High avidity drives the interaction between the streptococcal C1 phage endolysin, PlyC, with the cell surface carbohydrates of Group A Streptococcus

Broendum, S.S.; Williams, D.E.; Hayes, B.K.; Kraus, F.; Fodor, J.; Clifton, B.E.; ... ; McGowan, S.

Received: 12 December 2020 | Revised: 18 March 2021 | Accepted: 18 March 2021

DOI: 10.1111/mmi.14719

RESEARCH ARTICLE

WILEY

High avidity drives the interaction between the streptococcal C1 phage endolysin, PlyC, with the cell surface carbohydrates of Group A *Streptococcus*

Sebastian S. Broendum[1,2] | Daniel E. Williams[1] | Brooke K. Hayes[1] | Felix Kraus[2] | James Fodor[2,3] | Ben E. Clifton[4] | Anne Geert Volbeda[5] | Jeroen D. C. Codee[5] | Blake T. Riley[2,6] | Nyssa Drinkwater[1] | Kylie A. Farrow[1] | Kirill Tsyganov[2,7] | Ryan D. Heselpoth[8] | Daniel C. Nelson[8] | Colin J. Jackson[4] | Ashley M. Buckle[2] | Sheena McGowan[1]

[1]Biomedicine Discovery Institute, Department of Microbiology, Monash University, Melbourne, VIC, Australia

[2]Biomedicine Discovery Institute, Department of Biochemistry and Molecular Biology, Monash University, Melbourne, VIC, Australia

[3]Eccles Institute of Neuroscience, John Curtin School of Medical Research, The Australian National University, Canberra, ACT, Australia

[4]Research School of Chemistry, Australian National University, Acton, ACT, Australia

[5]Department of Bio-organic Synthesis, Leiden Institute of Chemistry, Leiden University, Leiden, The Netherlands

[6]Structural Biology Initiative, CUNY Advanced Science Research Center, New York, NY, USA

[7]Monash Bioinformatics Platform, Monash University, Melbourne, VIC, Australia

[8]Institute for Bioscience and Biotechnology Research, University of Maryland, College Park, Rockville, MD, USA

Correspondence
Sheena McGowan, Biomedicine Discovery Institute, Department of Microbiology, Monash University, Melbourne, VIC 3800, Australia.
Email: sheena.mcgowan@monash.edu

Present address
Felix Kraus, Department of Cell Biology, Blavatnik Institute, Harvard Medical School, Boston, MA, USA

Funding information
SSB and FK were supported by a Monash University Graduate Research Fellowship. DEW and BKH are supported by an RTP scholarship from the Australian Department of Education

Abstract

Endolysin enzymes from bacteriophage cause bacterial lysis by degrading the peptidoglycan cell wall. The streptococcal C1 phage endolysin PlyC, is the most potent endolysin described to date and can rapidly lyse group A, C, and E streptococci. PlyC is known to bind the Group A streptococcal cell wall, but the specific molecular target or the binding site within PlyC remain uncharacterized. Here we report for the first time, that the polyrhamnose backbone of the Group A streptococcal cell wall is the binding target of PlyC. We have also characterized the putative rhamnose binding groove of PlyC and found four key residues that were critical to either the folding or the cell wall binding action of PlyC. Based on our results, we suggest that the interaction between PlyC and the cell wall may not be a high-affinity interaction as previously proposed, but rather a high avidity one, allowing for PlyC's remarkable lytic activity. Resistance to our current antibiotics is reaching crisis levels and there is an urgent need to develop the antibacterial agents with new modes of action. A detailed understanding of this potent endolysin may facilitate future developments of PlyC as a tool against the rise of antibiotic resistance.

Abbreviations: bp, base pairs; CBD, cell wall binding domain; EAD, enzymatically active domain; GA, Gibson assembly; GAC, *Streptococcus* Lancefield group A carbohydrate; GAS, Group A *Streptococcus*; GlcNAc, *N*-acetylglucosamine; MurNAc, *N*-acetylmuramic acid; PG, peptidoglycan; wt, *wild type*.

Sebastian S. Broendum and Daniel E. Williams are the authors that have contributed equally.

KEYWORDS
antimicrobial resistance, avidity, bacteriophage, cell wall binding, endolysin, PlyC

1 | INTRODUCTION

Bacteriophage-encoded endolysin enzymes lyse their host bacterial cell at the end of the phage lytic reproductive cycle by degrading the peptidoglycan (PG) polymer in the bacterial cell wall (Young, 1992). This degradation of the cell wall polymer results in osmotic shock, cell rupture, and death of the bacterial host, (Fischetti, 2005). Exogenous application of purified endolysins to susceptible bacteria often results in rapid and specific elimination (Cheng et al., 2005; Loeffler et al., 2001; Nelson et al., 2001; Schmelcher et al., 2012). This discovery has formed the basis for the exploration of endolysins as novel antibacterial agents or "*enzybiotics*" (Nelson et al., 2001) for human and veterinary therapeutics (Briers et al., 2014; Czaplewski et al., 2016; Jun et al., 2014; Schuch et al., 2014), food quality (see review (Schmelcher & Loessner, 2016)), and primary industry (Hoopes et al., 2009; Schmelcher et al., 2015). The majority of endolysins-targeting Gram-positive bacteria are encoded on a single polypeptide chain and often consist of an enzymatically active domain (EAD) that confers the catalytic activity of the endolysin, and a cell wall binding domain (CBD) which enables the endolysin to selectively and specifically bind to the cell wall of target bacteria (Fenton et al., 2010; Schmelcher et al., 2012). Previously, these two components were thought to primarily consist of a single domain each, however, recent studies suggest that the situation may be more complex, as endolysins with two EADs are not uncommon and repetition within the CBD might be important for an endolysin's engagement of the bacterial cell wall (Broendum et al., 2018).

The streptococcal C1 phage endolysin, PlyC, is the most potent endolysin described to date, with a specific activity approximately 100-fold that of the next most active endolysin (Nelson et al., 2001). PlyC can rapidly lyse cultures of groups A, C, and E streptococci (Schmelcher et al., 2012). In mice that have been mucosally colonized by Group A *Streptococcus* (GAS), PlyC requires only nanogram quantities to reduce 10 million GAS cells by >6 logs only seconds after enzyme addition (Nelson et al., 2001). Furthermore, PlyC has been shown to be highly effective against GAS biofilms (Shen et al., 2013) and is capable of killing the intracellular GAS (Shen et al., 2016). These features make PlyC an attractive antibacterial agent for further development, with the potential for engineering of PlyC in order to generate superior clinical properties. To aid in these aims, an understanding of the molecular details that dictate the function and specificity of PlyC would be greatly advantageous.

PlyC is a unique endolysin in that it is an assembly of two distinct proteins, a single copy of PlyCA which contains the EADs, and eight copies of PlyCB that form the cell wall binding octamer (McGowan et al., 2012; Nelson et al., 2006). PlyCA is a 50 kDa protein consisting of two distinct catalytic domains that are connected via linkers to a central docking domain, which in turn interact with the homo-octameric PlyCB ring (Figure 1). The two catalytic domains within PlyCA work synergistically to cleave two different bonds in the PG substrate (Broendum et al., 2018; McGowan et al., 2012). The N-terminal glycosyl hydrolase 73 (GyH) domain within PlyCA has glycosidase activity and likely cleaves the glycan bond located on the reducing side of *N*-acetylglucosamine (GlcNAc) (Supplementary Figure S1). The C-terminal Cysteine, Histidine-dependent Amidohydrolase/Peptidase (CHAP) domain cleaves between *N*-acetylmuramic acid (MurNAc) and the L-alanine of the stem peptide within PG via its *N*-acetylmuramoyl-L-alanine amidase activity (Broendum et al., 2018; McGowan et al., 2012).

The PlyCB ring is made up of an octamer of 8 kDa subunits and is responsible for cell wall binding to target bacteria. The remarkable symmetry of the PlyCB octamer is broken by four copies of the PlyCB monomer that each contribute their first eight N-terminal residues to form a central β-sheet that extends up from the center of the PlyCB ring assembly, which in turn interacts with the docking domain of PlyCA. PlyCB is essential for lytic activity—cell wall binding is a prerequisite for cell lysis (Nelson et al., 2006). Based on site-directed mutagenesis, a groove between the C-terminal helix and the central β-sheet in PlyCB has been suggested as a putative cell wall binding site (McGowan et al., 2012). PlyC's bacterial specificity is likely a result of the selective and specific binding of PlyCB to the target bacterial cell wall, as deletion of PlyCB completely abolishes both the cell wall binding and lytic activity of PlyC (Nelson et al., 2006).

Addition of the *Streptococcus* Lancefield group A carbohydrate (GAC) partially inhibits the lytic activity of PlyC, indicating that this carbohydrate, or a part of it, might function as the target ligand for PlyC (Lood et al., 2014). The specificity of PlyC toward sensitive streptococci has previously been theorized due to a high-affinity binding domain that binds to the species-specific carbohydrate (Hoopes et al., 2009). The GAC is comprised of a polyrhamnose backbone with *N*-acetylglucosamine side chains (Coligan et al., 1978; McCarty, 1956) and it has been suggested that polyrhamnose is necessary for PlyC binding to the GAS cell wall (Supplementary Figure S1) (Fischetti, 2006; Lood et al., 2014). However, this interaction remains poorly understood as neither the target ligand nor the binding site within PlyC has been unambiguously identified.

In previous studies, a putative binding groove has been suggested where PlyCB may interact with its substrate. Based on analysis of the crystal structure of the PlyC holoenzyme, McGowan et al. identified a charged groove in each PlyCB monomer that was likely involved in cell wall binding (McGowan et al., 2012). The groove was formed by residues Y26, E36, R29, K59, and R66 (Figure 1c,d). Further studies by Shen et al identified the same shallow cleft as a prospective binding site for anionic lipids, such as the GAC—listing R29, K59, and R66 as key residues located in the groove (Shen et al., 2016). Mutation of R66 in both studies reduced the cell wall and/or anionic lipid binding while attempts to mutate R29 resulted in insoluble protein. These previous

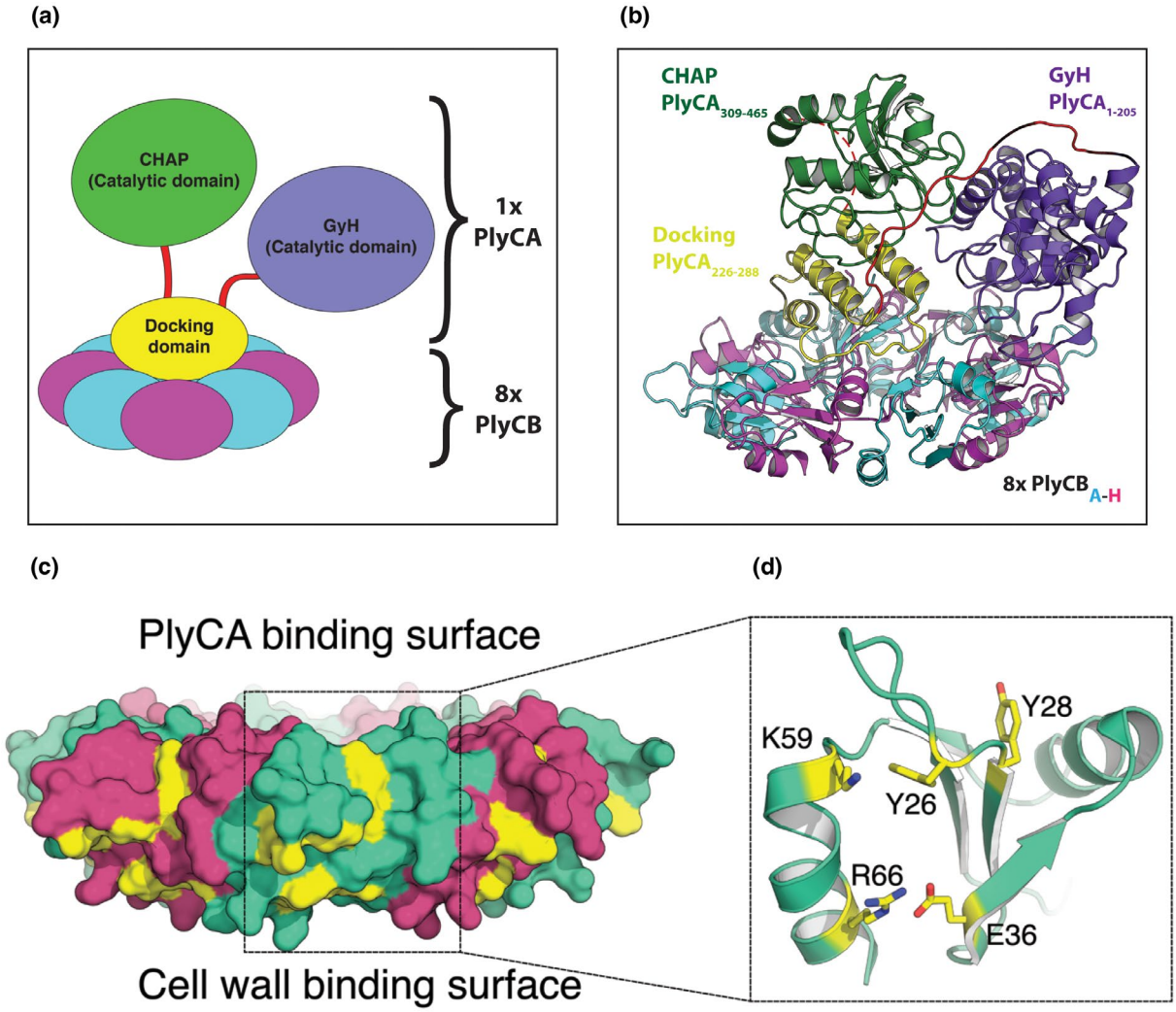

FIGURE 1 The structure of PlyC. (a). A schematic representation of PlyC's domain architecture. (b). X-ray crystal structure of PlyC (PDB ID: 4F88). The N-terminal GyH domain (residues 1–205) is shown in purple, the disordered linker 1 (residues 206–225) in red, the helical docking domain (residues 226–288) that docks PlyCA to PlyCB in yellow, the second disordered linker 2 (residues 289–308) is indicated as a dashed red line, and the CHAP domain (residues 309–465) is in green. PlyCB monomers (A–H) are colored in magenta and cyan alternatively. Coloring for A is the same as for B. (c). The crystal structure of PlyCB (PDB ID: 4F87) is shown as a surface representation with PlyCB subunits colored in alternating magenta and green. Residues targeted for site saturation mutagenesis are shown in yellow. (d). A single PlyCB subunit is shown in a cartoon representation. Residues targeted for site saturation mutagenesis are shown as sticks with yellow carbon atoms

studies formed the basis for our current investigation to probe the mutational tolerance of this groove and confirm its role in streptococcal cell wall binding. We selected Y26, E36, K59, and R66 for a site saturation mutagenesis study. R29 was not included in our study due to previous evidence suggesting it had a role in protein folding. We additionally also included Y28 as this residue is located at the end of the groove or cleft, allowing us to see if the binding interface extended beyond the originally identified groove (Figure 1c,d).

Here, we present the results of our investigation into the interaction between PlyC and polyrhamnose at the molecular level. We show that a synthetic tetrarhamnose molecule can compete with the natural ligand of PlyC in the GAS cell wall, strongly suggesting that PlyC binds to the polyrhamnose backbone in the GAC. Using site saturation mutagenesis, we identified four residues that are essential for the lytic activity, folding, and cell wall binding ability of PlyC. Based on this data, we propose a model for how PlyC binds to polyrhamnose in the bacterial cell wall, suggesting that the interaction is not a high-affinity interaction as previously proposed, but rather a high avidity one, allowing the enzyme the necessary mobility to navigate and digest the complex PG substrate.

2 | MATERIALS AND METHODS

2.1 | Reagents

Oligonucleotides were purchased from Bioneer Pacific. Tetrarhamnoside was synthesized as described previously (Geert

Volbeda et al., 2016). Chemicals were purchased from standard commercial suppliers.

2.2 | Protein expression and purification

PlyC variants were expressed and purified as previously described (McGowan et al., 2012; Nelson et al., 2006). Briefly, protein expression was induced with 0.4% arabinose for 4 hr. PlyC and variants were purified using ceramic hydroxyapatite (CHT) affinity chromatography using the CHT resin (Bio-Rad), followed by size-exclusion chromatography. The final purity of the protein was assessed using SDS-PAGE (Supplementary Figure S2).

2.3 | Growth of GAS B119

GAS B119 was typed as GAS using a streptococcal grouping kit (Thermo Fisher Scientific). GAS B119 was grown in Todd Hewitt broth supplemented with 1% (wt/vol) of yeast extract and incubated overnight in a capped bottle at 37°C with no shaking. The resulting culture was pelleted by centrifugation for 10 min at 1,000 × *g* and resuspended in 1× PBS pH 7.4 to the required optical density (OD).

2.4 | Lytic activity assays

Quantitative bacteriolytic assays were performed in triplicate in 96-well plates and carried out in 1 × PBS pH 7.4 at 37°C in a final volume of 100 μL. Activity was monitored for up to 1 hr. Standard lytic assays used 90 μL of bacteria at an OD_{595} of 1.3 mixed with 10 μL of the protein at various concentrations. Bacteriolysis was monitored by measuring the absorbance at 595 nm using a Fluostar Optima plate reader (BMG Labtech). The resulting absorbance curves were converted to percent cell lysis using activity from 10 μM *wild-type* PlyC to define 100% cell lysis, and samples containing no protein to define 0% cell lysis. EC_{50} values were calculated using GraphPad Prism v7.0 by fitting the data using the agonist concentration versus response with a variable slope. The tetrarhamnan competition assay used 2.5 nM *wild-type* PlyC, GAS B119 at an OD_{595} of 2 and varying concentrations of tetrarhamnan (0–500 μM). Prior to the addition of bacterial cells, *wild-type* PlyC and tetrarhamnan were incubated for 15 min at 37°C. Bacteriolysis was monitored over a 20-min period. GraphPad Prism was used to calculate the IC_{50} by fitting a log(inhibitor) versus response curve with a variable slope and four parameters to the relevant dataset.

2.5 | Generation of site saturation libraries

Site saturation libraries were generated using a degenerate codon approach combined with Gibson assembly (GA). For each library, three polymerase chain reaction (PCR) fragments were produced: (1). A pBAD24 vector fragment (*Vec*) extending from 50 base pairs (bp) downstream of the 3′ end of the PlyC operon to 50 bp upstream of the 5′ end of the PlyC operon (primers: vector forward (fw) and vector reverse (rv), Table 1). (2). Insert 1 (*Ins_1*) extended from 50 bp upstream of the 5′ prime end of the PlyC operon to the site targeted for mutation (primers: insert fw and XXX NNS rv, where XXX indicates interchangeably the variants of PlyC listed in Table 1). (3). Insert 2 (*Ins_2*) extended from the site targeted for mutation to 50 bp downstream of the 3′ end of the PlyC operon (XXX NNS fw and primers insert rv, Table 1). The fragments were amplified using Phusion Hotstart II polymerase (Thermo Fisher Scientific). The vector fragment was amplified using an annealing temperature of 65°C and an extension time of 3 min. All insert fragments were amplified using an annealing temperature of 66°C. Insert 1 fragments were amplified using an extension time of 15 s, whereas 30 s was used for all insert 2 fragments. The amplified fragments overlapped by ~30 bp to facilitate subsequent ligation by GA. Following amplification, fragments were run on a 1% agarose gel followed by in-gel purification using the Wizard SV Gel and PCR clean-up system (Promega). The three purified fragments (*Vec*, *Ins_1*, and *Ins_2*) from each library were ligated as described (Gibson et al., 2009), using a 1:5 molar ratio of vector to insert fragments, and an incubation time of 60 min. The ligated plasmids were purified using the FavorPrep plasmid DNA extraction kit (Favorgen) and 1 μL of the purified plasmid was electroporated into 50 μL of competent Top10 cells. Transformant colonies were resuspended in 6 ml of 2YT media, and spun down for 15 min at 3,000 ×*g* at 4°C. The resulting pellet was used for plasmid purification using the FavorPrep plasmid DNA extraction kit (Favorgen), resulting in the final plasmid library.

The second generation of site saturation libraries was generated by inoculating 300 μl of 2YT containing 100 μg/ml of ampicillin in a 2 ml 96-well plate with variants that had shown either *wild type*-like or compromised activity in the primary screen. The plate was incubated overnight at 37°C with 250 RPM shaking to allow bacterial growth. Fifty microliters from all inoculated wells were pooled and the cells were pelleted. The resulting pellet was used for plasmid purification using the FavorPrep plasmid DNA extraction kit (Favorgen), resulting in the second-generation plasmid library.

2.6 | Next-generation sequencing

Five nanograms of each plasmid library (first and second generation for each of the five libraries) was used to construct Illumina sequencing libraries using the Illumina Nextera XT Indexing chemistry (Illumina). A customized amplification procedure, derived from the technical note specified in manufacturer's instructions, was used and included 15 cycles for the primary (amplicon) PCR and 10 cycles for the secondary (index) PCR. The PCR primers used to produce the primary amplicon from genomic DNA can be found in Table 1. Libraries were quantitated using the Qubit DNA HS kit (Invitrogen). Libraries were sized and checked for adapter contamination using the Agilent Bioanalyzer 2100

TABLE 1 Plasmids and oligonucleotides used in this study

Plasmid	Name	Description	Origin
pSM83	PlyCΔΔ	PlyCA residues 1–225 + 289–465 deleted from *plyC* operon	This study
pSM85	PlyCB	*plyCB* gene alone	Nelson et al. (2006)
pSM86	*Wild type*	*plyC* operon (PlyC *wild-type* expression vector)	Nelson et al. (2006)
pSM93	R66E	*plyC* operon containing *plyCB(R66E)*	McGowan et al. (2012)
pSM185	R66A	*plyC* operon containing *plyCB(R66A)*	This study
pSM186	R66K	*plyC* operon containing *plyCB(R66K)*	This study
pSM187	Y28H	*plyC* operon containing *plyCB(Y28H)*	This study
pSM188	Y26A	*plyC* operon containing *plyCB(Y26A)*	This study
pSM189	Y26W	*plyC* operon containing *plyCB(Y26W)*	This study
pSM190	K59H	*plyC* operon containing *plyCB(K59H)*	This study
Primer name	Sequence (5′-3′)		
General sequencing			
fw seq	TTAGCGGATCCTACCTGACG		This study
rv seq	TTTTATCAGACCGCTTCTGC		This study
Site saturation mutagenesis and GA			
ins fw	CAGTGCCAAAGAAACTGCTAAATGTTTTAG		This study
ins rev	GGTTAGTTTGATAATGACACCATTCTAAGTTATG		This study
vector fw	CATAACTTAGAATGGTGTCATTATCAAACTAACC		This study
vector rv	CTAAAACATTTAGCAGTTTCTTTGGCACTG		This study
R66 NNS fw	GATGTAGAGGCTATCNNSAAGGCTATGAAAAAG		This study
R66 NNS rv	CTTTTCATAGCCTTNNSGATAGCCTCTACATC		This study
K59 NNS fw	GTCTATCAATATTAGTNNSTCTGATGTAGAGGC		This study
K59 NNS rv	GCCTCTACATCAGANNSACTAATATTGATAGAC		This study
Y28 NNS fw	GAAAAGAAAGTTACGGTNNSCGTGCTTTTATTAACG		This study
Y28 NNS rv	CGTTAATAAAAGCACGNNSACCGTAACTTTCTTTTC		This study
E36 NNS fw	CTTTTATTAACGGAGTTNNSATTGGTATTAAAGACATTG		This study
E36 NNS rv	CAATGTCTTTAATACCAATNNSAACTCCGTTAATAAAAG		This study
Y26 NNS fw	CATACCGATGGAAAAGAAAGTNNSGGTTATCGTGCTTTTATTAAC		This study
Y26 NNS rv	GTTAATAAAAGCACGATAACCNNSACTTTCTTTTCCATCGGTATG		This study
PlyCB site-directed mutagenesis			
PlyCB R66A fw	GTAAGTCTGATGTAGAGGCTATCGCAAAGGCTATGAA		This study
PlyCB R66A rv	TTCATAGCCTTTGCGATAGCCTCTACATCAGACTTAC		This study
PlyCB R66K fw	CTGATGTAGAGGCTATCAAAAAGGCTATGAA		This study
PlyCB R66K rv	TTCATAGCCTTTTTGATAGCCTCTACATCAG		This study
PlyCB Y26A fw	GATGGAAAAGAAAGTGCCGGTTATCGTGCTTT		This study
PlyCB Y26A rv	AAAGCACGATAACCGGCACTTTCTTTTCCATC		This study
PlyCB Y26W fw	CGATGGAAAAGAAAGTTGGGGTTATCGTGCTTT		This study
PlyCB Y26W rv	AAAGCACGATAACCCCAACTTTCTTTTCCATCG		This study
PlyCB Y28H fw	AAGAAAGTTACGGTCATCGTGCTTTTATTAAC		This study
PlyCB Y28H rv	GTTAATAAAAGCACGATGACCGTAACTTTCTT		This study

Note: All plasmids except pSM85 have a pBAD24 vector backbone and an ampicillin resistance marker. pSM85 is based on pBAD33 with chloramphenicol resistance.

microfluidics device, in conjunction with Agilent DNA HS kits (Agilent Technologies). After quantitation and sizing, libraries were diluted and pooled at equimolar concentrations. They were further diluted to 10 pM and denatured with 100 mM of sodium hydroxide. After denaturation, libraries were diluted to 8 pM followed by sequencing via an Illumina MiSeq Next-Generation Sequencer and sequencing by synthesis V3 chemistry according to the manufacturer's instructions (Illumina). The expected insert of 240 bp was sequenced from both ends of the fragment with 200 bp reads each, providing substantial overlap between paired-end reads. In order to get a full-length fragment, the paired-end reads were merged into single-end reads, using the PEAR tool (Zhang et al., 2014). The merged reads were mapped to the streptococcal C1 phage genome (GenBank Accession ID: AY212251) (Nelson et al., 2003) with a BWA-MEM (Li, 2013) aligner to obtain bam files. Bam files were coordinate sorted, indexed, and mapping rates checked with SAMtools (Li et al., 2009). An in-house Python script was used to count codons and an R script was used to aggregate and normalize for library size, providing the frequency of each amino acid in the respective libraries. These values were tabulated and the frequency in the first and second generation of the libraries was compared (Supplementary Table S1).

2.7 | Medium-throughput lytic activity screening assay

Fifty nanograms of each first-generation plasmid library (five total) was used to transform BL21 (DE3) *Escherichia coli* cells. Individual transformants were used to inoculate 300 μl of autoclaved 2YT media containing 100 μg/ml of ampicillin and 0.4% (wt/vol) of arabinose in a 2 ml V-bottom 96-well plate. As a positive control, wells A1, B3, C5, F7, G9, and H11 in each plate were inoculated with BL21 DE3 cells transformed with the plasmid pSM86 that encodes *wild-type* PlyC (Table 1). As a negative control, wells H1, G3, and F5 were inoculated with BL21 DE3 cells transformed with the plasmid pSM83, encoding the inactive mutant PlyCΔΔ, which lacks both catalytically active domains (Table 1). Wells A12, B10, and C8 were not inoculated in order to enable the detection of contamination and provide a no protein control. The deep 96-well plate was covered with a sterile lid and the plate was incubated for 20 hr at 37°C with 250 RPM shaking to allow for bacterial growth. A reference glycerol stock for each well was generated and stored. One milliliter of lysis buffer (25 mM HEPES pH 7.4, 125 mM NaCl, 2 mM EDTA, 1% (vol/vol) Triton X100, and cOmplete protease inhibitor cocktail) was added to the remaining 200 μl of cell culture. The plate was then sealed using a new sterile lid and incubated at −80°C for 3 hr, followed by 3 hr at 37°C with shaking at 250 RPM. This freeze/thaw process was repeated three times in total in order to lyse the bacterial cells. Following lysis, the plate was centrifuged at 3,700 ×*g* for 1 hr at 4°C to pellet insoluble cellular debris. Hundred microliters of the resulting supernatant was transferred into the equivalent well in a new sterile 96-well microcentrifuge plate containing 100 μl of sterile 50% glycerol, creating the soluble cell lysate used for testing the lytic activity of each library.

The lytic activity of the different variants from site saturation mutagenesis were measured similarly to the purified proteins described above. Briefly, 80 μl of bacterial cells were mixed with 20 μl of the previously prepared soluble cell lysate of each mutagenesis library in a sterile 96-well plate. Bacteriolysis was monitored by measuring the absorbance at 595 nm every 2 min using a Fluostar Optima plate reader (BMG Labtech). For each library, 168 individual variants were screened. This was performed using two separate 96-well plates each containing six positive controls (*wild-type* PlyC), six negative controls (3× PlyCΔΔ and 3× buffer only), and 84 samples. The resulting data were analyzed using a custom Python script. First, the raw data (absorbance as 595 nm as a function of time) were visualized and inspected. Subsequently, the average of the positive and negative controls was used to define 100% and 0% cell lysis, respectively. A confidence interval of ±2 standard deviations (*SD*) of the positive and negative controls were then calculated and variants were classified based on their lytic activity compared to these controls. Variants within ±2 *SD* of the negative control were classified as inactive, variants within ±2 *SD* of the positive control were classified as *wild type*-like, and variants in between were classified as compromised.

2.8 | Calculating probability of sampling

The probability of sampling a certain variant was calculated by means of a binomial probability distribution, using the variant frequency observed in the first generation of the site saturation mutagenesis libraries to define the probability of sampling that variant in a single trial. Due to the low abundance of several amino acids in the library, tens of thousands of variants would have to be tested to have a 95% probability of sampling all possible variants in each library (>3,000 variants from each library would have to be tested to have a 95% probability of sampling amino acids with a frequency of 0.1%). This is not feasible using a non-automated, 96-well plate-based screen. Instead, we opted to screen 168 variants from each library as this was compatible with the 96-well plate system (2 plates per library) and enabled us to sample amino acids with a frequency ≥1.3% with 90% probability, and amino acids with a frequency ≥1.7% with a 95% probability. These criteria allowed us to sample at least seven different amino acids in each library and amino acid residues with diverse chemical properties (Supplementary Table S1). In 168 samples assayed, only amino acids that had ≥85% chance of being sampled were analyzed further. Non-permissible residues were defined as a ≥60% reduction in frequency between the first- and second-generation libraries, tolerated as a 0–60% reduction in frequency, and permitted as no reduction in frequency.

2.9 | Confocal laser scanning microscopy

Based on previous studies (Nelson et al., 2006), PlyC variants were inactivated by overnight incubation at room temperature with

N-ethylmaleimide (NEM) at a 1:1,000 molar ratio. The inactivated protein was labeled with Alexa Fluor 488 NHS Ester (Thermo Fisher Scientific) at a 1:5 molar ratio for 1 hr at 4°C in the dark. Prior to the addition of the dye, the protein solution was adjusted to pH 8.5 using 1 M of sodium bicarbonate. The reaction was quenched for 1 hr at 4°C in the dark by the addition of Tris–HCl pH 8.0 to a final concentration of 50 mM. Excess Alexa Fluor 488 dye was removed using a 5 ml HiTrap desalting column (GE Healthcare). A final concentration of 2 μg of labeled protein was added to 1 ml of GAS B119 cells at an OD_{595} of 1 (in 1× PBS pH 7.4). The protein and cells were incubated for 5 min at room temperature. The mixture was then washed twice by centrifugation for 20 min at 1,000 ×*g*, followed by the supernatant being discarded and cells resuspended in 1 ml of 1× PBS pH 7.4. Two microliters of the resulting cell suspension was streaked onto a coverslip pre-coated with poly-L-lysine and allowed to dry. Ten microliters of Prolong Gold Antifade (Thermo Fisher Scientific) was added to the coverslip, prior to mounting on a glass slide. Coverslips were dried overnight and sealed using clear nail varnish. Confocal microscopy was performed on an inverted Leica TCS SP8 confocal microscope (containing 405, 488, 552, 647 nm diode lasers) equipped with HyD detectors and data were recorded using the Leica LAS X Life software (v3.5.5.19976). Images in all experimental groups were obtained using the same settings.

Images of the samples were obtained using a 63×/1.40 NA oil immersion objective (HC PLAPO, CS2, Leica Microsystems), 1% 488 nm laser power, 21.9% Smart Gain, and 6x line accumulation. Additional brightfield images were recorded in parallel using a conventional PMT, with 285 V gain and 3× line accumulation. Images were taken in a 1024 × 1024-pixel format, with bi-directional scanning at 600 Hz. When required, z-sectioning was performed using 250-nm slices. Leica .lif files were converted to multi-color.tiff composite stacks using custom-written Fiji/ImageJ macros. Images were analyzed using Fiji (Schindelin et al., 2012) (v1.52n) and custom-written macros. Briefly, the brightfield images were used to generate a cell-mask (Yen thresholding method) containing all bacteria in the image independent of protein binding. The Alexa Fluor 488 fluorescence intensity within this mask was then measured for each image, and the fluorescence intensity per square micron of mask was calculated. This was performed for five images per variant and all variants were analyzed in triplicates. All graphical representations and statistical analysis were carried out on GraphPad Prism (v7.0a) using Student's *t* tests.

2.10 | Crystallization, data collection, processing, and structural refinement of PlyCB mutants

Based on previous crystallization data of PlyCB (McGowan et al., 2012), hanging-drop vapor-diffusion crystal trays were set up manually in hand trays for the PlyCB mutants Y28H, K59H, and R66K. Each drop contained 1 μl of 17 mg/ml protein solution in 25 mM of Hepes, pH 6.0, 100 mM of NaCl, 2 mM of EDTA together with 1 μl of mother liquor suspended over 500 μl reservoirs. Each reservoir contained 500 μl of 0.1 M HEPES (pH 6.0), 0.2 M of sodium citrate, and 20%–25% methylpentanediol. Crystals were fully formed within a week at 293K, and were subsequently flash frozen in liquid nitrogen after being transferred onto a nylon loop (Hampton Research, California, USA). Diffraction data were collected at the MX2 beamline at the Australian Synchrotron (Aragão et al., 2018). This data were collected using a Dectris Eiger 16 M detector paired with a 100K nitrogen vapor stream, and the Blu-Ice control software (McPhillips et al., 2002). Diffraction images were integrated using XDS (Kabsch, 2010), with data reduction and scaling being performed using POINTLESS (Evans, 2006) and AIMLESS (Evans & Murshudov, 2013). The structure of each individual mutant was solved using molecular replacement, using a monomeric PlyCB subunit from the previously solved *wild-type* PlyCB structure (PDB ID: 4F87). This was performed using Phaser (McCoy et al., 2007) within the PHENIX software suite (Adams et al., 2010), using two search models. Structures were then built and refined using rounds of phenix.refine and manual model building in Coot (Emsley et al., 2010). Statistics are presented in Supplementary Table S2. Final structures were deposited to the PDB with accession IDs of: 7KWT (PlyCB Y28H), 7KWW (PlyCB K59H), and 7KWY (PlyCB R66K).

2.11 | Computational docking of tetrarhamnose to PlyCB

The structure of the rhamnose monomer was taken from PubChem Identifier 25310 (Kim et al., 2016), with the corresponding tetramer constructed using IQmol (http://iqmol.org). Monomer and tetramer forms of the PlyCB structure were obtained from PDB ID 4F87. Docking was performed using AutoDock Vina version 1.1.2 (Trott et al., 2009) and AutoDock Tools version 1.5.6 (Morris et al., 2009). A 20 $Å^3$ box was centered around the groove in the PlyCB monomer between the β-sheet and α-helix, in accordance with residues known to be involved in ligand binding (Y28, K59 and R66). The exhaustiveness parameter was set to 60, and the maximum of 19 bound conformations were generated.

3 | RESULTS

3.1 | A rhamnose tetramer can compete with PlyC's natural ligand

Previous studies had shown that the GAC was able to partially inhibit the lytic activity of PlyC (Lood et al., 2014), and it was hypothesized that this was due to PlyC binding to the polyrhamnose backbone in the GAC (Fischetti, 2006; Lood et al., 2014). To confirm that it was indeed polyrhamnose mediating binding to the bacteria, we performed a quantitative lytic activity assay in the presence of tetrarhamnan, a synthetic tetramer of polyrhamnose (Geert Volbeda et al., 2016). The results of the competition assay showed that tetrarhamnan can inhibit the lytic activity of PlyC in a dose-dependent

manner, with an IC_{50} of 76 ± 5 μM (Figure 2). Monomeric rhamnose was unable to inhibit PlyC, even at millimolar concentrations, suggesting rhamnose polymerization is essential for binding to PlyC.

3.2 | Site saturation mutagenesis explores the polyrhamnose binding site in PlyC

Previously, we identified a groove in PlyCB that appeared to form the cell wall binding site (Figure 1c,d) (McGowan et al., 2012). Mutagenesis of Y28, K59, and R66 to Ala, Glu, and Glu, respectively, resulted in a reduction in the ability of PlyCB to bind bacteria (McGowan et al., 2012). To map the detailed interaction between PlyCB and polyrhamnose, we used site saturation mutagenesis to probe the mutational tolerance at these specific residues. The residues Y26 and E36 line the same groove as Y28, K59, and R66 (Figure 1c,d), suggesting Y26 and E36 may also be involved in cell wall binding and thus these are included in our study. Therefore, a total of five residues in PlyCB were targeted for single-site saturation mutagenesis: Y26, Y28, E36, K59, and R66 (Figure 1c,d). To introduce mutations at the targeted residues, five separate single-site saturation libraries were generated using NNS codon mutagenesis, one for each target residue. Upon completion of the mutagenesis, the amino acid diversity in each of the five libraries was assessed using Illumina MiSeq Next-Generation sequencing. The sequencing results showed that all libraries contained a background level of *wild-type* PlyC, with the lowest background level observed in the K59 library (28%) and the highest in the E36 library (78%) (Supplementary Figure S3 and Supplementary Table S1). In addition to the persisting *wild-type* constructs, each library contained 7 to 11 different amino acids that existed with a frequency above 1.1% (Supplementary Table S1). Given the sampling size selected, a 1.1% frequency provides an 85% chance of sampling each variant in the primary screen. Residues with a frequency of greater than 1.1% represent a range of different chemical properties, including hydrophobic (Ala, Val, Leu, Pro, and Gly), charged (Arg, Glu, and Asp), polar (Thr, Ser, Gln, and His), and aromatic (Trp) residues (Supplementary Figure S3). The amino acid diversity in each library was, therefore, sufficient to determine the chemical and size requirements at each position.

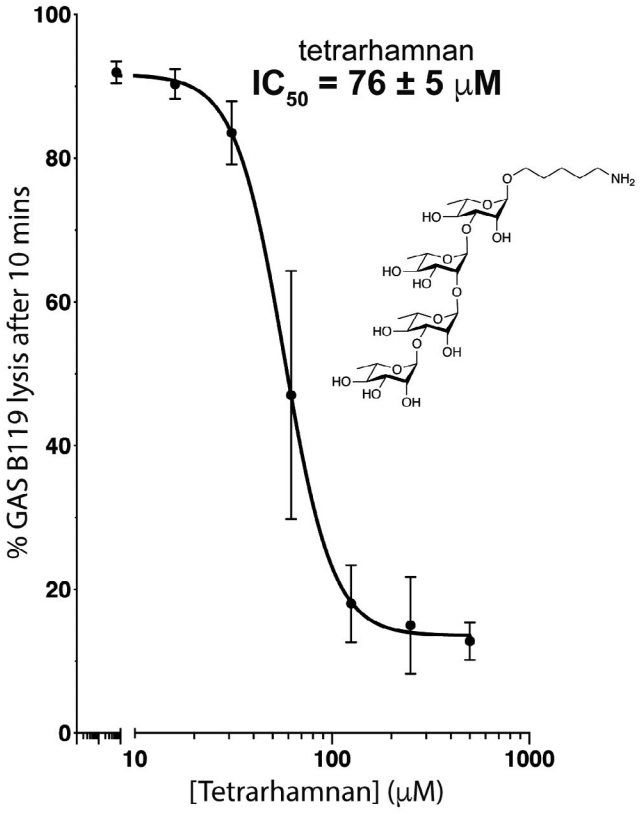

FIGURE 2 PlyC's lytic activity can be inhibited by tetrarhamnan. The lytic activity of 2.5 nM wild-type PlyC was tested across a range of different tetrarhamnan concentrations (0–500 μM). Each concentration was performed in triplicate. Error bars represent standard deviation. The insert shows the structure of the tetrarhamnan used for the competition experiments

3.3 | Development of a medium-throughput mutagenesis library screening assay

The challenge for saturation mutagenesis studies of this nature, is that screening the variants presents a bottleneck with regard to throughput. For an endolysin such as PlyC, the biological activity of the protein under scrutiny can be exploited. We chose to use the bacteriolytic activity of PlyC as a surrogate measure for cell wall binding to GAS bacteria. For PlyC to lyse bacteria from the "outside in," PlyC must bind to the bacterial cell wall (McGowan et al., 2012). Consequently, variants with impaired cell wall binding would also be expected to display reduced lytic activity. Exploiting this dependence, we developed a 96-well plate-based assay that allowed us to measure the bacteriolytic activity of each variant without having to purify individual mutant proteins (Figure 3). For this assay, each site saturation library was separately transformed into *E. coli* and individual transformants transferred for growth in a 96-well plate format (Figure 3a). Bacteria expressing different variants from the library were lysed using a freeze/thaw approach and the resulting cell lysate was tested for lytic activity against GAS B119 by monitoring the absorbance at 595 nm (Figure 3a–c). Each assay plate contained six positive controls (*wild-type* PlyC), as well as six negative controls (three containing no protein and three containing catalytically inactive PlyCΔΔ, which lack the EADs necessary for activity) (Figure 3c). The analysis of variants from this primary screen presented a practical problem as to how best to evaluate the activity of the variants in each library. We chose to group the variants from each library into three classes dependent on their activity: *wild type*-like, compromised, and inactive. Variants with a lytic activity within two standard deviations of the positive control were classified as *wild type*-like, and variants within two standard deviations of the negative control were classified as inactive (Figure 3d). Variants in between these two ranges were classified as compromised.

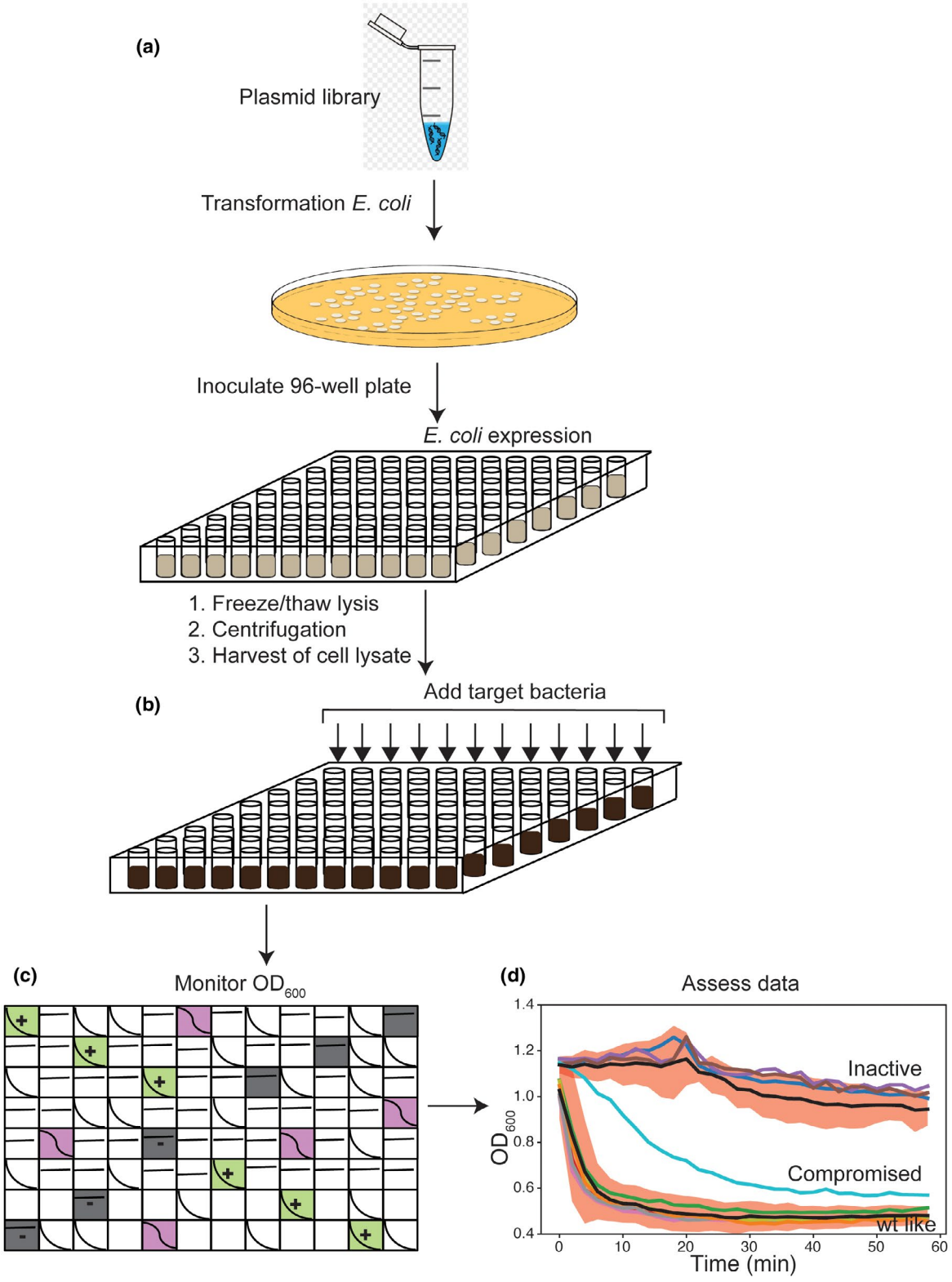

FIGURE 3 Screening mutagenesis libraries via measurement of bacteriolytic activity. (a). A plasmid library is used to transfer ampicillin resistance to *E. coli*. A single colony is used to inoculate a well in a deep 96-well plate. Protein expression occurs in a 96-well plate format and whole cell lysates are obtained through a freeze/thaw lysis process followed by centrifugation. A sample of the lysate is transferred to a sterile 96-well plate. (b). The target bacteria are added to all wells. (c). The plate is continuously monitored with temperature control. In our assays, we had positive controls (green squares with +) and negative controls (no protein, gray squares; inactive protein, gray squares with −) located throughout the plate. Raw data were grouped according to the activity observed. (d). Resultant data are then assessed—a representative figure showing the lytic activity over time is shown. The activity of a single row from one 96-well plate is shown as separate colored lines. The red shading represents ±2 *SD* for both negative and positive controls. Each 96-well plate contained six negative and six positive controls, the average of these is plotted in black

3.4 | Position 59 in PlyCB has a high mutational tolerance but cannot accept histidine

Our previous study showed that a K59E mutation reduced the PlyC GAS cell wall binding ability, suggesting that the charge at position 59 was important for cell wall binding (McGowan et al., 2012). Screening of our K59 library showed that 90% of the variants retained *wild type*-like activity, while 8% were compromised and 2% were completely inactive (Figure 4). Considering the *wild-type* background level in the K59 library was reasonably low at 28% (Figure 4 K59 and Supplementary Table S1), the high level of variants with *wild type*-like activity indicated that this position is not very important for the lytic activity of PlyC. Sequencing of the inactive and compromised variants as well as 26 randomly selected variants with *wild type*-like activity, identified K59E as compromised and K59H as inactive, whereas changing the position to Arg or a hydrophobic (Val, Leu, Pro, Gly, and Ala) amino acid had no effect on the lytic activity of PlyC. In addition, no polar (Gln, Met, and Thr) or aromatic (Trp) variants were identified as compromised or inactive—even though these were highly likely (>90%) to be sampled in the library—suggesting these variants retain *wild type*-like activity.

Given the high mutational tolerance of the K59 site, we would have expected the positive charge of histidine to be permissible. To investigate if the inactive K59H mutant was correctly folded and formed the holoenzyme, K59H was produced and purified, and its holoenzyme assembly was confirmed using size-exclusion chromatography (Supplementary Figure S4). We did not include K59E in this analysis as it has previously been shown to form the holoenzyme and display compromised lytic activity (McGowan et al., 2012). Quantitative analysis of the lytic activity of K59H showed that it was >1500-fold less active than the *wild-type* PlyC, indicating that histidine is not tolerated. In an attempt to explain this data, the K59H mutant was successfully crystallized. However, in the absence of a ligand bound structure, the PlyC(K59H) structure did not immediately clarify why K59H was not tolerated, as there were no significant changes to the local structure of the binding groove (Figure 5). There is some steric hindrance by the histidine that may preclude access of the polyrhamnose sugar to the binding cleft in the K59H variant. The lysine at position 59 in *wild-type* faces away from the binding groove and thus may not hinder access as histidine might. Taken together, these results suggest that the introduction of a histidine at position 59 inactivates PlyC, however this appears to be via an indirect influence on catalysis rather than a direct result of K59 acting as an important residue for binding to the rhamnose sugar.

3.5 | Residues Y26 and E36 contribute to the correct folding of PlyCB

Mutations at Y26 and E36 had a substantial impact on the lytic activity of PlyC. Comparisons between the *wild-type* background level in these libraries (Y26 = 60% and E36 = 78%) with the percent of variants displaying *wild type*-like activity (Y26 = 50% and E36 = 71%) suggest that the majority of variants with *wild type*-like activity correspond to *wild-type* constructs (Figure 4, Y26 and E36). The inactive variants in these two libraries were identified and removed to produce a second generation of the libraries. The second-generation libraries were subjected to the same high-throughput Illumina sequencing as the first-generation libraries, and the amino acid distribution was then compared to the original site saturation libraries (Supplementary Table S1). Due to the removal of inactive variants, the frequency of mutations that inactivate PlyC should decrease, whereas the frequency of mutations that resulted in either compromised or *wild type*-like activity should remain the same or increase. One caveat is that only amino acids that were sampled in the original library can be removed, therefore, the frequency in the first- and second-generation libraries were only compared for variants with a ≥85% chance of being sampled in the primary screen (Supplementary Table S1).

In the E36 library, a total of seven amino acids (Ala, Asp, Gln, Gly, His, Pro, and Val) were present at a frequency high enough to be sampled in our primary screen (Supplementary Table S1). The results indicate that replacement of E36 with a hydrophobic (Ala, Gly, Pro, and Val) or polar (Gln and His) residue abolishes the lytic activity of PlyC (Supplementary Figure S5 and Supplementary Table S1). In order to quantitate the effect of removing the longer, negatively charged side chain of glutamate, we used site-directed mutagenesis to produce *plyCB (E36A)* and attempted to purify the protein. The protein was produced in *E. coli* (Supplementary Figure S4), however, failed to bind to the CHT affinity column. Given only folded PlyCB will interact with the calcium phosphate resin of the affinity column, a lack of binding to the column suggests that the E36A mutant did not fold correctly (McGowan et al., 2012). In contrast, mutation to Asp was tolerated, suggesting it is the negative charge of E36 that is important for the lytic activity of PlyC rather than length.

For the Y26 library, a total of 11 amino acids (Ala, Arg, Asp, Glu, Gly, His, Leu, Pro, Ser, Trp, and Val) as well as a stop codon were present at a frequency high enough to be sampled in our primary screen (Supplementary Table S1). Mutation of Y26 to a stop codon was obviously non-permissible, and with the exception of Trp, all the other sampled amino acids were inactive. For Trp, no reduction in frequency between the first and second generation of libraries was observed (Supplementary Figure S5 and Supplementary Table S1). We used site-directed mutagenesis to produce Y26A and Y26W and attempted to purify both proteins in *E. coli*. This showed that Y26W was not produced in a soluble form in the *E. coli* host, and that Y26A (like E36A) failed to bind the CHT column, suggesting that this point mutant was not folded (Supplementary Figure S4). These results are indicative of problems with the folding pathway of PlyCB in the presence of these mutations.

3.6 | Y28 and R66 are important for cell wall binding and lytic activity

The Y28 and R66 libraries demonstrated that the alteration of either of these positions had a substantial impact on the lytic

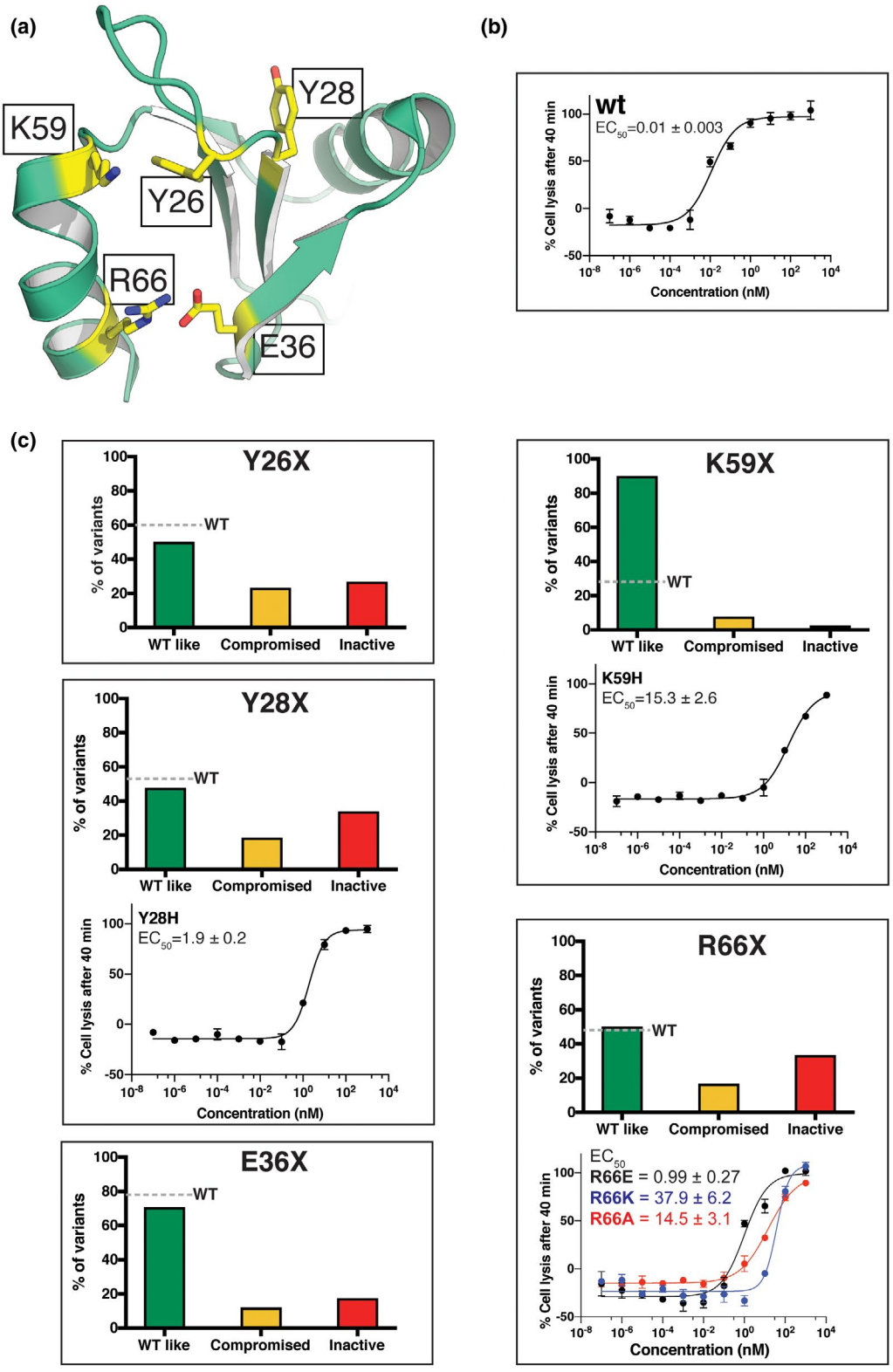

FIGURE 4 Lytic activity of the variants produced in site saturation libraries and single point mutations. (a). The position of the target residues for mutagenesis studies are shown on a cartoon representation of a PlyCB monomer. Residues of interest are marked in yellow. (b). The lytic activity of PlyC *wild-type* shown as a dose-response curve. (c). The results from the site saturation library screening are shown with boxes corresponding to the residue in the graph title. The percentage of variants in each library that showed *wild type*-like (green bars), compromised (yellow bars), and no activity (red bars) are indicated in each plot. The gray dotted line indicates the frequency of *wild-type* constructs in each library, in order to help compare the level of *wild type*-like activity against the number of actual *wild-type* constructs in the dataset. Where single point mutants were produced and purified, dose-response graphs are shown below. The dose-response graphs show the percent cell lysis after 40 min over an increasing concentration of protein. Error bars represent ±1 standard deviation, with experiments performed in triplicate. The EC_{50} is indicated in nM followed by the standard error

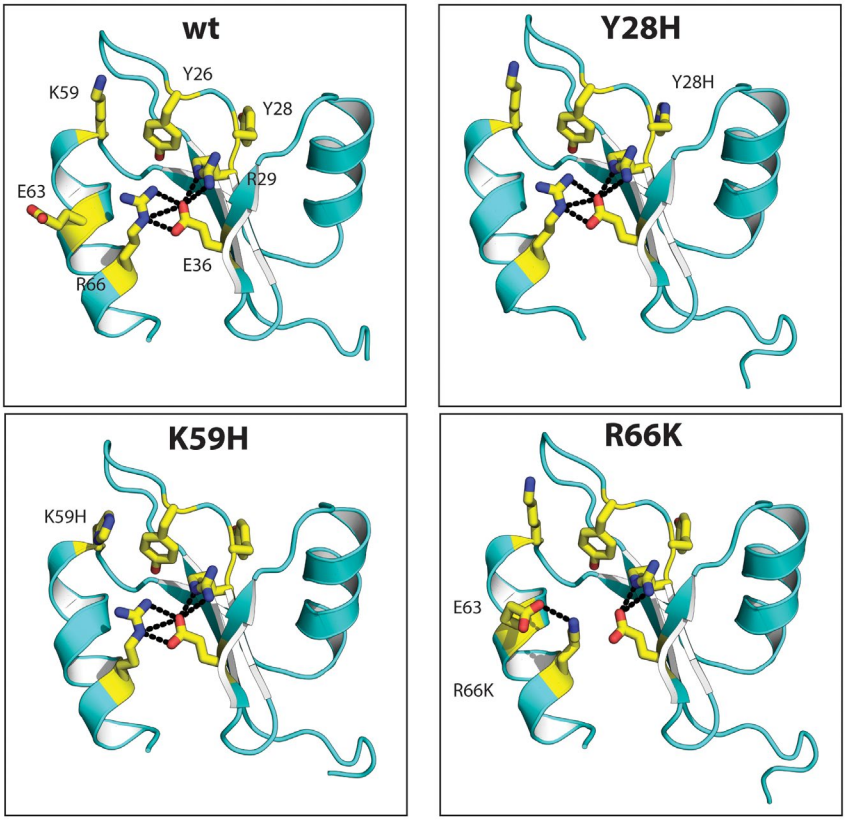

FIGURE 5 Crystal structures of the PlyCB mutants. Cartoon representations of PlyCB monomers from the X-ray crystal structures of *wild type* (PDB ID: 4F87), Y28H (PDB ID: 7KWT), K59H (PDB ID: 7KWW), and R66K (PDB ID: 7KWY). Residues of interest are shown as yellow sticks. Hydrogen bonds/salt bridges are indicated by black dashes

activity of PlyC. Similar to the Y26 and E36 cases, the *wild-type* background levels in these two libraries (Y28 = 54% and R66 = 48%) were similar to the number of variants displaying *wild type*-like activity (Y28 = 50% and R66 = 48%). This once again suggested that the majority of variants with *wild type*-like activity corresponded to *wild-type* constructs (Figure 4, Y28 and R66). We produced a second generation of these libraries using the same criteria as outlined above, and determined that 8 to 11 amino acids were sampled in each library and their frequency was compared (Supplementary Table S1).

In the R66 library, the amino acids Ala, Gly, Leu, Pro, Thr, Trp, and Val were all non-permissible with a reduction in frequency between the first and second generation of the library of >60% (Supplementary Table S1). This result suggests that mutation of R66 to a hydrophobic, aromatic or polar amino acid completely abolishes the lytic activity of PlyC. Given the apparent essential nature of R66, we decided to produce R66A, R66E, and R66K as individual mutants for further analysis. R66E and R66K were not sampled in the R66 site saturation library. R66E has previously been shown to reduce the lytic activity and cell wall binding of PlyC (McGowan et al., 2012) and, therefore, could provide an internal control for our analysis. Additionally, the R66K construct allowed us to probe the apparent requirement for a positive charge at position 66. Recombinant production of the single point mutants R66A, R66E, and R66K confirmed that the proteins were soluble, folded, and correctly formed the holoenzyme (Supplementary Figure S4). Quantitative analysis, in comparison to *wild-type* PlyC, showed a significant reduction in the lytic activity of all R66A, R66E, and R66K mutants (1450-fold, 99-fold and 3790-fold reduction in EC_{50}, respectively) (Figure 4, R66). R66 was thus found to be essential for the lytic activity of PlyC.

R66K was successfully crystallized in this study, while the structure of R66E was elucidated in previous work (Shen et al., 2016). R66 forms a network of bonds including salt bridges with R29 and E36 in *wild-type* PlyCB (Figure 5). However, the lysine of the R66K mutant faces outward and does not appear to make a direct bonding interaction with E36. In both monomers of the resolved dimer, R66K instead interacts with E63, thereby disrupting the R29 and E36 bonding network observed in the *wild-type* structure. The bonding network between this "salt-bridge triad" comprising residues R29, E36, and R66 in the *wild type* is therefore broken, and we suggest that all three residues are necessary for binding to the GAC. This salt-bridge triad is also broken in the structure of the previously solved R66E mutant, resulting in an almost complete loss of lytic activity. Given that lytic activity is a consequence of binding, these data taken together suggest that R66 is highly important for cell wall binding, and perhaps serves as an "anchor-point" for the polyrhamnose sugar, with lesser contributions of R29 and E36, guiding the interaction.

In the case of the Y28 library, similar results were observed as with R66: 11 different amino acids (Ala, Arg, Asp, Gln, Glu, Gly, His, Leu, Pro, Trp, and Val) and a stop codon were sampled. Of these, eight amino acids (Ala, Arg, Asp, Gln, Glu, Gly, Leu, and Pro) and the stop codon were non-permissible, whereas Trp and Val were tolerated (Supplementary Table S1). Interestingly, the mutation of Y28 to His was permitted, suggesting the aromatic phenol ring of Tyr can be substituted with the aromatic imidazole ring of His. We constructed and produced the Y28H mutant and showed, like the

R66 variants, that Y28H did express, purify and form the holoenzyme (Supplementary Figure S4). Quantitation of its activity showed a 190-fold reduction in lytic activity, compared to the *wild type* (Figure 4, Y28). In the crystal structure of Y28H, the side chain of Y28 faces away from the putative binding groove and as such it is not immediately obvious as to how the Y28H mutation interferes with rhamnose binding (Figure 5). In the *wild-type* structure, Y28 stacks with H19; repulsive forces between two positively charged histidine residues in the Y28H mutant may alter the local structure and result in reduced lytic activity under physiological conditions. Mutations at this site were generally poorly tolerated, suggesting that the mutation of Y28 leads to an indirect effect on polyrhamnose binding by altering the conformation of the polyrhamnose binding site. An alteration of local structure may perturb the binding groove, which is reflected in the low mutational tolerance and the reduction in activities in our mutagenesis studies. Alternatively, polyrhamnose in the GAC may be long enough to wrap around the molecule. This may indicate an as yet unpredicted binding mode, additionally explaining the low mutational tolerance of this site.

3.7 | Y28 and R66 are important for binding to the GAS cell wall

To test if the PlyC mutants' reduction in lytic activity was a result of reduced binding to the cell wall, we used confocal microscopy to quantitate cell wall binding. We fluorescently labeled *wild-type* PlyC as well as Y28H, K59H, R66A, R66E, and R66K with Alexa Fluor 488 via primary amines in each of the constructs. Prior to labeling and imaging, all proteins were inactivated with *N*-ethylmaleimide, which has previously been shown to render PlyC catalytically inactive (Nelson et al., 2006). Inactivated and labeled *wild-type* PlyC and mutants were incubated with GAS bacteria, washed twice to remove any unbound protein from the sample, and the resulting bacteria imaged for Alexa Fluor 488 fluorescence using confocal microscopy. For each sample, a brightfield and an Alexa Fluor 488 stack were collected to examine the phenotype of the bacteria and the amount of labeled protein bound. From the brightfield images, it is evident that GAS is present in all samples and displays the expected streptococcal phenotype of cocci that form chains (Figure 6). In the Alexa Fluor 488 stack, fluorescently labeled *wild-type* PlyC appears bright with the fluorescent signal forming defined structures on the GAS cells (Figure 6, *wild type*). In the R66E negative control, staining is more diffuse and the fluorescent signal is not observed as regularly (Figure 6, R66E). Previous studies using R66E show no binding of this protein to the GAS cell wall (McGowan et al., 2012; Shen et al., 2016). Our R66E images do display some binding to the GAS cell wall, however, we could visualize a significantly reduced fluorescence in comparison to *wild-type* PlyC images (Figure 6). To ensure that this observation was significant, we quantitated the fluorescence intensity per square micrometer of bacteria. The relative fluorescence intensity per square micrometer of GAS was significantly lower in the R66E (16,636 ± 2048 (s.d.)), compared to *wild-type* PlyC (23,269 ± 1942, Supplementary Figure S6). Therefore, R66E is considered the baseline for reduced or no binding.

The mutants produced in this study all showed an altered ability to bind to the Streptococcal cell wall compared to the wild type, with the previously characterized R66E used as a negative control (McGowan et al., 2012; Shen et al., 2016) (Figure 6, Supplementary Figure S6). Quantification of the normalized fluorescence intensity of each mutant, compared to the wild type, showed that Y28H, K59H, R66E, and R66A had significantly reduced cell wall binding. Interestingly, R66K did not meet the threshold for significance ($p = .052$, Supplementary Figure S6), however, visual inspection of the images clearly showed that R66K (and Y28H as well) visually resembled those of R66E. R66K appears to have some labeling heterogeneity resulting in a more diffuse, less punctated pattern of fluorescence which we speculate has resulted in the larger variance observed in these samples. Comparison of the fluorescence between R66K and the negative control R66E shows a very high *p* value of 0.815, suggesting that R66K behaves more similarly to the negative control than the wild type (Supplementary Figure S6). K59H and R66A both showed diffuse fluorescence scatter and appeared to have some nonspecific binding as fluorescence intensities could be observed that did align well with the brightfield images of the cocci cells (Supplementary Figure S6).

3.8 | R66 provides a tether allowing mobility to engage a complex substrate

Our attempts to solve the crystal structure of PlyCB in complex with tetrarhamnan using co-crystallization and/or soaking trials were unsuccessful. We suspect that our failure to produce this complex was a result of the relatively low affinity of the ligand (IC_{50} in the micromolar range). We, therefore, used our experimental data as constraints in a molecular docking approach to produce a model of PlyCB bound to tetrarhamnan. Defining the pharmacophore as the groove in the PlyCB monomer between the β-sheet and α-helix, in accordance with residues known to be located in the ligand-binding site, we produced 19 bound conformations (Supplementary Figure S7). Visual inspection of the 19 models reveals that, in all models, two to four pyranose rings bind within a cleft between R66 and K59 by forming at least one hydrogen bond to the side chain of R66 (Figure 7). In addition, 18 of 19 states involve a similar interaction by the formation of 1 to 5 hydrogen bonds to the guanidinium moiety of R29. In contrast, there are only two states that result in a bonding interaction between K59 and the tetrarhamnan molecule, which supports our experimental evidence that that this position is not particularly important for the lytic activity of PlyC.

More broadly outside of the R66 binding cleft, the tetrarhamnan molecule as a whole populates two major conformations (Figure 7a,b). One pose shows the sugar extending across the PlyCB ring parallel to its surface (Figure 7a), or alternatively the second pose shows the sugar wrapping around the PlyCB monomer opposite the PlyCA binding surface, and hence binding perpendicular to

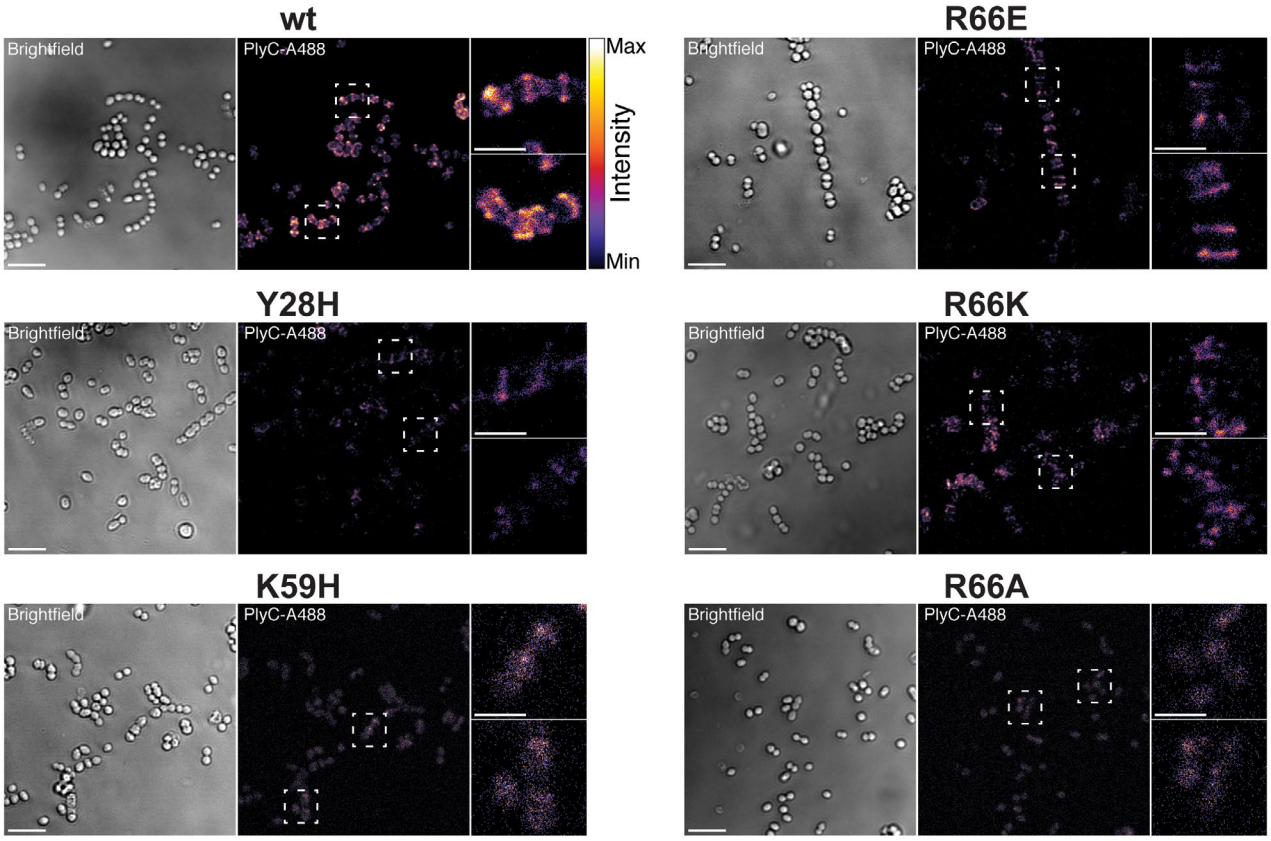

FIGURE 6 Binding of different PlyC variants to the GAS B119 cell wall. Confocal images of GAS B119 incubated with different Alexa Fluor 488-labeled PlyC variants. Each protein is indicated above each panel. The scale bar is equal to a distance of 5 μm in the brightfield (left) and fluorescence (center) images in each panel. Enlargements of the hatched boxes are depicted on the right-hand side of each image with the scale bar representing 2.5 μm

the surface of the PlyCB ring (Figure 7b). In approximately half of the models, the side chain of E36 forms either one or two hydrogen bonds with the tetrarhamnan sugar, suggesting a possible role for E36 in binding to the bacterial cell wall (Supplementary Figure S7). Unfortunately, as E36 has been determined to be an important site for folding, we cannot unambiguously suggest that it is also important for binding via our mutagenesis studies.

In either of the two major conformations, no contacts were observed between Y26 or Y28 and the tetrarhamnan molecule, suggesting these residues may not be directly involved in cell wall binding. While our current results indicate that Y26 is important for folding, our experimental data has not determined this site as being important for binding to tetrarhamnan. However, the results for Y28 are inconsistent with our experimental data that showed that Y28 is involved in binding to the GAC. It is possible that mutation of Y28 to histidine has perturbed the structure of the binding pocket, or that the polyrhamnose binding site extends beyond what is observed in our current model with a simplified ligand. Thus, changes in the local structure of the polyrhamnose binding site may be altered due to indirect effects as a result of mutations to Y28.

An unforeseen interaction outside the putative binding groove was identified in 4 of the 19 states modeled, involving a single hydrogen bonding interaction with K40 (Supplementary Figure S7, state 3, 13, 15, and 17). A single interaction was also detected between tetrarhamnan and E63 (Supplementary Figure S7, state 18). The interaction of K40 and E63 appear only in cases where the rhamnose sugar lies across the surface of the binding groove, and additionally these two residues lie just beyond the putative binding cleft. No other interaction was detected between tetrarhamnan and any other residue of PlyC. While the cases of K40 and E63 were unexpected, the tetrarhamnan docking data presented here suggest that the sugar has potentially great flexibility in acceptable binding poses—congruent with a large and mobile substrate such as the GAC—with the dominant interactions emanating from R29, R66, and E36. This is consistent with the experimental data. Together, between two and eight hydrogen bonding interactions were identified in each of the 19 binding states of PlyCB and tetrarhamnan, with the vast majority of these interactions involving these three key residues involved in the "salt-bridge triad" (93 out of 100 interactions).

4 | DISCUSSION

The primary determinant of endolysin specificity is its ability to recognize and bind to specific ligand molecules within the cell wall (Schmelcher et al., 2012). Lood et al showed that the addition of

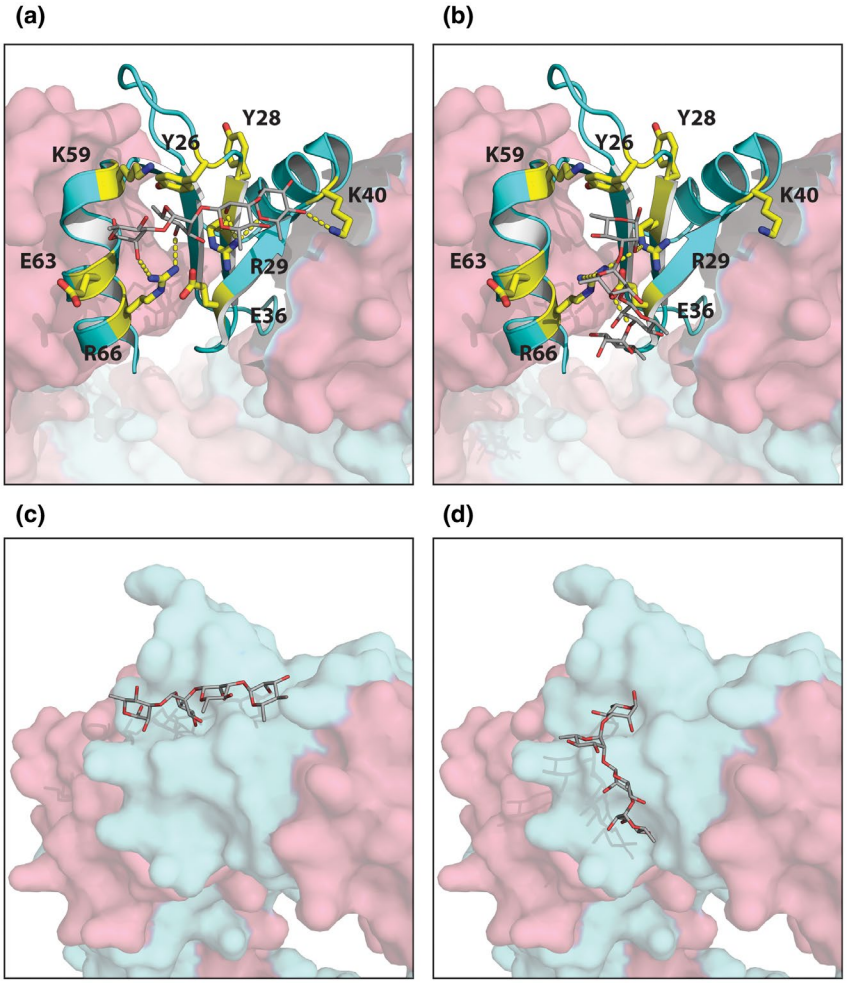

FIGURE 7 Docking of tetrarhamnan to PlyCB. A single PlyCB subunit is shown in a cartoon representation, whereas the remaining part of the ring is shown in a surface representation. Residues of interest are shown as yellow sticks. The tetrarhamnan molecule is shown in gray. The two binding modes identified are shown: (a). The tetrarhamnan sugar is extended across the PlyCB ring parallel to its surface. (b). The sugar is wrapped around the PlyCB monomer, opposite to the PlyCA binding surface, and has bound perpendicular to the surface of PlyCB. Hydrogen bonds were assumed to form if the relevant atoms were within 3.2 Å and are indicated by yellow dashes. (c). and (d). contain the same molecules as in a). and b)., respectively, with the PlyCB octamer in a surface representation. Alternating PlyCB monomers are shown in cyan and magenta. Displayed is the groove in which the tetrarhamnan may bind. All 19 states produced are depicted in Supplementary Figure S7

GAC to PlyC prior to incubation with GAS reduced PlyC's ability to lyse bacteria, suggesting PlyC binds to the GAC. However, previous work did not identify the constituents of the GAC involved (Lood et al., 2014). We have shown for the first time that the lytic activity of PlyC can be inhibited by a synthetic polyrhamnose tetramer, suggesting that the polyrhamnose backbone within the GAC is the cell wall binding target of PlyC. Based on the competition assay, the synthetic tetrarhamnan molecule had a low affinity for PlyC, with an IC50 in the micromolar range. While IC50 is not a direct indicator of affinity, the two are related for competitive inhibitors because inhibition is a consequence of binding. Interestingly, monomeric rhamnose was unable to inhibit PlyC, even at up to the 100 mM concentrations tested, suggesting that the ability to bind PlyC is dependent on the length of the rhamnose polymer.

A groove between the C-terminal helix and the central β-sheet in PlyCB has previously been suggested as a putative polyrhamnose binding site (McGowan et al., 2012). Five residues in this suspected binding groove were targeted for site saturation mutagenesis, to probe their importance for cell wall binding—and thus lytic activity—of PlyC. We discovered that K59 showed a high tolerance for alternate residues, as the majority of variants in the K59 library retained *wild type*-like activity. This indicated that this position is not important for the lytic activity of PlyC, despite this location likely residing in the binding site. However, mutation of K59 to either glutamate or histidine did reduce the lytic activity of PlyC (Figure 4), which is consistent with previous results (McGowan et al., 2012). Additionally, the cell wall of Gram-positive bacteria generally has a negative charge (Low et al., 2011), and thus the introduction of a negative charge at K59 may repel PlyC from the cell wall.

The investigation of the mutational tolerance of the selected mutants demonstrated that Y26 and E36 were required for PlyCB folding. All mutants at these positions generated in this study failed to fold. In the crystal structure of PlyCB, the side chain of E36 forms salt bridges with the side chains of both R29 and R66. The interaction with R66 is not essential for folding as the R66A mutant was folded and formed the PlyC holoenzyme. In contrast, PlyC R29E has previously been shown to be insoluble (McGowan et al., 2012; Shen et al., 2016), strongly suggesting that the salt bridge between E36 and R29 is essential for PlyCB folding. This is also consistent with E36D being tolerated as the salt bridge would likely be retained in this mutant. Because folding is a requirement for cell wall binding, the importance of Y26 and E36 for cell wall binding cannot be discerned from the experimental data. However, our computational model suggests that E36 may interact with polyrhamnose in several of the observed tetrarhamnan binding modes, whereas there is no apparent interaction between Y26 and the sugar.

The residues Y28 and R66 were found to be essential for the lytic activity of PlyC as substitution at either position with a panel of hydrophobic, polar, and aromatic residues rendered the resultant PlyC variants inactive. This was confirmed using purified R66A, R66E, R66K, and Y28H variants, which demonstrated substantially reduced lytic activity. For R66A and R66E, the reduction in lytic activity is most likely a result of decreased cell wall binding. In addition, we suggest that R66K may show decreased binding to the GAC as a result of interacting with E63 instead of the required ligand (Figure 5). The importance of R66 in binding to the GAC was reinforced in the computational models of PlyCB bound to tetrarhamnan, as the guanidinium nitrogens from the R66 side chain form between one and four hydrogen bonds to tetrarhamnan in all binding poses. This illustrates that the removal of this group would thus conflict with cell wall binding. Computational docking of PlyCB and tetrarhamnan manifested two major poses, with the sugar either extending across PlyCB parallel to its surface, or with the sugar wrapping around the PlyCB monomer. While R66 formed multiple hydrogen bonds to tetrarhamnan, Y26, Y28, and K59 did not. The case of Y28 was unexpected, as our experimental data suggests a role for Y28 in PlyC's lytic activity. One explanation for these discrepancies is our Y28H mutant may have an indirect effect on polyrhamnose binding by altering the conformation of the polyrhamnose binding site. In the crystal structure of PlyCB, Y28 forms a hydrogen bond and stacks with H19, and removal of these contacts may alter the local structure in this region.

The data presented here identifies that polyrhamnose is the molecular target of PlyCB, however, we must also consider that the GlcNAc side chains may play a part in the interaction between the GAC and PlyC. Given the context of our molecular docking data, these side chains have the capacity to face outward toward the solvent front, though we cannot rule out that these side chains may play some role in the interaction between the cell wall and residues in PlyC—especially given the variable and heterogeneous nature of the GAC. Notwithstanding this, we must consider that GlcNAc side chains occur on every second rhamnose sugar in the GAC, while in two other PlyC sensitive species of *Streptococcus* every second rhamnose is occupied by either a pair of *N*-Acetylgalactosamine (GalNAc) side chains (in the case of Group C *Streptococcus*), or a glucose (in the case of Group E *Streptococcus*) (Arditi et al., 1989; Wessman, 1986). Given their sensitivity to PlyC, it seems likely that the polyrhamnose backbone is the common moiety and thus may act as a "receptor" for PlyC. It has been noted that antibodies raised against non-GlcNAcylated GAC can recognize the native GAC (van Sorge et al., 2014), indicating that there are stretches of polyrhamnose within the GAC that do not contain these side chains, or perhaps the level of GlcNAc-ylation changes during different stages of cell wall growth. This raises interesting questions in regard to PlyC's highly specific activity—if PlyC can only bind stretches of polyrhamnose lacking side chains, may this explain why some strains in the same species have differing sensitivities to PlyC, or indeed explain the variable rates of PlyC lysis depending on the cell wall growth cycle? These hypotheses are beyond the scope of this study; however, this may provide additional context about the manner in which PlyC and the GAC interact.

It is clear from the experimental and computational evidence that R66 within PlyCB is critical for lytic activity via its role in binding substrate. The computational data taken together would suggest that a triad of residues (R29, E36, and R66) is probably the "anchor-point" of the substrate to PlyCB, with R66 playing a greater role as indicated by the experimental data that suggested binding to the GAC was almost entirely abolished in the various R66 mutants. Without R66, only a weakly bound tetrarhamnan is possible. The contributions of each of R29 and E36 to GAC binding cannot be unambiguously confirmed by mutational studies, since the R29–E36 salt bridge plays an important role in folding. Nevertheless, our data suggests all three residues in this triad are necessary for efficient binding to the target substrate.

In addition to providing an explanation for some of the experimental data, the PlyCB:tetrarhamnan computational model also provides intriguing insight into the possible binding mode between polyrhamnose and PlyCB. The model captured two binding conformations for tetrarhamnan, indicating that PlyCB may not bind to the polyrhamnose using a single binding pose. Instead, the polyrhamnose could form a "fuzzy complex" with PlyCB, similar to what has been observed for highly flexible proteins (Fuxreiter & Tompa, 2012). With recognition of the complexity of the system, we hypothesize that the association between polyrhamnose and R66, which is essential for cell wall binding, could function as an "anchor point," with the polyrhamnose otherwise remaining flexible and sampling multiple different binding states. The average length of the polyrhamnose backbone in the GAC is 18 repeating units or 36 rhamnosides (Kabanova et al., 2010; Rush et al., 2017). This length is much longer than the tetrarhamnan tested in this study and would be sufficient to extend across multiple PlyCB subunits, enabling one GAC to potentially interact with multiple PlyCB subunits in the homooctameric ring. Because the polyrhamnose backbone in the GAC is of sufficient length to extend across PlyCB, polyrhamnose could be anchored to R66 in several different PlyCB subunits concurrently, providing avidity to increase the overall affinity between PlyC and the GAS cell wall.

The ability of endolysins to bind the bacterial cell wall affects the activity of the endolysin (Schmelcher et al., 2012). It was previously believed that PlyC bound sensitive streptococci due to a high-affinity interaction between it and some then unknown component in the species-specific carbohydrate (Hoopes et al., 2009; Lood et al., 2014). While our data shows that individual PlyCB subunits display a relatively low affinity for polyrhamnose in the form of tetrarhamnan, polyrhamnose in the GAC is much longer in GAS cells. If indeed the GAC ligand is likely to interact with multiple PlyCB subunits simultaneously, the resulting avidity from multiple low affinity interactions could increase the functional avidity between the PlyCB ring and the GAS cell wall substantially. This combination of affinity and avidity could play an important role in PlyC's activity, as it would prevent PlyC from dissociating from the bacterial cell wall, yet provide it with sufficient mobility to move around on the cell surface. In this model, concurrent binding of multiple PlyCB subunits would ensure that even if one or two

PlyCB subunits detach from the GAC, PlyC remains bound. However, the relatively low affinity between the polyrhamnose sugar and the individual PlyCB subunits would allow each subunit to readily dissociate from polyrhamnose at any one time, which would enable PlyC to "walk" along the cell wall and thereby move to new PG substrate. This hypothesis is supported by studies showing that an increased affinity for the cell wall does not necessarily correlate with an increase in lytic activity. For example, Schmelcher et al. produced a Ply500 variant that had a ~50-fold increase in affinity toward its target bacteria, but displayed reduced lytic activity (Schmelcher et al., 2011). The authors suggest the decreased lytic activity is a result of decreased surface mobility. A recent study into Ply500 binding to the *Listeria* cell wall uncovered that increased avidity occurred as a result of multiple interactions with cell wall teichoic acids and Ply500, resulting in hydrolysis of the PG (Shen et al., 2021). This indicates that functional avidity can play an important role in the action of the CBD. It therefore appears that the cell wall affinity must be sufficient to allow efficient binding of the endolysin to the target bacterium, but low enough to retain surface mobility (Broendum et al., 2018; Schmelcher et al., 2011). The unique octameric structure of the PlyC CBD may enable it to achieve this compromise by combining multiple low affinity interactions with PlyCB monomers with overall high avidity.

Using avidity to obtain high functional affinity for the cell wall while maintaining surface mobility could be employed by several other endolysins, as repeating units within CBDs are common (Broendum et al., 2018). The Cpl-7 endolysin, encoded by the pneumococcal Cpl-7 phage, uses three repeats of a three-helix bundle to bind to choline in the bacterial cell wall (Diaz et al., 1991). Deletion of one of the repeats has no significant effect on the lytic activity of Cpl-7, but deletion of two repeats reduces the lytic activity substantially, and finally deletion of all three repeat renders Cpl-7 inactive (Bustamante et al., 2017). Similar results were observed for the CTP1L endolysin that targets *Clostridium tyrobutyricum*. CTP1L contains a secondary translation site within the endolysin gene, producing a truncated CBD. This domain interacts with the CBD of a full-length endolysin to produce the biologically functional assembly containing two copies of the CBD (Dunne et al., ,2014, 2016). Interestingly, disruption of the CBD dimer in CTP1L completely abolishes lytic activity, suggesting dimerization is essential for lytic activity (Dunne et al., 2014). Based on this and our data, having multiple cell wall binding repeats appear to be important for the lytic ability of some endolysins, and thus the previous canonical understanding of Gram-positive endolysins as containing one CBD and one EAD is outmoded.

Revealing the mechanism behind cell wall binding, the interplay between affinity and avidity, as well as the purpose of multiple CBDs are critical to future engineering efforts in developing endolysins as an adjunct to our traditional antibiotics ("enzybiotics") (Cheng et al., 2005; Loeffler et al., 2001; Nelson et al., 2001; Schmelcher et al., 2012). To date, one endolysin is in the clinic as a treatment for methicillin-resistant *Staphylococcus aureus* and several others are in clinical trials (Haddad Kashani et al., 2018). Our current understanding of what controls the activity and specificity of endolysins is still modest. Uncovering the molecular details of their function will be key for any future engineering efforts to modify these enzymes for specific purposes and to fully exploit their bacteriolytic potential. Considering the critical importance of cell wall binding for lytic activity, engineering endolysins with an optimal balance between cell wall affinity and surface mobility would be highly desirable. As mentioned earlier, Schmelcher *et al*. added an additional CBD to Ply500 resulting in a ~50-fold increase in overall affinity toward its target bacteria, but the endolysin displayed reduced lytic activity (Schmelcher et al., 2011). What if these engineering efforts had been combined with mutations that decreased the affinity between the individual CBDs and the cell wall, providing a trade-off between affinity and avidity? This decrease in affinity could be a key second engineering step to ensure that the endolysin retains sufficient surface mobility. Regardless, the further understanding of the interplay between affinity and avidity, as well as the mechanism behind cell wall binding, will prove critical for engineering PlyC and other endolysins for use in the clinical, veterinary, and agricultural spaces.

5 | CONCLUSIONS

Several key pieces of information arise from this work: namely, that we have established that PlyC binds to the polyrhamnose backbone in the GAC, and additionally we have further identified key residues in PlyC involved in this interaction. K59, despite its proximity to the polyrhamnose binding site, is not important for the lytic activity of PlyC. In contrast, Y26 and E36 are essential for the correct folding of PlyCB, while Y28 and R66 are critical in binding to the GAS cell wall. Based on these findings, a model for how PlyC binds to polyrhamnose has been proposed, suggesting that that the interaction is not high affinity as previously proposed, but rather high avidity. Thus, we provide here for the first time a detailed insight into the mechanism of how PlyC binds the bacterial cell wall. Understanding this interaction at the molecular level is key in our understanding of how PlyC interacts with the GAC, and as a consequence how this interaction contributes to the remarkable lytic activity of PlyC. Our data provides insight for future engineering efforts of PlyC and other endolysins in the clinical and industrial space.

ACKNOWLEDGMENTS

The authors thank the Australian Synchrotron (MX-2) and the beamline scientists for beamtime and for technical assistance. The authors also thank the Monash Molecular Crystallization Facility and the Monash Bioinformatics Platform for technical assistance. The graphical abstract was in part generated using BioRender.com.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest with the contents of this article.

AUTHOR CONTRIBUTIONS

S.S. Broendum and D.E. Williams performed experiments, analyzed data, co-wrote manuscript. B.K. Hayes, N. Drinkwater, F. Kraus,

J. Fodor, B.E. Clifton, J.D.D. Codee, B.T. Riley, K.A. Farrow, K. Tsyganov, and R.D. Heselpoth performed experiments, analyzed data. D.C. Nelson, C.J. Jackson, and A.M. Buckle concept, training, and provision of funding. S. McGowan concept, performed experiments, analyzed data, co-wrote manuscript, and provision of funding.

DATA AVAILABILITY STATEMENT

The coordinates for the X-ray crystal structures of PlyCB mutants can be found at the PDB ID 7KWW (K59H), 7KWY (R66K), and 7KWT (Y28H).

ORCID

Daniel E. Williams https://orcid.org/0000-0003-0577-2301
Blake T. Riley https://orcid.org/0000-0003-2176-0503
Daniel C. Nelson https://orcid.org/0000-0003-3248-4831
Colin J. Jackson https://orcid.org/0000-0001-6150-3822
Ashley M. Buckle https://orcid.org/0000-0003-2943-9044
Sheena McGowan https://orcid.org/0000-0001-6863-1106

REFERENCES

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N. et al (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D, Biological Crystallography*, *66*(Pt 2), 213–221.

Aragão, D., Aishima, J., Cherukuvada, H., Clarken, R., Clift, M., Cowieson, N.P. et al (2018) MX2: a high-flux undulator microfocus beamline serving both the chemical and macromolecular crystallography communities at the Australian Synchrotron. *Journal of Synchrotron Radiation*, *25*(Pt 3), 885–891. https://doi.org/10.1107/S1600577518003120

Arditi, M., Shulman, S.T., Davis, A.T. & Yogev, R. (1989) Group C beta-hemolytic streptococcal infections in children: nine pediatric cases and review. *Reviews of Infectious Diseases*, *11*(1), 34–45.

Briers, Y., Walmagh, M., Van Puyenbroeck, V., Cornelissen, A., Cenens, W., Aertsen, A. et al (2014) Engineered endolysin-based "Artilysins" to combat multidrug-resistant gram-negative pathogens. *MBio*, *5*(4), e01379-14. https://doi.org/10.1128/mBio.01379-14

Broendum, S.S., Buckle, A.M. & McGowan, S. (2018) Catalytic diversity and cell wall binding repeats in the phage-encoded endolysins. *Molecular Microbiology*, *110*(6), 879–896. https://doi.org/10.1111/mmi.14134

Bustamante, N., Iglesias-Bexiga, M., Bernardo-García, N., Silva-Martín, N., García, G., Campanero-Rhodes, M.A. et al (2017) Deciphering how Cpl-7 cell wall-binding repeats recognize the bacterial peptidoglycan. *Scientific Reports*, *7*(1), 16494. https://doi.org/10.1038/s41598-017-16392-4

Cheng, Q., Nelson, D., Zhu, S. & Fischetti, V.A. (2005) Removal of group B streptococci colonizing the vagina and oropharynx of mice with a bacteriophage lytic enzyme. *Antimicrobial Agents and Chemotherapy*, *49*(1), 111–117. https://doi.org/10.1128/AAC.49.1.111-117.2005

Coligan, J.E., Kindt, T.J. & Krause, R.M. (1978) Structure of the streptococcal groups A, A-variant and C carbohydrates. *Immunochemistry*, *15*(10–11), 755–760. https://doi.org/10.1016/0161-5890(78)90105-0

Czaplewski, L., Bax, R., Clokie, M., Dawson, M., Fairhead, H., Fischetti, V.A. et al (2016) Alternatives to antibiotics—a pipeline portfolio review. *The Lancet Infectious Diseases*, *16*(2), 239–251. https://doi.org/10.1016/S1473-3099(15)00466-1

Diaz, E., Lopez, R. & Garcia, J.L. (1991) Chimeric pneumococcal cell wall lytic enzymes reveal important physiological and evolutionary traits. *The Journal of Biological Chemistry*, *266*(9), 5464–5471. https://doi.org/10.1016/S0021-9258(19)67618-0

Dunne, M., Leicht, S., Krichel, B., Mertens, H.D.T., Thompson, A., Krijgsveld, J. et al (2016) Crystal structure of the CTP1L endolysin reveals how its activity is regulated by a secondary translation product. *The Journal of Biological Chemistry*, *291*(10), 4882–4893. https://doi.org/10.1074/jbc.M115.671172

Dunne, M., Mertens, H.D.T., Garefalaki, V., Jeffries, C.M., Thompson, A., Lemke, E.A. et al (2014) The CD27L and CTP1L endolysins targeting Clostridia contain a built-in trigger and release factor. *PLoS Pathogens*, *10*(7), e1004228. https://doi.org/10.1371/journal.ppat.1004228

Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. (2010) Features and development of Coot. *Acta Crystallographica Section D, Biological Crystallography*, *66*(Pt 4), 486–501.

Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallographica Section D, Biological Crystallography*, *62*(Pt 1), 72–82. https://doi.org/10.1107/S0907444905036693

Evans, P.R. & Murshudov, G.N. (2013) How good are my data and what is the resolution? *Acta Crystallographica Section D, Biological Crystallography*, *69*(Pt 7), 1204–1214. https://doi.org/10.1107/S0907444913000061

Fenton, M., Ross, P., McAuliffe, O., O'Mahony, J. & Coffey, A. (2010) Recombinant bacteriophage lysins as antibacterials. *Bioengineered Bugs*, *1*(1), 9–16. https://doi.org/10.4161/bbug.1.1.9818

Fischetti, V.A. (2005) Bacteriophage lytic enzymes: novel anti-infectives. *Trends in Microbiology*, *13*(10), 491–496. https://doi.org/10.1016/j.tim.2005.08.007

Fischetti, V.A. (2006) Using phage lytic enzymes to control pathogenic bacteria. *BMC Oral Health*, *6*(Suppl 1), S16. https://doi.org/10.1186/1472-6831-6-S1-S16

Fuxreiter, M. & Tompa, P. (2012) Fuzzy complexes: a more stochastic view of protein function. *Advances in Experimental Medicine and Biology*, *725*, 1–14.

Geert Volbeda, A., Reintjens, N.R.M., Overkleeft, H.S., van der Marel, G.A. & Codée, J.D.C. (2016) The cyanopivaloyl ester: a protecting group in the assembly of oligorhamnans. *European Journal of Organic Chemistry*, *2016*(31), 5282–5293. https://doi.org/10.1002/ejoc.201600956

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A. 3rd & Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*(5), 343–345. https://doi.org/10.1038/nmeth.1318

Haddad Kashani, H., Schmelcher, M., Sabzalipoor, H., Seyed Hosseini, E. & Moniri, R. (2018) Recombinant endolysins as potential therapeutics against antibiotic-resistant *Staphylococcus aureus*: current status of research and novel delivery strategies. *Clinical Microbiology Reviews*, *31*(1), e00071–17.

Hoopes, J.T., Stark, C.J., Kim, H.A., Sussman, D.J., Donovan, D.M. & Nelson, D.C. (2009) Use of a bacteriophage lysin, PlyC, as an enzyme disinfectant against *Streptococcus equi*. *Applied and Environment Microbiology*, *75*(5), 1388–1394. https://doi.org/10.1128/AEM.02195-08

Jun, S.Y., Jung, G.M., Yoon, S.J., Choi, Y.-J., Koh, W.S., Moon, K.S. et al (2014) Preclinical safety evaluation of intravenously administered SAL200 containing the recombinant phage endolysin SAL-1 as a pharmaceutical ingredient. *Antimicrobial Agents and Chemotherapy*, *58*(4), 2084–2088. https://doi.org/10.1128/AAC.02232-13

Kabanova, A., Margarit, I., Berti, F., Romano, M.R., Grandi, G., Bensi, G. et al (2010) Evaluation of a Group A *Streptococcus* synthetic oligosaccharide as vaccine candidate. *Vaccine*, *29*(1), 104–114. https://doi.org/10.1016/j.vaccine.2010.09.018

Kabsch, W.X.D.S. (2010) Acta crystallographica Section D. *Biological Crystallography*, *66*(Pt 2), 125–132.

Kim, S., Thiessen, P.A., Bolton, E.E., Chen, J., Fu, G., Gindulyte, A. et al (2016) PubChem substance and compound databases. *Nucleic Acids Research*, *44*(D1), D1202–D1213. https://doi.org/10.1093/nar/gkv951

Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:13033997.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. et al (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Loeffler, J.M., Nelson, D. & Fischetti, V.A. (2001) Rapid killing of *Streptococcus pneumoniae* with a bacteriophage cell wall hydrolase. *Science*, *294*(5549), 2170–2172. https://doi.org/10.1126/science.1066869

Lood, R., Raz, A., Molina, H., Euler, C.W. & Fischetti, V.A. (2014) A highly active and negatively charged *Streptococcus pyogenes* lysin with a rare D-alanyl-L-alanine endopeptidase activity protects mice against streptococcal bacteremia. *Antimicrobial Agents and Chemotherapy*, *58*(6), 3073–3084. https://doi.org/10.1128/AAC.00115-14

Low, L.Y., Yang, C., Perego, M., Osterman, A. & Liddington, R. (2011) Role of net charge on catalytic domain and influence of cell wall binding domain on bactericidal activity, specificity, and host range of phage lysins. *The Journal of Biological Chemistry*, *286*(39), 34391–34403. https://doi.org/10.1074/jbc.M111.244160

McCarty, M. (1956) Variation in the group-specific carbohydrate of group A streptococci. II. Studies on the chemical basis for serological specificity of the carbohydrates. *The Journal of Experimental Medicine*, *104*(5), 629–643. https://doi.org/10.1084/jem.104.5.629

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. & Read, R.J. (2007) Phaser crystallographic software. *Journal of Applied Crystallography*, *40*(4), 658–674.

McGowan, S., Buckle, A.M., Mitchell, M.S., Hoopes, J.T., Gallagher, D.T., Heselpoth, R.D. et al (2012) X-ray crystal structure of the streptococcal specific phage lysin PlyC. *Proceedings of the National Academy of Sciences USA.*, *109*(31), 12752–12757. https://doi.org/10.1073/pnas.1208424109

McPhillips, T.M., McPhillips, S.E., Chiu, H.J., Cohen, A.E., Deacon, A.M., Ellis, P.J. et al (2002) Blu-Ice and the distributed control system: software for data acquisition and instrument control at macromolecular crystallography beamlines. *Journal of Synchrotron Radiation*, *9*(Pt 6), 401–406.

Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S. et al (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, *30*(16), 2785–2791. https://doi.org/10.1002/jcc.21256

Nelson, D., Loomis, L. & Fischetti, V.A. (2001) Prevention and elimination of upper respiratory colonization of mice by group A streptococci by using a bacteriophage lytic enzyme. *Proceedings of the National Academy of Sciences USA*, *98*(7), 4107–4112. https://doi.org/10.1073/pnas.061038398

Nelson, D., Schuch, R., Chahales, P., Zhu, S. & PlyC, F.V.A. (2006) A multimeric bacteriophage lysin. *Proceedings of the National Academy of Sciences.*, *103*(28), 10765–10770. https://doi.org/10.1073/pnas.0604521103

Nelson, D., Schuch, R., Zhu, S., Tscherne, D.M. & Fischetti, V.A. (2003) Genomic sequence of C1, the first streptococcal phage. *Journal of Bacteriology*, *185*(11), 3325–3332. https://doi.org/10.1128/JB.185.11.3325-3332.2003

Rush, J.S., Edgar, R.J., Deng, P., Chen, J., Zhu, H., van Sorge, N.M. et al (2017) The molecular mechanism of N-acetylglucosamine side-chain attachment to the Lancefield group A carbohydrate in *Streptococcus pyogenes*. *The Journal of Biological Chemistry*, *292*(47), 19441–19457. https://doi.org/10.1074/jbc.M117.815910

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T. et al (2012) Fiji: an open-source platform for biological-image analysis. *Nature Methods*, *9*(7), 676–682. https://doi.org/10.1038/nmeth.2019

Schmelcher, M., Donovan, D.M. & Loessner, M.J. (2012) Bacteriophage endolysins as novel antimicrobials. *Future Microbiology*, *7*(10), 1147–1171. https://doi.org/10.2217/fmb.12.97

Schmelcher, M. & Loessner, M.J. (2016) Bacteriophage endolysins: applications for food safety. *Current Opinion in Biotechnology*, *37*, 76–87. https://doi.org/10.1016/j.copbio.2015.10.005

Schmelcher, M., Powell, A.M., Camp, M.J., Pohl, C.S. & Donovan, D.M. (2015) Synergistic streptococcal phage lambdaSA2 and B30 endolysins kill streptococci in cow milk and in a mouse model of mastitis. *Applied Microbiology and Biotechnology*, *99*(20), 8475–8486.

Schmelcher, M., Tchang, V.S. & Loessner, M.J. (2011) Domain shuffling and module engineering of Listeria phage endolysins for enhanced lytic activity and binding affinity. *Microbial Biotechnology*, *4*(5), 651–662.

Schuch, R., Lee, H.M., Schneider, B.C., Sauve, K.L., Law, C., Khan, B.K. et al (2014) Combination therapy with lysin CF-301 and antibiotic is superior to antibiotic alone for treating methicillin-resistant *Staphylococcus aureus*-induced murine bacteremia. *Journal of Infectious Diseases*, *209*(9), 1469–1478. https://doi.org/10.1093/infdis/jit637

Shen, Y., Barros, M., Vennemann, T., Gallagher, D.T., Yin, Y., Linden, S.B. et al (2016) A bacteriophage endolysin that eliminates intracellular streptococci. *Elife*, *5*, e13152. https://doi.org/10.7554/eLife.13152

Shen, Y., Kalograiaki, I., Prunotto, A., Dunne, M., Boulos, S., Taylor, N.M.I. et al (2021) Structural basis for recognition of bacterial cell wall teichoic acid by pseudo-symmetric SH3b-like repeats of a viral peptidoglycan hydrolase. *Chemical Science*, *12*(2), 576–589. https://doi.org/10.1039/D0SC04394J

Shen, Y., Köller, T., Kreikemeyer, B. & Nelson, D.C. (2013) Rapid degradation of *Streptococcus pyogenes* biofilms by PlyC, a bacteriophage-encoded endolysin. *Journal of Antimicrobial Chemotherapy*, *68*(8), 1818–1824. https://doi.org/10.1093/jac/dkt104

Trott, O. & AutoDock, O.A.J. (2009) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*. *31*, 455–461. https://doi.org/10.1002/jcc.21334

van Sorge, N., Cole, J., Kuipers, K., Henningham, A., Aziz, R., Kasirer-Friede, A. et al (2014) The classical lancefield antigen of Group A *Streptococcus* is a virulence determinant with implications for vaccine design. *Cell Host & Microbe*, *15*(6), 729–740. https://doi.org/10.1016/j.chom.2014.05.009

Wessman, G.E. (1986) Biology of the group E streptococci: a review. *Veterinary Microbiology*, *12*(4), 297–328. https://doi.org/10.1016/0378-1135(86)90081-7

Young, R. (1992) Bacteriophage lysis: mechanism and regulation. *Microbiological Reviews*, *56*(3), 430–481. https://doi.org/10.1128/MR.56.3.430-481.1992

Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, *30*(5), 614–620. https://doi.org/10.1093/bioinformatics/btt593

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section.

How to cite this article: Broendum SS, Williams DE, Hayes BK, et al. High avidity drives the interaction between the streptococcal C1 phage endolysin, PlyC, with the cell surface carbohydrates of Group A *Streptococcus*. *Mol Microbiol*. 2021;116:397–415. https://doi.org/10.1111/mmi.14719