The *Legionella* collagen-like protein employs a unique binding mechanism for the recognition of host glycosaminoglycans

Saima Rehman[1], Anna K. Antonovic[2], Ian E. McIntire[3], Huaixin Zheng[3], Leanne Cleaver[1], Carlton O. Adams[3], Theo Portlock[1,4], Katherine Richardson[4], Rosie Shaw[4], Alain Oregioni[5], Giulia Mastroianni[2], Sara B-M. Whittaker[6], Geoff Kelly[5], Arianna Fornili[2*], Nicholas P. Cianciotto[3*] & James A. Garnett[1*]

[1]*Centre for Host-Microbiome Interactions, Faculty of Dental, Oral & Craniofacial Sciences, King's College London, London, UK*

[2]*School of Physical and Chemical Sciences, Queen Mary University of London, London, UK*

[3]*Department of Microbiology and Immunology, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA*

[4]*School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK*

[5]*The Medical Research Council Biomedical NMR Centre, the Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK*

[6]*School of Cancer Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK*

**Correspondence to:*

James Garnett, Centre for Host-Microbiome Interactions, Faculty of Dental, Oral & Craniofacial Sciences, King's College London, London, UK, email: james.garnett@kcl.ac.uk

Nicholas Cianciotto, Department of Microbiology and Immunology, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA, email: n-cianciotto@northwestern.edu

Arianna Fornili, School of Physical and Chemical Sciences, Queen Mary University of London, London, UK, email: a.fornili@qmul.ac.uk

Abstract

Bacterial adhesion is a fundamental process which enables colonisation of niche environments and is key for infection. However, in *Legionella pneumophila*, the causative agent of Legionnaires' disease, these processes are not well understood. The *Legionella* collagen-like protein (Lcl) is an extracellular peripheral membrane protein that recognises sulphated glycosaminoglycans (GAGs) on the surface of eukaryotic cells, but also stimulates bacterial aggregation in response to divalent cations. Here we report the crystal structure of the Lcl C-terminal domain (Lcl-CTD) and present a model for intact Lcl. Our data reveal that Lcl-CTD forms an unusual dynamic trimer arrangement with a positively charged external surface and a negatively charged solvent exposed internal cavity. Through Molecular Dynamics (MD) simulations, we show how the GAG chondroitin-4-sulphate associates with the Lcl-CTD surface via unique binding modes. Our findings show that Lcl homologs are present across both the Pseudomonadota and Fibrobacterota-Chlorobiota-Bacteroidota phyla and suggest that Lcl may represent a versatile carbohydrate binding mechanism.

Keywords: Lcl, *Legionella pneumophila*, adhesin, GAG, glycosaminoglycan, chondroitin-4-sulphate

Introduction

Legionella pneumophila is a Gram-negative bacterium that inhabits both natural and artificial freshwater systems. It thrives within a complex aquatic microbiome, which includes other biofilm associated bacterial species and cyanobacteria[1, 2]. It infects and replicates within amoebae and ciliates[3] but as an opportunistic pathogen it can also infect the lungs and causes Legionnaires' disease, and the self-limiting and milder Pontiac fever[4]. Infection occurs via inhalation of water droplets from contaminated sources, where it invades macrophages in the lungs and replicates intracellularly, resulting in pneumonia[5]. During infection *L. pneumophila* first binds the eukaryotic cell-surface, then after cell entry, it evades degradation through the formation of a specialised membrane bound replicative compartment, the *Legionella* containing vacuole (LCV)[6]. *L. pneumophila* utilises a type IV secretion system (T4SS/Dot/Icm) to transport >300 effectors directly into the host cytoplasm, which are key factors that drive LCV biogenesis and bacterial replication[7, 8]. In addition, *L. pneumophila* employs a type II secretion system (T2SS/Lsp) to export >25 substates/effectors out of the bacterium, and these play major roles in supporting the early stages of infection and during extracellular survival[9, 10, 11, 12, 13, 14, 15, 16, 17, 18].

We initially identified the *Legionella* collagen-like protein (Lcl) in a proteomic study of type-II dependent secretion in *L. pneumophila* strain 130b[11]. Subsequently, the *lcl* gene was detected in >500 other *L. pneumophila* strains examined, indicating that Lcl expression is a conserved trait of the *L. pneumophila* species[19, 20, 21]. Although limited in its broader prevalence within the *Legionella* genus, relative to that of other T2SS substrates, the *lcl* gene occurs in five out of 57 other *Legionella* species examined (i.e., *Legionella oakridgensis*, *Legionella nagasakiensis*, *Legionella hackeliae*, *Legionella quateirensis* and *Legionella tucsonensis*) and the majority of these are associated with human infection[9]. In addition to being detected in culture supernatants on multiple occasions[11, 19, 22], Lcl appears to also be a peripheral membrane bound protein and upon its secretion from *L. pneumophila* it is targeted to the bacterial surface[19, 23] and found in outer membrane vesicles[22]. Lcl is important for *L. pneumophila* auto-aggregation and biofilm formation[20, 23, 24, 25], although this precise mechanism

remains unclear. However, Lcl can also facilitate adhesion and cell entry of *L. pneumophila* to human lung epithelial (A549), lung mucoepidermoid (NCI-H292), and macrophage (U937) cell lines, and this indicates that Lcl has a fundamental role during infection of lung tissue[19, 20].

Lcl contains both an N-terminal region composed of collagen-like repeat (CLR) sequences, which are variable in length between different strains, and a C-terminal region with no overall sequence homology outside of the *Legionella* genus[11, 19, 26]. The C-terminal region of Lcl binds a range of sulphated glycosaminoglycan (GAG) polysaccharides that are present within the lung[20], including heparin and chondroitin-4-sulphate, while the collagen-like region has been shown to bind fucoidan[21], a heavily sulphated GAG found in many species of brown seaweed. GAGs are diverse linear carbohydrate structures that are formed from repeating disaccharide units of an amino sugar (*N*-acetylglucosamine or *N*-acetylgalactosamine) and glucuronic acid or galactose[27]. Sulphated GAGs exist as protein conjugates in the plasma membrane of nucleated cells and secreted into the extracellular matrix, and many bacterial pathogens including *L. pneumophila* use host GAGs as a means of adhesion during infection[28, 29, 30]. However, this is not well understood in *L. pneumophila* and there is a general lack in our structural and mechanistic understanding of cellular adhesion across the *Legionella* genus, which is a key step during colonisation and host invasion.

In this study, we report a structural model for full-length Lcl based on X-ray crystallographic, *in silico* modelling and nuclear magnetic resonance (NMR) spectroscopic data. We show that Lcl is also targeted to the surface of *L. pneumophila* 130b strain after its secretion, and this is mediated by its N-terminus, which forms an amphipathic helix. We present the crystal structure of the Lcl C-terminal domain (Lcl-CTD) which reveals an unusual trimer arrangement, and our structural and biochemical studies demonstrate a unique GAG binding mechanism. Our work provides a molecular understanding of how Lcl can recognise and interact with a broad range of GAG ligands and provides strong evidence for the role of Lcl in facilitating direct recognition of glycosaminoglycans in host tissue during *L. pneumophila* infection.

Results

T2SS-dependent Lcl is expressed on the surface of *L. pneumophila* strain 130b. Previously, immunoblot analysis identified Lcl in an outer membrane fraction of wild-type strain Philadelphia-1[19], and an immunofluorescence assay detected the protein on the surface of strain Lp02, which is a lab-generated derivative of Philadelphia-1[23]. Since Lp02 contains multiple point mutations and a large (~ 45-kb) deletion in the bacterial chromosome[31], we began this study by determining whether Lcl is also surface-exposed in wild-type strain 130b in addition to being present within its culture supernatants. To that end, Lcl from strain 130b (numbered 1 to 401 from the mature protein; ORF *lpw28961*; *lpg2644* in strain Philadelphia-1, *lpp2697* in strain Paris)[9, 11] was expressed recombinantly in *Escherichia coli*, purified, and then used to generate polyclonal anti-Lcl antibodies. In confirmation of our earlier work[11], immunoblot analysis revealed Lcl in the culture supernatants of strain 130b but not in the supernatants of either a T2SS (*lspF*) mutant or two newly created *lcl* mutants (**Fig. 1a**). By utilizing a whole-cell enzyme-linked immunosorbent assay (ELISA) method that had previously examined the location of another T2SS-dependent protein, ChiA[18], we determined that Lcl is in fact present on the surface of wild-type strain 130b but not the *lspF* mutant or *lcl* mutants (**Fig. 1b**).

Overall architecture of Lcl. We next turned our attention to the structural characterisation of Lcl. Using Multi-Angle Light Scattering (MALS), we determined a molecular mass of 123.4 ± 0.2 kDa (theoretical mass 42.3 kDa) for recombinant Lcl (**Fig. 2a)**, which supported Lcl being a trimer in solution. Inspection of the Lcl sequence from strain 130b indicated three defined regions: a collagen-like repeat (CLR) region (consensus repeat: GPQGLPGPKGD(K/R)GEA) and C-terminal region (CTD) which contains a domain of unknown function (DUF1566)[32], but also a 30 residue N-terminal helical region (N)[33] (**Fig. 2b, Supplementary Table 1**). Recombinant Lcl was analysed by rotary shadowing electron microscopy and inspection of the micrographs revealed a clear "lollipop-shaped" structure with a globular head and a stalk, consistent with a trimer of C-terminal domains and a triple helical collagen-like region, respectively (**Fig. 2c, Supplementary Fig. 1**). We also

examined the N-terminal region (Lcl-N) using neural network-based modelling and solution ^{1}H NMR nuclear Overhauser effect spectroscopy (NOESY), which suggested that this forms an amphipathic helix[34] (**Fig. 2b,d, Supplementary Fig. 2**). On addition of sodium dodecyl sulphate (SDS) peak broadening was observed in a NOESY spectrum, and this indicated that Lcl-N can bind micelles and this region may act as an extracellular membrane anchor for Lcl after its secretion (**Fig. 2b,e**).

We next probed the quaternary structure of the collagen-like region using standard multidimensional NMR. We designed a peptide that encompassed a consensus repeat sequence (Lcl-CLR) and observed a monomeric species that formed a polyproline II conformation in solution at 15°C (**Fig. 2b, Supplementary Fig. 3**). When an Lcl-CLR peptide containing uniformly ^{15}N labelled glycine residues was studied at 2°C, a ^{1}H-^{15}N heteronuclear single quantum correlation (HSQC) spectrum showed six glycine cross-peaks from the monomeric peptide, but also a higher molecular weight species containing at least 16 distinct glycine residues (**Fig. 2f**). This suggested that the Lcl-CLR peptide is in equilibrium between a monomeric and pseudo-symmetric trimer state under these conditions, containing six and 18 glycine residues, respectively. This was further supported by the comparison of cross-peaks in NOESY and rotating frame Overhauser effect spectroscopy (ROESY) spectra, where there were significant differences between NOE/ROE patterns for the higher molecular weight species at 2°C, which disappeared at 37°C (**Supplementary Fig. 4**). Analysis of intact Lcl using circular dichroism (CD) spectroscopy showed negative and positive peaks at 199 nm and 222 nm, respectively, which is in line with previous reports for Lcl from the Lp02 strain[21], and is indicative of a collagen-like structure (**Supplementary Fig. 5**). Furthermore, while monitoring the peak at 199 nm over increasing temperatures, we observed a two-stage unfolding process with T_m values of 38°C and 44°C. Together this further supports the CLR region of Lcl forming a triple helical structure in solution.

Overall structure of Lcl-CTD. As anticipated, analysis of the Lcl C-terminal domain (Lcl-CTD, residues 252 to 401) with MALS again revealed a stable trimer in solution (55.0 ± 0.1 kDa;

theoretical mass 18.6 kDa) and crystallographic studies were initiated. The structure of Lcl-CTD was determined using selenomethionine single wavelength anomalous dispersion (Se-SAD) phasing, with electron density maps refined to 1.9 Å (**Supplementary Table 2**). Lcl-CTD is composed of a trimer with disordered N-termini (Glu252 to Val270) that could not be modelled, with each domain having an identical conformation formed from two α-helices (H2, H3), two 3_{10}-helices ($3_{10}1$, $3_{10}2$) and nine β-strands (S1-S9) (**Fig. 3a,b, Supplementary Fig. 6**). Using small angle X-ray scattering (SAXS) we confirmed that the crystal structure is consistent with solution measurements and that the N-termini are highly dynamic, with an R_g value of 2.7 nm and a D_{max} of 9.7 nm (**Supplementary Fig. 7,8, Supplementary Tables 3,4**).

The Lcl-CTD structure is stabilised through the burial of an unusually small surface area per subunit (~14,000 $Å^2$) and this is due to a solvent accessible cavity permeating from the underside into the core of the trimer (**Fig. 3c**). Inter-subunit interactions are mediated by charge complementary (e.g. Asp316, Asp319, Arg342) and hydrophobic residues (e.g. Trp315, Ile321, Phe343) and while the internal surface contains negatively charged patches (e.g. Asp336, Glu368), the upper surface displays strong positive charge (e.g. Arg342, Lys369, Lys380, Lys385 and Lys391) (**Fig. 3d,e**). Using the DALI server[35] we established that the Lcl-CTD monomer is similar to C-type lectin-like domains found in snake venom toxins and bacterial invasins/intimins[36, 37, 38, 39]. However, Lcl-CTD lacks the expected motifs required for Ca^{2+}/carbohydrate and integrin/Tir binding (**Supplementary Fig. 9**) and we could not identify any trimeric structures that share tertiary homology.

The DUF1566/pfam07603[32] domain is found in diverse proteins from a wide range of prokaryotes. DUF1566 is also located between residues 314 to 399 of Lcl-CTD and is composed of the H2, $3_{10}1$, H3 and $3_{10}2$ helices, and S5-S9 strands (**Supplementary Fig. 10**). With truncation of the S1-S4 β-strands almost all inter-subunit interactions are still maintained, but with the depth of the internal cavity of Lcl-CTD greatly reduced. Highly conserved residues in the DUF1566 domain are largely located within the core of Lcl-CTD, with just three residues located on the surface: Trp315 at the inter-domain interface, and Arg338 and Glu344, which form an intra-domain salt

bridge within the internal cavity. While a generic role for the DUF1566 domain is not clear, based on Lcl it could act in carbohydrate recognition and/or promote trimer formation. Further examination of DUF1566 containing proteins that also possess a collagen-like repeat region (gly_rich_SclB superfamily) shows Lcl actually belongs to a larger family, with homologs identified in *Legionella bononiensis* and *Legionella longbeachae* from the *Legionella* genus, but also in species across the Pseudomonadota phylum (*Comamonas sp.*, *Methylomonas paludism*, *Methylobacter sp.*, *Thiocystis minor*), and the Fibrobacterota-Chlorobiota-Bacteroidota (FCB) superphylum (*Bacteroidetes bacterium*, *Bacteroidia bacterium*, *Candidatus Fluviicola riflensis*, *Chitinophagaceae bacterium*, *Flavobacteria bacterium*, *Formosa sp.*, *Fluviicola sp.*, *Nonlabens sp.*, *Oceanihabitans sediminis*, *Psychroflexus planctonicus*, *Winogradskyella pacifica*, and *Winogradskyella wichelsiae*) (**Supplementary Table 5**).

GAGs bind the charged surface of Lcl-CTD. Chondroitin is composed of repeating disaccharide units of [-4)GlcA(*β*1-3)GalNAc(*β*1-]$_n$ (GlcA: D-glucuronate; GalNAc: *N*-acetyl-D-galactosamine), with C4S sulphated at the C4 position of GalNAc[27]. Heparin is formed of repeating disaccharide units of [-4)IdoA(β1-4)GlcN(β1-]$_n$ (IdoA: L-iduronate; GlcN: D-glucosamine) and is highly sulphated, with sulphation at the 2O position of IdoA (IdoA(2S)) and the 6O and N positions of GalNAc (GlcNS(6S)) being the most common form[27]. Both C4S and heparin are abundant in the lung[40] and have variable molecular weights that range between ~5-50 kDa, which equates to ~15-135 disaccharide repeats in each GAG chain. Intact Lcl was previously shown to recognise a range of variable length commercially prepared sulphated GAGs, including chondroitin-4-sulphate (C4S) and heparin (**Fig. 4a**), with the isolated C-terminal domain of Lcl also showing binding to heparin[20]. We therefore attempted to crystallise Lcl-CTD in the presence of defined C4S (GlcA/GalNAc(4S)) and heparin (IdoA(2S)/GlcNS(6S)) fragments with 4, 6 and 8 disaccharide repeats (degree of polymerisation; dp4, dp6, dp8, respectively) but were unsuccessful. Nonetheless, we did identify a new crystal form of Lcl-CTD grown from high concentrations of ammonium sulphate and solved its structure using molecular replacement and refined electron density maps to 1.9 Å (**Supplementary Table 2**). The two trimer structures are highly similar (Root Mean Square Deviation (RMSD) over all C_α atoms of 0.3)

(**Supplementary Fig. 11**) but in the new form, two sulphate ions were also observed on the surface bound to residues Lys369 and Lys391 (**Fig. 3e**). As GAG binding sites are usually formed from clefts or relatively flat positively charged patches[41], we speculated that Lys369 and Lys391, along with the adjacent Arg342, Lys380 and Lys385 residues, may recognise the negatively charged sulphate groups that decorate GAG polymers.

To assess this GAG binding model, we created constructs carrying R342A, K369A, K380A, K385A or K391A mutations (Lcl-CTD^{R342A}, Lcl-CTD^{K369A}, Lcl-CTD^{K380A}, Lcl-CTD^{K385A} and Lcl-CTD^{K391A}, respectively) which we anticipated would abrogate binding to GAGs. In addition, we also created constructs carrying a D386A mutation (Lcl-CTD^{D386A}) located on the Lcl-CTD surface, and a E368A mutation (Lcl-CTD^{E368A}) within the internal cavity, which we expected would not affect binding. Using SAXS, all constructs except for Lcl-CTD^{R342A} produced scattering profiles like wild-type Lcl-CTD, confirming that they were still correctly folded (**Supplementary Fig. 12, Supplementary Table 3**). The R342A mutation resulted in dissociation of the trimer, and it was therefore not used for subsequent analysis. We then assessed the ability of the correctly folded mutants to bind immobilised commercially prepared C4S and heparin extracted from bovine trachea and porcine intestinal mucosa, respectively, using an ELISA method. As anticipated, constructs carrying the K369A, K380A, K385A or K391A mutations all displayed a significant reduction in their ability to bind these GAGs when compared with wild-type Lcl-CTD, while the D386 mutation showed no difference. However, the E368 mutation resulted in higher binding capacity (**Fig. 4b**).

C4S binds Lcl-CTD across multiple domains. Although the structures of C4S and dermatan sulphate differ in just the location of hydroxyl and carboxyl groups at the C2 and C5 positions of D-glucuronate and D-iduronate, respectively, Lcl does not bind dermatan sulphate[20, 42]. In an attempt to understand this specificity, we started by using solution NMR spectroscopy to investigate interactions between Lcl-CTD and C4S. Using a partially deuterated sample and multidimensional transverse relaxation-optimised spectroscopy (TROSY) NMR we were able to assign 61% of the potential amide backbone resonances of Lcl-CTD (**Supplementary Fig. 13)**. Most missing assignments were

238 located at the N-terminus (Glu252 to Ser275), the H2 helix (Asp316 to Asn323; positioned at the
239 inter-domain interface), and the adjacent S3-S4 loop and S4 strand (Val303 to Ser311) (**Fig. 5a**).
240 Furthermore, many peaks displayed variable intensity and ~10% of residues were present in
241 multiple conformational states (**Supplementary Fig. 13)**. Together, this indicated that the Lcl-CTD
242 trimer experiences conformational exchange and is a dynamic structure. We then compared ^{1}H-^{15}N
243 TROSY spectra of Lcl-CTD titrated against increasing concentrations of commercially prepared
244 C4S and observed significant broadening that approached saturation at 0.5 mg/ml C4S
245 (**Supplementary Fig. 14**). Although no reliable data could be measured for Lys369, Lys380, Lys385
246 and Lys391 due to spectral overlap, significant peak broadening was observed for the neighbouring
247 residues Thr381 and Thr392 (**Fig. 5b,c,d**). Moreover, substantial broadening was also detected for
248 residues adjacent to Lys369 (Tyr292, Thr313, Trp315, His326, Arg342, Met350).

249 We next carried out molecular docking with HADDOCK[43, 44], using monomeric and trimeric Lcl-CTD
250 and dp4, dp6, dp8 and dp10 C4S oligosaccharides as starting structures, and ambiguous interaction
251 restraints (AIRs) derived from the GAG binding ELISA and NMR chemical shift perturbations (CSP).
252 Docking between monomeric Lcl-CTD and C4S dp8 (cluster one of the three major clusters)
253 produced models consistent with the experimental data (**Supplementary Fig. 15**), and a trimer bound
254 with one molecule of C4S dp8 was then created (HM model) and further examined using MD (**Fig.**
255 **6a, Supplementary Fig. 16**). Docking between trimeric Lcl-CTD (HT1 and HT2 models) caused
256 changes to the trimer interface and they were unstable during Molecular Dynamics (MD) simulations
257 (**Supplementary Fig. 16**), and these were not taken forward for further analysis. MD simulations
258 were also run on Lcl-CTD (residue 271 to 401) alone (**Supplementary Fig. 17**). Analysis of the Root
259 Mean Square Fluctuation (RMSF) profiles indicated a high flexibility for the first 7 residues of each
260 monomer (Asp217 to Ile223), consistent with the disordered nature of the adjoining N-terminal part
261 of the chain. However, the overall trimeric structure of the domain was stable throughout the
262 simulations, with RMSD from the initial structure quickly reaching a plateau and staying below 2 Å
263 in all the replicas.

While the overall structure of Lcl-CTD was found to be stable in all replicas of simulations run on the HM model, C4S displayed highly dynamic binding to the Lcl-CTD surface (**Supplementary Fig. 16**).

Although during the simulations C4S remained in contact with Lcl-CTD for much of the time, the different components of the glycan frequently detached and then reattached to different regions of the Lcl-CTD surface. From its starting position, during the simulations the polysaccharide either remained in the same region or explored other parts of the top surface of the protein. As indicated by the spatial distribution function (sdf) of C4S sulphur atoms (**Fig. 6b**) and the frequency of contacts between C4S and the Lcl-CTD trimer, (**Fig. 6c**), C4S more often bound to the central region of the top surface.

Structures from all the replicas were clustered using an optimised cut-off of 17.5 Å on the pairwise C_α RMSD values, the high value reflecting the variety of binding poses explored by C4S in the different replicas. Three major C4S binding modes were identified, and although they had a relatively high RMSD, they broadly reflected a preference for Lcl-CTD surface localisation of the bound glycan chain, although the clusters did not reflect a preference in orientation (**Fig. 7a**). After considering the 3-fold symmetry of the system, the first and third mode were found to be closely related, with C4S dp8 showing a similar position in the two modes, and these were therefore combined. The first and major binding mode is the most frequently observed (M-BM; 63% frequency) and represents C4S binding to the top, central region of Lcl-CTD, along the chain A/C interface. The second and minor binding mode (m-BM; 20% frequency) represents C4S binding primarily to Lcl-CTD chain A and resembles the initial input HM model (**Fig. 6a**).

A more detailed analysis of the distance and interactions between C4S and Lcl-CTD highlighted that C4S can binds across one (36% frequency), two (35% frequency) or all three (28% frequency) Lcl-CTD chains (**Fig. 7b, Supplementary Fig. 18, Supplementary Table 6,7**). While binding of C4S to a single chain of Lcl-CTD is observed in both the major and minor binding modes, binding across

multiple Lcl-CTD chains largely reflects the major binding mode alone. Scrutiny of these different complex formations indicates that the Lcl-CTD residues found to be most frequently involved in hydrogen bonding with C4S are Ser371, Ser390 and K391, which are located at the central region of the top surface (**Fig. 6c, Supplementary Table 7**). On the other hand, we observed that C4S dp8 binds to Lcl-CTD using 4 to 6 saccharide units (GalNAc(4S) and GlcA), either as a continuous stretch or with the glycan looped out in the middle of the chain, and forms hydrogen bonding interactions through its carboxylates, amides, sulphates, and hydroxyl groups. Moreover, examination of the representative structures of C4S dp8 binding across 1-, 2-, and 3-chains of Lcl-CTD suggests that the replacement of D-glucuronate with D-iduronate would result in the disruption of some hydrogen bond interactions (**Fig. 7b**). It would also require changes in the glycan conformation to avoid clashes within dermatan sulphate and between dermatan sulphate and Lcl-CTD, and together this provides at least some explanation for the selectivity of Lcl-CTD for different GAGs.

Discussion

Adhesion is a fundamental process in bacteria and adhesin proteins often work in synergy to enable colonisation of niche environments. Several adhesins have been identified in *L. pneumophila* that play important roles in the recognition of eukaryotic hosts, and these include the type IV pilus (T4P)[45] and its associated PilY1 pilus tip adhesin[46, 47], Hsp60[48], RtxA[49], MOMP[50], LaiA[51], and Lcl[19]. We have determined that Lcl is a trimeric structure formed of three regions: an N-terminal helix, an elongated collagen-like region, and a DUF1566 containing C-terminal region. We previously observed Lcl secreted in bacterial culture supernatants of *L. pneumophila* strain 130b [11]. Subsequently Lcl was detected on the bacterial surface in strains Philadelphia-1[19] and Lp02[23], and we have now shown this in 130b. The *L. pneumophila* T2SS exports >25 proteins, and three of these associate with host organelles and/or the bacterial surface upon their secretion (i.e., ChiA, Lcl, ProA); we previously observed ChiA and ProA tethered to the LCV membrane[52] and ChiA to the outer membrane surface[18].

Although the mode of ProA membrane binding is unclear, ChiA binds the *L. pneumophila* surface using its N3 domain, formed of a fibronectin III module domain-like fold[18]. In addition, we have shown that NttA binds phosphatidylinositol-3,5-bisphosphate (PtdIns(3,5)P_2) and other phosphorylated phosphoinositides, which indicates that NttA may also be targeted to host organelles during infection[17]. In this study we have revealed that Lcl is likely binds the bacterial surface via an amphipathic helix motif at its N-terminus, which represents a new mechanism that has not been observed for other T2SS substrates (**Fig. 2d,e**).

Bacterial collagen-like proteins have been identified in a wide range of gram-positive and gram-negative bacteria[53]. A defining feature of collagen is the presence of Gly-X-Y repeats, where in eukaryotes X and Y are often proline and hydroxyproline, respectively, with hydroxyproline mediating inter-chain hydrogen bonding to stabilise the triple helical structure. However, bacteria cannot make hydroxyproline, and their collagen-like structures do not have a requirement for proline. Instead, they contain a higher proportion of charged/polar residues, and these are predicted to interact across different chains[54, 55]. As both bacterial and eukaryotic collagens display similar thermal stabilities (T_m ~35-39°C and ~37°C, respectively)[54], it is unclear why eukaryotic systems do not produce bacterial-like collagen, although these structures may be unfavourable for the formation of higher-order fibrils, which are not observed in prokaryotes.

From our examination, Lcl from *L. pneumophila* strain 130b contains 12 repeats of a consensus 15 residue sequence (GPQGLPGPKGD(K/R)GEA) within its collagen-like region. Analysis of Lcl from other strains isolated from clinical samples, the environment, and hot springs, however, has demonstrated a high polymorphism within this region, with Lcl from Philadelphia-1 containing 19 repeats[19]. While hot spring isolates (≥ 40°C) displayed a preference for 13 repeats, clinical and environmental isolates (≤ 37°C) were bimodal with a preference for both 8 and 13/14 repeats. Using a 19-residue peptide encompassing a single Lcl CLR consensus sequence, we observed this peptide to be largely monomeric, but able to form triple helical structures with a reduction in temperature. Together this suggests that variability in Lcl repeats may reflect the minimal length of collagen

required for Lcl to retain its folding under different environmental temperatures. Furthermore, our CD spectroscopy analysis showed that Lcl has two melting temperatures of 38°C and 44°C (**Supplementary Fig. 5**), which are similar to other reported bacterial and eukaryotic collagens[54], but also highlight the additional role of the C-terminal domain in stabilising the collagen-like region.

Lcl has been shown to mediate adhesion/invasion of *L. pneumophila* to a range of host cell types. In one study, a *lcl* mutant displayed a ~30% reduction in binding to NCI-H292 lung mucoepidermoid cells, compared with the wild-type Lp02[20]. In another study, incubation of wild-type Philadelphia-1 with Lcl antibodies resulted in a ~50% drop in binding to A549 lung epithelial cells, and 0-30% drop in binding to U937 macrophage cells, although no difference in binding was observed with the amoeba *Acanthamoeba castellanii*[19]. Specifically, Lcl binds to sulphated GAGs on the surface of host cells and both the collagen-like and C-terminal regions have been implicated here[19, 20, 21]. When *lcl* containing 14 or 19 repeats was expressed in Philadelphia-1, there was an increase in adhesion/invasion of A549 cells with 14 compared with 19 repeats, but the opposite was observed with U937 cells[19]. Fucoidan has also been shown to bind the collagen-like repeat region of Lcl from Lp02 with a higher affinity than to the C-terminal domain, and increasing the number of CLRs has been correlated with tighter binding[21]. Together this suggests that the C-terminal domain plays a general role in the recruitment of ligands, but at least for some GAGs, synergistic binding along the collagen-like region can provide further increases in overall affinity.

The C-terminal domain of Lcl is highly conserved (>97%) across different strains of *L. pneumophila* and we have determined that it forms a unique trimer structure with a deep negatively charged internal cavity and positively charged external surface (**Fig. 3**). Intact Lcl from Lp02 binds fucoidan with 10-fold higher affinity than C4S (K_D 18 nM and 173 nM, respectively)[42], and this likely reflects the increased level of sulphation in fucoidan. Using mixed chain length heparin and C4S, we have demonstrated that the strong positive charge on the Lcl-CTD surface is important for glycan recognition (**Fig. 4b**). Furthermore, using MD simulations, and focussing on binding to a dp8 structure of C4S, we have identified two predominant binding modes for this ligand (**Fig. 7a**). A

major mode (M-BM) which runs across the middle of the Lcl-CTD trimer, and a minor mode (m-BM) which is largely localised to a single chain of the trimer. Only Lys385 and Lys391 form hydrogen bonds with C4S during the simulations, with relatively low frequency, while Lys369, Lys380, Lys385 and Lys391 are all within proximity (**Fig. 6b**, **Supplementary Table 6,7**). This indicates that the primary role of these lysine residues is to provide general electrostatic attraction for the GAGs, rather than specific recognition. Furthermore, during the simulations the majority of the Lcl-CTD upper surface is involved in sampling dp8 C4S, mainly through serine, threonine, and asparagine residues (**Fig. 6c**), and this may reflect a mechanism that enables recognition of a broad range of ligands. Nonetheless, residues at the inter-trimer interface (Ser371, Ala372, Lys391) also appear to have a more targeted role in binding, although it is not clear whether these are specific for C4S or represents a more general GAG binding mode. As C4S and other GAGs can contain up to ~135 disaccharide repeats it is feasible that one or more glycan chains could bind simultaneously at multiple sites on the Lcl-CTD surface (**Fig. 7c**). Using NMR with mixed chain length heparin and C4S, we also observed major CSPs for residues on the side of the Lcl-CTD trimer (i.e., Arg342, Met350) (**Fig. 5d**), although this binding was not observed during the MD simulations using C4S dp8. Arg342 and Met350 could represent a lower affinity site that is only occupied once GAGs have bound to the top surface of Lcl-CTD but may facilitate single GAG chain binding between the C-terminal domain and the collagen-like repeat region. Alternatively, as these residues are located at the domain interface and our NMR analysis has highlighted conformation exchange broadening at the inter-trimer interfaces (**Fig. 5a**), these CSPs may instead reflect indirect binding events such as stabilisation of the Lcl-CTD trimer upon association with GAGs.

We have now identified the *lcl* gene in eight *Legionella* species, however, except for *L. quateirensis* we see little conservation in the surface lysine residues that are present in 130b (**Supplementary Fig. 19**). Although this suggests that Lcl-CTD from these other species will not exhibit a large positively charged surface, we do see conservation of other key C4S binding residues mainly located at the inter-trimer interface (i.e. Ser370, Ser371, Ala372, Asn373, Asn378). This may indicate that the Lcl C-terminal domain has different glycan specificity outside of *L. pneumophila*. Lcl is also known to

mediate auto-aggregation and biofilm formation of *L. pneumophila* in the presence of divalent cations[20, 25] and it has been suggested that trimeric Lcl from Lp02 can form higher-order structures which could function in clumping adjacent bacteria[21]. However, these observations were independent of divalent cations, and our biophysical characterisation of Lcl from 130b shows that it is extremely stable and homogenous (**Fig. 2a,c, Supplementary Fig. 5**). Glu368 is highly conserved across the *Legionella* genus and is located within the internal cavity, where three residues are in proximity perpendicular to the trimer 3-fold axis (**Fig. 3c, Supplementary Fig. 19**). When Glu368 was mutated to alanine, we observed a significant increase in binding of Lcl-CTD to both heparin and C4S (**Fig. 3c**), which could be explained by this mutation stabilising the trimer and priming it for GAG recognition. However, we also speculate that Glu368 may bind divalent cations and have a role in modulating the biofilm activity of Lcl, but further studies are needed.

Materials and Methods

Bacterial strains and media. All bacterial strains, plasmids and primers used in this study are listed in **Supplementary Table 8**. *L. pneumophila* strain 130b (American Type Culture Collection [ATCC] strain BAA-74; also known as strain AA100 or Wadsworth) served as wild type and parent for all mutants[56]. The *L. pneumophila lspF* mutant NU275 used here was also previously described[57]. These strains and all newly made mutants (below) were routinely grown at 37°C on buffered charcoal yeast extract (BCYE) agar or in buffered yeast extract (BYE) broth[58]. Isotopically defined M9 minimal medium (pH 7.4) contained (per litre) 6.0 g $Na_2HPO_4{\cdot}7H_2O$, 3 g KH_2PO_4, 0.5 g NaCl, 0.12 g $MgSO_4{\cdot}7H_2O$, 22 µg $CaCl_2$, 40 µg thiamine, 8.3 mg $FeCl_3{\cdot}6H_2O$, 0.5 mg $ZnCl_2$, 0.1 mg $CuCl_2$, 0.1 mg $CoCl_2{\cdot}6H_2O$, 0.1 mg H_3BO_3 and 13.5 mg $MnCl_2{\cdot}6H_2O$, supplemented with 2 g [U-$^{13}C_6$]glucose and/or 0.7 g $^{15}NH_4Cl$. M9 media was made up in deuterium oxide for the production of perdeuterated protein samples and pH was adjusted using 1 M NaOH solution. All NMR isotopes were from Sigma (UK).

Mutant construction. To make *L. pneumophila* mutants (i.e., strains NU468, NU469) that have a nonpolar, unmarked deletion within the *lcl* gene, we employed overlap extension PCR (OE-PCR) followed by allelic exchange, as before[13, 59]. DNA fragments of the 5' and 3' regions flanking the *lcl* ORF were PCR-amplified from 130b DNA using the primer pairs *lcl*-UpF and *lcl*-UpR for 5' *lcl*, and *lcl*-DownF and *lcl*-DownR for 3' *lcl*. A kanamycin (Kn)-resistance gene flanked by Flp recombination target sites was PCR-amplified from the vector pKD4[60] using the primers *lcl*-P1 and *lcl*-P2. We then performed two-step OE-PCR to combine the 5' and 3' regions of *lcl* with the respective Kn-resistance cassette. A PCR product matching the correct target size was gel purified and ligated into pGEM-T Easy (Promega) to yield plcl::Kn. After transforming strain 130b with the newly made plasmid, bacteria containing an inactivated *lcl* gene was obtained by plating on BCYE agar containing Kn. Confirmation of the mutated *lcl* gene were done by PCR using the above-mentioned primers. Following transformation with pBSFLP, which encodes a Flp recombinase along with a gentamicin-resistance marker[60], mutants harboring the desired unmarked deletion and lacking pBSFLP were recovered by plating on BCYE agar containing 5% (w/v) sucrose and scored for loss of resistance to both Kn and gentamicin, as before[61]. The mutants were verified by sequencing of PCR amplicons.

Immunoblot analysis of bacterial culture supernatants. Wild-type and mutant *L. pneumophila* strains that had been grown for three days on BCYE agar were suspended into BYE broth to an OD_{660} of 0.3 and grown overnight at 37°C to an OD_{660} of ~1.5. Supernatants were obtained by centrifugation, sterilized by passage through 0.2-µm filters (EMD Millipore), and then concentrated, as before[52]. Following dilution in SDS–loading buffer, the samples were subjected to PAGE and immunoblot analysis. To that end, purified recombinant Lcl protein (above) was submitted to Lampire Biological Laboratories (Pipersville, PA) at a concentration of 2 mg/ml for the production of rabbit polyclonal antisera, analogously to what we had been done before for other secreted proteins of *L. pneumophila*[52]. Following an overnight incubation at 4°C in 5% BSA (w/v) - Tris-buffered saline (TBST), the blot was incubated overnight at 4°C with the primary anti-Lcl antiserum at 1:1,000 in 5% BSA-TBST. After four, 10-min washes with the TBST buffer, the membrane was further incubated

for 1 h at room temperature with secondary goat anti-rabbit horseradish peroxidase antibody (Cell Signaling Technology) at 1:10,000 in 5% BSA-TBST. Finally, after another series of washes, the blot was developed using Amersham ECL Prime reagent and exposed to X-ray film, as before[13, 18].

Bacterial whole-cell ELISA. The assay for detecting protein on the surface of *L. pneumophila* was done as previously described for the detection of ChiA, another substrate of the *L. pneumophila* T2SS[18]. In the present case, the bacterial strains were grown on BCYE agar for 3 days at 37°C, and the whole-cell ELISA utilised anti-Lcl antibodies at 1:1,000 dilution.

Construction of recombinant expression plasmids. Intact *lcl* (residues 1–401) and its C-terminal fragment (residues 252–401) were amplified by PCR from *L. pneumophila* 130b gDNA using primer pairs RLC1/ RLC2 and RLC3/ RLC4, respectively (**Supplementary Table 8**). These were then cloned into the pET-46 Ek/LIC vector (Novagen) using ligation independent cloning. Synthetic genes gRLCm1 to gRLCm7 were synthesised by Synbio Technologies (USA) and cloned into pET28b vector using NcoI and XhoI restriction sites to create plasmids pRLCm1 to pRLCm7, respectively (**Supplementary Tables 8,9**).

Protein purification. Intact Lcl and Lcl-CTD were expressed in *E. coli* strain BL21(DE3) (New England Biolabs) grown in either LB media, minimal media supplemented with selenomethionine (Molecular Dimensions), minimal media containing 0.07% (w/v) $^{15}NH_4Cl_2$ (Cambridge Isotope Laboratories), 100% (v/v) D_2O (Sigma) or minimal media containing 0.07% (w/v) $^{15}NH_4Cl_2$, 0.2% (w/v) [^{13}C]glucose (Cambridge Isotope Laboratories), 100% (v/v) D_2O. Expression was induced with 0.5 mM isopropyl-d-1-thiogalactopyranoside (IPTG) at an OD_{600nm} of 0.6 and cells were harvested after growth overnight at 18°C. Samples were purified using nickel-affinity chromatography followed by gel filtration using a Superdex-200 gel-filtration column (GE Healthcare), equilibrated in 20 mM Tris-HCl pH 8.0, 200 mM NaCl. To ensure efficient back exchange of amide protons, perdeuterated Lcl-CTD samples were initially purified in the presence of 8 M urea and then after nickel-affinity chromatography they were refolded by dialysis against 20 mM Tris–HCl pH 8, 200 mM NaCl, 1 M urea, 5 mM ethylenediaminetetraacetic acid (EDTA) and then 20 mM Tris–HCl pH 8, 200 mM NaCl.

Engineered Lcl-CTD carrying R342A, E368A, K369A, K380A, K385A, D386A and K391A mutations in the *lcl-CTD* gene (Lcl-CTD^{H326A}, Lcl-CTD^{R342A}, Lcl-CTD^{R368A}, Lcl-CTD^{K369A}, Lcl-CTD^{K380A}, Lcl-CTD^{K385A}, Lcl-CTD^{D386A} and Lcl-CTD^{K391A}, respectively) were purified as wild-type Lcl-CTD.

SEC-MALS. Lcl or Lcl-CTD were injected onto a Superose 6 Increase 10/300 column (GE Healthcare) coupled to a Wyatt Technology system and run in 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.5, 200 mM NaCl. BSA was run as a monodisperse reference protein. A dn/dc value of 0.185 ml/g was used for molecular weight calculations and data analysis was performed with Astra V software.

Rotary shadowing electron microscopy. The overall structure of Lcl was analysed using transmission electron microscopy after rotary shadowing using an adapted mica sandwich technique[62, 63]. Five µl of Lcl in 20 mM HEPES pH 7.5 (5 µg/ml) was sprayed on a freshly cleaved mica sheet, allowed to adsorb, and then washed with ultrapure water. The mica was mounted on the stage of a Polaron Freeze fracture instrument and then freeze dried at -100°C. The temperature was lowered to -150°C for shadowing with Pt/C on a low angle (5°) and a carbon backing layer was added for support. These were removed from the mica in distilled water and placed on 400 mesh copper grids. Micrographs were taken with a JEM 1230 transmission electron microscope operated at 80 kV.

Peptide modelling. Modelling of Lcl-N was carried out using the sequence for Lcl residues 1-30 from *L. pneumophila* 130b strain with AlphaFold2[64]. Sequence alignments and templates were generated through MMseqs2[65] and HHsearch[66], and run through the ColabFold notebook[67]. No prior template information was provided, and sequences used during modelling were both paired from the same species and unpaired from multiple sequence alignment.

Peptide NMR. All peptides were synthesised by ThermoScientific to >95% purity. Unlabelled Lcl-N peptide (KSNPASQAYVDGKVSELKNELTNKINSIPS-NH_2) was resuspended to 1 mM in

25 mM $NaPO_4$ pH 6.5, 100 mM NaCl, 10% (v/v) D_2O with or without 80 mM perdeuterated d_{25}-SDS, and ^{1}H-^{1}H NOESY spectra (200 ms mixing time) were recorded at 25°C on a 700 MHz Bruker Avance III HD equipped with cryoprobe. Unlabelled Lcl-CLR peptide (Ac-EAGPQGLPGPKGDRGEAGP-NH_2) and Lcl-CLR peptide containing uniformly ^{15}N labelled glycine residues (Ac-EAGPQGLPGPKGDRGEAGP-NH_2; labelled positions underlined) were resuspended to 3 mM in 20 mM HEPES pH 6.0, 50 mM NaCl, 10% (v/v) D_2O. Peptides were then incubated at 90°C for 15 min and then 4°C for one week. Full backbone and side chain assignments for the monomeric unlabelled peptide was achieved using standard double-resonance peptide assignment experiments (^{1}H-^{15}N HSQC, ^{1}H-^{13}C HSQC, ^{1}H-^{13}C total correlation spectroscopy (TOCSY), ^{1}H-^{1}H TOCSY, ^{1}H-^{1}H correlation spectroscopy (COSY), ^{1}H-^{1}H ROESY with 200 ms mixing time) recorded at 15°C on a 700 MHz Bruker Avance III HD equipped with cryoprobe. In addition, ^{1}H-^{1}H ROESY (200 ms mixing time), ^{1}H-^{1}H NOESY (240 ms mixing time) and ^{1}H-^{15}N HSQC spectra were recorded at 2°C, and a ^{1}H-^{1}H NOESY spectrum (240 ms mixing time) was recorded at 37°C, on an 800 MHz Bruker Avance III HD equipped with cryoprobe. All spectra were processed using NMRPipe[68] and analysed using ANALYSIS[69]. Secondary structure propensity of the monomeric Lcl-CLR peptide at 15°C were calculated using the δ2D server, providing C_α, C_β, H_α, N, NH backbone chemical shifts[70].

Circular dichroism. Far-UV CD spectra were measured in a Chirascan (Applied Photophysics) spectropolarimeter thermostated at 10°C. Spectra for Lcl (0.05 mg/ml) in 10 mM HEPES pH 8.0 was recorded from 260 to 195 nm, at 0.5 nm intervals, 1 nm bandwidth, and a scan speed of 10 nm/min. Three accumulations were averaged for each spectrum. For thermal denaturation experiments, Lcl (0.05 mg/ml) in 10 mM HEPES pH 8.0 was recorded at 199 nm between 10°C and 75°C in 1°C increments. Each increment was recorded in triplicate and then averaged.

Crystal structure determination. Selenomethionine labelled Lcl-CTD (15 mg/ml) and native Lcl-CTD (20 mg/ml) in 20 mM Tris-HCl pH 8.0, 200 mM NaCl, 20 mM EDTA were crystallised using the sitting-drop vapour-diffusion method grown at 20°C in either 2.0 M $(NH_4)_2SO_4$, 0.1 M Bis-Tris pH 6.5 or 20% (v/v) glycerol, 20% (w/v) polyethylene glycol (PEG) 4000, 30 mM $NaNO_3$, 30 mM

Na_2HPO_4, 30 mM $(NH_4)_2SO_4$, 100 mM Bicine, 100 mM Tris pH 8.5, respectively. Crystals were briefly soaked in well solution complemented with additional 30% or 10% (v/v) glycerol, respectively, before flash freezing in liquid nitrogen. Diffraction data were collected at 100K at station I03 of the Diamond Light Source (DLS), United Kingdom. Data were processed using XDS and scaled with AIMLESS, within the XIA2 pipeline[71, 72, 73]. For labelled Lcl-CTD, two selenomethionine sites were located in each Lcl-CTD molecule using SHELXD[74] and then phases were calculated using autoSHARP[75]. After automated model building with ARPWARP[76], the remaining structure was manually built within Coot[77]. Refinement was carried out with REFMAC[78] using non-crystallographic symmetry (NCS) and translation-libration-screw (TLS) groups, and 5% of the reflections were omitted for cross-validation. For native Lcl-CTD, molecular replacement was carried out in PHASER[79] using a single chain of Lcl-CTD as the search model. Refinement was again carried out with REFMAC[78] using non-crystallographic symmetry (NCS) and translation-libration-screw (TLS) groups, and 5% of the reflections were omitted for cross-validation. Both structures were run through PDBREDO[80] as a final step of refinement. Processing and refinement statistics of the final model can be found in **Supplementary Table 2**.

SEC-SAXS. Data were collected on beamline at the Diamond Light Source (DLS), UK[81]. 60 µl of WT Lcl-CTD, Lcl-CTD^{R342A}, Lcl-CTD^{E368A}, Lcl-CTD^{K369A}, Lcl-CTD^{K380A}, Lcl-CTD^{K385A}, Lcl-CTD^{D386A} and Lcl-CTD^{K391A} (5 mg/ml) in 20 mM Tris–HCl pH 8, 200 mM NaCl, 5 mM EDTA were applied to a Shodex KW403-4F column at 0.16 ml/min and SAXS data were measured over a momentum transfer range of $0.004<q<0.44$ Å^{-1}. Peak integration and buffer subtraction were performed in CHROMIXS[82]. The radius of gyration (R_g) and scattering at zero angle (I(0)) were calculated from the analysis of the Guinier region by AUTORG[83]. The distance distribution function (P(r)) was subsequently obtained using GNOM[83], yielding the maximum particle dimension (D_{max}). The disordered N-terminus of the Lcl-CTD crystal structure was built using MODELLER[84] and refinement of the N-terminus in the intact model against the corresponding SAXS curve was carried out with EOM2[85] with fixing of the ordered domains in the trimer. Processing and refinement statistics can be found in **Supplementary Tables 3,4**.

GAG binding ELISA. Immulon 2-HB 96-well plates (VWR) were coated overnight at 4°C with either 50 µl of heparin from porcine intestinal mucosa (Sigma) or chondroitin-4-sulfate from bovine trachea (Sigma) at 100 µg/ml in 50 mM carbonate/bicarbonate pH 9.6. Wells were blocked for 1 hr at 25°C with 200 µl of 0.1% (w/v) bovine serum albumin (BSA) in PBS–0.05% (v/v) Tween 20 and then washed once with 200 µl of incubation buffer (0.05% (w/v) BSA in PBS–0.05% (v/v) Tween 20). Wells were then incubated for 3 hrs at 25°C with 50 µl of WT Lcl-CTD, Lcl-CTD^{E368A}, Lcl-CTD^{K369A}, Lcl-CTD^{K380A}, Lcl-CTD^{K385A}, Lcl-CTD^{D386A} or Lcl-CTD^{K391A} at 10 µM in incubation buffer. This was followed by four washes with 200 µl of incubation buffer and incubation with 50 µl of anti-His-HRP antibody (Sigma), diluted 1:2000 in incubation buffer for 1 hr at room temperature. After four washes with 200 µl of incubation buffer, 150 µl of *o*-Phenylenediamine dihydrochloride (Sigma) was added for 30 min and then data was recorded at 450 nm.

TROSY NMR. Measurements were performed at 37°C on a $^{2}H^{15}N^{13}C$-labelled Lcl-CTD sample (0.5 mM) in 20 mM HEPES pH 7.0, 50 mM NaCl, 5 mM EDTA, 10% D_2O on a cryoprobe-equipped Bruker Avance III HD spectrometer with 900 MHz Oxford Instruments magnet. Backbone assignments for 61% of Lcl-CTD (not including the N-terminal His-tag and proline residues) was achieved using standard double- and triple-assignment methods. NMR titration experiments were carried out on a Bruker Avance III HD 800 MHz spectrometer equipped cryoprobe. $^{2}H^{15}N$-labelled Lcl-CTD (0.2 mM) in 20 mM HEPES pH 7.0, 50 mM NaCl, 10% D_2O with the addition of 0, 10, 20, 50, 100 and 500 µg/ml chondroitin-4-sulphate from bovine trachea (Sigma) was used to measure $^{1}H^{15}N$ TROSY spectra at 37°C. All spectra were processed using NMRPipe[68] and analysed using the program NMRVIEW[86]. Residues that displayed spectral overlap were not analysed for changes in peak intensity between different spectra.

Experimental driven docking. Molecular docking of C4S oligosaccharides to Lcl-CTD monomer and trimer was carried out with HADDOCK[43, 44, 87] with a modified approach previously used to dock heparin oligosaccharides[87]. Oligosaccharides dp4, dp6, dp8 and dp10 were generated by the GAG Builder server[88]. Active and passive residues were chosen based on the CSPs and ELISA-based

mutational analysis. Topology and parameter files for the C4S oligosaccharides were generated using the PRODRG server[89]. Docking of dp2, dp4, dp8 and dp14 were performed for a 1:1 and 3:1 Lcl-CTD:C4S complex for the Lcl-CTD monomer and trimer, respectively. During initial rigid body docking a total of 1000 structures were generated, and then semi-flexible simulated annealing (SA) was performed on the best 200 structures followed by explicit solvent refinement. The final structures were clustered using a RMSD cut-off value of 7.5 Å and the clusters were sorted using RMSD and the HADDOCK score.

Molecular Dynamics. MD simulations were carried out starting from the crystallographic structure of Lcl-CTD and from three HADDOCK derived models of the complex between C4S and Lcl-CTD. Two models (HT1 and HT2) were obtained from HADDOCK where one molecule of C4S dp8 was docked against a trimer of Lcl-CTD (3:1 Lcl-CTD:C4S). The final model (HM) was obtained from HADDOCK with one molecule of C4S dp8 docked against a monomer of Lcl-CTD (1:1 Lcl-CTD:C4S) but then reconstituted as a timer (3:1 Lcl-CTD:C4S) based on the crystal structure. Simulations were performed using GROMACS 2020[90], with the Amber99SB*-ILDN[91] force field for the Lcl-CTD and GLYCAM-06j for C4S[92]. A truncated octahedral box of TIP3P[93] water molecules was used to solvate the systems, setting a minimum distance of 12 Å between the protein and the edges of the box. Residues with ionisable groups were set to their standard protonation states at pH 7. Counterions were added to neutralise the system and reach an ionic strength of 100 mM, leading to a total of ~50,100 atoms. Periodic boundary conditions were applied. The equations of motion were integrated using the leap-frog method with a 2-fs time step. The LINCS[94] algorithm was used to constrain all covalent bonds in the protein, while SETTLE[95] was used for water molecules. Electrostatic interactions were evaluated with the Particle Mesh Ewald (PME) method[96] using a 9 Å distance cut-off for the direct space sums, a 1.2 Å FFT grid spacing and a 4-order interpolation polynomial for the reciprocal space sums. A 9 Å cut off was set for van der Waals interactions and long-range corrections to the dispersion energy were included.

A previously used protocol[18] was followed for minimisation, equilibration, and production. Longer equilibrations were run for Lcl-CTD in the presence of C4S (45 ns in total for all the steps) compared to Lcl-CTD alone (6.5 ns), to allow for relaxation of the GAG binding pose. Lcl-CTD simulations were run in three replicas (400 ns for each production, for a total production time of 1.2 ms). For the Lcl-CTD/C4S system, preliminary 50 ns production runs were first carried out. The trimeric structure of Lcl-CTD was very stable with the HM model, while unbinding of monomers from the rest of the protein was observed for MT1 and HT2, so only the former model was retained for subsequent simulations. A total of 21 replicas were run for HM (150 ns production for each replica). The glycan remained in contact with Lcl-CTD in all but one replica, which was not considered for subsequent analysis, so that the overall production simulation time for Lcl-CTD+C4S was 3 μs.

Contacts between C4S and the protein were analysed with the bio3D[97] R-package. A residue was considered in contact with C4S if the minimum distance calculated over all pairs of non-hydrogen side chain atoms was lower than 4 Å. Frames sampled every 100 ps were analysed. The frequency of occurrence of a given contact was calculated as the percentage of frames in which that contact was observed. The highest occurring contact between any part of C4S and a given protein residue was calculated and averaged over all 20 replicas to give the final value. Cluster analyses were performed using the gromos[98] method implemented in GROMACS on the pseudo-trajectories generated by concatenating all the replicas for a given system (production only; 3 × 400 ns for Lcl-CTD and 20 × 150 ns for Lcl-CTD/C4S), with frames sampled every 1000 ps. For both calculations, the Lcl-CTD C_α atoms, not including the flexible residues 271 to 277, were first fitted to the coordinates of the initial minimised structure. The distance between structures was calculated as the RMSD of the C_α atoms (271 to 277 excluded) for the Lcl-CTD simulations and the RMSD of all C4S atoms for the Lcl-CTD/C4S simulations. Cut-off values were determined to optimise the clustering for each system and were set to 1.1 Å for Lcl-CTD and 17.5 Å for Lcl-CTD/C4S. The large value for C4S reflects the variety of binding poses explored by the GAG in the different replicas. The structure with the highest number of neighbours for each cluster (central structure) was selected as cluster representative. The population of each cluster was adjusted to consider the 3-fold symmetry of the system. For a given

cluster, each frame in the cluster was first rotated by ~ 120^{o} in both directions. Rotation was carried out by superimposing symmetrically equivalent monomers. If after rotation the C4S structure in the frame was found to be closer to a cluster representative different from the original cluster (as measured by the C4S RMSD), the frame was reassigned to that cluster. The spatial distribution function[99] (sdf) of C4S sulphur atoms around the protein was calculated by running the GROMACS *gmx spatial* tool on the pseudo-trajectory of concatenated replicas (production only), with frames sampled every 1 ps. Each frame was first fitted to the minimised starting structure using best-fit superposition of C_{α} atoms (271 to 277 excluded). A grid spacing of 0.5 Å was used for the sdf calculation. The average of non-null sdf values was calculated and the isosurface connecting points with sdf =20 × average was analysed. Each frame of the Lcl-CTD/C4S simulations was classified into one of three binding categories (1-chain, 2-chains, or 3-chains) by calculating the number of Lcl-CTD chains within 3 Å of C4S (with the distance calculated as minimum distance between all possible pairs of non-hydrogen atoms from C4S and Lcl-CTD).

Data availability. The authors declare that all the data supporting the findings of this study are available within the paper and its Supplementary Information or are available from the corresponding author on request. Atomic coordinates and structure factors files have been deposited in the Protein Data Bank under accession codes (Lcl-CTD: 8Q4E; Lcl-CTD/SO_4: 8QK8).

References

1. Tison DL, Pope DH, Cherry WB, Fliermans CB. Growth of Legionella pneumophila in association with blue-green algae (cyanobacteria). *Appl Environ Microbiol* **39**, 456-459 (1980).

2. Stewart CR, Muthye V, Cianciotto NP. Legionella pneumophila persists within biofilms formed by Klebsiella pneumoniae, Flavobacterium sp., and Pseudomonas fluorescens under dynamic flow conditions. *PLoS One* **7**, e50560 (2012).

3. Rowbotham TJ. Preliminary report on the pathogenicity of Legionella pneumophila for freshwater and soil amoebae. *J Clin Pathol* **33**, 1179-1183 (1980).

4. Fields BS, Benson RF, Besser RE. Legionella and Legionnaires' disease: 25 years of investigation. *Clin Microbiol Rev* **15**, 506-526 (2002).

5. Steinert M, Heuner K, Buchrieser C, Albert-Weissenberger C, Glockner G. Legionella pathogenicity: genome structure, regulatory networks and the host cell response. *Int J Med Microbiol* **297**, 577-587 (2007).

6. Isberg RR, O'Connor TJ, Heidtman M. The Legionella pneumophila replication vacuole: making a cosy niche inside host cells. *Nat Rev Microbiol* **7**, 13-24 (2009).

7. Hubber A, Roy CR. Modulation of host cell function by Legionella pneumophila type IV effectors. *Annu Rev Cell Dev Biol* **26**, 261-283 (2010).

8. Schroeder GN. The Toolbox for Uncovering the Functions of Legionella Dot/Icm Type IVb Secretion System Effectors: Current State and Future Directions. *Front Cell Infect Microbiol* **7**, 528 (2017).

9. White RC, Cianciotto NP. Assessing the impact, genomics and evolution of type II secretion across a large, medically important genus: the Legionella type II secretion paradigm. *Microb Genom* **5**, (2019).

10. Soderberg MA, Rossier O, Cianciotto NP. The type II protein secretion system of Legionella pneumophila promotes growth at low temperatures. *J Bacteriol* **186**, 3712-3720 (2004).

11. DebRoy S, Dao J, Soderberg M, Rossier O, Cianciotto NP. Legionella pneumophila type II secretome reveals unique exoproteins and a chitinase that promotes bacterial persistence in the lung. *Proc Natl Acad Sci U S A* **103**, 19146-19151 (2006).

12. McCoy-Simandle K, *et al.* Legionella pneumophila type II secretion dampens the cytokine response of infected macrophages and epithelia. *Infect Immun* **79**, 1984-1997 (2011).

13. White RC, Cianciotto NP. Type II Secretion Is Necessary for Optimal Association of the Legionella-Containing Vacuole with Macrophage Rab1B but Enhances Intracellular Replication Mainly by Rab1B-Independent Mechanisms. *Infect Immun* **84**, 3313-3327 (2016).

14. White RC, *et al.* Type II Secretion-Dependent Aminopeptidase LapA and Acyltransferase PlaC Are Redundant for Nutrient Acquisition during Legionella pneumophila Intracellular Infection of Amoebas. *MBio* **9**, (2018).

15. White RC, Truchan HK, Zheng H, Tyson JY, Cianciotto NP. Type II Secretion Promotes Bacterial Growth within the Legionella-Containing Vacuole in Infected Amoebae. *Infect Immun* **87**, (2019).

16. Mallama CA, McCoy-Simandle K, Cianciotto NP. The Type II Secretion System of Legionella pneumophila Dampens the MyD88 and Toll-Like Receptor 2 Signaling Pathway in Infected Human Macrophages. *Infect Immun* **85**, (2017).

17. Portlock TJ, *et al.* Structure, Dynamics and Cellular Insight Into Novel Substrates of the Legionella pneumophila Type II Secretion System. *Front Mol Biosci* **7**, 112 (2020).

18. Rehman S, *et al.* Structure and functional analysis of the Legionella pneumophila chitinase ChiA reveals a novel mechanism of metal-dependent mucin degradation. *PLoS Pathog* **16**, e1008342 (2020).

19. Vandersmissen L, De Buck E, Saels V, Coil DA, Anne J. A Legionella pneumophila collagen-like protein encoded by a gene with a variable number of tandem repeats is involved in the adherence and invasion of host cells. *FEMS Microbiol Lett* **306**, 168-176 (2010).

20. Duncan C, *et al.* Lcl of Legionella pneumophila is an immunogenic GAG binding adhesin that promotes interactions with lung epithelial cells and plays a crucial role in biofilm formation. *Infect Immun* **79**, 2168-2181 (2011).

21. Abdel-Nour M, *et al.* Polymorphisms of a Collagen-Like Adhesin Contributes to Legionella pneumophila Adhesion, Biofilm Formation Capacity and Clinical Prevalence. *Front Microbiol* **10**, 604 (2019).

22. Galka F, *et al.* Proteomic characterization of the whole secretome of Legionella pneumophila and functional analysis of outer membrane vesicles. *Infect Immun* **76**, 1825-1836 (2008).

23. Mallegol J, *et al.* Essential roles and regulation of the Legionella pneumophila collagen-like adhesin during biofilm formation. *PLoS One* **7**, e46462 (2012).

24. Chatfield CH, Zaia J, Sauer C. Legionella pneumophila Attachment to Biofilms of an Acidovorax Isolate from a Drinking Water-Consortium Requires the Lcl-Adhesin Protein. *Int Microbiol* **23**, 597-605 (2020).

25. Abdel-Nour M, *et al.* The Legionella pneumophila collagen-like protein mediates sedimentation, autoaggregation, and pathogen-phagocyte interactions. *Appl Environ Microbiol* **80**, 1441-1454 (2014).

26. Coil DA, Vandersmissen L, Ginevra C, Jarraud S, Lammertyn E, Anne J. Intragenic tandem repeat variation between Legionella pneumophila strains. *BMC Microbiol* **8**, 218 (2008).

27. Gandhi NS, Mancera RL. The structure of glycosaminoglycans and their interactions with proteins. *Chem Biol Drug Des* **72**, 455-482 (2008).

28. Jinno A, Park PW. Role of glycosaminoglycans in infectious disease. *Methods Mol Biol* **1229**, 567-585 (2015).

29. Thomas R, Brooks T. Common oligosaccharide moieties inhibit the adherence of typical and atypical respiratory pathogens. *J Med Microbiol* **53**, 833-840 (2004).

30. Yaradou DF, *et al.* Zinc-dependent cytoadherence of Legionella pneumophila to human alveolar epithelial cells in vitro. *Microb Pathog* **43**, 234-242 (2007).

31. Rao C, Benhabib H, Ensminger AW. Phylogenetic reconstruction of the Legionella pneumophila Philadelphia-1 laboratory strains through comparative genomics. *PLoS One* **8**, e64129 (2013).

32. Lu S, *et al.* CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res* **48**, D265-D268 (2020).

33. Buchan DWA, Jones DT. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res* **47**, W402-W407 (2019).

34. Gautier R, Douguet D, Antonny B, Drin G. HELIQUEST: a web server to screen sequences with specific alpha-helical properties. *Bioinformatics* **24**, 2101-2102 (2008).

35. Holm L. DALI and the persistence of protein shape. *Protein Sci* **29**, 128-140 (2020).

36. Huang KF, Ko TP, Hung CC, Chu J, Wang AH, Chiou SH. Crystal structure of a platelet-agglutinating factor isolated from the venom of Taiwan habu (Trimeresurus mucrosquamatus). *Biochem J* **378**, 399-407 (2004).

37. Horii K, Okuda D, Morita T, Mizuno H. Crystal structure of EMS16 in complex with the integrin alpha2-I domain. *J Mol Biol* **341**, 519-527 (2004).

38. Luo Y, *et al.* Crystal structure of enteropathogenic Escherichia coli intimin-receptor complex. *Nature* **405**, 1073-1077 (2000).

39. Hamburger ZA, Brown MS, Isberg RR, Bjorkman PJ. Crystal structure of invasin: a bacterial integrin-binding protein. *Science* **286**, 291-295 (1999).

40. Papakonstantinou E, Karakiulakis G. The 'sweet' and 'bitter' involvement of glycosaminoglycans in lung diseases: pharmacotherapeutic relevance. *Br J Pharmacol* **157**, 1111-1127 (2009).

41. In: *Essentials of Glycobiology* (eds Varki A, *et al.*). Cold Spring Harbor Laboratory Press

(2022).

42. Su H, Li S, Terebiznik M, Guyard C, Kerman K. Biosensors for the Detection of Interaction between Legionella pneumophila Collagen-Like Protein and Glycosaminoglycans. *Sensors (Basel)* **18**, (2018).

43. van Zundert GCP, *et al.* The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J Mol Biol* **428**, 720-725 (2016).

798 44. Dominguez C, Boelens R, Bonvin AM. HADDOCK: a protein-protein docking approach
799 based on biochemical or biophysical information. *J Am Chem Soc* **125**, 1731-1737 (2003).
800
801 45. Stone BJ, Abu Kwaik Y. Expression of multiple pili by Legionella pneumophila:
802 identification and characterization of a type IV pilin gene and its role in adherence to
803 mammalian and protozoan cells. *Infect Immun* **66**, 1768-1775 (1998).
804
805 46. Hoppe J, *et al.* PilY1 Promotes Legionella pneumophila Infection of Human Lung Tissue
806 Explants and Contributes to Bacterial Adhesion, Host Cell Invasion, and Twitching Motility.
807 *Front Cell Infect Microbiol* **7**, 63 (2017).
808
809 47. Treuner-Lange A, *et al.* PilY1 and minor pilins form a complex priming the type IVa pilus in
810 Myxococcus xanthus. *Nat Commun* **11**, 5054 (2020).
811
812 48. Garduno RA, Garduno E, Hoffman PS. Surface-associated hsp60 chaperonin of Legionella
813 pneumophila mediates invasion in a HeLa cell model. *Infect Immun* **66**, 4602-4610 (1998).
814
815 49. Cirillo SL, Bermudez LE, El-Etr SH, Duhamel GE, Cirillo JD. Legionella pneumophila entry
816 gene rtxA is involved in virulence. *Infect Immun* **69**, 508-517 (2001).
817
818 50. Bellinger-Kawahara C, Horwitz MA. Complement component C3 fixes selectively to the
819 major outer membrane protein (MOMP) of Legionella pneumophila and mediates
820 phagocytosis of liposome-MOMP complexes by human monocytes. *J Exp Med* **172**, 1201-
821 1210 (1990).
822
823 51. Chang B, Kura F, Amemura-Maekawa J, Koizumi N, Watanabe H. Identification of a novel
824 adhesion molecule involved in the virulence of Legionella pneumophila. *Infect Immun* **73**,
825 4272-4280 (2005).
826
827 52. Truchan HK, Christman HD, White RC, Rutledge NS, Cianciotto NP. Type II secretion
828 substrates of *Legionella pneumophila* translocate out of the pathogen-occupied vauole via a
829 semi-permeable membrane. *mBio* **8**, e00870-00817 (2017).
830
831 53. Qiu Y, Zhai C, Chen L, Liu X, Yeo J. Current Insights on the Diverse Structures and
832 Functions in Bacterial Collagen-like Proteins. *ACS Biomater Sci Eng* **9**, 3778-3795 (2023).
833

54. Zhuoxin Y, Bo A, John AMR, Barbara B. Bacterial collagen-like proteins that form triple-helical structures. *Journal of Structural Biology* **186**, 451-461 (2014).

55. Han R, *et al.* Assessment of prokaryotic collagen-like sequences derived from streptococcal Scl1 and Scl2 proteins as a source of recombinant GXY polymers. *Appl Microbiol Biotechnol* **72**, 109-115 (2006).

56. Stewart CR, Burnside DM, Cianciotto NP. The surfactant of Legionella pneumophila Is secreted in a TolC-dependent manner and is antagonistic toward other Legionella species. *J Bacteriol* **193**, 5971-5984 (2011).

57. Rossier O, Starkenburg SR, Cianciotto NP. Legionella pneumophila type II protein secretion promotes virulence in the A/J mouse model of Legionnaires' disease pneumonia. *Infect Immun* **72**, 310-321 (2004).

58. Chatfield CH, Cianciotto NP. Culturing, media, and handling of legionella. *Methods Mol Biol* **954**, 151-162 (2013).

59. Campbell JA, Cianciotto NP. Legionella pneumophila Cas2 Promotes the Expression of Small Heat Shock Protein C2 That Is Required for Thermal Tolerance and Optimal Intracellular Infection. *Infect Immun* **90**, e0036922 (2022).

60. Bryan A, Harada K, Swanson MS. Efficient generation of unmarked deletions in Legionella pneumophila. *Appl Environ Microbiol* **77**, 2545-2548 (2011).

61. Bryan A, Abbott ZD, Swanson MS. Constructing unmarked gene deletions in *Legionella pneumophila*. *Methods Mol Biol* **954**, 197-212 (2013).

62. Mould AP, Holmes DF, Kadler KE, Chapman JA. Mica sandwich technique for preparing macromolecules for rotary shadowing. *J Ultrastruct Res* **91**, 66-76 (1985).

63. Ghosh N, *et al.* Collagen-like proteins in pathogenic E. coli strains. *PLoS One* **7**, e37872 (2012).

64. Jumper J, *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).

65. Steinegger M, Soding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* **35**, 1026-1028 (2017).

66. Steinegger M, Meier M, Mirdita M, Vohringer H, Haunsberger SJ, Soding J. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20**, 473 (2019).

67. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold - Making protein folding accessible to all. *bioRxiv*, 2021.2008.2015.456425 (2021).

68. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-293 (1995).

69. Vranken WF, *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687-696 (2005).

70. Sormanni P, Camilloni C, Fariselli P, Vendruscolo M. The s2D method: simultaneous sequence-based prediction of the statistical populations of ordered and disordered regions in proteins. *J Mol Biol* **427**, 982-996 (2015).

71. Kabsch W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132 (2010).

72. Evans PR, Murshudov GN. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204-1214 (2013).

73. Winter G, Lobley CM, Prince SM. Decision making in xia2. *Acta Crystallogr D Biol Crystallogr* **69**, 1260-1273 (2013).

74. Sheldrick GM. A short history of SHELX. *Acta Crystallogr A* **64**, 112-122 (2008).

75. Vonrhein C, Blanc E, Roversi P, Bricogne G. Automated structure solution with autoSHARP. *Methods Mol Biol* **364**, 215-230 (2007).

76. Langer G, Cohen SX, Lamzin VS, Perrakis A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* **3**, 1171-1179 (2008).

77. Casanal A, Lohkamp B, Emsley P. Current developments in Coot for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. *Protein Sci* **29**, 1069-1078 (2020).

78. Winn MD, Murshudov GN, Papiz MZ. Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol* **374**, 300-321 (2003).

79. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).

80. de Vries I, *et al.* New restraints and validation approaches for nucleic acid structures in PDB-REDO. *Acta Crystallogr D Struct Biol* **77**, 1127-1141 (2021).

81. Cowieson NP, *et al.* Beamline B21: high-throughput small-angle X-ray scattering at Diamond Light Source. *J Synchrotron Radiat* **27**, 1438-1446 (2020).

82. Panjkovich A, Svergun DI. CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics* **34**, 1944-1946 (2018).

83. Franke D, *et al.* ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J Appl Crystallogr* **50**, 1212-1225 (2017).

84. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**, 779-815 (1993).

85. Tria G, Mertens HD, Kachala M, Svergun DI. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* **2**, 207-217 (2015).

86. Johnson BA, Blevins RA. NMR View: A computer program for the visualization and analysis of NMR data. *J Biomol NMR* **4**, 603-614 (1994).

87. Sepuru KM, Nagarajan B, Desai UR, Rajarathnam K. Molecular Basis of Chemokine CXCL5-Glycosaminoglycan Interactions. *J Biol Chem* **291**, 20539-20550 (2016).

88. Singh A, Montgomery D, Xue X, Foley BL, Woods RJ. GAG Builder: a web-tool for modeling 3D structures of glycosaminoglycans. *Glycobiology* **29**, 515-518 (2019).

89. van Aalten DM, Bywater R, Findlay JB, Hendlich M, Hooft RW, Vriend G. PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. *J Comput Aided Mol Des* **10**, 255-262 (1996).

90. Abraham MJ, *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19-25 (2015).

91. Lindorff-Larsen K, Maragakis P, Piana S, Eastwood MP, Dror RO, Shaw DE. Systematic validation of protein force fields against experimental data. *PLoS One* **7**, e32131 (2012).

92. Kirschner KN, *et al.* GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J Comput Chem* **29**, 622-655 (2008).

93. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926-935 (1983).

94. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18**, 1463-1472 (1997).

95. Miyamoto S, Kollman PA. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry* **13**, 952-962 (1992).

96. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **103**, 8577-8593 (1995).

97. Skjærven L, Yao X-Q, Scarabelli G, Grant BJ. Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC Bioinformatics* **15**, 399 (2014).

98. Daura X, Gademann K, Jaun B, Seebach D, van Gunsteren WF, Mark AE. Peptide Folding: When Simulation Meets Experiment. *Angewandte Chemie International Edition* **38**, 236-240 (1999).

99. Fornili A, Autore F, Chakroun N, Martinez P, Fraternali F. Protein–Water Interactions in MD Simulations: POPS/POPSCOMP Solvent Accessibility Analysis, Solvation Forces and

Hydration Sites. In: *Computational Drug Discovery and Design* (ed Baron R). Springer New York (2012).

100. Adasme MF, *et al.* PLIP 2021: expanding the scope of the protein–ligand interaction profiler to DNA and RNA. *Nucleic Acids Research* **49**, W530-W534 (2021).

Acknowledgements.

This work was supported by the MRC (grants MR/M009920/1, MR/R017662/1, and MR/W000814/1), the NIH (AI043987 and AI175460) and an EPSRC studentship. Support in part was also from the Wellcome Trust (grant number 099185/Z/12/Z), and we thank HWB-NMR staff at the University of Birmingham for providing open access to their Wellcome Trust-funded 900 MHz spectrometer. In addition, this work was supported by the Francis Crick Institute through provision of access to the MRC Biomedical NMR Centre. The Francis Crick Institute receives its core funding from Cancer Research UK (FC001029), the UK Medical Research Council (CC1078), and the Wellcome Trust (CC1078). We also thank the Centre for Biomolecular Spectroscopy at King's College London for additional NMR access, funded by the Wellcome Trust and British Heart Foundation (ref. 202767/Z/16/Z and IG/16/2/32273, respectively). This work made use of time on HPC granted via the UK High-End Computing Consortium for Biomolecular Simulation, HECBioSim (http://hecbiosim.ac.uk), supported by EPSRC (grant no. EP/R029407/1). We thank the beamline scientists at I03 and BL21 of the Diamond Light Source, United Kingdom. We would also like to thank Prof. Krishna Rajarathnam and Dr. K. Mohan Sepuru (UTMB) for their guidance in creating C4S input files for use in HADDOCK.

Author contributions

Conceived and designed the experiments: SR, AA, IM, HZ, CA, AO, GM, SW, GK, AF, NC, JG. Performed the experiments: SR, AA, IM, HZ, CA, TP, KR, RS, AO, GM, SW, GK, AF, JG. Analyzed the data: SR, AA, IM, HZ, CA, TP, KR, RS, AF, NC, JG. Contributed reagents/materials/analysis tools: AF, AA, NC, JG. Wrote the paper: SR, AA, LC, CA, AF, NC, JG.

Additional information

Competing interests**.** The authors declare no competing interests.

Figure Legends

Figure 1 | *L. pneumophila* 130b surface association of Lcl. a, Analysis of Lcl secretion from BYE culture supernatants of wild-type 130b, *lcl* mutants NU468 and NU469, and *lspF* mutant NU275 reacted with anti-Lcl antibodies. Results are representative of at least two independent experiments. **b,** Detection of bacterial surface binding. Whole cell ELISA of wild-type 130b, *lcl* mutants NU468 and NU469, and *lspF* mutant NU275 detected with anti-Lcl antibodies. *** $P < 0.0001$; versus WT control, Student's t-test. Data are the means and standard deviations from technical triplicates and are representative of at least three experiments.

Figure 2 | Global characterisation of Lcl. a, Size-Exclusion Chromatography coupled to Multi-Angle Light Scattering (SEC-MALS) profile of recombinant Lcl and Lcl-CTD, using a Superose 6 Increase 10/300 column. Normalised refractive index (grey and green line) and average molecular weight calculated across the elution profile (orange and gold line) are shown for Lcl and Lcl-CTD, respectively. **b,** Upper: schematic of the Lcl domains with residue numbering based on mature

sequence shown below. SS: periplasmic signal sequence; N: N-terminal helix; CLR: collagen-like repeat region; CTD: C-terminal domain. Lower: constructs used in this study with position of His_6 affinity tags shown. Peptides modifications are annotated (Ac: acylation; NH_2: amidation) along with sequence and numbering. Right: Schematic of the Lcl trimer presented on the bacterial surface. **c,** Micrograph showing lollipop-shaped structures of Lcl trimers. The concentration of Lcl was 5 µg/ml. Scale bar = 50 nm. The globular shapes correspond to trimeric C-terminal domains (green arrow), while the stalks contain trimeric collagen-like region (grey arrow). **d,** Helical wheel diagram and structural model of the Lcl H1 helix (Lcl-N) generated by the HELIQUEST[34] and AlphaFold2[64], respectively, with terminal residues numbered. Yellow/grey: large/small hydrophobic, pink/purple: large/small polar, blue/red: positively/ negatively charged. **e,** ^{1}H-^{1}H NOESY spectra of Lcl-N peptide in the presence/absence of 80 mM perdeutotared SDS, highlighting the amide region. **f,** ^{1}H-^{15}N HSQC spectrum of ^{15}N-glycine labelled Lcl-CLR peptide showing resonances for monomeric (m) and trimeric (t) states. Assignment of specifici glycines residues in monomeric Lcl-CLR is shown with peak positions for trimeric Lcl-CLR glycines residues numbered from left to right in subscript.

Fig. 3 | Crystal structure of Lcl -CTD. a, Monomer of Lcl-CTD shown as cartoon and rotated by 180°. **b,** Trimer of Lcl-CTD shown from the top as cartoon. **c,** Trimer of Lcl-CTD shown from the side as a cut-away electrostatic surface highlighting the internal charged cavity. Position of Asp336 and Glu368 in two chains is shown. **d,** Monomer of Lcl-CTD shown as electrostatic surface and rotated by 180°, with the inter-trimer interface highlighted with a yellow outline. Inter-trimer residues and charged surface residues are highlighted. **e,** Crystal structure of trimeric Lcl-CTD/SO_4 shown as electrostatic surface and highlighting the bound sulphate ions (yellow spheres) and charged surface residues.

Fig. 4 | GAG binding to Lcl -CTD. a, Chemical structure of chondroitin-4-sulphate (C4S) and heparin. GlcA: D-glucuronate; GalNAc(4S): *N*-acetyl-D-galactosamine-4-*O*-sulphate; IdoA(2S): α-L-iduronate-2-*O*-sulphate; GlcNS(6S): 6-*O*-sulpho-2-(sulphoamino)-D-glucosamine. **b,** ELISA analysis of binding between immobilised mixed length C4S or heparin and

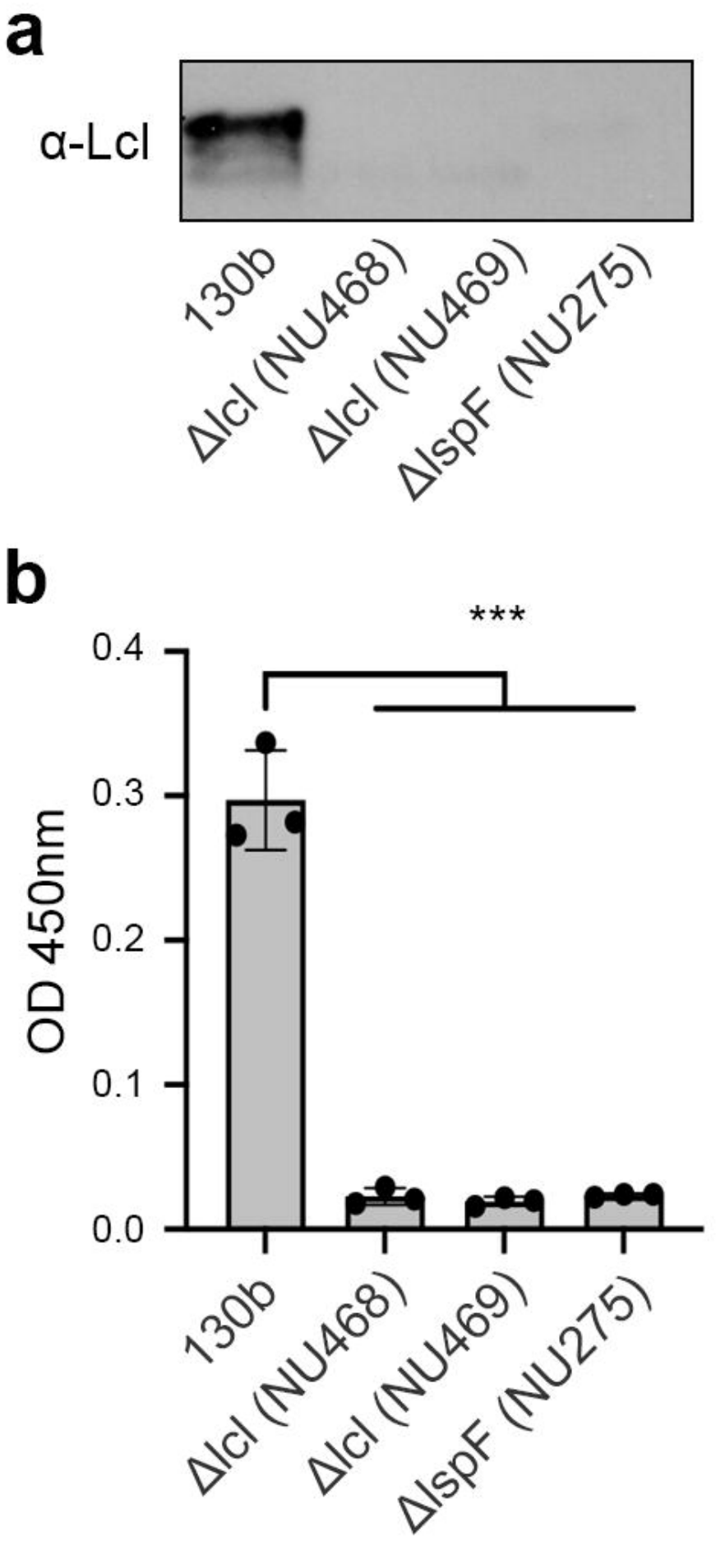
wild-type (WT) and mutant (E368A, K369A, K380A, K385A, D386A, K391A) His-tagged Lcl-CTD. BSA-coated wells were used as controls. ** $P < 0.01$, *** $P < 0.001$; verses the respective WT by two-tailed Student's test. Data are the means and standard deviations from four technical repeats and are representative of at least two experiments.

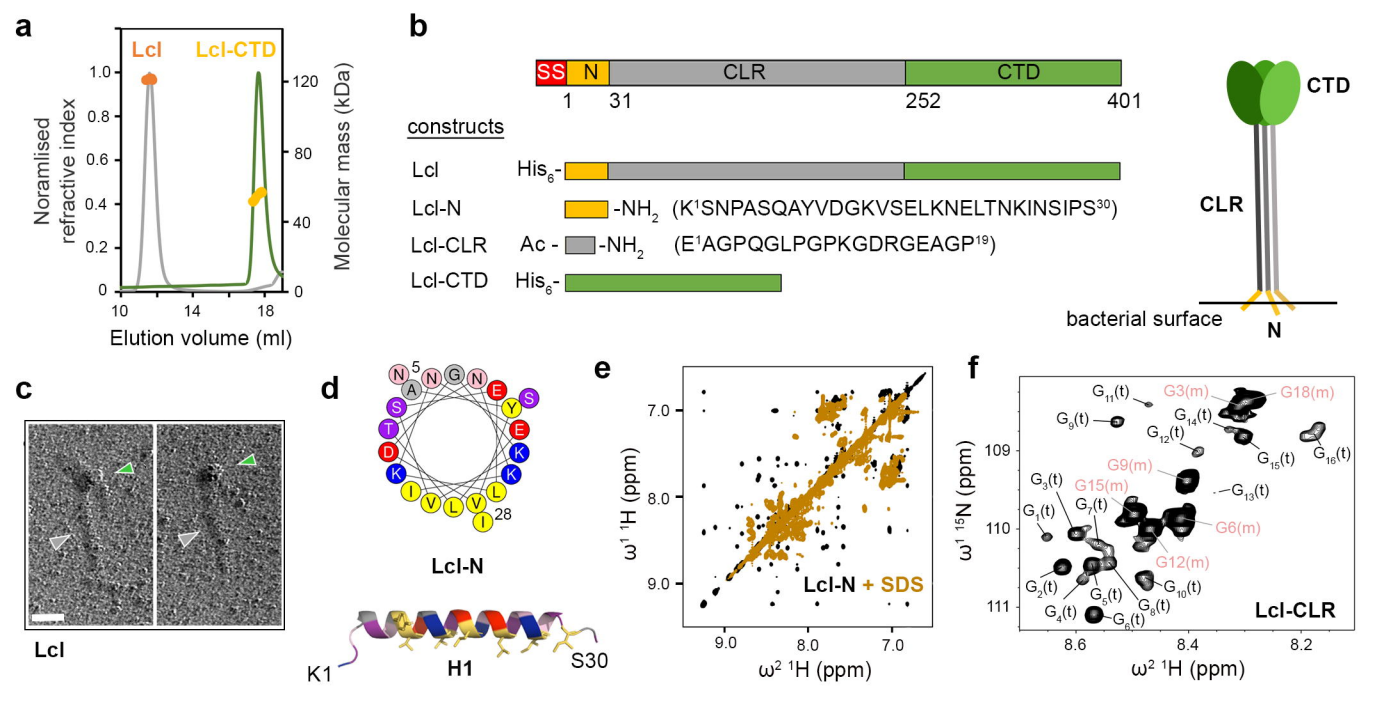
Fig. 5 | Solution NMR spectroscopy analyses of C4S binding to Lcl-CTD. a, Trimer of Lcl-CTD shown as surface representation with residues whose amides could be assigned coloured green, and those that could not be assigned coloured purple. **b,** NMR $^{1}H–^{15}N$ TROSY spectrum of Lcl-CTD in presence (right) or absence (left) of 0.5 mg/ml mixed length C4S. Chemical shifts that have disappeared after addition of C4S are highlighted in red, and those that display significant broadening (reduction of >85% peak intensity) are highlighted in orange. **c,** Same information as (**b**) shown as a bar graph with orange bars highlighting significant peak broadening on addition of C4S. Missing assignments have a value of zero and those where peaks disappear on addition of C4S are highlighted with red circles. **d,** As (**c**) and (**b**) but mapped onto the surface trimer of Lcl-CTD.

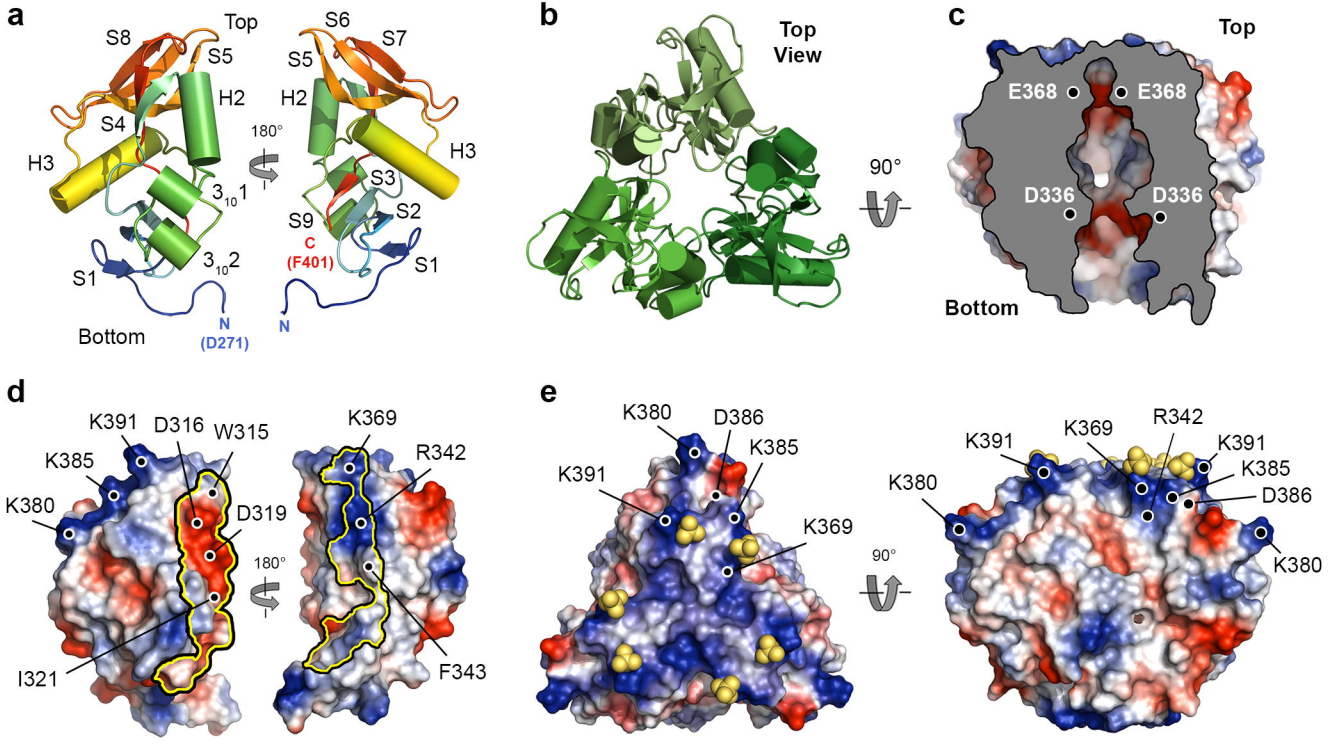
Fig. 6 | Molecular Dynamics analyses of C4S dp8 binding to Lcl-CTD. a. Modified HADDOCK model (HM: C4S dp8 docked against a monomer of Lcl-CTD and reconstituted as a timer) used as a starting structure for MD simulations. Surface residues in close contact of C4S are annotated and coloured red (reduction >100% peak intensity by NMR), orange (reduction >85% peak intensity) or blue (lysine residues identified by ELISA). **b,** Spatial distribution function (sdf) of the sulphur atoms of C4S dp8 during the simulations. The purple isosurface connects the points with sdf = 20 × average value. The protein surface (initial MD structure) is represented in white with the positions of Lys369, Lys 380, Lys 385 and Lys391 coloured blue. **c,** Frequency of occurrence (occupancy) of contacts between C4S dp8 and the Lcl-CTD during the simulations colour-mapped onto the protein surface (initial MD structure) from white (0-10%) to orange (40.9%). Residues with an occupancy >10% in chain A are annotated as (**a**) or black (newly identified from MD).

Fig. 7 | Binding modes of C4S dp8 to Lcl-CTD. a, Representative C4S dp8 structures (lowest RMSD within the cluster) of the first (MD cluster 1, population=43%), second (MD cluster 2,

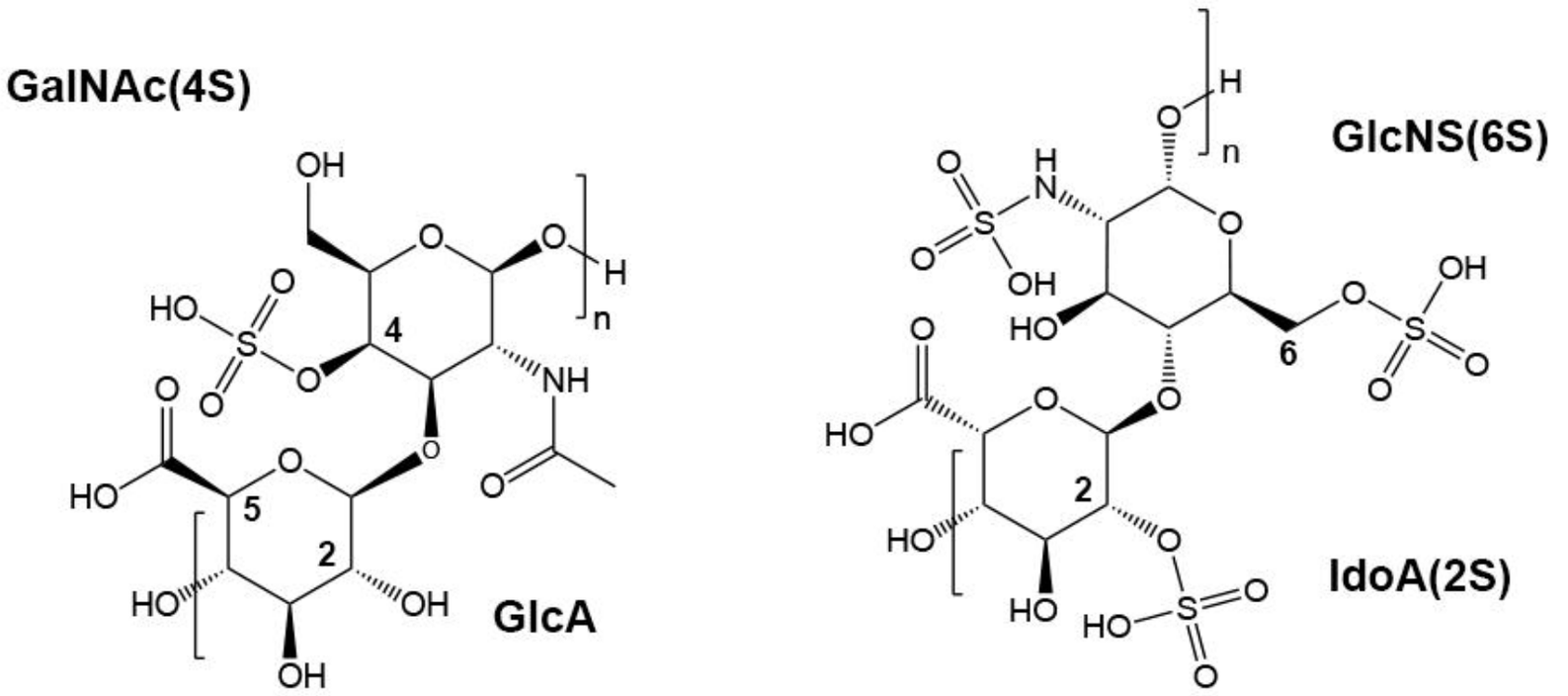
population=21%), and third (MD cluster 3, population=20%) most populated clusters are shown as sticks, together with the initial protein structure (green surface). The position and orientation of cluster 3 is like that of cluster 1 when 3-fold rotational symmetry is considered. These therefore represent two major binding modes: clusters 1 and 3 (M-BM; major binding mode, population=63%) and cluster 2 (m-BM; minor binding mode, population=21%). **b,** Representative MD structures of C4S dp8 bound to Lcl-CTD selected to illustrate binding across 1-chain (population=36%; derived from m-BM and M-BN), 2-chains (population=35%; primarily derived from M-BM) and 3-chains (population=28%; primarily derived from M-BM). Structures were selected from replica 7, 11 and 6, respectively. Hydrogen bonding interactions between C4S and Lcl-CTD detected by PLIP[100] are shown as dashed red lines. The protein residues involved in the interactions are labelled. Cyan spheres indicate the location of C1 hydroxyl and C5 carboxyl groups within C4S D-glucuronate residues, which if switched would perturb binding. **c,** Models of GAG binding to the Lcl-CTD trimer. Schematics of the general major and minor binding mode of C4S are shown with bound glycan chain as an arrow, which can bind in either direction. The Lcl-CTD surface could support simultaneous binding to GAGs from one continuous chain (black connected arrows) and/or from multiple chains (olive and wheat arrows).


a
α-Lcl
130b
ΔIcl (NU468)
ΔIcl (NU469)
ΔIspF (NU275)
b
***
OD 450nm
0.4
0.3
0.2
0.1
0.0
130b
ΔIcl (NU468)
ΔIcl (NU469)
ΔIspF (NU275)


a
Lcl
Lcl-CTD
Noramlised refractive index
Molecular mass (kDa)
Elution volume (ml)
b
SS
N
CLR
CTD
1
31
252
401
constructs
Lcl
Lcl-N
Lcl-CLR
Lcl-CTD
His6-
-NH2
(K1SNPASQAYVDGKVSELKNELTNKINSIPS30)
Ac -
(E1AGPQGLPGPKGDRGEAGP19)
CTD
CLR
bacterial surface
N
c
Lcl
d
Lcl-N
K1
H1
S30
e
Lcl-N + SDS
ω1 1H (ppm)
ω2 1H (ppm)
f
Lcl-CLR
ω1 15N (ppm)
ω2 1H (ppm)

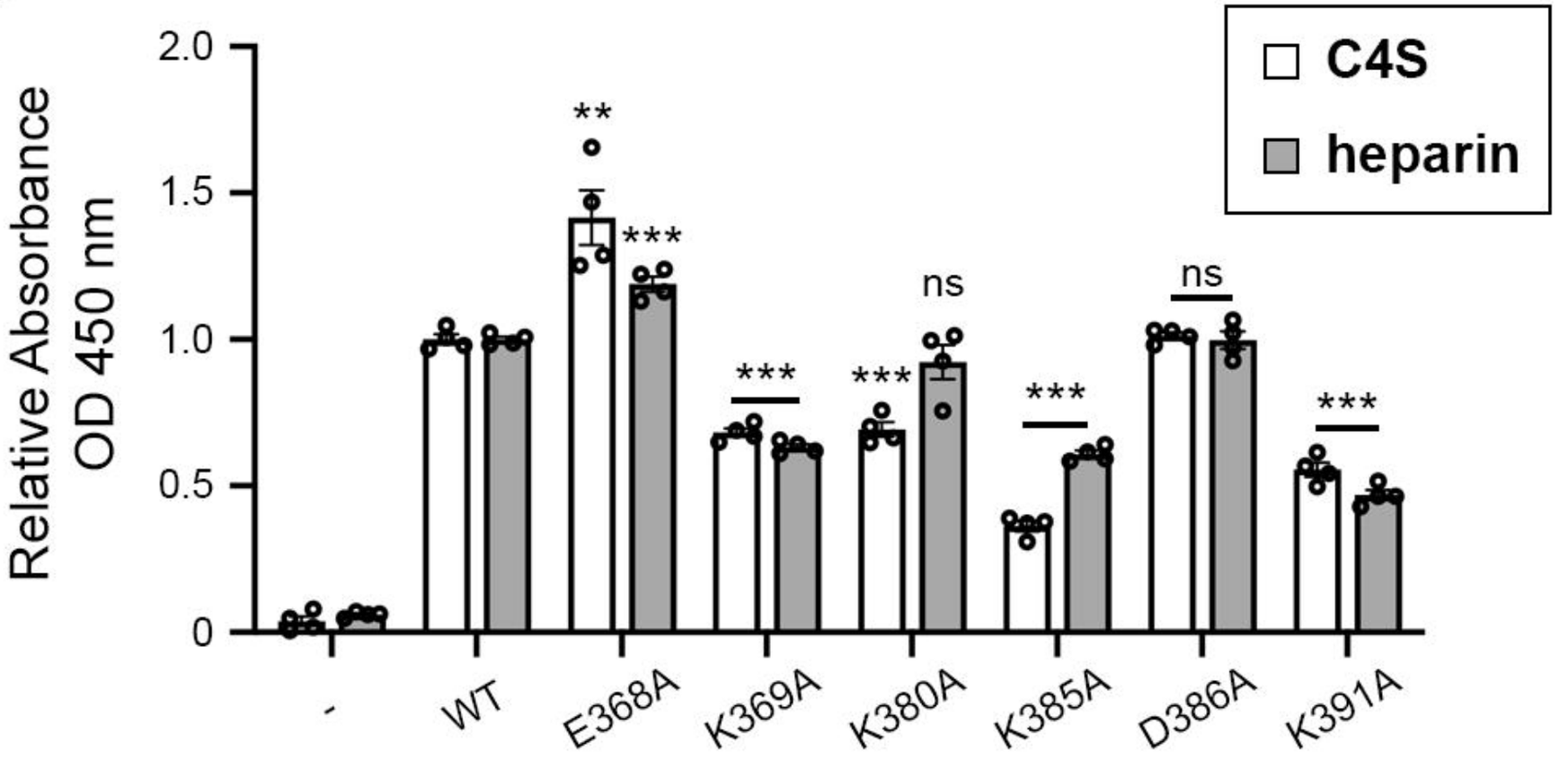

a
Top
S8
S5
H2
S4
H3
$3_{10}1$
$3_{10}2$
S1
Bottom
N
(D271)
180°
S6
S7
S5
H2
H3
S3
S2
S9
C
(F401)
S1
N
b
Top
View
90°
c
Top
E368
E368
D336
D336
Bottom
d
K391
D316
W315
K385
K380
D319
I321
180°
K369
R342
F343
e
K380
D386
K385
K391
K369
90°
K391
K369
R342
K391
K385
K380
D386
K380

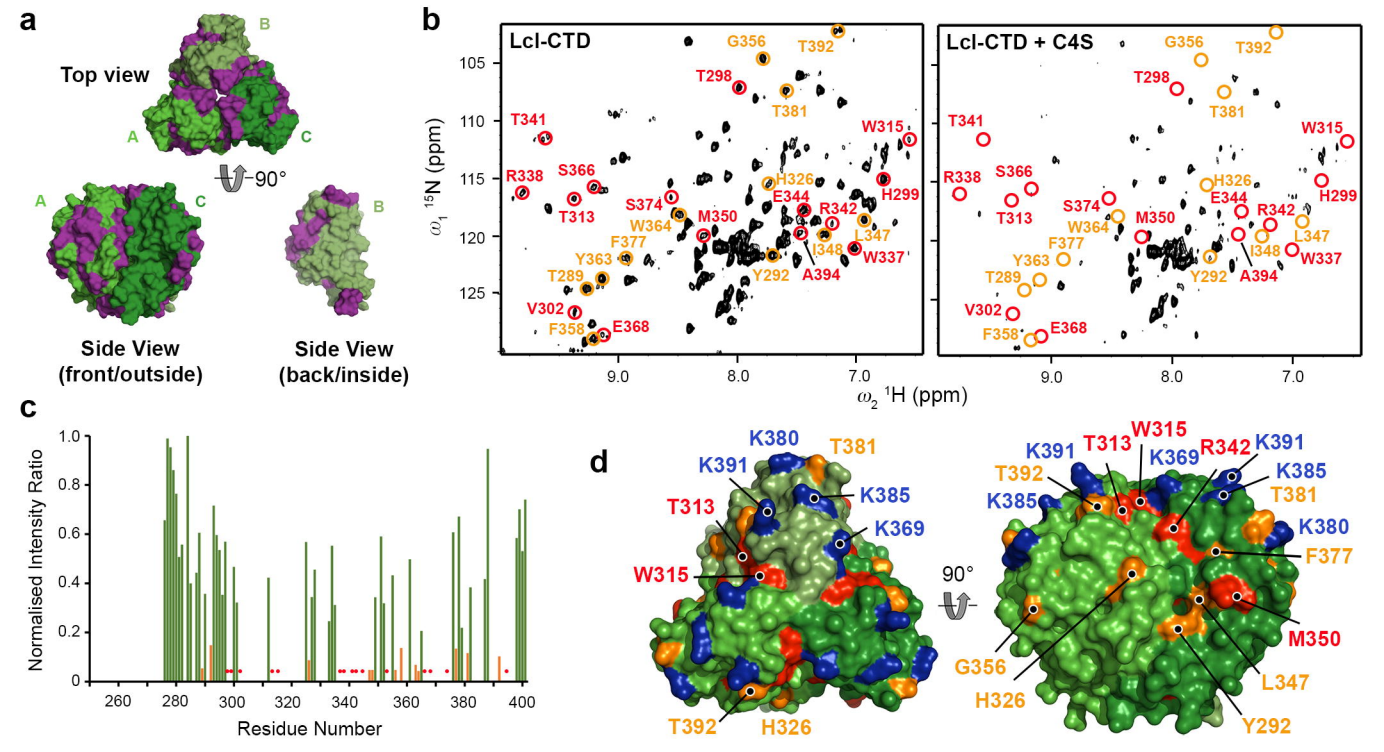
a

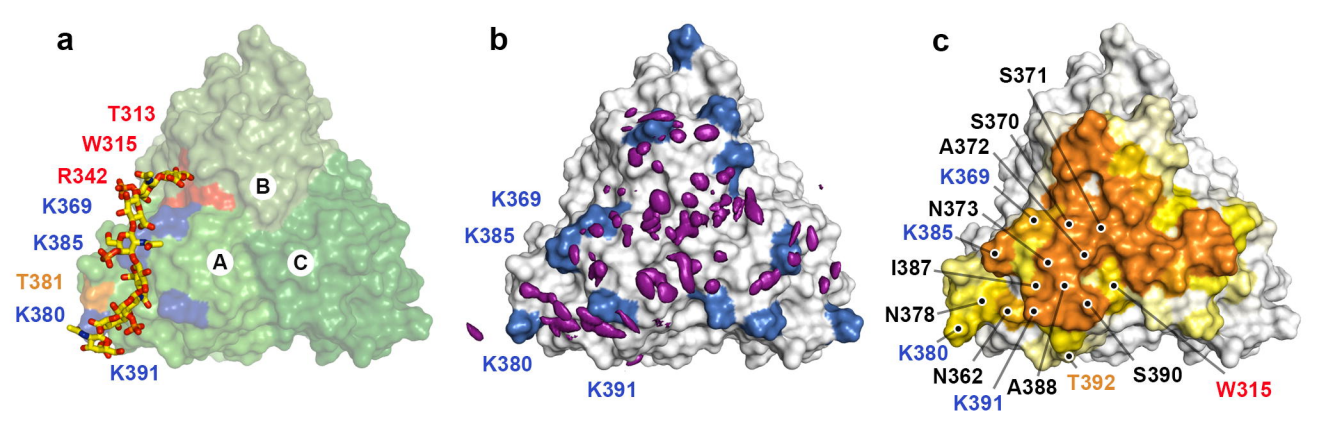


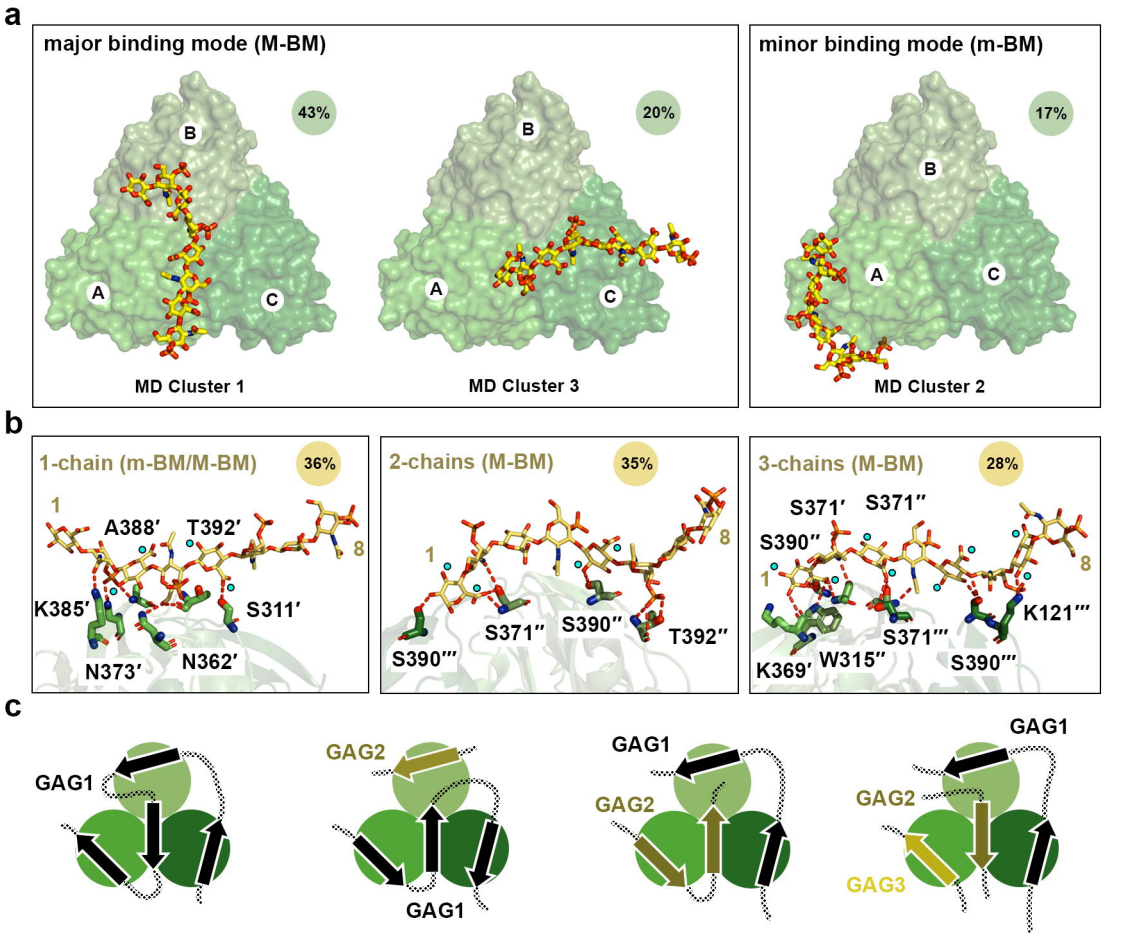
chondroitin-4-sulfate (C4S) **heparin**

b




a
Top view
A
B
C
90°
Side View
(front/outside)
Side View
(back/inside)
b
LcI-CTD
LcI-CTD + C4S
ω_1 ^{15}N (ppm)
ω_2 ^{1}H (ppm)
c
Normalised Intensity Ratio
Residue Number
d
90°


a
T313
W315
R342
K369
K385
T381
K380
K391
B
A
C
b
K369
K385
K380
K391
c
S371
S370
A372
K369
N373
K385
I387
N378
K380
N362
A388
K391
T392
S390
W315


a
major binding mode (M-BM)
43%
B
A
C
MD Cluster 1
20%
B
A
C
MD Cluster 3
minor binding mode (m-BM)
17%
B
A
C
MD Cluster 2
b
1-chain (m-BM/M-BM)
36%
1
8
A388′
T392′
K385′
S311′
N373′
N362′
2-chains (M-BM)
35%
1
8
S390‴
S371″
S390″
T392″
3-chains (M-BM)
28%
S371′
S371″
S390″
1
8
K121‴
S371‴
K369′
W315″
S390‴
c
GAG1
GAG2
GAG1
GAG1
GAG2
GAG1
GAG2
GAG3