EDGE ARTICLE

Check for updates

Structural basis for recognition of bacterial cell wall teichoic acid by pseudo-symmetric SH3b-like repeats of a viral peptidoglycan hydrolase†

Yang Shen, ‡*a Ioanna Kalograiaki, ‡bc Alessio Prunotto,d Matthew Dunne,a Samy Boulos,e Nicholas M. I. Taylor,f Eric T. Sumrall,a Marcel R. Eugster,a Rebecca Martin,a Alicia Julian-Rodero,a Benjamin Gerber,a Petr G. Leiman,g Margarita Menéndez,ch Matteo Dal Peraro,d Francisco Javier Cañada bc and Martin J. Loessnera

Endolysins are bacteriophage-encoded peptidoglycan hydrolases targeting the cell wall of host bacteria via their cell wall-binding domains (CBDs). The molecular basis for selective recognition of surface carbohydrate ligands by CBDs remains elusive. Here, we describe, in atomic detail, the interaction between the *Listeria* phage endolysin domain CBD500 and its cell wall teichoic acid (WTA) ligands. We show that 3′O-acetylated GlcNAc residues integrated into the WTA polymer chain are the key epitope recognized by a CBD binding cavity located at the interface of tandem copies of beta-barrel, pseudo-symmetric SH3b-like repeats. This cavity consists of multiple aromatic residues making extensive interactions with two GlcNAc acetyl groups via hydrogen bonds and van der Waals contacts, while permitting the docking of the diastereomorphic ligands. Our multidisciplinary approach tackled an extremely challenging protein–glycopolymer complex and delineated a previously unknown recognition mechanism by which a phage endolysin specifically recognizes and targets WTA, suggesting an adaptable model for regulation of endolysin specificity.

aInstitute of Food, Nutrition and Health, ETH Zurich, Schmelzbergstrasse 7, 8092 Zurich, Switzerland. E-mail: yang.shen@hest.ethz.ch

bCentro de Investigaciones Biológicas, Margarita Salas, Consejo Superior de Investigaciones Científicas, Ramiro de Maeztu 9, 28040 Madrid, Spain

cCentro de Investigación Biomédica en Red-Enfermedades Respiratorias (CIBERES), Avenida de Monforte de Lemos 3-5, 28029 Madrid, Spain

dLaboratory for Biomolecular Modeling, EPFL, IBI-SV Station 19, 1015 Lausanne, Switzerland

eInstitute of Food, Nutrition and Health, ETH Zurich, Schmelzbergstrasse 9, 8092 Zurich, Switzerland

fStructural Biology of Molecular Machines Group, Protein Structure & Function Programme, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3B, Copenhagen 2200, Denmark

gUniversity of Texas Medical Branch, Department of Biochemistry and Molecular Biology, Sealy Center for Structural Biology and Molecular Biophysics, 301 University Blvd, Galveston, TX 77555-0647, USA

hInstituto de Química-Física Rocasolano, Consejo Superior de Investigaciones Científicas, Serrano 119, 28006 Madrid, Spain

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0sc04394j

‡ These authors contributed equally to this work.

Introduction

Bacteriophage-encoded endolysins are peptidoglycan hydrolases that digest the bacterial cell wall and induce bacterial lysis at the final stage of the phage lytic cycle.[1] Although they naturally work from within the infected cells, exogenous application of recombinant enzyme to susceptible bacteria can likewise lead to rapid degradation of their cell wall, termed lysis from-without.[2,3] Due to their exceptional specificity and efficacy, phage endolysins are increasingly considered as alternatives to antibiotics for combating multidrug-resistant bacteria.[4–6]

The vast majority of endolysins from bacteriophages infecting Gram-positive bacteria possess modular architectures, encompassed in at least one enzymatically-active domain (EAD) and one cell wall-binding domain (CBD).[7] The N-terminal EAD is responsible for cleaving a particular peptidoglycan (PG) bond, and the C-terminal CBD usually holds an independent regulatory function and determines the lytic spectrum upon targeting specific cell-wall elements distributed in a genus-, species- or strain-specific manner. The binding ligands for CBDs are usually of peptidoglycan or carbohydrate nature.[8–10] Given their binding specificity, endolysins and in particular the CBDs constitute highly attractive scaffolds for antibacterial protein engineering.[11,12]

Listeria phage A500 encodes the endolysin Ply500, which comprises L-alanoyl-D-glutamate endopeptidase activity against *Listeria* spp. serovars 4, 5, and 6.[10] Notably, Ply500 is able to rapidly lyse all serovar 4b strains, which are responsible for the vast majority of listeriosis outbreaks,[13,14] rendering Ply500 an

appealing anti-*Listeria* agent to use in food safety applications.[15–17] Ply500 displays a modular architecture comprising largely independent CBD and EAD moieties connected *via* a short and flexible linker peptide (Fig. 1a).[18] Although both modules seem to be required for full activity of this endolysin on its natural murein substrate,[10,18] the isolated C-terminal
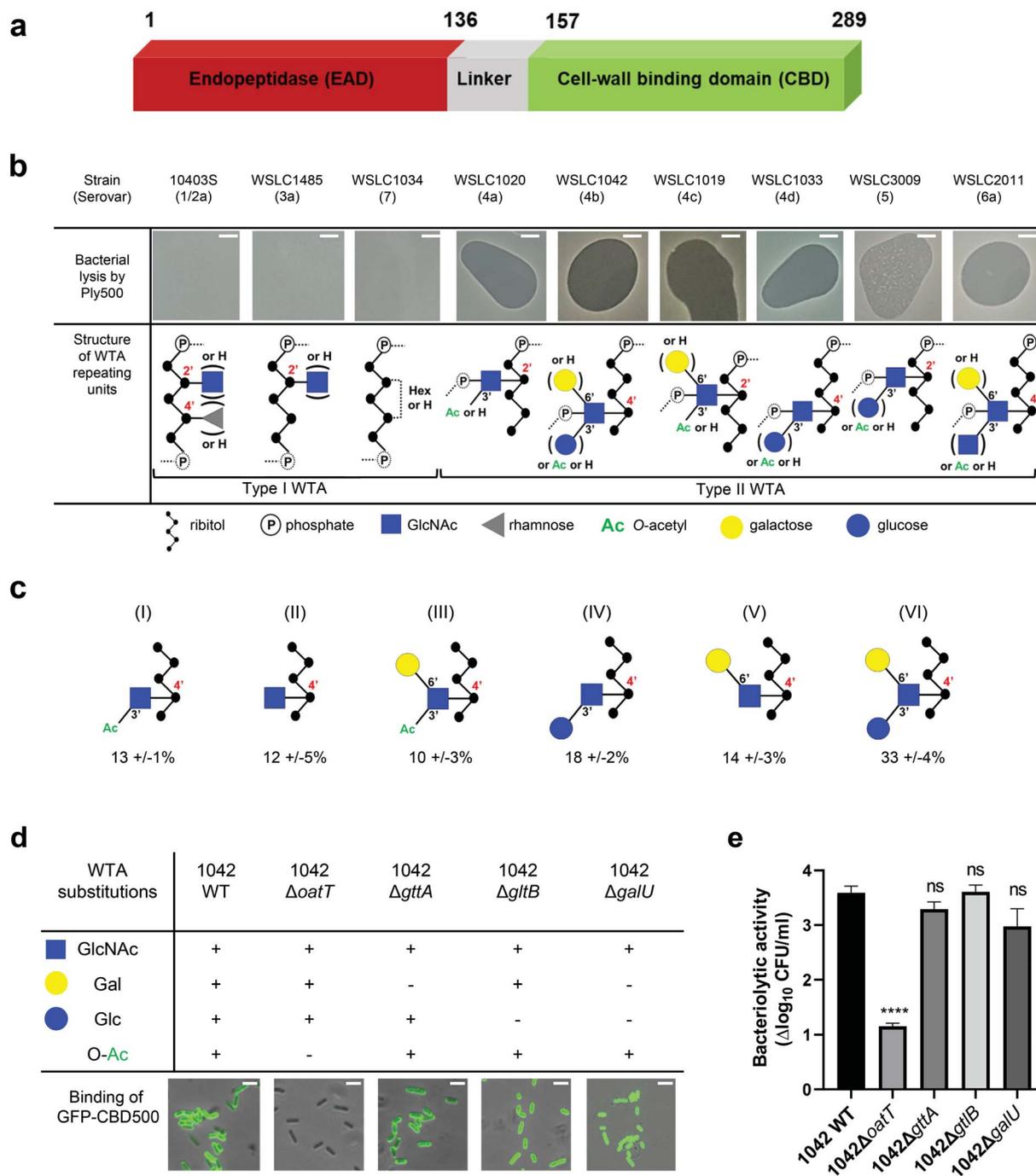

Fig. 1 Correlation of *Listeria* WTA structure with the bacteriolytic activity of Ply500 endolysin and the binding activity of its C-terminal domain CBD500. (a) Modular architecture of the full-length Ply500. (b) Lytic activity of Ply500 detected by spot-on-lawn assay (upper panel; scale bar 300 μm) and corresponding types of WTA reported in the genus *Listeria*.[20] Images are representative of three experiments, and contrast was adjusted for clarity. (c) Relative abundance (%) of the repeating unit variants within *Listeria* WTAs derived from strain WSLC1042 (WTA$_{1042}$), determined by UPLC-MS analysis. Latin numerals (I–VI) denote the repeating unit species of WTA$_{1042}$ identified. Data represent means ± SEM; $n = 3$. (d) Binding of GFP-CBD500 to wild-type WSLC1042 and mutants lacking specific substituents in the backbone of the WTA repeating unit. Scale bar represents 1 μm. Images are representative of three experiments, and contrast was adjusted for clarity. (e) Reduction of bacterial colony counts (Δlog$_{10}$ CFU ml^{-1}) after 1 h incubation of WSLC1042 and selected WTA mutants in the presence of 100 μg ml^{-1} of Ply500. Results are the mean ± SEM of three experiments. Statistical significance was assessed using one-way ANOVA test (****$p < 0.0001$; ns: not significant).

domain (termed CBD500) binds live bacteria cells,[10,19] as well as to purified wall teichoic acid (WTA)[20] with high affinity.

In *Listeria*, WTA polymers account for up to 60% of the dry weight of the cell-wall, and are covalently bound to *N*-acetylmuramic acid (MurNAc) of the PG by a phosphodiester bond through a conserved linkage unit.[21] The WTA chains extend beyond the PG layer, comprising 20 to 40 repeating units that are highly variable due to further glycosidic substitutions or modifications.[22] Depending on their backbone units, two types of WTA exist (Fig. 1b): type I WTAs possess poly-[5]-Rbo-1-P-(O→]$_n$ chains of ribitol phosphate (Rbo-P) units, while type II WTAs feature repeating units of GlcNAc-Rbo-P, following the pattern [4]-GlcNAc-(β1 → 2/4)-Rbo-1-P-(O→]$_n$. WTAs represent the O-antigen determinants in *Listeria* spp., and confer serovar diversity along with some H-antigens.[23] We previously demonstrated that fluorescently labelled CBDs with distinct recognition and binding pattern could differentiate *Listeria* cells down to the serovar level.[24,25] In particular, CBD500 was shown to specifically decorate *Listeria* strains carrying type II WTAs equipped with 3′*O*-acetylated GlcNAc, while tolerating further hexose substitution of the integrated GlcNAc moiety (Fig. 1b).[20] Yet, the underlying recognition mechanism remains elusive with respect to polymer length, GlcNAc/ribitol connectivity, inherent flexibility and microheterogeneity (due to non-uniform, variable decoration of the repeating units). Thus, the attainment of an atomic-scale model required the combination of multiplex ligand- and receptor-based methods.

We used a state-of-the-art multidisciplinary approach to elucidate the mechanism underpinning the recognition of *Listeria* WTAs by CBD500 from both the ligand and protein perspectives. Employing X-ray crystallography, we solved the structure of CBD500 to 1.6 Å resolution, which comprises two copies of beta-barrel SH3b-like repeats held together by means of swapped beta-strands. Saturation Transfer Difference (STD) NMR[26] studies of CBD500 in complex with full length WTAs derived from a selected panel of *Listeria* strains and mutants, highlighted the prerequisite of GlcNAc 3′*O*-acetylation for interaction with CBD500. "Blind" molecular docking on the solved structure suggested three putative binding pockets whose involvement in binding was furtherly assessed by site-directed mutagenesis and functionality assays. STD NMR-driven molecular dynamics simulations and isothermal titration calorimetry (ITC) aided to model the interaction, which is driven by hydrogen bond networks and van der Waals contacts. For the first time, we provide the atomic evidence that swapped SH3b-like repeats found in a *Listeria* phage endolysin permit the selective recognition of carbohydrate ligands,[10,27] instead of PG-related peptides.[28–30] This pioneering study fosters the existing knowledge on WTA recognition by endolysins and provides evidence on the diversification of the SH3b domain folding and binding site.

Results

Ply500 targets *Listeria* WTAs by recognizing *O*-acetylated GlcNAc residues

Spot-on-lawn assays confirmed previous findings that Ply500 specifically targets bacterial cell walls displaying type II WTAs

(Fig. 1b), all of which feature *O*-acetylated GlcNAc integrated in the polymeric ribitol chain.[10,20] To dissect the role of each respective modification of WTA in CBD500 binding, a series of WTA gene deletion mutants[31] derived from the *L. monocytogenes* serovar 4b strain WSLC 1042 were employed. In the WT strain, WTA$_{1042}$ features an unequal ratio of six monomer variants, determined by UPLC-MS analysis (Fig. 1c). Knock-out of *oatT* resulted in loss of WTA *O*-acetylation (ESI Fig. 1a†) and CBD500 binding, whereas deletion of *gttA* (led to a loss of Gal decoration) did not change the binding (Fig. 1d). Interestingly, loss of Glc in both *galU* and *gltB* knock-outs led to slightly enhanced binding, possibly owing to the higher degree of *O*-acetylation found in their WTAs (ESI Fig. 1b†). Consistent with the loss of CBD500 binding, 1042Δ*oatT* mutant cells were found to be significantly less sensitive to Ply500-mediated lysis as opposed to other WTA mutants (Fig. 1e).

Next, we sought to validate bacterial binding using CBD500 alone and extracted WTA polymers. As such, we immobilized CBD500 onto a surface plasmon resonance (SPR) chip, and used wild type and mutant WTA as the analytes. The SPR analysis revealed that the association of WTA with CBD500 is abolished for WTA lacking the *O*-acetylation (ESI Fig. 1c†). Likewise, binding of CBD500 to *Listeria ivanovii* WSLC 3009 (SV 5) was also prevented upon loss of GlcNAc 3′*O*-acetylation (ESI Fig. 1d and e†). In summary, these results indicate that CBD500 specifically recognizes *O*-acetylated WTA as its ligand.

Identification of the carbohydrate epitope recognized by CBD500 with STD-NMR

First, the NMR signals of WTAs derived from WT strains and mutants were assigned (ESI Fig. S2a–c and Tables 1 and 2†), and the mean molecular size of WTA$_{1042}$ and WTA$_{1020}$ was estimated by diffusion ordered spectroscopy[32] (DOSY) (14.7 and 17.7 kDa) respectively, (ESI Fig. S2d and e†) to facilitate the calculation of the molar concentration of repeating units taking into account the HPLC-MS estimates of relative monomer abundance. Next, their recognition by CBD500 was investigated by STD-NMR. This technique is based on selective irradiation of protein protons and subsequent detection of magnetization transfer to the ligand, allowing for uncovering of middle-range affinity binding, ligand screening, and epitope mapping.[33] Measurements carried out in the presence of CBD500 gave rise to STD peaks for both WTA$_{1042}$ and WTA$_{1020}$ (Fig. 2a and ESI Table 3†), confirming specific binding of CBD500 to these structures. Resonances with chemical shifts (δ) in the range 2–2.13 ppm were unambiguously assigned to the methyl protons of GlcNAc units. For both WTA$_{1042}$ and WTA$_{1020}$, the strongest STD signals were observed for the *O*-Ac- and *N*-Ac-methyl protons of the 3′*O*Ac-GlcNAc unit with up to 19 and 25% of spin saturation compared to the off-resonance spectra for *O*-Ac in WTA$_{1042}$ and *N*-Ac in WTA$_{1020}$ respectively, that were normalized to 100% for epitope mapping in each case (see ESI Table 3†), demonstrating that this unit is the primary docking point for the protein (Fig. 2a). Although several proton resonances of the 3′*O*Ac-GlcNAc moiety are located in the crowded spectral region between 3.5–4.3 ppm, where partial resonance overlap
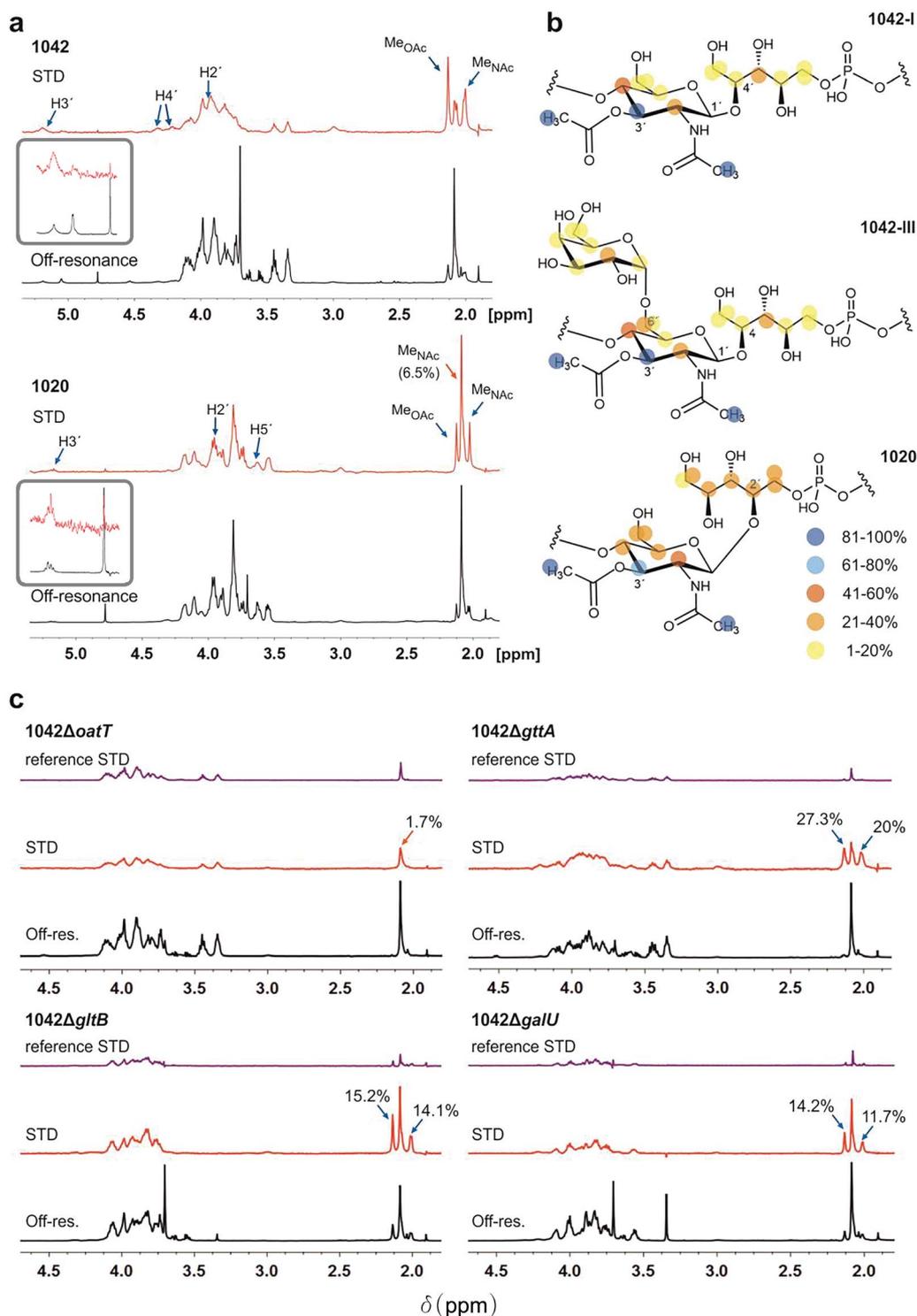
Fig. 2 WTA$_{1042}$ and WTA$_{1020}$ epitopes recognized by CBD500. (a) Representative Saturation Transfer Difference (STD) NMR (red, scaled 16×) and off-resonance (black) spectra of WTA$_{1042}$ and WTA$_{1020}$ recorded in presence of CBD500 upon irradiation at 7.1 and 100 ppm, respectively (1 : 25 CBD : WTA mixtures containing 30 μM CBD500 and approx. 0.75 mM WTA repeating units). Select STD signals are highlighted. Blue arrows point to O-acetylated GlcNAc protons and orange arrows to GlcNAc protons, bearing −OH or −OGlc at C3′, (inserts correspond to zoomed H3′ signals of the 3′-O-acetylated units). (b) STD epitopes for WTA$_{1042-I}$, WTA$_{1042-III}$ and WTA$_{1020}$ O-acetylated GlcNAc moieties as derived from at least three independent experiments; blue and yellow spheres indicate higher and lower saturation transfer intensities. (c) Representative STD (red, scaled 16×) and off-resonance (black) spectra of CBD500−WTA$_{1042}$ complexes where WTA were derived from Lm 1042 mutants ΔoatT ΔgttA, ΔgltB and ΔgalU as indicated; reference STD spectra acquired in absence of CBD500 are also depicted for all mutants in purple (ligand-only control, same scale as STD in presence of protein). Select STD signals are highlighted and absolute STD values are given for the methyl-group protons of relevance. All spectra were acquired as above and the chemical shifts are measured in ppm. Detailed information on STD absolute and normalized values is summarized in ESI Tables 3 and 4.†

complicates the assignment (ESI Table 3†), peaks belonging to H3′ of the 3′O-acetylated unit (~5.2 ppm) were clearly distinguished in the STD difference spectrum, claiming for a strong engagement of the GlcNAc ring in binding. For WTA_{1042}, and to a lesser extent for WTA_{1020}, high magnetization transfer was observed for the 3′OAc-GlcNAc H4′ protons (Fig. 2a and ESI Table 3†) located on the opposite face of the pyranose ring. In striking contrast, the αGal and Rbo protons solely received medium to low saturation transfer, suggesting larger distances to the protein surface (Fig. 2b), again pointing to the GlcNAc moiety as the key interaction residue. Only negligible transfer was observed to the N-Ac-methyl protons of GlcNAc lacking 3′O-acetylation (less than 2% absolute values for WTA_{1042}) discarding a significant involvement of these units in complex formation.

To establish the minimum WTA epitope required for CBD500–WTA interaction and investigate the relevant contributions of 3′O-acetylation or hexose (3′-OGlc or 6′-OGal) substitutions of the GlcNAc moiety, STD-NMR experiments were also performed using WTA isolated from defined WTA mutants (Fig. 1d). Lack of 3′O-acetylation completely abolished recognition by CBD500 (1042ΔoatT, Fig. 2c), in line with the previous observations (ESI Fig. 1c†). Indeed, the absolute STD value measured for the GlcNAc N-Ac-methyl protons was equivalent to the non-specific ligand saturation observed in absence of CBD500 (reference STD, Fig. 2c). For all three variant WTAs derived from the 1042ΔgttA, ΔgltB or ΔgalU mutants,

magnetization transfer values were of similar intensity (ESI Table 4†), suggesting that 3′O-Glc and 6′O-Gal modifications are not essential for CBD500 recognition. Overall, findings obtained by NMR established 3′OAc-GlcNAc moiety as the key epitope for CBD500, and, in the case of WTA_{1042}, indicate that Gal substitution at position 6′ is well tolerated.

Crystal structure of CBD500

To further our understanding of the molecular interactions between CBD500 and WTA, we determined the crystal structure of His-CBD500 (aa 133–289) at 1.6 Å resolution by molecular replacement using the CBD of PlyPSA (PDB ID 1XOV) as a search model (structure solution statistics provided in ESI Table 5†). As described for Ply PSA,[27] CBD500 consists of two adjoined SH3b-like (pfam PF08239) domains connected through structurally swapped β-sheets (Fig. 3a). The first β-strand (β1p/β1′d) forms a part of the proximal domain before connecting to the distal domain (β2d to β8d/β1′p) with the end of β8d connected to β2p of the proximal domain, which concludes with β6p. Only the distal domain features the α-helix turn (α1) and an additional β-strand (β7) as found in other SH3b domains.[34,35] The two RT loops[36] (connecting β2p-β3p and β2d-β3d) that are distinctive features of SH3 and SH3b repeats are positioned facing each other at the interface between the domains. For a simplified view, a topological diagram of the CBD500 secondary structure is shown in Fig. 3b.
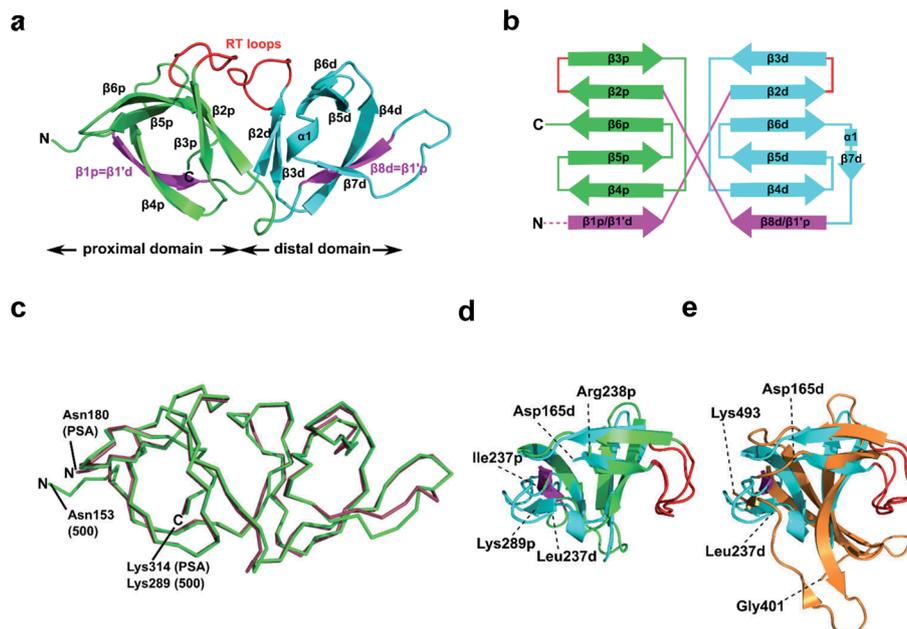

Fig. 3 Structural features of CBD500. (a) Cartoon representation of the proximal (p; green, Cα atoms 153–164 and 238–289) and distal (d; cyan, Cα atoms 165–237) SH3b-like repeats of CBD500 (number corresponds to the full length Ply500). The β-strands β1p and β8d (magenta) are structurally swapped between the two repeats, thus corresponding to β1′d and β1′p, respectively. Loops connecting β2p-β3p and β2d-β3d closely resemble the ligand-binding RT loops[44] of canonical SH3 domains (red). (b) Topology diagram of CBD500 with secondary structure elements shown as cylinders for α-helices and arrows for β-strands. (c) Wire representation of CBD500 and CBDPSA (PDB ID 1XOV, Cα atoms 180–314) structures upon alignment (RMSD of 0.39 Å for 120 Cα atoms). (d) Superposition of the pseudo-symmetric (proximal and distal) repeats of CBD500 (RMSD of 1.23 Å, for 34 Cα atoms). (e) Superposition of the distal repeat of CBD500 (cyan) and the single SH3b domain of lysostaphin calculated by the DALI server[37] (PDB ID 5LEO; RMSD of 1.7 Å, Cα atoms 401–493, orange) with the RT loops colored red.

A DALI search[37] of the PDB with full-length CBD500 identified the CBD of PlyPSA[27] as a structural homolog (RMSD 0.9 Å for 135 residues), which features the same structurally swapped β-strands connecting its SH3b repeats (Fig. 3c). Interestingly, as individual repeats, the proximal and distal subdomains of CBD500 feature high structural similarity (Fig. 3d), despite only 17% sequence identity. Both domains also present high structural similarity to *staphylococcal* bacteriocin CBDs consisting of single SH3b domains, such as ALE-1 (PDB ID 1R77; RMSD 2.0 Å for 54 Cα atoms)[38] and lysostaphin (PDB ID 5LEO; RMSD 2.1 Å

for 58 Cα atoms),[35] including similar positioning of their individual RT loops (Fig. 3e).

Identification of the WTA-binding site in CBD500 by blind docking and site-directed mutagenesis

To precisely identify the WTA binding site, the 3-*O*-acetylated repeating unit of WTA_{1020} (GlcNAc-(β1 → 2)-Rbo-1-PO_4) was docked onto the crystal structure of CBD500. Three putative cavities with hydrophobic patches were selected. Two of them were also
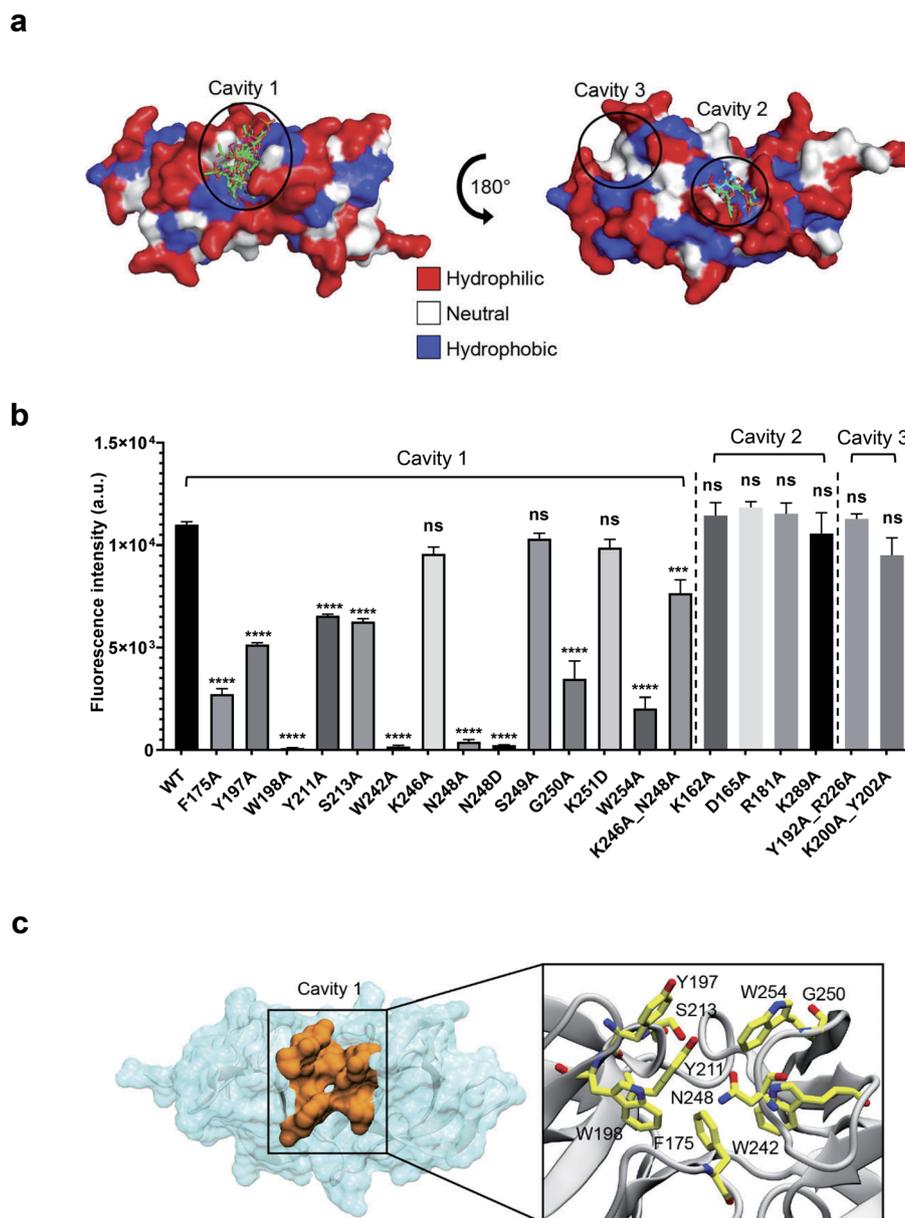

Fig. 4 Identification of the ligand binding cavity by blind docking and site-directed mutagenesis. (a) Putative binding cavities on the surface of CBD500 predicted from blind docking analysis. All ligand poses identified through blind docking are represented in sticks. The CBD500 surface is colored by residue hydrophobicity: hydrophilic (red), neutral (white), and hydrophobic (blue) regions. (b) Relative binding of GFP-CBD500 mutants to *Listeria* 1042 cells compared to WT, quantified by fluorescence spectroscopy. Results are the mean ± SEM of three experiments with different protein batches. Statistical significance was assessed using one-way ANOVA test (***$p < 0.001$; ****$p < 0.0001$; ns: not significant). (c) Close-up view of the CBD500 binding cavity 1 (orange). The enlarged view of the square region, shown on the left, depicts the residues (stick representation) shown to be important for binding of CBD500 to *Listeria* surfaces, based on mutational studies.

identified by a blind docking carried out with AutoDock Vina[39] using a grid box comprising the whole CBD500 surface (Fig. 4a), with cavity 1 preferentially occupied by ligand. A panel of alanine-scanning mutants within residues of these three cavities were produced and tagged with GFP, to enable quantitative determination of their relative binding capacity to the WSLC 1042 *Listeria* surface by fluorescence spectroscopy (see Methods†). Changing of W198, W242 and N248 to alanine almost completely abolished the binding, closely followed by W254A, F175A and G250A mutants (Fig. 4b). Other mutants showing a significant decrease in bacterial surface binding were Y197A, Y211A and S213A, thus inferring a contribution of the mutated residues to either conservation of the protein conformation or the complex formation. All affected residues are within the cavity 1, confirming the output prediction from the docking study.

The molecular weight, chain integrity, and conservation of the protein fold for all GFP-CBD500 mutants were confirmed by SDS-PAGE, analytic gel filtration, and circular dichroism spectroscopy, respectively (ESI Fig. 3a–c†). Only in the case of N248A a size reduction was observed, consistent with a 20 aa truncation at C-terminus identified by MS peptide mapping (data not shown). Another mutant (N248D) did not reveal any sign of truncation, yet still exhibited abolished binding (Fig. 3b). Interestingly, the double mutant (K246A_N248A) was able to restore the binding, albeit not to WT level. Furthermore, to reduce the influence of GFP tag, we examined the conformation of several non-tagged CBD500 mutants by analytical gel filtration. The elution profiles of all the mutants were similar to that of the WT (ESI Fig. 3d and e†).

We next tested if the lack of cell wall recognition by select mutants (W198A, W242A, N248A, and W254A) had an effect on the lytic ability of full-length Ply500 endolysin. Ply500 mutants were purified and tested against purified cell wall substrate (ESI Fig. S3f and g†). As expected, all Ply500 mutants bearing an inactivated CBD showed significantly reduced activity,[18] confirming the importance of host recognition by the CBD and of the residues within cavity 1 for modulating the activity of Ply500.

Collectively, the site-directed mutagenesis experiments identified a tweezer-like site in the cavity 1 that possesses a distinctive patch of several hydrophobic and polar residues adjacent to the two RT-loops (Fig. 4c).

Modelling of $3'OAc$-βGlcNAc-Rbo-PO$_4$ in complex with CBD500

To identify the feasible interactions between CBD500 and WTA upon complex formation, STD-guided Molecular Dynamics (MD) simulations were run. First, the structure of each complex (CBD500 : WTA$_{1020}$, CBD500 : WTA$_{1042-I}$ and CBD500 : WTA$_{1042-III}$) was modelled by docking of the corresponding $3'OAc$-bearing ligands onto the cavity 1 of CBD500. The MD starting structures were selected based on a combined criteria including favorable binding energy estimation, chain continuity, as well as best fitting between the experimental STD and the theoretically calculated values by applying CORCEMA-ST analysis (see Methods†).[40]

For the WTA$_{1020}$ monomer, the CORCEMA-ST analysis pointed towards a single docking conformation for β-GlcNAc-(1 → 2)-Rbo in line with experimental STD values (Fig. 2a, ESI Table 3†). This docking output was used as input for MD simulations as described in the Methods section.† Throughout the simulation the monomeric ligand remained strongly bound at the interface of the RT loops, while the loops and the overall protein structure did not suffer notable distortions (Fig. 5a), therefore advocating for the establishment of a single WTA$_{1020}$ repeating unit as the minimal epitope recognized by CBD500. The complete MD 5000-frame trajectory of WTA$_{1020}$ was subjected to CORCEMA-ST analysis. The profile of the mean theoretical STD values was found to be analogous to the absolute STD experimental values (Fig. 5b), further supporting the use of a single repeating unit in the complex modelling. Protons that received the highest saturation transfer were the methyl protons of O- and N-acetyl groups of $3'OAc$-GlcNAc moiety, followed by protons H3′ and H5′ of the sugar ring (Fig. 5b). The MD output was screened for experimentally conformist poses using CORCEMA-ST, yielding two representative monomeric structures for which the calculated absolute STD profile of the sugar ring protons reproduced experimental mean values (Fig. 5b). The position of the sugar ring remains unchanged at the binding cavity surrounded by several tryptophan residues (Fig. 5c). Both selected structures, belonging to the same and most populated group generated by geographical clustering (see Methods section†), were found to occupy a position close to the centroid of the cluster (Fig. 5d). Even considering with high confidence the GlcNAc ring anchored through the two acetyl groups, the Rbo protons exhibit significant motility throughout the MD (Fig. 5b and c), which could be attributed at least in part to the inherent flexibility of Rbo, as expected for a monomeric ligand with open edges.

Docking of WTA$_{1042}$ I and III monomers onto CBD500 yielded two feasible solutions (modes A and B) per complex. Notably, they were up-down opposite orientations (ESI Fig. 4a and b†), where both acetyl groups of $2'$-NAc and $3'$-OAc of the GlcNAc exchange their relative positions, while still pointing inside the cavity, thus resulting in an inverted orientation of the polymer chain. MD simulations were run for mode A, with the equivalent orientation to the WTA$_{1020}$ monomer pose. Although the ligand remained strongly bound throughout the MD trajectory, the mean theoretical STD profile for both WTA$_{1042}$ monomers did not totally fit with the experimental values obtained with the WT polymer (ESI Fig. 4c and d†). While the calculation predicts the highest saturations for both, N- and O-acetyl proton groups, larger deviations originate from the calculated values for H4′ and H5′ protons in the opposite faces of GlcNAc pyranose ring, in contrast with MD results of WTA$_{1020}$. As done for WTA$_{1020}$, two representative structures for each monomer of WTA$_{1042}$ I and III were selected after the clustering analysis. After superimposition into the binding site of CBD500, increased motility was observed for both WTA$_{1042}$ monomers compared to that of WTA$_{1020}$ (Fig. 5e and f). In fact, the two solutions of WTA$_{1042-I}$ showed a considerable shift of the GlcNAc ring, which was significantly reduced by the presence of Gal at position 6′ in WTA$_{1042-III}$. In the context of
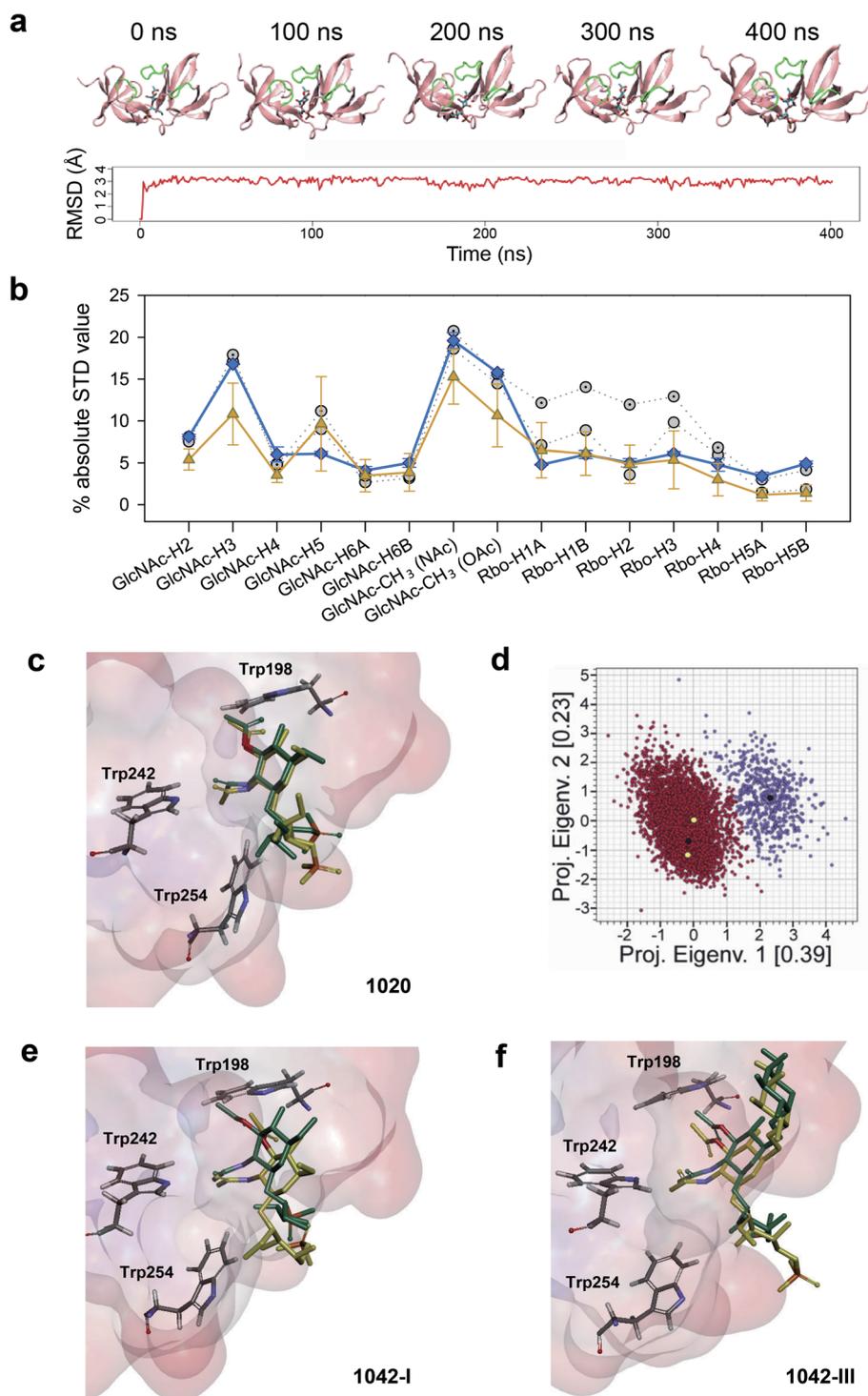
Fig. 5 STD NMR-driven MD simulations and structure selection for complex modeling. (a) Time evolution of the complex of CBD500 with the 3′OAc-bearing WTA$_{1020}$ repeating unit by MD simulations. Representative time-frames (top) and root mean square deviation of the monomeric ligand (bottom). The WTA monomer is represented in sticks, and the CBD500 loops that conform the binding pocket are colored green. (b) Absolute STD values per proton for the recognized WTA$_{1020}$ repeating unit. The mean ± SD values per proton for the 5000 structures (80 ps intervals), extracted from the MD trajectory (orange) and calculated using CORCEMA, is plotted along with the average experimental values from three independent STD-NMR experiments (blue) and the calculated values for the two structures selected on the basis of best correlation to the experimental results (grey). (c) Superposition of the two selected structures for the CBD500 : WTA$_{1020}$ monomer complex. The ligand and relevant Trp residues for reference are represented in sticks and the CBD500 binding pocket in the surface hydrophobicity mode (scale as in Fig. 4, panel a). (d) Eigenvalues projection map of the 5000 structures (80 ps intervals) extracted from the 400 ns MD trajectory of WTA$_{1020}$. The two identified clusters are represented in red and violet respectively, and the black dots represent their centroids. The yellow dots represent the complex structures that were selected on the basis of correlation with experimental results. (e and f) Superposition of the two selected structures for the CBD500 : WTA$_{1042-I}$ and CBD500 : WTA$_{1042-III}$ complexes (mode A), respectively. Ligand and protein representation as in (c).

a polymeric chain, the longer monomer chain and increased rotational freedom provided by ribitol in βGlcNAc-(1 $\rightarrow$ 4)-Rbo backboned WTAs could also account for a differential response in their recognition by CBD500, compared to β(1 $\rightarrow$ 2) in WTA$_{1020}$.

Thermodynamics of the interaction of CBD500 and mutants with WTA polymers

Next, the binding of CBD500 WT and selected mutants to WTA polymers was examined (Fig. 6a) by isothermal titration

calorimetry (ITC) with emphasis on WTA$_{1042}$ to investigate the binding energetics of complex formation. As often in carbohydrate–protein interactions,[41] the complex formation takes place with a favorable enthalpy change, partially compensated by unfavorable entropic contributions (Table 1 and ESI Table 6†). The affinity of CBD500 for WTA$_{1020}$ is about three times higher than for WTA$_{1042}$, due to its less unfavorable entropy of binding. This fact could denote a higher loss of conformational freedom upon complex formation related to differences in the conformational behavior of their respective repeats. In agreement with
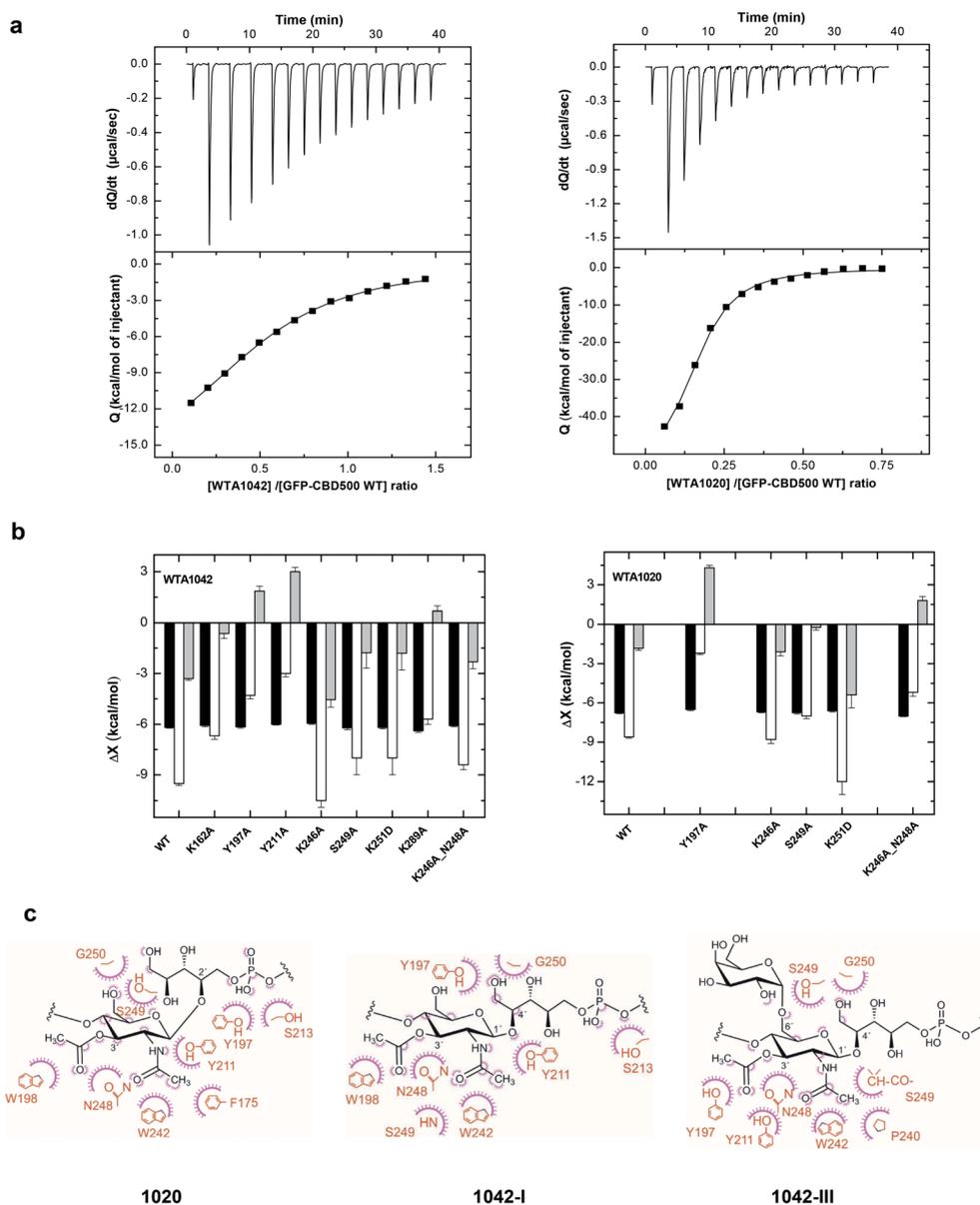

Fig. 6 Binding thermodynamics of CBD500 variants to WTA$_{1020}$ and WTA$_{1042}$ and models of complexes. (a) ITC titration curves of wild-type GFP-CBD500 with WTA$_{1042}$ (left panels) and WTA$_{1020}$ (right panels). Top panels show the raw data, and bottom panels depict the dependence of the heat evolved per mol of ligand injected with various ligand/protein molar ratios. Solid lines correspond to the theoretical fit of the experimental data (■). See Methods† for details. (b) Thermodynamic dissection of the binding energetics ($\Delta G = \Delta H - T\Delta S$) of wild-type GFP-CBD500 and its active-binding mutants to WTA$_{1042}$ (left) and WTA$_{1020}$ (right). Black bars, ΔG; white bars, ΔH; grey bars, $T\Delta S$. (c) Feasible key non hydrogen-bonded contacts of CBD500 with 3′OAc-βGlcNAc-Rbo-PO$_4$ as derived from the MD analyses (1020 in left panel, 1042-I middle panel and 1042-III right panel). Contacts below 4 Å are indicated schematically; the amino-acid residue (in orange) contacts are represented by a purple arc with spokes radiating towards the ligand atoms they contact. The contacted ligand atoms are shown with purple spokes radiating back.

Table 1 *Listeria* cell wall binding capacity of GFP-CBD500 alanine mutants and ITC binding-thermodynamics to WTA$_{1042}$

Constructs	Binding to *Listeria*[a]	N[b]	K_b (M^{-1})	$-\Delta H$ (kcal mol^{-1})	$-\Delta G$ (kcal mol^{-1})	$T\Delta S$ (kcal mol^{-1})
WT	+	1.88 ± 0.03	$(3.5 \pm 0.2) \times 10^4$	9.5 ± 0.1	6.20 ± 0.03	-3.3 ± 0.1
K162A	+	2.1 ± 0.2	$(2.8 \pm 0.4) \times 10^4$	6.7 ± 0.2	6.07 ± 0.08	-0.6 ± 0.3
Y197A	(+)	1.8 ± 0.1	$(3.3 \pm 0.4) \times 10^4$	4.3 ± 0.2	6.16 ± 0.07	1.9 ± 0.3
Y211A	(+)	2.2 ± 0.2	$(2.5 \pm 0.2) \times 10^4$	3.0 ± 0.2	6.00 ± 0.05	3.0 ± 0.25
K246A	+	2.0 ± 0.1	$(2.3 \pm 0.2) \times 10^4$	10.5 ± 0.2	5.95 ± 0.05	-4.55 ± 0.45
S249A	(+)	1.5 ± 0.1	$(3.6 \pm 0.5) \times 10^4$	8.2 ± 0.2	6.30 ± 0.03	-1.9 ± 0.2
K251D	+	1.0 ± 0.1	$(3.5 \pm 0.5) \times 10^4$	8 ± 1	6.20 ± 0.08	-1.8 ± 1
K289A	+	2.0 ± 0.2	$(4.8 \pm 0.7) \times 10^4$	5.7 ± 0.2	6.39 ± 0.09	0.7 ± 0.3
K246A_N248A	+	2.3 ± 0.5	$(2.9 \pm 0.3) \times 10^4$	8.4 ± 0.3	6.09 ± 0.06	-2.3 ± 0.4
I252A	(+)	Protein aggregation prevented quantification of ITC data				
W254A	—	Extremely weak affinity not quantifiable by ITC				
F175A[c]	—	No binding detected by ITC				
W198A	—	Extremely weak affinity not quantifiable by ITC				
W242A[c]	—	No binding detected by ITC				
N248A[d]	—	Extremely weak affinity not quantifiable by ITC				
GFP	—	No binding detected by ITC				

[a] Binding was visualized by fluorescence microscopy using the GFP-CBD500 construct and its variants: '+', strong binding; '(+)', weak binding; '—', no binding. [b] N is the number of CBD molecules bound per WTA chain, K_b, is the binding constant per monomer of CBD bound to WTA as measured by ITC, and ΔG ($-RT \ln K_b$), ΔH and ΔS represent the corresponding thermodynamic parameters. [c] ITC measured with CBD500 mutants. [d] ITC measured with CBD500 and GFP-CBD500 mutants.

fluorescence binding studies, those negative-binding mutants practically abolished the interaction with WTA, whereas the combination of K246A and N248A restored the WTA binding capacity lost by N248A. The binding energetics of CBD500 WT and mutants K246A, S249A, K251D and K246A_N248A to WTA$_{1042}$ were similar. Equivalent results were obtained for WTA$_{1020}$ titration into K246A, S249A and K251D mutants, although significant compensating differences in the enthalpy and entropy of binding of the two latter were perceived with respect to WT. Mutants Y197A, Y211A and K246A_N248A showed a different behavior, as both the enthalpy and the entropy change favored their binding to WTA$_{1042}$ and/or WTA$_{1020}$. This indicates a reduction of hydrogen bonds and/or van der Waals interactions, and is consistent with the binding modes inferred from STD and MD results (see below). Besides, positive entropy changes also support the participation of hydrophobic interactions in the complex formation. Replacement of K162 or K289 with alanine unexpectedly affected both the enthalpy and the entropy of binding, despite their distance to cavity 1. This behavior may denote some alteration in the pi coordination of aromatic rings mediated by Phe175 (see below), since K289 is connected to Phe175 *via* Tyr286, and to K162 by triple H-bonding. Remarkably, the stoichiometry of binding for CBD500 WT and the positive-binding mutants was comparable, with approximately 50–60% of all available 3'OAc-GlcNAc units in a WTA polymer chains interacting with a CBD monomer each, and making feasible an avidity increase due to the high local concentration of WTA chains at the bacterial wall surface.

Atomic interaction of CBD500 with 3′OAc-βGlcNAc-Rbo-PO$_4$

The models derived by the STD-driven MD studies for the three feasible complexes (CBD500 : WTA$_{1020}$, CBD500 : WTA$_{1042-I}$ and CBD500 : WTA$_{1042-III}$) displayed both *N*- and *O*-acetyl groups of the GlcNAc unit deeply docked into the binding site, whereas the Gal unit (WTA$_{1042-III}$) sat at a shallower location, near the CBD500 surface. Notably, the amino acid residues inferred, from the functional and thermodynamic data, to be involved in GlcNAc recognition were well conserved in the final selected models based in STD data and in MD clustering analysis based on topographical criteria (ESI Fig. 5a–c and Table 7†). The GlcNAc ring was consistently involved in bidentate hydrogen bonding through atoms of the acetyl groups, mainly O2N and O1A, and van der Waals contacts mediated by their methyl groups, further supporting the essential role of 3′-OAc decoration in WTA for CBD500 recognition. Trp242 seemed to establish an undisrupted H-bond with GlcNAc atom O2N through its NE1 proton (RMSD between atoms kept at ∼3 Å). Additionally, two main patterns of GlcNAc gripping by Trp198 and Tyr211 were discerned out of the three 400 ns simulations and, remarkably, they appear to be mutually exclusive (ESI Fig. 5b,† right panels). Thus, when Trp198 is hydrogen-bonded to O1A of the 3′-OAc group, the N2 atom of 2′-NAc group is captured by neighboring Tyr211, as predominantly observed in WTA$_{1020}$ recognition (ESI Fig. 5a†). Positioning of the Rbo-PO$_4$ chain in these cases also favored the establishment of H-bonds between the side chain of Ser213 and phosphate oxygen atoms. Alternatively, GlcNAc:O1A can act as acceptor of the OH group of Tyr211 instead of Trp198:NE1, while the distance between Ser213 and the phosphate sharply increases. This pattern was primarily observed throughout the simulations run with WTA$_{1042}$ (ESI Fig. 5, c and Table 7†). Out of the two WTA$_{1042-I}$ poses selected on the basis of correlation to the experimental STD values for the Δ*gttA* polymer (lacking 6′ Gal, Fig. 5e), each showed a distinct recognition pattern – GlcNAc:N2 to Tyr211:OH or Tyr211:OH to GlcNAc:O1A (ESI Fig. 5b,† upper and lower structures respectively) and belonged to separate, comparably populated clusters (data not shown). In contrast, the two selected poses of WTA$_{1042-III}$ (Fig. 5f) belonged to the

same and most populated cluster (data not shown), and presented the same H-bond pattern with Tyr211:OH acting as donor (ESI Fig. 5c†).

Apart from the aforementioned key residues, additional amino acids participated in an intricate H bond network (ESI Fig. 5b†, right panel), such as Ser249, whose main chain gripped either GlcNAc:O2N (WTA$_{1042}$) or Rbo:HO-5 (WTA$_{1020}$). The neighboring amino acid, Gly250, frequently captured Rbo:O5 in all repeating units, yet atom distance showed high deviation (ESI Table 7†), probably due to the Rbo inherent flexibility. Taken together, the prevalent establishment of a distinct H-bonding pattern in WTA$_{1042}$ compared to WTA$_{1020}$ advocates for a distinct, comprehensive CBD500 recognition mode in a global perspective.

In addition to H bonding, van der Waals interactions[41] and carbohydrate–aromatic interactions[42] were expected to be particularly relevant in the formation of CBD500-complex (Fig. 6c), thanks to methyl-protons sets of C–H moieties and the GlcNAc pyranose ring on the ground of a binding site with high density of aromatic residues. Interestingly, Tyr197 appeared as a good candidate to stack with GlcNAc pyranose ring in some docking poses. However, the interaction was not maintained throughout the 400 ns MD considering the distance and the angle between aromatic and pyranose rings. Alanine scanning indicated that Tyr197 switching to Ala did not induce significant affinity loss when the Y197A mutant was titrated with WTA$_{1042}$ (Table 1), despite some alteration in thermodynamic behavior of the complex. Furthermore, the quantified fluorescent binding of Y197A mutant labelled with GFP to the *Listeria* cells was in the same range as for the WT (Fig. 4b). Altogether, these data suggest that pyranose-aromatic geometrical stacking does not play an essential role in the binding. On the other side, MD simulations with the WTA$_{1020}$ monomer also indicated the establishment of van der Waals contacts between the 3′*O*Ac-methyl group of the ligand and the side chain of Asn248, whose loss might also alter the interactions mediated by the Ser249 main chain in the N248A mutant. Phe175, buried in the bottom of the binding groove, was not found to interact directly with WTA$_{1042}$ as with WTA$_{1020}$, but it may play an important structural role through pi coordination of aromatic rings. Therefore, our data suggest that both H bonding and van der Waals interactions greatly contribute to the specificity of the carbohydrate-binding CBD500, allowing for ligand selection with structural heterogeneity.

Discussion

Bacteriophage-encoded endolysins represent a highly attractive class of potent antibacterials, due to their narrow target spectrum, rapid and strong lytic activity. Accordingly, the molecular basis of their cell wall recognition by the endolysin CBD domains is important for development and design of such novel therapeutics. Toward this end, we employed a multidisciplinary strategy, including genetics, biochemical characterization, mutational screening, X-ray crystallography, ligand-based NMR, thermodynamics of binding and computational modelling, to elucidate the structure–function relationship of a bacteriophage endolysin in complex with its WTA ligand.

However, structural studies of WTAs are extremely challenging due to their inherent microheterogeneity and flexibility. Yet, the here described multidisciplinary strategy allows to pinpoint the mechanism by which CBD500 specifically recognizes the 3′*O*Ac-GlcNAc moiety in the context of a Rbo backbone *via* a surface-exposed cavity located at the interface of SH3b-like repeats. This cavity comprises multiple aromatic residues that are commonly involved in protein–carbohydrate recognition,[43] despite no direct pyranose-aromatic stacking was identified. The residues located at the cavity make extensive interactions with the two acetyl groups *via* hydrogen bonds and van der Waals contacts, while permitting the docking of the WTA ligand regardless of the diastereomorphic connectivity (C2 *vs.* C4) on Rbo and hexose substitution.

We proposed a model for the CBD500–WTA complex (ESI Fig. S6a†), which is new and distinct from current knowledge of ligand interaction with other SH3b-like domains. For instance, the single SH3b of lysostaphin has been shown to recognize the stem peptide and pentaglycine bridge of PG using a two binding-site mechanism[9] (ESI Fig. S6b†). The CBD of *Bacillus* endolysin LysPBC15[30] was shown to bind the peptidoglycan through separate binding sites located on opposite sides of its SH3b domains (ESI Fig. S6c†). Similarly, the CBD of *Clostridium* endolysin PhiSM101[29] features putative peptide-binding sites facing in opposite directions of 'pseudo dimer'-like SH3b repeats (ESI Fig. S6d†). This bivalent mode of binding is in strong contrast to the single binding site of CBD500 at the interface between the pseudo-symmetric SH3b repeats. Based on these observations, it appears that the SH3b domains are not homogenous, but have evolved to produce diverse binding sites for recognition of different peptide and carbohydrate ligands. In addition, we found that the SH3b repeats of CBD500 are joined together through β-strand swapping, while the SH3b repeats of both Psm and LysPBC5 fold separately and connect in a more conservative fashion *via* a short linker. The difference in topology may also contribute to the different mechanisms of ligand binding.

The RT loops appear to be fundamental for binding by SH3 and SH3b domains. In eukaryotic systems, SH3 domains feature a canonical binding site involving the region between the RT and Src loops, which typically binds through protein–protein interactions.[44] In fact, this closely resembles the properties of CBD500 and the lysostaphin SH3b domain that interacts with PG through their RT loops.[35] Despite most *Listeria* phage CBDs feature similar domain architectures, some of them, *e.g.*, CBDP35 and CBD025, appear to recognize different types of WTA polymers.[19,20] Further research is required to dissect this subtle structural disparity and understand how phage endolysins evolved diverse binding strategies to different ligands at the atomic level. Interestingly, three known *Listeria* virulence factors, InlB, autolysin-like IspC and P60 also feature multiple SH3b repeats at their C-terminus, which have been demonstrated to adhere to glycosyl-polyol teichoic acids.[31,45,46] Our data may provide guidance on the target specificity and support

future studies of the role of other SH3b-like repeats in virulence and immune defense.

In addition to SH3b, several cell wall-binding modules and their recognizing ligands have been described in structural detail. For instance, the multimodular LysM motifs bind to the PG disaccharide β-N-acetylmuramic acid (1 → 4)-β-N-acetylglucosamine in a 'beads on a string' fashion,[47] or form a binding groove through the intrachain dimerization when interacting with N-acetylglucosamine-containing chitin.[48] Intriguingly, the molecular diversity of LysM is reminiscent of what we found in diverse SH3b-like domains. The three tandem α-helical repeats of Cpl-7 that recognize the PG MurNAc-L-Ala-D-isoGln muropeptide;[8] the CBDs of Bacillus phage endolysin PlyL and PlyG that binds to the secondary cell wall polysaccharides;[49] and the choline binding repeats described in Streptococcus pneumoniae PG hydrolases that interact with the phosphorylcholine residues on teichoic acids through a left-handed β-solenoid domain.[50] These variable cell-wall binding modules provide a pool of scaffolds that may be engineered as molecular tools or payloads to target specific bacteria for diagnostic and therapeutic applications.

O-Acetylation of the MurNAc residues in PG is known to modulate cell wall biosynthesis in S. pneumoniae.[51] In a large number of Gram-positive pathogens, this peptidoglycan modification has been shown to confer resistance to lysozyme, a key effector of the innate immunity.[52] However, the presence of O-acetylation on other cell wall components has only been uncovered recently, such as in Listeria WTA,[20,53] Mesorhizobium loti LPS,[54] and the GlcNAc residue of Lactobacillus plantarum PG.[55] These findings underline the important roles of O-acetylation in modulation of autolysin activity, interaction with protozoans, and even virulence.[31] Interestingly, O-acetylation of other bacterial surface polysaccharides has also been identified as essential modifications for antibody recognition, and thus needs to be considered in vaccine development.[56,57] Here, we identified another eminent function of O-acetylation of WTA in the recognition by phage endolysins, which are important for phage–bacteria interaction. Surprisingly, the binding affinity (ca. $K_D = 10^{-4}$ M measured in solution by ITC analysis) appears significantly lower than the previously reported nanomolar affinities determined by immobilizing whole cells on the SPR surfaces.[10] It is conceivable that weaker local biomolecular interaction may allow the endolysins to better diffuse within the cell wall based on higher dissociation and re-association, and results in an increased avidity through WTA multivalent presentation at cell surface, thereby potentiating the hydrolysis of PG bonds. An intriguing hypothesis is that the mutual adaptability between the binding site and the ligand not only determines the binding specificity, but also boosts the activity of endolysin. The atomic insights and dynamic structures shown in this study will pave new ways towards research in this direction.

Data availability

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author contributions

Conceptualization, Y. S., I. K., A. P., M. M., M. D. P., F. J. C., and M. J. L.; methodology, Y. S., I. K., A. P., M. D., S. B. and M. M.; investigation, Y. S., I. K., A. P., M. D., S. B., M. M., N. M. I. T., E. T. S., M. R. E., R. M., and A. J. R.; writing—original draft, Y. S., and I. K.; writing—review & editing, Y. S., I. K., A. P., M. D., M. M., M. D. P., F. J. C., and M. J. L.; visualization, Y. S., I. K., A. P., M. D., S. B., B. G. and M. M.; resources, M. M., P. L., M. D. P., F. J. C., and M. J. L.; funding acquisition, M. M., M. D. P., F. J. C., and M. J. L.; project administration, Y. S.; supervision, Y. S., I. K., M. M., M. D. P., F. J. C., and M. J. L.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgements

References

1 R. Young, Microbiol. Rev., 1992, 56, 430–481.

2 M. J. Loessner, Curr. Opin. Microbiol., 2005, 8, 480–487.

3 M. Pastagia, R. Schuch, V. A. Fischetti and D. B. Huang, J. Med. Microbiol., 2013, 62, 1506–1516.

4 V. A. Fischetti, Viruses, 2018, 10, 310.

5 M. J. Love, D. Bhandari, R. C. J. Dobson and C. Billington, Antibiotics, 2018, 7, 17.

6 L. Czaplewski, R. Bax, M. Clokie, M. Dawson, H. Fairhead, V. A. Fischetti, S. Foster, B. F. Gilmore, R. E. Hancock, D. Harper, I. R. Henderson, K. Hilpert, B. V. Jones, A. Kadioglu, D. Knowles, S. Olafsdottir, D. Payne, S. Projan, S. Shaunak, J. Silverman, C. M. Thomas, T. J. Trust, P. Warn and J. H. Rex, Lancet Infect. Dis., 2016, 16, 239–251.

7 M. Schmelcher, D. M. Donovan and M. J. Loessner, Future Microbiol., 2012, 7, 1147–1171.

8 N. Bustamante, M. Iglesias-Bexiga, N. Bernardo-Garcia, N. Silva-Martin, G. Garcia, M. A. Campanero-Rhodes, E. Garcia, I. Uson, R. M. Buey, P. Garcia, J. A. Hermoso, M. Bruix and M. Menendez, Sci. Rep., 2017, 7, 16494.

9 L. S. Gonzalez-Delgado, H. Walters-Morgan, B. Salamaga, A. J. Robertson, A. M. Hounslow, E. Jagielska, I. Sabala, M. P. Williamson, A. L. Lovering and S. Mesnage, *Nat. Chem. Biol.*, 2020, **16**, 24–30.

10 M. J. Loessner, K. Kramer, F. Ebel and S. Scherer, *Mol. Microbiol.*, 2002, **44**, 335–349.

11 V. A. Fischetti, *Int. J. Med. Microbiol.*, 2010, **300**, 357–362.

12 D. Kim, S. J. Kwon, X. Wu, J. Sauve, I. Lee, J. Nam, J. Kim and J. S. Dordick, *ACS Appl. Mater. Interfaces*, 2018, **10**, 13317–13324.

13 J. McLauchlin, *Eur. J. Clin. Microbiol. Infect. Dis.*, 1990, **9**, 210–213.

14 C. Buchrieser, R. Brosch, B. Catimel and J. Rocourt, *Can. J. Microbiol.*, 1993, **39**, 395–401.

15 M. Schmelcher and M. J. Loessner, *Curr. Opin. Biotechnol.*, 2016, **37**, 76–87.

16 K. Solanki, N. Grover, P. Downs, E. E. Paskaleva, K. K. Mehta, L. Lee, L. S. Schadler, R. S. Kane and J. S. Dordick, *Sci. Rep.*, 2013, **3**, 1584.

17 M. L. Van Tassell, M. Angela Daum, J. S. Kim and M. J. Miller, *Curr. Opin. Biotechnol.*, 2016, **37**, 88–96.

18 I. P. Korndorfer, A. Kanitz, J. Danzer, M. Zimmer, M. J. Loessner and A. Skerra, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2008, **64**, 644–650.

19 M. Schmelcher, V. S. Tchang and M. J. Loessner, *Microb. Biotechnol.*, 2011, **4**, 651–662.

20 Y. Shen, S. Boulos, E. Sumrall, B. Gerber, A. Julian-Rodero, M. R. Eugster, L. Fieseler, L. Nystrom, M. O. Ebert and M. J. Loessner, *J. Biol. Chem.*, 2017, **292**, 17832–17844.

21 F. Fiedler, J. Seger, A. Schrettenbrunner and H. P. R. Seeliger, *Syst. Appl. Microbiol.*, 1984, **5**, 360–376.

22 K. Uchikawa, I. Sekiawa and I. Azuma, *J. Biochem.*, 1986, **99**, 315–327.

23 K. Kamisango, H. Fujii, H. Okumura, I. Saiki, Y. Araki, Y. Yamamura and I. Azuma, *J. Biochem.*, 1983, **93**, 1401–1409.

24 M. Schmelcher, T. Shabarova, M. R. Eugster, F. Eichenseher, V. S. Tchang, M. Banz and M. J. Loessner, *Appl. Environ. Microbiol.*, 2010, **76**, 5745–5756.

25 E. T. Sumrall, C. Rohrig, M. Hupfeld, L. Selvakumar, J. Du, M. Dunne, M. Schmelcher, Y. Shen and M. J. Loessner, *Appl. Environ. Microbiol.*, 2020, **86**, 00612–00620.

26 T. Haselhorst, A. C. Lamerz and M. Itzstein, *Methods Mol. Biol.*, 2009, **534**, 375–386.

27 I. P. Korndorfer, J. Danzer, M. Schmelcher, M. Zimmer, A. Skerra and M. J. Loessner, *J. Mol. Biol.*, 2006, **364**, 678–689.

28 S. S. Broendum, A. M. Buckle and S. McGowan, *Mol. Microbiol.*, 2018, **110**, 879–896.

29 E. Tamai, H. Yoshida, H. Sekiya, H. Nariya, S. Miyata, A. Okabe, T. Kuwahara, J. Maki and S. Kamitori, *Mol. Microbiol.*, 2014, **92**, 326–337.

30 K. O. Lee, M. Kong, I. Kim, J. Bai, S. Cha, B. Kim, K. S. Ryu, S. Ryu and J. Y. Suh, *Structure*, 2019, **27**, 1355–1365.

31 E. T. Sumrall, Y. Shen, A. P. Keller, J. Rismondo, M. Pavlou, M. R. Eugster, S. Boulos, O. Disson, P. Thouvenot, S. Kilcher, B. Wollscheid, D. Cabanes, M. Lecuit, A. Grundling and M. J. Loessner, *PLoS Pathog.*, 2019, **15**, e1008032.

32 P. Groves, M. O. Rasmussen, M. D. Molero, E. Samain, F. J. Canada, H. Driguez and J. Jimenez-Barbero, *Glycobiology*, 2004, **14**, 451–456.

33 M. Mayer and B. Meyer, *J. Am. Chem. Soc.*, 2001, **123**, 6108–6117.

34 J. Gu, R. Lu, X. Liu, W. Han, L. Lei, Y. Gao, H. Zhao, Y. Li and Y. Diao, *Curr. Microbiol.*, 2011, **63**, 538–542.

35 P. Mitkowski, E. Jagielska, E. Nowak, J. M. Bujnicki, F. Stefaniak, D. Niedzialek, M. Bochtler and I. Sabala, *Sci. Rep.*, 2019, **9**, 5965.

36 D. Zhao, X. Wang, J. Peng, C. Wang, F. Li, Q. Sun, Y. Zhang, J. Zhang, G. Cai, X. Zuo, J. Wu, Y. Shi, Z. Zhang and Q. Gong, *J. Struct. Biol.*, 2014, **187**, 194–205.

37 L. Holm and P. Rosenstrom, *Nucleic Acids Res.*, 2010, **38**, W545–W549.

38 J. Z. Lu, T. Fujiwara, H. Komatsuzawa, M. Sugai and J. Sakon, *J. Biol. Chem.*, 2006, **281**, 549–558.

39 O. Trott and A. J. Olson, *J. Comput. Chem.*, 2010, **31**, 455–461.

40 N. R. Krishna and V. Jayalakshmi, *Top. Curr. Chem.*, 2008, **273**, 15–54.

41 D. Solís and H.-J. Gabius, *Protein-carbohydrate interactions: basic concepts and methods for analysis*, Wiley-VCH, 2009, pp. 233–224.

42 J. L. Asensio, A. Arda, F. J. Canada and J. Jimenez-Barbero, *Acc. Chem. Res.*, 2013, **46**, 946–954.

43 A. B. Boraston, D. N. Bolam, H. J. Gilbert and G. J. Davies, *Biochem. J.*, 2004, **382**, 769–781.

44 N. Kurochkina and U. Guha, *Biophys. Rev.*, 2013, **5**, 29–39.

45 M. Marino, M. Banerjee, R. Jonquieres, P. Cossart and P. Ghosh, *EMBO J.*, 2002, **21**, 5623–5634.

46 F. Carvalho, S. Sousa and D. Cabanes, *Environ. Microbiol.*, 2018, **20**, 3941–3951.

47 S. Mesnage, M. Dellarole, N. J. Baxter, J. B. Rouget, J. D. Dimitrov, N. Wang, Y. Fujimoto, A. M. Hounslow, S. Lacroix-Desmazes, K. Fukase, S. J. Foster and M. P. Williamson, *Nat. Commun.*, 2014, **5**, 4269.

48 A. Sanchez-Vallet, R. Saleem-Batcha, A. Kombrink, G. Hansen, D. J. Valkenburg, B. P. Thomma and J. R. Mesters, *Elife*, 2013, **2**, e00790.

49 J. Ganguly, L. Y. Low, N. Kamal, E. Saile, L. S. Forsberg, G. Gutierrez-Sanchez, A. R. Hoffmaster, R. Liddington, C. P. Quinn, R. W. Carlson and E. L. Kannenberg, *Glycobiology*, 2013, **23**, 820–832.

50 J. A. Hermoso, L. Lagartera, A. Gonzalez, M. Stelter, P. Garcia, M. Martinez-Ripoll, J. L. Garcia and M. Menendez, *Nat. Struct. Mol. Biol.*, 2005, **12**, 533–538.

51 J. Bonnet, C. Durmort, M. Jacq, I. Mortier-Barriere, N. Campo, M. S. VanNieuwenhze, Y. V. Brun, C. Arthaud, B. Gallet, C. Moriscot, C. Morlot, T. Vernet and A. M. Di Guilmi, *Mol. Microbiol.*, 2017, **106**, 832–846.

52 S. A. Ragland and A. K. Criss, *PLoS Pathog.*, 2017, **13**, e1006512.

53 P. A. Spears, E. A. Havell, T. S. Hamrick, J. B. Goforth, A. L. Levine, S. T. Abraham, C. Heiss, P. Azadi and P. E. Orndorff, *Mol. Microbiol.*, 2016, **101**, 714–730.

54 M. A. Karas, A. Turska-Szewczuk, M. Marczak, M. Jaszek, M. Janczarek, K. Dworaczek, D. Stefaniuk and J. Wydrych, *Int. J. Mol. Sci.*, 2018, **19**, 3510.

55 E. Bernard, T. Rolain, P. Courtin, A. Guillot, P. Langella, P. Hols and M. P. Chapot-Chartier, *J. Biol. Chem.*, 2011, **286**, 23950–23958.

56 J. Boutet, P. Blasco, C. Guerreiro, F. Thouron, S. Dartevelle, F. Nato, F. J. Canada, A. Arda, A. Phalipon, J. Jimenez-Barbero and L. A. Mulard, *Chemistry*, 2016, **22**, 10892–10911.

57 R. Marchetti, M. J. Dillon, M. N. Burtnick, M. A. Hubbard, M. T. Kenfack, Y. Bleriot, C. Gauthier, P. J. Brett, D. P. AuCoin, R. Lanzetta, A. Silipo and A. Molinaro, *ACS Chem. Biol.*, 2015, **10**, 2295–2302.