the Desulfovibrionaceae. No Ldt_{Go2} homologs were found in the model organisms included in our analysis (M. tuberculosis, C. difficile, E. coli, and E. faecium). Since active site geometry is thought to be a key determinant of L,Dtranspeptidase substrate preference and activity, a further search for structural homologues was conducted using Foldseek and an AlphaFold model of the Ldt_{Go2} catalytic domain (27, 28). Setting an E-value threshold of \leq 2e-2 (selecting for matches better than Bacillus subtilis' prototypical YkuD domain), led to 147 hits. The results of this structural search largely validated the results of the sequence-based BlastP search, with 22% of hits coming from the Acetobacteraceae and 39% from the Burkholderiaceae, but a much larger number of hits (23%) now fell outside of the families found by BLAST.

Overall, these analyses establish that L,D-transpeptidases associated with 1-3 cross-linking contain a catalytic domain related to the canonical YkuD transpeptidase domain but form a distinct enzymatic subfamily.

Discussion

In this study, we determine the high-resolution structure of *G. oxydans* peptidoglycan using a version of PGFinder that can generate dynamic databases containing 1-3 cross-linked multimers. We show that *G. oxydans* peptidoglycan contains a high proportion of dipeptide stems with unusual amino acids at their *C*-terminus, leading us to propose that the enzyme forming 1-3 cross-links uses dipeptide stems as a donor substrate. We identify 2 *G. oxydans* enzymes distantly related to L,D-transpeptidases making 3-3 cross-links. Based on the characterization of a transposon mutant and heterologous

1-3 peptidoglycan cross-links in Acetobacteraceae



Figure 5. Heterologous protein synthesis of Ldt_{Go2} in *E. coli* BL21(DE3) increases the proportion of 1-3 L,D-transpeptidation products. *E. coli* BL21(DE3) transformed with the control pET2818 plasmid or pET2818 encoding Ldt_{Go2} was grown in auto-induction medium overnight and peptidoglycan from both cultures were purified. The muropeptide profile are shown in (*A*); *bottom* profile is from the control strain; *bottom* panel is from *E. coli* (pET-ldt_{Go2}). Two major peaks containing muropeptides of interest resulting from 1-3 L,D-transpeptidation are indicated. *B*, PGFinder analysis of *E. coli* control strain (transformed with the empty plasmid, *E. coli* (pET2818)) and expressing Ldt_{Go2} (*E. coli* (pET-Ldt_{Go2})). Only monomers validated by Byonic based on MS/MS data were search as well as their 1-3 transpeptidation products. The monomers and dimers resulting from 1-3 L,D-transpeptidation are indicated in *red*.

expression experiments, we demonstrate that one of these two candidates (Ldt_{Go2}) catalyses the formation of 1-3 cross-links.

This work demonstrated that Ldt_{Go2} plays a predominant role in the formation of 1-3 cross-links in *G. oxydans*. The role of Ldt_{Go1} remains unclear since the inactivation of the corresponding gene is associated with only marginal changes in the peptidoglycan composition (Table 1). Our attempts to express recombinant Ldt_{Go1} and Ldt_{Go2} in *E. coli* as his-tagged or maltose-binding fusion proteins remained unsuccessful and both proteins were systematically found in the insoluble fraction, irrespective of the expression strains and conditions tested. Further experiments are therefore required to produce and purify these recombinant proteins to examine their activity *in vitro* more closely.

The formation of 3-3 cross-links in Enterococci is controlled by the availability of disaccharide-tetrapeptides used as donor substrate. In *E. faecium*, L,D-transpeptidation can bypass the D,D-transpeptidation following the activation of a cryptic D,D-carboxypeptidase (29). How the disaccharide-dipeptide substrates are generated in *G. oxydans* remains unknown. *G. oxydans* encodes two potential endopeptidases containing a CHAP domain that could generate Ldt_{Go2} substrates (GOX_RS06930 and GOX_RS07380 in *G. oxydans* 621H). The transposon inactivation of each gene was tested

but did not abolish the production of 1-3 cross-links (data not shown), indicating that these genes do not play a predominant role in the formation of dipeptide stems or are functionally redundant. The inactivation of both genes simultaneously will be required to further investigate the Ldt_{Go2} partners that contribute to the formation of unusual cross-links.

Interestingly, three Ldt_{Go2} homologs, those found in Roseomonas gilardii, Acidocella facilis, and Acidomonas methanolica, were not annotated with any YkuD-like (CDD: cd16913) catalytic domains by InterProScan. Although 1-3 cross-links have only been reported in A. methanolica (22), it would be worth revisiting the peptidoglycan in the two other species to confirm the absence of 1-3 cross-links using PGFinder. This work confirms previous studies which showed that this software is a powerful tool to elucidate the highresolution of bacterial peptidoglycan structures and their quantification. Unlike peptidoglycan analyses based on UV quantification, PGFinder allows the systematic and unbiased identification of low-abundance muropeptides accounting for less than 0.01% of all muropeptides. A striking result illustrating the low detection threshold provided by PGFinder is the identification of 1-3 crosslinks and gm-A muropeptides in E. coli, indicating that these organisms' L,D-transpeptidases can also form unusual reactions.

1-3 peptidoglycan cross-links in Acetobacteraceae



Figure 6. Ldt_{Go2} represents a distinct L,D-transpeptidase subclade with a divergent catalytic domain that is found primarily in the Acetobacteracea and Burkholderiaceae. A, an unrooted phylogenetic tree of putative L,D-transpeptidases throughout Alphaproteobacteria and characterized enzymes (Table S5) reveals that Ldt_{Go1} and Ldt_{Go2} homologues form distinct transpeptidase subfamilies. Ldt_{Go1} and Ldt_{Go2} are labelled with asterisks (*). Previously characterised L,D-transpeptidases from *Escherichia coli* (LdtA-F), Clostridioides difficile (LdtCd1-3), Mycobacterium tuberculosis (LdtM1-5) and *Enterococcus faecium* (Ldt_{Efm}) have also been labelled. *B*, a phylogram of Ldt_{Go1}, Ldt_{Go2}, and their homologues were annotated only with the CDD YkuD_like domain (44), and the rest lacked domain annotation entirely. The Ldt_{Go2} homologues highlighted in *orange* are found in bacterial species where no 1-3 cross-links could be detected (22). *C*, an expanded search for Ldt_{Go2} homologues beyond the Alphaproteobacteria reveals the presence of this subfamily throughout the Burkholderiales and *Desulfovibrionaceae*. Structural homologues located using Foldseek (28) show a similar evolutionary distribution as those located using BlastP (26).

Moving from sequence to structural analysis, the predicted fold of Ldt_{Go2} revealed the presence of a much more open, bowl-like active site (Fig. 7). Given that this enzyme uses a

shorter peptide stem as a donor substrate, it is likely that the catalytic site does not require canonical cleft or trapping loops to accommodate the substrate. Instead, the open conformation



Figure 7. Structural analysis of *G. oxydans* **Ldt**_{Go2}. *A*, predicted fold of Ldt_{Go2}, inclusive of well-modeled residues 79 to 336, taken from EBI Alphafold repository (27). The catalytic residue (C264) shown in stick form with SH sidechain coloured *yellow*. *B*, surface representation of Ldt_{Go2}, demonstrating flat bowl-like active site surrounding C264 (*yellow*). *C*, superimposition of Ldt_{Go2} (*grey*) with *Vibrio cholerae* LdtA (RCSB entry 7AJO, unreleased, *blue*; bound reaction intermediate at C444 shown in stick form), reveals the relatively more closed/capped cleft of 3-3 cross-link forming enzymes and outlines that potential donor and acceptor substrates of Ldt_{Go2} will likely be less constrained.

of the catalytic site could ensure that the bulky sugar moieties of a dipeptide substrate don't limit access to the catalytic cysteine residue responsible for the formation of 1-3 crosslinks.

The discovery of enzymes forming 1-3 cross-links reaffirmed that the catalytic reactions carried out by domains belonging to the YkuD family are very diverse. This work expands our knowledge of peptidoglycan polymerization and opens new avenues to study how remodeling contributes to the maintenance of cell envelope integrity (20). *Acetobacteraceae* (also called acetic acid bacteria) are important for the food industry and are key organisms have a high capacity to oxidize ethanol as well as various sugars to form acetic acid and display resistance to high concentrations of acetic acid released into the fermentative medium. It is tempting to speculate that the formation of 1-3 cross-links contributes to the maintenance of cell envelope integrity in these harsh conditions.

Experimental procedures

Bacterial strains, plasmids, oligonucleotides, and growth conditions

Bacterial strains, plasmids, and oligonucleotides are described in Table S6. G. oxydans B58 (ATCC NRLL-BR8) and isogenic derivatives were grown in yeast peptone mannitol (YPM; 5 g/l yeast extract, 3 g/l peptone, 25 g/l mannitol) broth or agar at 30 °C under agitation (200 rpm). G. oxydans cultures were inoculated with an overnight preculture at an OD_{600nm}=0.05 and grown for 36 h to stationary phase. G. oxydans transposon mutants were grown in the presence of kanamycin (100 μ g/ml) and gentamicin (10 μ g/ml) for complementation experiments Ldt_{Go2} expression in G. oxydans was induced by adding 100 ng/ml anhydrotetracycline to the media at an OD_{600nm}=0.5. For heterologous expression, E. coli was grown in an auto-induction medium (31) at 30 $^\circ\mathrm{C}$ under agitation (200 rpm) supplemented with 100 µg/ml ampicillin.

Plasmid constructions

Plasmid pBBR-TetR-Go2227 used to complement the transposon insertion in the ldt_{Go2} is a derivative of pBBR1MCS-5-TgdhM-tetR-mNG allowing the inducible expression of proteins in G. oxydans under the control of the tetracycline promoter (25). pBBR-TetR-Go2227 was built using Golden Gate assembly. Three PCR fragments corresponding to (i) the gentamicin cassette (1632 bp), (ii) the pBBR1 origin of replication + the TetR gene (4029 bp), and (iii) the ldt_{Go2} full length sequence (1021 bp) were amplified using oligos SM_0729 + SM_0730, SM_0725 + SM_0726 and SM_0727 + SM_0728, respectively using pBBR1MCS-5-TgdhM-tetR-mNG or G. oxydans chromosomal DNA as templates. The PCR products were purified by gel extraction, mixed in an equimolar ratio, and assembled using the NEBridge Golden Gate Assembly Kit (BsaI-HF v2) according to the manufacturer's instructions. Recombinant plasmids were screened by PCR and plasmid candidates were fully sequenced by Plasmidsaurus (Plsamidsaurus.com) to confirm the absence of mutations.

pET-Go2227, a pET2818 derivative expressing the full-length Ldt_{Go2} enzyme was built using a synthetic DNA fragment with optimized codon usage for *E. coli* provided by Genewiz. The synthetic open reading frame corresponding to the full-length Ldt_{Go2} gene (with a stop codon) was cloned into pET2818 as a NcoI-XhoI fragment.

Preparation of G. oxydans competent cells and transformation

G. oxydans was grown in 100 ml of YPM to an OD_{600nm} of 0.9 and spun for 10 min at 4000g at 4 °C. After three washes in 1 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer (pH7.0), cells were resuspended in 250 µl. Electroporation was carried out in 1 mm cuvettes using 50 µl of electrocompetent cells and 100 ng of plasmid in a volume of 1-2 µl; parameters for electroporation were 2 kV, 25 µF and 200 Ω . After the pulse, 800 µl of YPM media supplemented with 0.25% (m/v) MgSO₄ and 0.15% (m/v) CaCl₂ was added to

the cells that were left to recover under agitation for 16 h before plating on YPM media supplemented with kanamycin and gentamicin.

Peptidoglycan extraction

G. oxydans and *E. coli* strains were grown until the stationary phase in YPM or auto-induction medium, respectively. Cells were pelleted, supernatant discarded, and cell pellet snap frozen in liquid nitrogen. The cell pellet was resuspended in 20 ml of boiling MilliQ water (MQ) before the addition of sodium dodecyl sulfate (SDS) at a final concentration of 4% (m/v). After 30 min at 100 °C, the cells were cooled down to room temperature. Peptidoglycan was pelleted at 150,000g for 1 h, washed five times using warm MQ water, freeze-dried and resuspended at a final concentration of 10 mg/ml.

Preparation of soluble muropeptides

2 mg of purified peptidoglycan was digested for 16 h in 20 mM phosphate buffer (pH 5.5) supplemented with 250 Units of mutanolysin (Sigma) in a final volume of 200 µl. Following heat inactivation of mutanolysin (5 min at 100 °C), soluble disaccharide peptides were mixed with an equal volume of 250 mM borate buffer (pH 9.25) and reduced with 0.2% (m/v) sodium borohydride. After 20 min at room temperature, the pH was adjusted to 5.0 using phosphoric acid. Reduced muropeptides were analyzed by HPLC using a C18 analytical column (Hypersil Gold aQ, 1.9 µm particles, 150 × 2.1 mm; Thermo Fisher Scientific) at a temperature of 50 °C. Muropeptide elution was performed at 0.3 ml/min by applying a mixture of solvent A (water, 0.1% [v/v] formic acid) and solvent B (acetonitrile, 0.1% [v/v] formic acid). LC conditions were 0 to 12.5% B for 25 min increasing to 20% B for 10 min. After 5 min at 95%, the column was re-equilibrated for 10 min with 100% buffer A. UV absorbance at 202 nm was used to check the quality of samples and determine the volume to inject for LC-MS. A volume of sample with an intensity of the most abundant monomer of 1500 mAU was used, giving an ion intensity of approximately 5.109.

LC-MS/MS

An Ultimate 3000 High-Performance Liquid Chromatography (HPLC; Dionex/Thermo Fisher Scientific) system coupled with a high-resolution Orbitrap Exploris 240 mass spectrometer (Thermo Fisher Scientific) was used for LC-MS analysis. Muropeptides were separated using a C18 analytical column (Hypersil Gold aQ, 1.9 µm particles, 150 × 2.1 mm; Thermo Fisher Scientific) at a temperature of 50 °C. Muropeptide elution was performed as described in the previous paragraph. The Orbitrap Exploris 240 was operated under electrospray ionization (H-ESI high flow)-positive mode, full scan (m/z 150–2250) at resolution 120,000 (FWHM) at m/z200, with normalized AGC Target 100%, and automated maximum ion injection time (IT). Data-dependent MS/MS were acquired on a 'Top 5' data-dependent mode using the following parameters: resolution 30,000; AGC 100%, automated IT, with normalized collision energy 25%.

Analysis of peptidoglycan structure

LC-MS datasets were deconvoluted with the Byos software v3.11 (Protein Metrics). Sequential searches were carried out with PGFinder v1.1.1, with default settings (10 ppm tolerance, 0.5 min cleanup window) following the strategy described in Figs. S1 and S3. Data from individual matching output was consolidated as previously described to calculate average intensities, retention times, observed monoisotopic masses, and ppm differences. The output from individual searches and consolidated data are described in Files S1 and S2). Crosslinking index and glycan chain length were determined as described previously (32). The cross-linking index is defined as 0.5 * (% of dimers) + 0.33 * (% of trimers); glycan chain length was inferred from the abundance of anhydroMurNAc groups, which are found at the end of glycan chains. It is defined as 1/(% of AnhydroMurNAc monomers + 0.5 * (% of AnhydroMurNAc dimers) + 0.33 * (% of AnhydroMurNAc trimers).

Comparative genomics and bioinformatic analysis

Reference genomes and protein sequences (Table S5) were downloaded from NCBI Datasets (v15.25.0), and protein sequences were annotated locally using InterProScan (v5.64-96.0) (33, 34). A custom Julia (35) script was then used to search the produced GFF3 files for YkuD-containing proteins and to extract their catalytic domains. Ldt_{Go1} and Ldt_{Go2} homologues were located by running a PSI-BLAST on the RefSeq Select database restricted to taxa in (Table S5) and iterating until no new hits were returned (26). Extracted YkuD proteins and Ldt_{Go1/2} homologues were aligned using Muscle (v5.1) (36), and maximum likelihood trees were constructed using IQ-TREE (v2.2.2.7) with ModelFinder (which selected WAG+R7 for Fig. 6A and WAG + F + G4 for Fig. 6B) and 1000 UFBoot replicates enabled (37-39). Trees were visualised and annotated using iTOL (v6.8.1) (40) with finishing touches applied in Inkscape (v1.3). ColabFold's AlphaFold2_batch notebook (v1.5.2) was used with the default settings and relaxation enabled to obtain predicted structures for Ldt_{Go1}, Ldt_{Go2}, and their respective catalytic domains (27, 41). Finally, Foldseek (v8-ef4e960) was used to search the AFDB50 database for structural homologues of Ldt_{Go2} (27, 42).

Data availability

LC-MS/MS datasets have been deposited in the GLYCO-POST repository (GPST000377). All plasmid sequences are available upon request. *G. oxydans* NRRL B58 genome has been deposited at DDBJ/ENA/GenBank under the accession JAIPVW000000000.

Code availability

The script for PGFinder v1.1.0 is available at https://github.com/Mesnage-Org.

Supporting information—This article contains supporting information (Figs. S1–S3, Tables S1–S6 and Supporting files S1–S2).

Acknowledgments—We thank Tino Polen (Jülich Research Institute) for the gift of pBBR1-TetR plasmid to carry out complementation experiments.

Author contributions—S. M. conceptualization; M. G. A.-Z., B. J. R., and S. M. data curation; M. G. A.-Z., and S. M. formal analysis; S. M. funding acquisition; M. G. A.-Z., B. J. R., C. A. E., B. P., D. G., S. M.-O., I. D. E. A. L., A. L. L., and S. M. investigation; B. J. R., C. A. E., M. J. D., I. D. E. A. L., B. M. B., and S. M. methodology; B. M. B. and S. M. project administration; B. M. B., I. D. E. A. L., and S. M. supervision; B. J. R. software; M. G. A.-Z., C. A. E., and S. M. validation; M. G. A.-Z. and S. M. visualization; S. M. writing – original draft; M. G. A.-Z., B. J. R., C. A. E., B. P., D. G., S. M.-O., M. J. D., I. D. E. A. L., L. L., B. M. B., and S. M. writing – review & editing.

Funding and additional information—The work in B. M. B.'s lab was supported by Cornell University startup funds, an Academic Venture Fund award from the Atkinson Center for Sustainability at Cornell University, a Career Award at the Scientific Interface from the Burroughs Welcome Fund, ARPA-E Award DE-AR0001341, and a gift from Mary Fernando Conrad and Tony Conrad. S. M. and A. L. L. were supported by an MRC grant (MR/S009272/1). M. G. A.-Z is funded by a Mexican Government PhD scholarship (CONAHCYT, 2021-000007-01EXTF-00221). B. J. R. is the recipient of a NERC PhD studentship (NERC ACCE NE/S00713X/1).

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: DAP, diaminopimelic acid; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; MQ, MilliQ water; MS/MS, tandem mass spectrometry; MurNAC, *N*-acetylmuramic acid; PBP, Penicillin Binding Protein; SDS, sodium dodecyl sulfate; TIC, Total Ion Chromatograms; YPM, yeast peptone mannitol.

References

SASBMB

- Weidel, W., and Pelzer, H. (1964) Bagshaped macromolecules a new outlook on bacterial cell walls. *Adv. Enzymol. Relat. Subj. Biochem.* 26, 193–232
- Vollmer, W., Blanot, D., and de Pedro, M. A. (2008) Peptidoglycan structure and architecture. *FEMS Microbiol. Rev.* 32, 149–167
- Ciak, J., and Hahn, F. E. (1957) Penicillin-induced lysis of *Escherichia coli*. Science 125, 119–120
- Park, J. T., and Strominger, J. L. (1957) Mode of action of penicillin. Science 125, 99–101
- Tipper, D. J., and Strominger, J. L. (1965) Mechanism of action of penicillins: a proposal based on their structural similarity to acyl-D-alanyl-Dalanine. *Proc. Natl. Acad. Sci. U. S. A.* 54, 1133–1141
- Cho, H., Uehara, T., and Bernhardt, T. G. (2014) Beta-lactam antibiotics induce a lethal malfunctioning of the bacterial cell wall synthesis machinery. *Cell* 159, 1300–1311
- Wietzerbin, J., Das, B. C., Petit, J. F., Lederer, E., Leyh-Bouille, M., and Ghuysen, J. M. (1974) Occurrence of D-alanyl-(D)-meso-diaminopimelic acid and meso-diaminopimelyl-meso-diaminopimelic acid interpeptide linkages in the peptidoglycan of Mycobacteria. *Biochemistry* 13, 3471–3476
- Lavollay, M., Arthur, M., Fourgeaud, M., Dubost, L., Marie, A., Veziris, N., et al. (2008) The peptidoglycan of stationary-phase *Mycobacterium* tuberculosis predominantly contains cross-links generated by L,D-transpeptidation. J. Bacteriol. 190, 4360–4366
- Mahapatra, S., Crick, D. C., McNeil, M. R., and Brennan, P. J. (2008) Unique structural features of the peptidoglycan of *Mycobacterium leprae*. *J. Bacteriol.* 190, 655–661

1-3 peptidoglycan cross-links in Acetobacteraceae

- Peltier, J., Courtin, P., El Meouche, I., Lemee, L., Chapot-Chartier, M. P., and Pons, J. L. (2011) *Clostridium difficile* has an original peptidoglycan structure with a high level of N-acetylglucosamine deacetylation and mainly 3-3 cross-links. *J. Biol. Chem.* 286, 29053–29062
- Mainardi, J. L., Fourgeaud, M., Hugonnet, J. E., Dubost, L., Brouard, J. P., Ouazzani, J., *et al.* (2005) A novel peptidoglycan cross-linking enzyme for a beta-lactam-resistant transpeptidation pathway. *J. Biol. Chem.* 280, 38146–38152
- Kim, H. S., Im, H. N., An, D. R., Yoon, J. Y., Jang, J. Y., Mobashery, S., et al. (2015) The cell shape-determining Csd6 protein from *Helicobacter* pylori constitutes a new family of L,D-Carboxypeptidase. J. Biol. Chem. 290, 25103–25117
- Godessart, P., Lannoy, A., Dieu, M., Van der Verren, S. E., Soumillion, P., Collet, J. F., *et al.* (2021) beta-Barrels covalently link peptidoglycan and the outer membrane in the alpha-proteobacterium *Brucella abortus. Nat. Microbiol.* 6, 27–33
- Sandoz, K. M., Moore, R. A., Beare, P. A., Patel, A. V., Smith, R. E., Bern, M., *et al.* (2021) beta-Barrel proteins tether the outer membrane in many Gram-negative bacteria. *Nat. Microbiol.* 6, 19–26
- Bahadur, R., Chodisetti, P. K., and Reddy, M. (2021) Cleavage of Braun's lipoprotein Lpp from the bacterial peptidoglycan by a paralog of l,d-transpeptidases, LdtF. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2101989118
- 16. Galley, N. F., Greetham, D., Alamán-Zárate, M. G., Williamson, M. P., Evans, C. A., Spittal, W. D., *et al.* (2023) *Clostridioides difficile* canonical L,D-transpeptidases catalyse a novel type of peptidoglycan cross-links and are not required for beta-lactam resistance. *J. Biol. Chem.*, 105529
- Winkle, M., Hernandez-Rocamora, V. M., Pullela, K., Goodall, E. C. A., Martorana, A. M., Gray, J., et al. (2021) DpaA detaches braun's lipoprotein from peptidoglycan. *mBio* 12, e00836-00821
- Bern, M., Beniston, R., and Mesnage, S. (2017) Towards an automated analysis of bacterial peptidoglycan structure. *Anal. Bioanal. Chem.* 409, 551–560
- Sutterlin, L., Edoo, Z., Hugonnet, J. E., Mainardi, J. L., and Arthur, M. (2018) Peptidoglycan cross-linking activity of L,D-Transpeptidases from *Clostridium difficile* and inactivation of these enzymes by beta-lactams. *Antimicrob. Agents Chemother.* 62, e01607–01617
- More, N., Martorana, A. M., Biboy, J., Otten, C., Winkle, M., Serrano, C. K. G., et al. (2019) Peptidoglycan remodeling enables *Escherichia coli* to survive severe outer membrane assembly defect. mBio 10, e02729-18
- Hernandez, S. B., Castanheira, S., Pucciarelli, M. G., Cestero, J. J., Rico-Perez, G., Paradela, A., et al. (2022) Peptidoglycan editing in nonproliferating intracellular Salmonella as source of interference with immune signaling. *PLoS Pathog.* 18, e1010241
- 22. Espaillat, A., Forsmo, O., El Biari, K., Bjork, R., Lemaitre, B., Trygg, J., et al. (2016) Chemometric analysis of bacterial peptidoglycan reveals atypical modifications that empower the cell wall against predatory enzymes and fly innate immunity. J. Am. Chem. Soc. 138, 9193–9204
- 23. Patel, A. V., Turner, R. D., Rifflet, A., Acosta-Martin, A. E., Nichols, A., Awad, M. M., et al. (2021) PGFinder, a novel analysis pipeline for the consistent, reproducible, and high-resolution structural analysis of bacterial peptidoglycans. *Elife* 10, e70597
- 24. Schmitz, A. M., Pian, B., Medin, S., Reid, M. C., Wu, M., Gazel, E., et al. (2021) Generation of a Gluconobacter oxydans knockout collection for improved extraction of rare earth elements. Nat. Commun. 12, 6693
- 25. Fricke, P. M., Lurkens, M., Hunnefeld, M., Sonntag, C. K., Bott, M., Davari, M. D., et al. (2021) Highly tunable TetR-dependent target gene expression in the acetic acid bacterium *Gluconobacter oxydans. Appl. Microbiol. Biotechnol.* 105, 6835–6852
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009) BLAST+: architecture and applications. BMC Bioinformatics 10, 421
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589
- 28. van Kempen, M., Kim, S. S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C. L. M., et al. (2023) Fast and accurate protein structure search

J. Biol. Chem. (2024) 300(1) 105494 11

248

with Foldseek. Nat. Biotechnol. https://doi.org/10.1038/s41587-023-01773-0

- 29. Sacco, E., Hugonnet, J. E., Josseaume, N., Cremniter, J., Dubost, L., Marie, A., et al. (2010) Activation of the L,D-transpeptidation peptidoglycan cross-linking pathway by a metallo-D,D-carboxypeptidase in *Enterococcus* faecium. Mol. Microbiol. 75, 874–885
- 30. Gomes, R. J., Borges, M. F., Rosa, M. F., Castro-Gomez, R. J. H., and Spinosa, W. A. (2018) Acetic acid bacteria in the food industry: systematics, characteristics and applications. *Food Technol. Biotechnol.* 56, 139–151
- Studier, F. W. (2018) T7 expression systems for inducible production of proteins from cloned genes in *Escherichia coli. Curr. Protoc. Mol. Biol.* 124, e63
- Glauner, B. (1988) Separation and quantification of muropeptides with high-performance liquid chromatography. *Anal. Biochem.* 172, 451–464
- 33. Blum, M., Chang, H. Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., et al. (2021) The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. 49, D344–D354
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., et al. (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240
- Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. (2017) Julia: a fresh approach to numerical computing. *SIAM Rev.* 59, 65–98
- Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797

- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018) UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., *et al.* (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534
- Letunic, I., and Bork, P. (2021) Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296
- Mirdita, M., Schutze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022) ColabFold: making protein folding accessible to all. *Nat. Methods* 19, 679–682
- 42. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., et al. (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with highaccuracy models. Nucleic Acids Res. 50, D439–D444
- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2021) Pfam: the protein families database in 2021. Nucleic Acids Res. 49, D412–D419
- 44. Wang, J., Chitsaz, F., Derbyshire, M. K., Gonzales, N. R., Gwadz, M., Lu, S., et al. (2023) The conserved domain database in 2023. Nucleic Acids Res. 51, D384–D388

Appendix 3

This appendix represents my last co-authorship produced during my thesis study.

Alamán-Zárate, M. G^{*}., Rady, B. J.*, Ledermann, R., Shephard, N., Evans, C.A., Dickman, M.J., Turner, R.D., Rifflet, A., Patel, A.V., Gomperts Boneca, I., Poole, P.S., Bern, M., & Mesnage, S. (2025). A software tool and strategy for peptidoglycomics, the high-resolution analysis of bacterial peptidoglycan via LC-MS/MS. Communications Chemistry, 8 (1), 91. <u>https://doi.org/10.1038/s42004-025-01490-6</u>

*First co-authors

A Nature Portfolio journal

Article

9

https://doi.org/10.1038/s42004-025-01490-6

A software tool and strategy for peptidoglycomics, the high-resolution analysis of bacterial peptidoglycans via LC-MS/MS

Check for updates

Marcel G. Alamán-Zárate ^{1,7}, Brooks J. Rady ^{1,7}, Raphael Ledermann², Neil Shephard ³, Caroline A. Evans ⁴, Mark J. Dickman ⁴, Robert D. Turner³, Aline Rifflet⁵, Ankur V. Patel¹, Ivo Gomperts Boneca ⁵, Philip S. Poole², Marshall Bern ⁶ & Stéphane Mesnage ¹⊠

Peptidoglycan is an essential component of the bacterial cell envelope — a mesh-like macromolecule that protects the bacterium from osmotic stress and its internal turgor pressure. The composition and architecture of peptidoglycan is heterogeneous and changes as bacteria grow, divide, and respond to their environment. Though peptidoglycan has long been studied via LC-MS/MS, the analysis of this data remains challenging as peptidoglycan's unusual composition and branching can't be handled by proteomics software. Here we describe user-friendly open-source tools and a web interface for building peptidoglycan databases, performing MS searches, and predicting the MS/MS fragmentation of muropeptides. We then use *Rhizobium leguminosarum* to describe a step-by-step strategy for the high-resolution analysis of peptidoglycan. The unprecedented detail of *R. leguminosarum*'s peptidoglycan composition (>250 muropeptides) reveals even the subtlest remodelling between growth conditions. These new and easier to use tools enable more systematic analyses of peptidoglycan dynamics.

Peptidoglycan (PG) is a ubiquitous and essential component of the bacterial cell envelope, which forms a single bag-shaped macromolecule (or sacculus) around the cell¹. PG synthesis has been extensively studied since many antibiotics work by disrupting it, including widely used β -lactam antibiotics (like penicillin) and last resort antibiotics such as vancomycin^{2,3}. The composition and remodelling dynamics of PG during growth, division, and differentiation can be critical for maintaining cell viability in response to changing environmental conditions. During this remodelling, PG fragments are naturally released into the environment; those released by the microbiota are important microbe-associated molecular patterns (MAMPs) recognised by the innate immune system⁴. They can contribute to acute or chronic inflammatory diseases and are thought to be key signalling molecules in the gut-brain axis^{5,6}. PG fragments have also been shown to mediate more unusual symbiotic relationships, as in the case of the Hawaiian bobtail squid, where bioluminescent *Vibrio fischeri* provide the host with nocturnal

camouflage $\bar{}$, PG's unique role in bacterial adaptation, pathogenesis, and symbiosis makes it an essential molecule to study.

Whilst the overall structure of PG and its building blocks are well conserved, it is continually restructured and modified as bacteria grow and divide, introducing vast and often subtle complexity. Monitoring PG structural dynamics, therefore, requires automated, robust, and sensitive tools. Most analyses currently involve a biased identification of major peaks in UV absorbance chromatograms⁸ that precludes the identification of low abundance or co-eluting muropeptides. The limited number of muropeptides commonly described this way (usually 10–25, even for so-called "high-resolution analyses")^{9,10} does not provide enough detail to track the variation in important muropeptides like those corresponding to covalent protein anchoring. To achieve this greater level of detail, other studies have made use of LC-MS/MS—even proposing the term "peptidoglycomics" for the discipline in 2013¹¹. Despite this, a lack of software tools and published search

¹Molecular Microbiology, School of Biosciences, University of Sheffield, Sheffield, UK. ²Department of Biology, University of Oxford, Oxford, UK. ³Research Software Engineer team, University of Sheffield, Sheffield, UK. ⁴Department of Chemical and Biological Engineering, ChELSI Institute, University of Sheffield, Sheffield, UK. ⁵Institut Pasteur, Université Paris Cité, INSERM U1306, CNRS UMR6047, Biology and genetics of the bacterial cell wall Unit, Paris, France. ⁶Protein Metrics, 225 Franklin Street, Boston, USA. ⁷These authors contributed equally: Marcel G. Alamán-Zárate, Brooks J. Rady. email: s.mesnage@sheffield.ac.uk

Communications Chemistry | (2025)8:91

strategies means that the LC-MS/MS analysis of PG has remained a tediously manual, error-prone, and inconsistent process.

To address this, several more comprehensive tools for peptidoglycomics have recently been developed^{12–15}. Whilst these tools have vastly improved the consistency and throughput of LC-MS/MS analysis, they remain either inflexible, incomplete, or difficult to use.

Our previously described tool, PGFinder^{15,16}, focused on ease-of-use and the quantification of muropeptides in LC-MS datasets but left room for improvement. Here, we build on PGFinder in two key ways: (i) by improving the usability and capability of the existing MS tool, and (ii) by including new modules that automate additional analysis steps in the LC-MS/MS pipeline. Highlights include PGFinder's new, user-friendly web interface (https://mesnage-org.github.io/pgfinder/) and PGLang, a formal language for the concise description of muropeptides that enables both automated mass calculation and MS/MS fragment prediction. Finally, to demonstrate how these improvements fit into a complete analysis pipeline, we describe a step-by-step strategy that we use to characterise the changes in Rhizobium leguminosarum's PG composition when grown on minimal (as opposed to rich) media. Empowered by this approach, we report unprecedented PG complexity (>250 muropeptides) and accurately monitor subtle changes in the PG, laying the groundwork for more systemic analyses of muropeptide composition, cross-linking, and protein anchoring in the future.

Results

Enhancing PGFinder's existing functionality with an improved MS output and web interface

The first published version of PGFinder ($v0.02^{15}$;) offered automated MS analysis but search outputs still required post-processing in Excel to build a final table of muropeptides. Processing involved Δppm calculation and

consolidation of intensities (sum) across retention times. Now, in version 1.3.2, PGFinder automatically picks the best match according to its Δ ppm and consolidates search results into a table of muropeptides sorted by abundance. Finally, a new metadata column makes it possible to keep track of the data analysed, parameters, and PGFinder version used to generate each output. Taken together, these changes are a major step towards reducing the amount of manual processing required.

To use PGFinder v0.02 without installation, we previously provided an interactive Jupyter notebook hosted on MyBinder¹⁷. This made PGFinder significantly easier to set up and use than similar tools, but the resource limitations imposed by MyBinder regularly made loading our notebook slow (or even impossible). After loading, users also needed to ensure that all cells were run, in order, exactly once and needed to manually reset the notebook between each search. To circumvent these usability issues, we built a new, intuitive web interface that makes running an MS search as simple as uploading your deconvoluted data, picking a mass database, and clicking "Run Analysis" (Fig. 1). Since all computation is now done on the client-side (via WebAssembly), we no longer require hosting services like MyBinder and loading/computation times have been dramatically reduced. Moving to this interface has also allowed for bulk processing and made it trivial to add new modules like the Mass Calculator and Fragment Generator (Fig. 1). PGFinder is now easier to pick up than ever, further encouraging its adoption by others in the field.

Condensing complex muropeptide structures into PGLang, a concise formal language

Before PGFinder could be expanded to handle tasks like mass calculation or MS/MS fragment prediction, we needed a way to model muropeptide chemistry in software. PG building blocks (muropeptides) are made of *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc)



Fig. 1 | PGFinder's web interface makes MS analysis easier than ever and enables new functionality. The interface includes the original MS analysis module for identifying PG fragments from deconvoluted LC-MS data, as well as two newly developed modules: the mass calculator for building PGFinder-compatible mass databases, and the Fragment Generator for predicting MS/MS fragment ions.

Communications Chemistry | (2025)8:91

disaccharides linked to a short (possibly branched) peptide stem containing both L and D-amino acids¹ (Fig. 2A). In PGFinder, these muropeptides are represented as monosaccharide and amino-acid residues that can be decorated with various modifications and bonded together to form a directed graph (Fig. 2B). Each residue contains distinct functional groups that are either free, modified, or donate/accept a particular bond (Fig. 2C). These rules ensure that every muropeptide is chemically valid, and tracking each muropeptide's free groups makes it possible to automatically identify potential modification sites and cross-linking positions.

To represent these muropeptide graphs compactly, we developed a language called PGLang with a minimal and straightforward syntax (Fig. 3). Each monomer is partitioned into a glycan chain (represented by lowercase letters) and a stem peptide (represented by uppercase letters). Lateral chains can be attached to any diamino or dicarboxylic amino acids using square brackets, and any residue can be modified using round brackets. Monomers can be connected via their glycan chain (~), or via cross-linked stem peptides (=). When monomers are connected via cross-linked peptides, the structure is followed by a bracketed list of cross-link descriptors: 3-3, 4–3, etc. A complete syntax diagram for PGLang is available in Fig. S1, and tables detailing the currently available monosaccharides, amino acids, and modifications are provided in Tables S1 and S2. Finally, to close the loop and move backwards from PGLang to a full molecular structure (including stereochemical information), we've included a PGLang to SMILES translator in PGFinder's new Mass Calculator module.

Expanding PGFinder to automate the mass calculation and MS/ MS fragment prediction

Once muropeptides described using PGLang have been translated into their chemical graph representations, implementing a number of new features becomes straightforward. Here that means automating two additional parts of the analysis pipeline that were previously out of scope for PGFinder: monoisotopic mass calculation and MS/MS fragment prediction. The mass of any given muropeptide is simply the sum of its residue, modification, and bond masses, and fragment prediction is a three-step process involving

bond cleavage, ion formation (depending on the bond broken, the acceptor and donor fragments may gain or lose a particular chemical group), and mass-charge ratio (m/z) calculation (all ions are currently $[M + H]^+$ adducts). Exposing this functionality are two new UI modules: the Mass Calculator, which generates mass databases that can be fed directly into the MS Analysis module, and the Fragment Generator, which produces a list of ions with PGLang-like descriptions that make it clear to users what fragment each ion represents. To get users started quickly, the Mass Calculator also includes several PGLang databases for common model organisms that can be easily downloaded and adapted using a text editor.

Rhizobium leguminosarum as a model system for describing a five-step PG analysis strategy

The *R. leguminosarum* genome encodes many D,D, and L,D-transpeptidases, so its PG structure is expected to be complex, making it a good model organism for testing our new PG analysis tools. To prepare some sample datasets, triplicate cultures of *R. leguminosarum* were grown in both minimal and rich media, and their PG was analysed via UHPLC-MS/MS (Fig. 4).

The chromatograms confirmed that *R. leguminosarum's* muropeptide profile was complex and revealed obvious differences between the two media conditions. Consequentially, the corresponding LC-MS/MS datasets were ideal for showcasing our comparative PG analysis strategy. Four sequential searches focus on monomers (step 1), modifications (step 2), PGanchored proteins (step 3) and multimers (step 4) that inform a final, fifth search producing a comprehensive muropeptide quantification that can be used in statistical comparisons (Fig. 5).

Step 1: Identifying PG monomers using MS and MS/MS. An initial, unbiased search of the TY datasets was performed using PGFinder's "MS Analysis" module. The monomer database (DB_1; Table S3) contained 223 disaccharide peptides with stem lengths ranging from one to five amino acids. A total of 131 unique matches were found within 20 ppm of the observed masses; 78 of which were found in all three datasets, 22 in only two, and 31 in just one (Table S4 and File S1).



Fig. 2 | Muropeptides are represented as a set of chemically linked, optionally modified residues. A An example muropeptide showcasing several modifications (yellow) and peptide branches (red). B The same muropeptide converted to its chemical graph representation. Free functional groups (those that remain unmodified and unbonded) are shown as black dots on each residue, bonds are shown as

labelled arrows pointing from donor to acceptor, and modifications are shown as yellow flags. **C** The five functional groups of *N*-Acetylmuramic acid are shown in detail, including the modifications each can have and the bonds they can donate/ accept.

Fig. 3 | PGLang is a simple language for describing potentially modified, branched, or cross-linked muropeptides. Each cartoon PG fragment corresponds to a colour-coded PGLang structure. Monosaccharides forming glycan chains are represented by hexagons whilst amino acids are represented by circles. Modifications are shown as flags protruding from the residue they modify (Am, Amidation; Anh, anhydroMurNAc; Glyc, Glycolylation). J represents *meso*-diaminopimelic acid.



Several matches had very unusual compositions (e.g., gm-AEJCC or gm-AEJYS) and relatively high Appm values (8.8 and 9.1, respectively), suggesting that these identifications resulted from mass coincidences. A manual inspection of MS spectra confirmed this hypothesis; many ions matching the theoretical m/z of these unusual muropeptides did not contain the signature ion corresponding to the loss of GlcNAc due to in-source fragmentation. To screen out these mass coincidences and resolve the structure of any isomers, we confirmed each monomer via MS/MS. We carried out a search of the TY datasets using the ByonicTM module of the Byos® software that can automatically analyse and score MS/MS spectra. The list of monomers to search for contained 423 muropeptides with stem lengths ranging from one to five amino acids, including disaccharidetetrapeptides with all possible residues in position four (gm-AEJX) and disaccharide-pentapeptides with all possible combinations in positions four and five (gm-AEJXX; File S2). Based on the automatic scoring of MS/MS spectra, Byonic[™] confirmed 39 monomers (Fig. S2). Further, manual inspection of the Byonic[™] output allowed us to validate an additional 29 monomers that satisfied two criteria: they were fragmented in at least one replicate and contained at least half of the expected b-ions and half of the expected y-ions. The validated muropeptides contained many tetra- and pentapeptide stems with unusual residues in their final position-indicative of D,D or L,D-transpeptidase exchange activity. Two structures with identical masses, however, could not be differentiated with certainty (gm-AEJAG and gm-AEJQ, both 998.429165 Da). Seven monomers did not meet the criteria for validation (only one MS/MS spectrum across three biological replicates or less than half of the expected b- or y-ions). The lack of b- or y-ions for the three monomers (gm-AEJQE, gm-AEJEQ, and gm-AEJWW) prompted us to explore the corresponding MS/MS spectra and revealed that ByonicTM's automatic identification of these monomers was incorrect. The gm-AEJQE and gm-AEJEQ monomers were found to really be gm-AEJADA, whilst gm-AEJWW was, in fact, gm-AEJ = AEJ (3-3) (a dimer of tripeptides missing a disaccharide moiety). The 29 validated monomers that were present in all three TY replicates became the monomer database DB_2 (Table S5).

Step 2: Identifying PG modifications. Database DB_2 was then used to determine which monomers were the most abundant using PGFinder. A subset of 11 muropeptides, accounting for >90% of the monomer abundance (Table S6 and File S3), were selected to create a third database called DB_3 (Table S7). DB_3 was then used to search for six different

modifications using PGFinder. The modifications considered are listed in Table 1 and can be sorted into glycan modifications (de-acetylation, *O*-acetylation, and 1,6-anhydroMurNAc), peptide modifications (amidation), and hydrolysis products resulting from Glucosaminidase (loss of GlcNAc) or amidase activity (presence of an extra GlcNAc-MurNAc). For each modification, matches were consolidated (summing the intensities of matches found at different retention times), and matches absent from any of the three replicates were discarded. Three additional criteria were then used to validate the modified muropeptide matches: (i) a retention time consistently higher or lower (depending on the modification considered) than the unmodified muropeptide; (ii) the presence of signature ions corresponding to each modification; (iii) a similar relative abundance of modified and unmodified muropeptides (Table 1).

As an example, Fig. 6A shows how nine putative AnhydroMurNAccontaining muropeptides were identified. Most modified muropeptides were present in all three replicates and had a consistently higher retention time than their unmodified counterparts. Next, we manually searched the MS/MS data for signature fragment ions predicted by PGFinder's "Fragment Generator" module. Fig. 6B summarises the MS/MS analysis of the gm(Anh)-AEJA monomer. Out of the 20 predicted fragment ions, 12 were present, including five out of nine possible signature ions (highlighted in red in Fig. 6B). The presence of these fragment ions ultimately contributes to validating the gm(Anh)-AEJA monomer.

On average, AnhydroMurNAc-modified muropeptides were 10% as abundant as their unmodified counterparts, with a particularly high proportion of the disaccharide-tripeptide gm-AEJ being modified (48% of its unmodified intensity) (Fig. 6A). In the end, only the anhydro versions of the three most abundant monomers (gm-AEJA, gm-AEJG, and gm-AEJ) could be confirmed (Fig. 6A), though this was in part due to a lack of MS/MS data for the other matches. Five other modifications (Table 1) were searched for using the same strategy, but none of these modifications could be confirmed via MS/MS (Fig. S3).

Step 3: Identifying outer membrane proteins covalently anchored to the PG. The covalent anchoring of outer membrane β -barrel proteins to the PG helps maintain cell envelope integrity in Alphaproteobacteria^[18,19]. We identified ten putative β -barrel proteins encoded by the genome of *R. leguminosarum* bv. *viciae* (strain 3841) (Fig. S4) and investigated if any of them were anchored to the PG.

Communications Chemistry | (2025)8:91



Fig. 4 | R. leguminosarum's PG composition is complex and varies between growth conditions. Total ion chromatograms (TICs) show reduced R. leguminosarum muropeptides. PG was extracted from cells grown in either rich (TY) or minimal (MM) media. The TICs corresponding to each triplicate are shown.

A fourth database (DB_4, Table S8) was created to search for any amino acid scars left behind by β -barrel proteins that had been attached to the PG. Since tetrapeptide stems are thought to act as the donors during L,D-transpeptidase mediated protein anchoring, DB_4 contained muropeptides comprised of disaccharide tripeptides (gm-AEJ) followed by residues corresponding to the N-terminal of each anchored porin (with its signal peptide removed; Fig. S5A). Although trypsin digestion is expected to generate porin "scars" with a basic residue at the C-terminal, previous analyses revealed that muropeptides containing non-canonical and missed cleavages are common. To avoid missing any of these non-canonical "scars", the muropeptides in DB_4 contained every N-terminal porin sequence from 2 to 17 amino acids in length.

A total of 29 masses matching PG-anchored β -barrel proteins were found in all three TY replicates (File S4), though MS/MS data was only available to confirm muropeptides associated with RopA1,2,3, RopB, and pRL90069 (Fig. S5B). Collectively, muropeptides corresponding to these proteins accounted a total intensity of 3.51E + 08, representing 13.3% of all monomer intensity (2.64E + 09) (Table 2).

Step 4: Identifying PG multimers and confirming their structure. PGFinder's "MS Analysis" module was used to identify dimers and trimers resulting from both D,D- and L,D-transpeptidation using the monomers previously validated (DB_2; Table S6) as acceptors (apart from gm-AE). A total of 88 multimeric masses (41 dimers and 47 trimers) were present across all three replicates (File S5 and Table S9).

Out of the 41 dimers, 28 could be unambiguously classified as products of D,D or L,D-transpeptidation. This was possible because muropeptides like gm-AEJ=gm-AEJX (3-3) could only be formed via L,D-transpeptidation, and gm-AEJA=gm-AEJAX (4–3) structures could only be formed via D,D-transpeptidation. To differentiate the remaining 13 muropeptides, MS/ MS analysis was necessary. They were either (i) isomers—with the same residues but distinct cross-linking; e.g., gm-AEJA=gm-AEJ (4–3) vs gm-AEJ=gm-AEJA (3-3)—or (ii) mass coincidences—with distinct compositions but the same chemical formula; e.g., gm-AEJ=gm-AEJQ (3-3) vs gm-AEJ=gm-AEJAG (3-3). In the Q vs AG case, MS/MS did not allow us to discriminate between the two structures, as a fragmentation between the C-terminal alanine and glycine was never observed. It was possible,

Communications Chemistry | (2025)8:91



Fig. 5 | An end-to-end strategy for PG structural analysis via LC-MS/MS. Sequential searches were performed using PGFinder and Byonic[™]. The monomer database DB_2 was built based on MS/MS analysis. The most abundant monomers were then used to build DB_3, which was used to identify modified muropeptides. DB_4 contained muropeptides with a gm-AEJ stem followed by N-terminal porin sequences (with signal peptides removed). DB_2 was used to identify dimers and trimers, and then MS/MS data from matching output was manually inspected to differentiate between mass coincidences. A final search was carried out with DB_5, which combines muropeptides from DB_2 (monomers) and the muropeptides from DB_4 corresponding to the MS/MS confirmed RopA1,2,3, RopB, and pRL90069 porins. The final PGFinder search, with anhydroMurNAc modifications and 3–3/ 4–3 multimers enabled, was carried out with a 5 ppm tolerance. The final search output was manually inspected and modified to remove any known mass coincidences.

Table 1 | A list of the PG modifications searched for, and the strategy used to validate them

Modification		Description	Retention time	Mass change	Signature ion <i>m/z</i> ([M+H] ⁺)	Relative abundance	Present
Glycan modification	าร					1) Similar % of modified monomers for	
AnhydroMurNAc (Anh)		1,6-anhydroMurNAc	Increased	-20.026	258.096 (AnhMurNAc)	all monomers 2) Belative abundance of modified	YES
Deacetylation	(DeAc)	Loss of acetyl group	Decreased	-42.010	162.077 (GlcN); 236.113 (MurN ^{red})	monomers should be similar to the relative abundance of unmodified	NO
O-acetylation	(Ac)	Gain of acetyl group	ain of acetyl group Increased +42.010 246.098 (GlcNAc+Ac) monomers	- monomers	NO		
					320.134 (MurNAc ^{red} +Ac)	-	
Peptide modifications						-	
Amidation	(Am)	Glu or <i>m</i> Dap amidation	Decreased	-0.984	172.109 (mDAP _{NH2}); 129.066 (Gln)	-	NO
Hydrolysis products						-	
Loss of g		Glucosaminidase	Decreased	-203.079	278.124 (MurNAc ^{red})	_	NO
		product			206.087 (MurNAc ^{red} -Lactyl)	_	
Extra gm		Partial muramidase digestion	Increased	+478.180	276.108 (MurNAc)	-	NO

however, to assign one of the two possible isomeric structures—gm-AEJA=gm-AEJ (4–3) or gm-AEJ=gm-AEJA (3-3) to each of the three peaks in the extracted ion chromatogram (m/z = 897.88; Fig. 7A).

As a first step, we predicted a list of fragment ions using PGFinder's "Fragment Generator" for each of the isomeric structures (71 fragments for the 3-3 dimer and 55 for the 4–3 dimer, coming together to form a set of 59 unique m/z values; Table S10). As expected, a large proportion of the predicted ions (68%, 59%, and 58%) were detected in each peak. To assign each of the three peaks in the extracted ion chromatogram to a 3-3 or a 4–3 dimer,

we computed a list of signature ions for each structure (8 for the 4–3 dimer and 16 for the 3-3 dimer; Fig. 7B) and recorded their presence and intensity in each peak. The intensity associated with each set of signature ions allowed us to conclude that peak 1 was comprised of mostly 4–3 dimer whilst peaks 2 and 3 were mostly 3-3 dimer. The remaining ambiguous dimers were then analysed using the same strategy. Out of 46 total dimers identified, 21 contained 3-3 cross-links, and 25 contained 4–3 cross-links (Table S11).

Amongst the 47 trimers identified (Table S10), 26 matched only a single structure, Given the complexity of each MS/MS analysis and the prior

Communications Chemistry | (2025)8:91

Α

Characture a	Interation	0/ Ambudan	DT (min)	AT (min)	Pressent of simultance is a
Structure	Intensity	% Annyaro	RT (min)	ΔI (min)	Presence of signature ions
gm-AEJA 1	1.1E+09	3.5%	9.50 ± 0.03	+ 6 62	Ves: validated with 1+ and 2+ ions (TV1 dataset)
gm(Anh)-AEJA 1	3.7E+07	3.570	16.12 ± 0.00		res, validated with 1+ and 2+ lons (111 dataset)
gm-AEJG 1	2.4E+08	7 7%	6.85 ± 0.04	+ 7 42	Voc: validated with 1+ and 2+ ions (TV1 datacet)
gm(Anh)-AEJG 1	1.8E+07	1.270	14.27 ± 0.01	+ 7.42	res, validated with 1+ and 2+ lons (111 dataset)
gm-AEJ 1	1.6E+08	19 00/	5.48 ± 0.03	+ 9 10	Voc: validated with 1+ and 2+ ions (TV1 datacet)
gm(Anh)-AEJ 1	7.9E+07	40.0%	13.58 ± 0.00	+ 8.10	res, validated with 1+ and 2+ lons (1+1 dataset)
gm-AEJAI 1	1.5E+08	0.6%	21.40 ± 0.00		No fragmented ion available
gm(Anh)-AEJAI 1	8.8E+05	0.0%	28.39 ± 0.02	+ 0.99	No tragmented for available
gm-AEJF 1	1.5E+08	2.0%	21.62 ± 0.00	. 7.09	No frequented ion quailable
gm(Anm)-AEJF 1	4.5E+06	3.0%	3.0% 28.70 ± 0.00		No tragmented for available
gm-AEJAA 1	6.5E+07	0.0%	10.86 ± 0.02	+ 6.40	Only present in 2 of 2 datasets
gm(Anh)-AEJAA 1	5.9E+05	0.576	17.26 ± 0.02	+ 0.40	Only present in 2 or 5 datasets
gm-AEJK 1	4.6E+07	0.4%	7.71 ± 0.06		No frequented ion quailable
gm(Anh)-AEJK 1	4.3E+06	9.4%	14.12 ± 0.01	+ 0.41	No tragmented fon available
gm-AEJAG 1, gm-AEJQ 1	4.6E+07	1 40/	8.80 ± 0.05	1 11 15	No cignoture ions found
gm(Anh)-AEJAG 1, gm(Anh)-AEJQ 1	6.5E+05	1.470	19.95 ± 6.37	+ 11.15	No signature foris round.
gm-AEJAD 1	2.5E+07	11 70/	9.61 ± 0.04	+ 6 21	No fragmented ion available
gm(Anh)-AEJAD 1	2.9E+06	11.7%	15.82 ± 0.01	+ 0.21	No magnement on available
gm-AEJAF 1	5.0E+07	0.0%	15.4 ± 0.0	-	No modification found

^a Non modified monomers are sorted by abundance (most abundant first)

 $^{\rm b}~\Delta T$ is defined as the difference (in min) between the average RT of the unmodified and modified muropeptide



Fig. 6 | The existence of AnhydroMurNAc-modified monomers can be confirmed via MS/MS. A Summary of unmodified monomers and their anhydroMurNAc counterparts identified by PGFinder. For each muropeptide, the intensity and abundance of the AnhydroMurNAc modification are provided, alongside the retention time. The shift in retention time (Δ T (min)) and the presence of signature ions are indicated. **B** Example MS/MS spectrum showing the identified signature ions in red for a singly charged ([M + H]⁺) ion corresponding to the gm(Anh)-AEJA muropeptide. RT retention time.

validation of dimers, we chose to simply assign ambiguous trimer matches to the structure built from the most abundant dimer linked to the most abundant donor. For example, structures like gm-AEJA=gm-AEJA=gm-AEJA (4–3, 4–3) were chosen over structures like gm-AEJA-gm-AEJ-gm-AEJAA (4–3, 3-3). The final list of 121 muropeptides, including monomers, dimers, porins, and their modified counterparts is described in Table S12.

Step 5: Final quantification of muropeptides and comparison of growth conditions. Growing *R. leguminosarum* in minimal media (MM) as opposed to rich TY media leads to changes in the muropeptide profile, suggesting that PG remodelling occurs under these conditions. We sought to apply the strategy described (and summarised in Fig. 5) to compare the PG structure of *R.* leguminosarum grown in rich and MM.

To perform a final quantification, we combined the monomers from DB_2 with DB_4's porin muropeptides from RopA1,2,3, RopB, and pRL90069 to generate the database DB_5 (Table S13). This database was then used to perform a "one off" search using PGFinder's new bulk processing feature. All three TY and MM datasets were searched with a low mass tolerance (5 ppm) and anhydroMurNAc modifications and 3-3/4–3 multimers enabled. Individual search outputs were consolidated and manually checked wherever retention times had a standard deviation of

more than 0.5 min. Dimer and trimer ambiguities were resolved using the strategy described in Step 4.

The final list of muropeptides contains 255 structures found across all three biological replicates of either condition: 65 monomers, 97 dimers, and 93 trimers. 111 muropeptides were exclusively found in the TY datasets, and 25 were exclusively found in MM (File S6). Comparing the two conditions reveals subtle differences in PG remodelling (Table 3 and Fig. 8). Growth in MM was associated with a significantly lower cross-linking index $(28.5\% \pm 0.4\% \text{ vs } 31.3\% \pm 0.5\%; P = 0.002)$ and a significant increase in glycan chain length (18.4 vs 21.1 residues; P = 0.011). Interestingly, we found a moderate but significant increase of 3-3 cross-links in the MM samples (64.7% ± 0.3% vs 62.6% ± 0.4%; P = 0.003). Whilst 3-3 cross-linking increased for dimers and trimers, L,D-transpeptidase-mediated exchange activity (which leads to non-canonical residues in the fourth position) drastically dropped in the MM samples: non-canonical AEJX peptide stems represented only $2.18\% \pm 0.4\%$ of all muropeptides as compared to 26.6% \pm 0.5% in the TY samples (P > 0.001). A significant decrease in the proportion non-canonical AEJAX peptide stems was also found in the MM samples (16.6% ± 0.3% vs 5.7% ± 0.4%; P < 0.001). Finally, a significant increase in the proportion of PG-bound porin peptides was observed in the MM datasets (5.2% \pm 0.9% vs 9.1% \pm 2.0%; P = 0.037).

Communications Chemistry | (2025)8:91

Table 2	PGFinder identification of N-terminal	peptides from	RopA1,2,3, RopI	3, and pRL90069 a	nchored to the disa	ccharide-
tripepti	de gm-AEJ					

	Structure	Av. Intensity	TY1	TY2	түз	Sum
RopA1,2,3, RopB or PRL90069	gm-AEJAD	2.52E+07	3.66E+07	2.05E+07	1.86E+07	2.52E+07
RopA1,2,3 or RopB	gm-AEJADA	1.69E+08	2.35E+08	1.66E+08	1.07E+08	1.69E+08
RopA1,2,3	gm-AEJADAIVA	8.03E+07	7.96E+07	6.78E+07	9.36E+07	
	gm-AEJADAIVAA	1.14E+07	1.87E+07	9.81E+06	5.74E+06	
	gm-AEJADAIVAAEPEPVE	8.19E+06	9.59E+06	4.72E+06	1.03E+07	
	gm-AEJADAIV	7.24E+06	8.39E+06	7.55E+06	5.77E+06	
	gm-AEJADAIVAAEPEPV	2.41E+06	3.28E+06	1.14E+06	2.79E+06	
	gm-AEJADAI	1.24E+06	1.54E+06	1.14E+06	1.04E+06	
	gm-AEJADAIVAAE	9.13E+05	9.10E+05	9.26E+05	9.02E+05	
	gm-AEJADAIVAAEPEP	6.88E+04	6.88E+04	ND	ND	1.12E+08
RopB or pRL90069	gm-AEJADAV/gm-AEJADLG	3.01E+06	3.91E+06	2.84E+06	2.28E+06	3.01E+06
RopB	gm-AEJADAVDQVPEAPVAQ	1.69E+07	1.44E+07	1.29E+07	2.35E+07	
	gm-AEJADAVDQVPEAPVAQE	2.50E+06	2.96E+06	2.31E+06	2.23E+06	
	gm-AEJADAVD	2.28E+06	3.18E+06	2.33E+06	1.32E+06	
	gm-AEJADAVDQVPEAP	1.40E+06	1.48E+06	1.61E+06	1.12E+06	2.31E+07
pRL90069	gm-AEJADLGTRTYEEPDLRNGV	8.12E+06	1.56E+07	4.60E+06	4.17E+06	
	gm-AEJADL	5.55E+06	5.43E+06	6.16E+06	5.08E+06	
	gm-AEJADLGTR	5.13E+06	6.56E+06	4.55E+06	4.29E+06	1.88E+07
						3.51E+08



Theo	Signature ions						
m/z	gm-AEJA=gm-AEJ (4-3)	gm-AEJ=gm-AEJA (3-3)					
90.0550		A					
191.1026	L						
244.1292	AL						
244.1292	EJ						
345.1769		J=J (3-3)					
444.2089	AEJA						
474.2195		J=EJ (3-3)					
545.2566		J=AEJ (3-3)					
503.2620		EJ=EJ (3-3)					
568.2985	m(r)-AEJ						
574.2992		EJ=AEJ (3-3)					
721.3250	m(r)-AEJA						
745.3363		AEJ=AEJ (3-3)					
822.3727		m(r)-AEJ=J (3-3)					
871.3779	gm(r)-AEJ						
924.4044	gm(r)-AEJA						
951.4153		m(r)-AEJ=EJ (3-3)					
.022.4524		m(r)-AEJ=AEJ (3-3)					
025.4521		gm(r)-AEJ=J (3-3)					
154.4947		gm(r)-AEJ=EJ (3-3)					
225.5318		gm(r)-AEJ=AEJ (3-3)					
299.5686		m(r)-AEJ=m(r)-AEJ (3-3)					
502.6480		m(r)-AEJ=gm(r)-AEJ (3-3)					
705.7273		gm(r)-AEJ=gm(r)-AEJ (3-3)					

Fig. 7 | Validation of gm-AEJ=gm-AEJA and gm-AEJA=gm-AEJ mass coincidence. A Predicted list of signature ions for gm-AEJ=gm-AEJA (3–4 cross-link) and gm-AEJA=gm-AEJ (3-3 cross-link). B MS data corresponding to 897.8918 m/z ion (top panel) and extracted ion chromatogram showing 3 peaks at consecutive retention times (bottom).

Discussion

The numerous functionality improvements to PGFinder's existing MS tool, as well as the new, PGLang-enabled mass calculation and MS/MS fragment prediction features, were key to our PG analysis strategy. By making these improved tools accessible through an easy-to-use web interface and laying out our approach in step-by-step tutorials, we hope to encourage others to adopt our rigorous and reproducible LC-MS/MS pipeline. Though PG structural analysis remains challenging, we feel that the improvements made to PGFinder throughout this work are a significant step towards the eventual elimination of labor-intensive and error-prone manual analysis.

Since PGFinder was first described, two other tools dedicated to the LC-MS analysis of PG have been published^{13,14}. Every existing tool, PGFinder included, comes with its own trade-offs and rank differently when it comes to flexibility, completeness, and ease-of-use. HAMA, for example, is one of the more complete tools, covering the whole LC-MS/MS pipeline, but is written in MATLAB and lacks a GitHub repository. This makes it impossible to adapt or extend without a MATLAB licence and difficult to contribute those improvements back to HAMA. Additionally, HAMA doesn't currently build 3-3 cross-linked dimers, because they are often confused with 4–3 dimers due to the similarity of their fragmentation

Communications Chemistry | (2025)8:91

Table 3 | Comparative muropeptide analysis of PG extracted from cells grown in TY or MM

	TY Average SD	MM Average SD	Unpaired <i>t</i> test <i>P</i> value (significance)
Monomers (inc. porins)	30.72% ±1.06%	35.57% ±1.17%	
Dimers (inc. porins)	49.65% ±0.85%	42.94% ±0.31%	
Trimers (inc. porins)	19.63% ±0.22%	21.49% ±0.90%	
3–3	62.56% ±0.45%	64.68% ±0.31%	0.003 (**)
4–3	37.44% ±0.45%	35.32% ±0.30%	0.002 (**)
Cross-linking index	31.30% ±0.49%	28.53% ±0.43%	0.002 (**)
AEJX monomers (excl. AEJA)	4.86% ±0.21%	1.56% ±0.34%	
AEJX dimers (excl. AEJA)	16.01% ±0.50%	0.50% ±0.06%	
AEJX trimers (excl. AEJA)	5.70% ±0.05%	0.12% ±0.01%	
All AEJX	26.58% ±0.52%	2.18% ±0.41%	<0.001 (***)
AEJAX monomers (excl. AEJAA)	4.01% ±0.09%	3.20% ±0.17%	
AEJAX dimers (excl. AEJAA)	5.28% ±0.11%	1.78% ±0.27%	
AEJAX trimers (excl. AEJAA)	7.31% ±0.26%	0.67% ±0.14%	
All AEJAX	16.59% ±0.29%	5.65% ±0.41%	<0.001 (***)
Anh monomers (inc. porins)	2.18% ±0.08%	2.26% ±0.15%	
Anh dimers (inc. porins)	5.95% ±0.17%	4.98% ±0.16%	
Anh trimers (inc. porins)	0.89% ±0.02%	0.07% ±0.02%	
All anhydro	9.02% ±0.14%	7.77% ±0.32%	
Chain length	18.37±0- .23	21.08±1 03	0.011 (*)
Porin monomers	3.83% ±0.63%	5.28% ±1.26%	
Porin dimers	1.17% ±0.23%	3.08% ±0.60%	
Porin trimers	0.23% ±0.04%	0.75% ±0.17%	
All porins	5.23% ±0.87%	9.12% ±2.01%	0.037 (*)

spectra¹³. In contrast, PGFinder is written entirely using open-source programming languages, follows software best practices, and is hosted on GitHub where anyone can easily contribute improvements back to the project. Though PGFinder doesn't yet automate MS/MS analysis, our signature ion approach makes it possible to know for certain if 3-3 or 4–3 crosslinks are present, a critical distinction when it comes to assessing things like antibiotic resistance or L,D-transpeptidase activity. Another powerful tool for the in silico fragmentation of muropeptides, PGN_MS2, suffers from an incomplete description of its Python dependencies, making the installation process challenging. Additionally, by operating entirely at the atomic level, the fragment generation process is slow, and fragments can only be

Communications Chemistry | (2025)8:91

described in SMILES, making it difficult to tell at a glance which fragment came from which part of the muropeptide. Finally, though its fragment prediction is currently more complete than either HAMA or PGFinder's, it's limited to this task only; users will need to use another tool for the actual MS and MS/MS analysis¹⁴. PGFinder, on the other hand, requires no installation whatsoever, and its residue-graph abstraction makes generating fragments orders of magnitude faster than PGN_MS2 whilst giving them each useful PGLang-like names. Additionally, MS analysis can be done within PGFinder itself, only requiring additional software for data deconvolution/ feature extraction. The next steps for PGFinder are clear: more of the LC-MS/MS analysis pipeline can be covered by further automating tasks like cross-replicate consolidation, summary statistic generation, and MS/MS analysis/disambiguation. These changes, along with incorporating data deconvolution into PGFinder directly (eliminating the need for MaxQuant or Byos), would bring PGFinder closer to being a true, one-click muropeptide analysis tool.

We believe that the wider adoption of PGLang could help address the inconsistency of muropeptide descriptions throughout the literature. This inconsistency can make it difficult to understand the composition of many muropeptides; for example, the monomer GlcNAc-MurNAc-Ala-GlumDAP-Gly (gm-AEJG in PGLang) has been described in many ways: GM-Tripeptide + Gly²⁰, (NAG)(NAM)-AemG¹⁴, AEmG²¹, Tri-Gly²², DS-TP-Gly²³, M3G²⁴, B-M-l(-A-E-H-G)¹³, or even as numbers originally defined in other publications²⁵. The description of dimers, trimers, and modifications are likewise inconsistent. By building on the existing intuition of those in the field, PGLang aims to remain intuitive whilst striking the right balance between concision and unambiguity. The automated monoisotopic mass calculation of PGLang structures will also help to address a surprising inconsistency in masses reported by the literature. For example, the theoretical monoisotopic mass of the major reduced monomer in E. coli (gm-AEJA, 941.407702 Da) is reported variably as: 941.4099 in ref. 26 (Δppm=2.3), 941.4030 in ref.²¹ (Δppm=5.0), 941.41 in ref. 27 (Δppm=2.4) or 941.4064 in ref. 28 ($\Delta ppm = 1.4$). Additionally, the PGLang to SMILES translator can be used to get stereoisomer-resolved structures that make obtaining a chemical formula, chemical drawing, or protein-ligand docking trivial (using a tool like Boltz-1²⁹;). Note that whilst PGN_MS2 does output SMILES structures for the muropeptides it fragments¹⁴, these structures do not contain stereoisomer information and are therefore unsuitable for docking into stereospecific enzymes like L.D- and D.D-transpeptidases. By designing PGLang to be easy for humans to read and by including a number of useful tools for its translation and manipulation, we hope that it can become a standard nomenclature capable of improving consistency throughout the field.

Our proof-of-concept study describing PGFinder v0.02¹⁵ was largely limited to a description of the software. The step-by-step strategy laid out in this paper allows any user with a basic understanding of PG structure to perform comprehensive structural analyses. Using *R. leguminosarum* as a model system, we identified 265 muropeptides, which represents (by far) the most comprehensive PG analysis to date and the first PG characterisation of this organism. This work provides a solid foundation for exploring cell envelope remodelling occurring throughout the rhizobial life cycle: from a free-living soil-dwelling bacterium to a terminally differentiated bacteroid that can fix atmospheric nitrogen. The remodelling of PG has been associated with morphogenetic changes during growth and exposure to various stressors, but how specific enzymes contribute to this adaptation remains poorly understood. The level of detail of our analysis will allow us to more easily investigate the roles played by these PG remodelling enzymes in the future.

Another aspect of PG analysis that requires highly sensitive tools is the description of covalently anchored proteins. In *R. leguminosarum*, we demonstrated that a large proportion of muropeptides contain N-terminal residues from β -barrel proteins. This covalent anchoring of β -barrel proteins is known to tether the outer membrane in the closely related genera *Coxiella* and *Brucella* and play a role in maintaining cell envelope integrity^{18,19}. The increase in the proportion of PG fragments with β -barrel

Article

Fig. 8 | *R. leguminosarum* grown in minimal media (MM) has a different PG composition compared to those grown in rich media (TY). The Sankey diagram shows the total PG composition broken down first by oligomerisation state, then by stem peptides. Branch size is proportional to percentage, and only peptides stems are represented. A, L- or Dalanine; E, γ-D-glutamic acid; J, *meso*-diaminopimelic acid; X, any residue except Alanine.



"scars" in MM may, therefore, be indicative of an increase in envelope stress. Further studying the dynamics of this process and establishing if distinct βbarrel proteins are preferentially anchored under different conditions, as has been shown in Coxiella burnetii¹⁹, would provide a valuable insight into rhizobial adaptation and symbiosis. An increase in 3-3 cross-linking by L,Dtranspeptidases has also previously been implicated in stress resistance and cell envelope homoeostasis³⁰. In the case of Rhizobium, growth in MM has a significant (albeit subtle) impact on the abundance of 3-3 cross-links, but a dramatic impact on the abundance of gm-AEJX muropeptides. Looking at these specific L,D-transpeptidation products will be useful for better understanding the role that individual L,D-transpeptidases play in PG remodelling and how they contribute to cellular fitness. Finally, we demonstrated that our PG analysis strategy can uncover unexpected muropeptides like those containing unusual amino acids in the fifth position of their peptide stem (gm-AEJAX in Tables 3, 16.6% of total muropeptides). The biological significance of these unusual stems and the enzymes responsible remain unknown, but warrant investigation, as unusual residues in the fifth position are likely to impact PBP-mediated PG polymerisation.

Overall, the granularity of the PG analysis described by this work makes it possible to monitor minor changes in PG structure and composition like never before and transforms how we study the bacterial cell wall and the role it plays in helping species like *R. leguminosarum* thrive in a highly dynamic environment.

Methods

Bacterial strains and growth conditions

R. leguminosarum bv. *viciae* strain 3841^{31} was grown at 28 °C in TY (5 g/L Tryptone + 3 g/L Yeast Extract + 1.3 g/L CaCl₂.6H₂O) broth or agar (15 g/L). The recipe for minimal medium is described in ref. 31. Liquid cultures were grown in 2 L flasks under agitation (180 rpm).

PG extraction and muropeptide preparation

Cells corresponding to 500 mL of culture spun and resuspended in 20 mL of boiling Milli-Q water prior to the addition of SDS 5% (w/v) final. After 30 min at 100 °C, PG was recovered by centrifugation (2 h at 125,000 x g, room temperature) and washed three times in Milli-Q water. Samples were treated with trypsin (100 μ g/mL) for 4 h at 37 °C in 50 mM Tris-HCl (pH 7.5). Trypsin was heat-inactivated (10 min at 65 °C) and removed by washes in Milli-Q water. The material was freeze-dried and resuspended at a concentration of 10 mg/mL.

LC-MS/MS analysis

2 mg of purified PG was digested for 16 h using 250 units of mutanolysin (Sigma-Aldrich) in 20 mM phosphate buffer (pH 5.5) in a final volume of 200 μ L. Following heat inactivation (5 min at 100 °C), the soluble

Communications Chemistry | (2025)8:91

disaccharide peptides were mixed with an equal volume of 250 mM borate buffer (pH 9.25) and reduced via the addition of 25 µL of a sodium borohydride solution at 25 mg/mL. After 20 min at room temperature, the pH was adjusted to 5.0 using phosphoric acid. An Ultimate 3000 Ultra High-Performance Chromatography (UHPLC; Dionex/Thermo Fisher Scientific) system coupled with a high-resolution Q Exactive Focus mass spectrometer (Thermo Fisher Scientific) was used for LC-MS/MS analysis. Muropeptides were separated using a C18 analytical column (Hypersil Gold aQ, 1.9 µm particles, 150 × 2.1 mm; Thermo Fisher Scientific), at a temperature of 50 °C. Muropeptide elution was performed by applying a mixture of solvent A (water, 0.1% (v/v) formic acid) and solvent B (acetonitrile, 0.1% (v/v) formic acid). Following a 10 µL sample injection, MS/MS spectra were recorded over a 40 min gradient: 0-12.5% B for 25 min; 12.5-20% B for 5 min; held at 20% B for 5 min, followed by column re-equilibration for 10 min under the initial conditions. The Q Exactive Focus was operated under electrospray ionisation (H-ESI II) in positive mode. Full scan (m/z 150-2250) used resolution 70,000 (FWHM) at m/z 200, with an automatic gain control (AGC) target of 1×10^6 ions and an automated maximum ion injection time (IT). MS/MS spectra were recorded in "Top 3" data-dependent mode using the following parameters: resolution 17,500; AGC 1×10^5 ions, maximum IT 50 ms, NCE 25%, and a dynamic exclusion time of 5 seconds.

Determination of glycan chain length and cross-linking index

Cross-linking index and glycan chain length were calculated based on the formulae described previously³². The cross-linking was calculated as:

$$\frac{1}{2}$$
 (% of all dimers) + $\frac{2}{3}$ (% of all trimers)

No glycosidically-linked multimers were identified, so all dimers and trimers included in this calculation were peptide cross-linked.

Glycan chain length was inferred from the abundance of anhydroMurNAc groups, which are found at the ends of glycan chains:

 $(\% \text{ of anhydro monomers}) + \frac{1}{2}(\% \text{ of anhydro dimers}) + \frac{1}{3}(\% \text{ of anhydro trimers})$

Because no di-anhydro muropeptides were included in the search process, they have also been excluded from the formula above.

Byos[®] searches

Unbiased searches were performed using Byonic v5.2.5. For monomer searches, a FASTA file containing each peptide stem was used, and glycan moieties (gm, 480.1955 Da) were added as N-terminal modifications. For PG-anchored proteins, searches were performed against the entire *R*.

leguminosarum proteome. Modified peptides with a mass of 852.3600 Da (gm-AEJ) permitted once per peptide on any residue within the peptide were searched using non-specific cleavage parameters. Precursor mass tolerance was set at 8 ppm and fragment mass tolerance was set to 20 ppm for HCD fragmentation. Spectra corresponding to peptides containing an N-terminal disaccharide-tripeptide were examined manually.

Reporting summary

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

Data availability

LC-MS/MS datasets have been deposited in the GLYCOPOST repository (GPST000405). All databases and search outputs are available in Files S1–S6.

Code availability

Code for the latest version of PGFinder can be found here: https://github. com/Mesnage-Org/pgfinder. The exact version described in this manuscript is archived here: https://github.com/Mesnage-Org/pgfinder/releases/ tag/v1.3.2-ncc; https://doi.org/10.5281/zenodo.14946462.

Received: 8 October 2024; Accepted: 10 March 2025; Published online: 26 March 2025

References

- Vollmer, W., Blanot, D. & de Pedro, M. A. Peptidoglycan structure and architecture. *FEMS Microbiol. Rev.* 32, 149–167 (2008).
- Poole, K. Resistance to beta-lactam antibiotics. Cell Mol. Life Sci. 61, 2200–2223 (2004).
- Zapun, A., Vernet, T. & Pinho, M. G. The different shapes of cocci. FEMS Microbiol. Rev. 32, 345–360 (2008).
- Chu, H. & Mazmanian, S. K. Innate immune recognition of the microbiota promotes host-microbial symbiosis. *Nat. Immunol.* 14, 668–675 (2013).
- Gonzalez-Santana, A. & Diaz Heijtz, R. Bacterial peptidoglycans from microbiota in neurodevelopment and behavior. *Trends Mol. Med.* 26, 729–743 (2020).
- Huang, Z. et al. Antibody neutralization of microbiota-derived circulating peptidoglycan dampens inflammation and ameliorates autoimmunity. *Nat. Microbiol.* 4, 766–773 (2019).
- Nyholm, S. V. & McFall-Ngai, M. J. A lasting symbiosis: how the Hawaiian bobtail squid finds and keeps its bioluminescent bacterial partner. *Nat. Rev. Microbiol.* **19**, 666–679 (2021).
- Alvarez, L. et al. Control of bacterial cell wall autolysins by peptidoglycan crosslinking mode. *Nat. Commun.* 15, 7937 (2024).
- Rimal, B. et al. Peptidoglycan compositional analysis of Mycobacterium smegmatis using high-resolution LC-MS. Sci. Rep. 12, 11061 (2022).
- van der Aart, L. T. et al. High-resolution analysis of the peptidoglycan composition in *Streptomyces coelicolor*. J. Bacteriol. 200, e00290-18 (2018).
- Wheeler, R., Chevalier, G., Eberl, G. & Gomperts Boneca, I. The biology of bacterial peptidoglycans and their impact on host immunity and physiology. *Cell Microbiol.* 16, 1014–1023 (2014).
- Anderson, E. M., Greenwood, N. A., Brewer, D. & Khursigara, C. M. Semi-quantitative analysis of peptidoglycan by liquid chromatography mass spectrometry and bioinformatics. J. Vis. Exp. 13 (2020).
- Hsu, Y.-C., Su, P.-R., Huang, L.-J., Cheng, K.-Y., Chen, C. & Hsu, C.-C. High-throughput automated muropeptide analysis (HAMA) reveals peptidoglycan composition of gut microbial cell walls. *eLife* 12, RP88491 (2024).
- Kwan, J. M. C. et al. In silico MS/MS prediction for peptidoglycan profiling uncovers novel anti-inflammatory peptidoglycan fragments of the gut microbiota. *Chem. Sci.* 15, 1846–1859 (2024).

- Patel, A. V. et al. PGFinder, a novel analysis pipeline for the consistent, reproducible and high-resolution structural analysis of bacterial peptidoglycans. *Elife* 0, e70597 (2021).
- Rady, B. J. & Mesnage, S. PGFinder, an open-source software for peptidoglycomics: the structural analysis of bacterial peptidoglycan by LC-MS. *Methods Mol. Biol.* 2836, 111–132 (2024).
- Jupyter, P. et al. Binder 2.0 Reproducible, interactive, sharable environments for science at scale. In: *Proceedings of the 17th Python in science conference*, 113–120 (2018).
- Godessart, P. et al. beta-Barrels covalently link peptidoglycan and the outer membrane in the alpha-proteobacterium *Brucella abortus*. *Nat. Microbiol.* 6, 27–33 (2021).
- Sandoz, K. M. et al. beta-Barrel proteins tether the outer membrane in many Gram-negative bacteria. Nat. Microbiol. 6, 19–26 (2021).
- Coullon, H. et al. Peptidoglycan analysis reveals that synergistic deacetylase activity in vegetative *Clostridium difficile* impacts the host response. *J. Biol. Chem.* 295, 16785–16796 (2020).
- Anderson, E. M. et al. Peptidoglycomics reveals compositional changes in peptidoglycan between biofilm- and planktonic-derived *Pseudomonas aeruginosa. J. Biol. Chem.* **295**, 504–516 (2020).
- Peltier, J. et al. *Clostridium difficile* has an original peptidoglycan structure with a high level of *N*-acetylglucosamine deacetylation and mainly 3-3 cross-links. *J. Biol. Chem.* **286**, 29053–29062 (2011).
- Popham, D. L., Meador-Parton, J., Costello, C. E. & Setlow, P. Spore peptidoglycan structure in a *cwID dacB* double mutant of *Bacillus subtilis*. *J. Bacteriol.* **181**, 6205–6209 (1999).
- Desmarais, S. M. et al. High-throughput, highly sensitive analyses of bacterial morphogenesis using ultra performance liquid chromatography. J. Biol. Chem. 290, 31090–31100 (2015).
- de Jonge, B. L., Chang, Y. S., Xu, N. & Gage, D. Effect of exogenous glycine on peptidoglycan composition and resistance in a methicillinresistant *Staphylococcus aureus* strain. *Antimicrob. Agents Chemother.* 40, 1498–1503 (1996).
- Kuhner, D., Stahl, M., Demircioglu, D. D. & Bertsche, U. From cells to muropeptide structures in 24 h: peptidoglycan mapping by UPLC-MS. Sci. Rep. 4, 7494 (2014).
- Hernandez, S. B. et al. Peptidoglycan editing in non-proliferating intracellular *Salmonella* as source of interference with immune signaling. *PLoS Pathog.* 18, e1010241 (2022).
- Bui, N. K. et al. The peptidoglycan sacculus of *Myxococcus xanthus* has unusual structural features and is degraded during glycerolinduced myxospore development. *J. Bacteriol.* **191**, 494–505 (2009).
- Wohlwend, J. et al. Boltz-1 democratizing biomolecular interaction modeling. *bioRxiv* https://www.biorxiv.org/content/10.1101/2024. 11.19.624167v1 (2024).
- Moré, N. et al. Peptidoglycan remodeling enables *Escherichia coli* to survive severe outer membrane assembly defect. *mBio* 10, e02729–18 (2019).
- Wheatley, R. M. et al. Role of O₂ in the growth of *Rhizobium* leguminosarum bv. viciae 3841 on glucose and succinate. *J. Bacteriol.* 199, e00572–16 (2017).
- Glauner, B. Separation and quantification of muropeptides with highperformance liquid chromatography. *Anal. Biochem.* 172, 451–464 (1988).

Acknowledgements

This work was funded by a BBSRC grant (BB/W013800/1 to S.M. and P.S.P.). M.G.A.Z. is funded by a Mexican government PhD scholarship (CONAHCYT, 2021-000007-01EXTF-00221). B.J.R. is the recipient of a NERC PhD iCASE studentship (NERC ACCE NE/S00713X/1). AR and IGB were supported by grants ANR Peptimet (ANR-18-CE15-0018) and IntraBacWall (ANR-16-IFEC-0004).

Author contributions

Conceptualisation: M.G.A.Z., B.J.R., M.B., S.M.; methodology: M.G.A.Z., B.J.R., N.S., C.A.E., R.D.T., S.M.; software: B.J.R., A.V.P., N.S, R.D.T.;

Communications Chemistry | (2025)8:91

validation: M.G.A.Z., S.M.; formal analysis: M.G.A.Z., B.J.R., S.M.; investigation: M.G.A.Z., B.J.R., R.L., C.A.E., A.R.; resources: R.L., P.S.P.; data curation: M.G.A.Z., C.A.E., B.J.R., N.S., R.D.T., S.M.; writing—original draft: M.G.A.Z., B.J.R., S.M.; writing—review & editing: M.G.A.Z., R.L., R.D.T., B.J.R., M.B., S.M.; visualisation: M.G.A.Z., B.J.R., S.M.; supervision: N.S., R.D.T., M.J.D., I.G.B., P.S.P., M.B., S.M.; project administration: R.D.T., M.J.D., I.G.B., P.S.P., M.B., S.M.; funding acquisition: P.S.P., S.M.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42004-025-01490-6.

Correspondence and requests for materials should be addressed to Stéphane Mesnage.

Peer review information *Communications Chemistry* thanks Frederique Lisacek and the other, anonymous, reviewers for their contribution to the peer review of this work. Peer review reports are available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2025

Communications Chemistry | (2025)8:91