Identification of the receptor-binding protein of *Clostridium difficile* phage CDHS-1 reveals a new class of receptor-binding domains.

Ahmed S. A. Dowah[1], Guoqing Xia[3], Ali Abdul Kareem Ali[1], Anisha M. Thanki[1], Jinyu Shan[1], Russell Wallis[2], Martha R. J. Clokie[1].

Department of Genetics and Genome Biology, University of Leicester, Leicester, LE1 7RH, UK[1].

Department of Molecular and Cell Biology, University of Leicester, Leicester, LE1 7RH[2]

Division of Infection, Immunity & Respiratory Medicine, University of Manchester, Oxford Rd, Manchester M13 9PL[23]

Introductory Paragraph:

Phage-bacterial recognition is species-specific, determined by interactions between phage receptor-binding proteins (RBPs) and corresponding bacterial receptors. RBPs are diverse and we present data demonstrating the identification and characterisation of a novel *C. difficile* phage RBP. Putative RBP were identified for CDHS-1 and overexpressed, purified, and polyclonal antibodies were raised and used in phage neutralization assays. Anti-gp22 neutralised CDHS-1, indicating it is the RBP. Immunogold-labelling and transmission electron microscopy confirmed this enabling visualization of the protein locations. A detailed structural understanding was obtained from determining the three-dimensional structure of gp22 by X-ray crystallography. gp22 is a new RBP class consisting of an N-terminal L-shaped α-helical superhelix domain and a C-terminal Mg^{2+}-binding domain. The protein is a stable homodimer in solution mediated via reciprocal contacts between an α-helical hairpin, located within the superhelix domain and additional asymmetrical contacts between the ends of the short arm of each L-shaped protomer. The dimer resembles U-shape with a crossbar formed from the hairpin of each partner. *C. difficile* binding is Mg^{2+}-dependent. CDHS-1 could not infect a *C. difficile* S-layer mutant suggesting the bacterial receptors are within the S-layer. These findings provide novel insights into phage biology and extend our knowledge of RBPs.

Introduction

Clostridium difficile is an anaerobic, Gram-positive, spore-forming bacterium responsible for gastrointestinal disease in humans (Vardakas *et al.*, 2012). *C. difficile* infection (CDI) is often associated with antibiotic treatment and disruption of the host gut microbiota that allows overgrowth of *C. difficile* (Burke and Lamont, 2014; Hargreaves and Clokie, 2014). CDI treatment is problematic because *C. difficile* is naturally resistant to most antibiotics. Alternative treatments for CDI are urgently required (Nale *et al.*, 2016). One such alternative is bacteriophages (phages), viruses that kill bacteria. Phages are often species-specific and consequently cause minimal disruption to the gut microbiota, yet they can self-amplify at the infection site. Phage therapy against CDI has been used successfully in animal models to reduce bacterial carriage and increase survival time following infection (Nale *et al.*, 2016).

For phage therapeutic development it is important to have a detailed understanding of the physical interactions between phages and their bacterial hosts and yet this is only known about for a relatively small number of phages. Not knowing the phage receptor binding

proteins limits our ability to develop phages as therapeutics in an informed way. As there is so little structural similarity between different phages the only robust way to identify RPBs is by using biochemical and structural approaches.

The infection process starts when phages attach to one or more receptor on the bacterial cell, and penetrate the cell membranes (Rakhuba *et al.*, 2010). This attachment process is mediated via interactions between proteins located at the end of the phage tail called receptor binding proteins (RBPs) and ligands (receptors) on the bacterial surface (Mahony and van Sinderen, 2012).

Phage RBPs are structurally diverse, and many bacterial cell surface molecules can act as receptors including surface proteins, carbohydrates, or wall teichoic acids (Dowah and Clokie, 2018). RBPs are vital determinants of phage host range (Le *et al.*, 2013) and knowing what they are and how they work can inform novel diagnostics and therapeutics (Waseh *et al.*, 2010; Javed *et al.*, 2013; Simpson, Sacher and Szymanski, 2016).

Although sequence similarity is often limited between phages, comparative genomic analysis of phages infecting Gram-positive bacteria including *B. subtilis, L. lactis, L. monocytogenes* and *S. aureus,* previously revealed that genes encoding phage tail proteins are often located in the tail module between *tmp* gene encoding the tape measure protein (TMP), and the holin and endolysin proteins (Bielmann *et al.*, 2015).

Relatively few phages that infect Firmicutes have known, well characterized RBPs and the best known examples are *Staphylococcus aureus* phage φ11 and *Listeria monocytogenes* phages A118 and P35 (Bielmann *et al.*, 2015; Li *et al.*, 2016). Despite the fact that several phages that target clinically relevant *C. difficile* strains have been isolated, (Shan *et al.*, 2012; Nale *et al.*, 2016) no studies have identified their RBPs and little information is known about their bacterial receptors.

Recently, Kirk et al. showed that the S-layer protein of *C. difficile* acts as a receptor for Avidocin-CDs (are genetically modified versions of natural R-type bacteriocins that specifically kill *C. difficile* ribotype 027 strains). The *C. difficile* specificity of Avidocin-CD was acquired from fusing RBP phages such phi-027b, phi-123, phi147 and phi-242.c1and Phi-68.4 (Kirk et al., 2017). There is no sequence similarity between these RBP and Gp22 amino acid sequence.

Here we report the identification of the RBPs from CDHS-1 and describe its three-dimensional structure. CDHS-1 is particularly effective on the epidemic 027 ribotype (Nale *et al.*, 2016). The capsid diameter of CDHS-1 is (70nm), and the tail length is 270 nm (Nale *et al.*, 2016).

Gp22 represents a new class of RBPs; it is a stable homodimer consisting of an N-terminal L-shaped α-helical superhelix domain and a C-terminal Mg^{2+}-binding domain. Protomers of the dimer are asymmetrical with the short arm of one polypeptide interacting with its partner. Additional reciprocal contacts are mediated via an α-helical hairpin within the super helical domain. This is the first study to identify the structure of RBPs for a *C. difficile* phage, and the work provides a mechanistic understanding of host recognition for these phages. It also identifies a novel class of RBPs thus expanding our comprehension of 'Phage receptor binding protein space'.

Results

Putative functions of the CDHS-1 tail proteins

To identifying the RBP's of CDHS-1, the tail putative tail module was analysed using bioinformatic approaches. A sequence analysis of the *C. difficile* CDHS-1 tail module identified four open reading frames that encode putative tail proteins: Gp18, Gp19, Gp21, and Gp22. The functional role of these proteins was investigated using prediction tools HHpred and Phyre2. The *in-silico* analysis revealed that Gp18 is homologous to the Gp19.1 protein from the *B. Subtilis* SPP1 phage (Figure S1). This *B. subtilis* Gp19.1 is a distal tail protein (Dit) (PDB 2X 8K), that forms the central hub of the SPP1phage baseplate, therefore it is likely Gp18 is the CDHS-1 Dit.

The HHpred analysis showed that CDHS-1 Gp19 is homologous to the gp18 protein from *L. Monocytogenes* phage A118 (PDB 3GS9) (Figure S1) and the Gp 44 protein of enteric phage Mu which are all tail-associated lysin-like proteins (Tal). The main function of these proteins is to help phages inject their genetic material into the bacteria by degrading the peptidoglycan layer (Bielmann *et al.*, 2015). Consistent with this putative role, a CDHS-1 Gp19 blastp search revealed that this protein has an endopeptidase domain (residues 7- 378).

CDHS-1 gp21 is homologous with the N-terminal domain of the ORF48 Upper baseplate protein (BppU) (PDB 3UH8) of phage TP901-1 of *L. lactis* (Figure S1). This protein links the RBP to the central baseplate core of *Lactococcus* phage TP901-1 so likely has a similar role in CDHS-1.

Finally, *in silico* analysis of CDHS-1 Gp22 revealed that there was no similarity to any protein in the Protein Data Bank (PDB), thus rendering it of particular interest as the potential RBP for this page. Of note, the blastp search showed that Gp22 is almost identical to the putative tail fibre proteins of other *C. difficile* phages from the database such as PhiCD38.2 and phiCD146 only have a single amino acid difference, and homologous to the putative tail proteins of other phages infecting *C. difficile* phiCD111 has 82% similarity. Interestingly, the C-terminal 89 residues are less well-conserved (64% match) than the N-terminal 529 residues (86% match) (Figure S2). This is a similar pattern to well-studied RBPs in Lactococcal phages: sk1, TP901-1, and bIL170 phages. In each case, the receptor-binding sites are within the C-terminal (non-conserved) region of RBPs.

Phage neutralization using polyclonal antibodies

To determine if Gp22 does indeed determine phage infection, polyclonal antibodies were raised against recombinant Gp22 proteins, and the ability of it to neutralize CDHS-1 infection of C. *difficile* strain CD105LC1 was investigated. The putative upper baseplate protein Gp21 was also overexpressed as a control and to establish if it's position could be determined on the phage particle. Antibodies were added to the phages and used in neutralisation assays. However, pre-incubation with anti-gp22 serum led to complete inhibition of the phage infection, up to a dilution of 1:10000 (Figure 1b). Thus, Gp22 plays an essential role in phage CDHS-1 infection of *C. difficile*. In contrast, pre-incubation of anti-Gp21 serum with CDHS-1 had no impact on phage infection (Figure 1a).

The interactions of recombinant Gp22 protein with CD105LC1 strain

To confirm that Gp22 is the CDHS-1 RBP, *C. difficile* CD105LC1 cells were incubated with glutathione-Sepharose 4B beads (Biosciences, Sweden) coated with either GST alone or Gp22 tagged with GST, and viewed using confocal microscopy under phase contrast. The bacteria cells attached to beads coated with GST-tagged Gp22 (Figure 2 a) but not to the beads coated with GST alone (Figure 2 b), indicating that Gp22 binds directly to *C. difficile*.

When increasing concentrations of recombinant Gp22 were pre-incubated with *C. difficile* CD105LC1 strain and in the presence of CDHS-1 phage, the phage adsorption was inhibited in a dose-dependent manner (Figure 2 c). Thus, it is clear that the Gp22 blocks the binding site on *C. difficile* for phage CDHS-1, confirming that Gp22 is the RBP. Unfortunately, attempts to produce a truncated folded form of Gp22 containing only the C-terminal region were unsuccessful because the truncated Gp22 was consistently degraded during the purification process.

TEM Localisation of Gp22 and Gp21

To identify the precise locations of Gp22 and Gp21 on the phage particle, CDHS-1 was pre-incubated with rabbit anti-Gp21 and anti-Gp22 sera. This was incubated with goat anti-rabbit secondary antibodies coupled with 12 nm gold colloids. The gold-labelled antibodies were seen as black spots under TEM, showing the location of each protein on the phage particle. CDHS-1 was incubated with the pre-immune sera from the rabbits as a negative control. The TEM showed that Gp21 and Gp22 are located on the tail baseplate of CDHS-1 (Figure 3a and b). As expected, only a random scattering of black spots was observed on the grids containing CDHS-1 incubated with the pre-immune sera (Figure 3d and e).

Structure of Gp22 (RBP)

To determine the 3D structure of Gp22 protein. Selenomethionine-enriched Gp22 protein was produced in *E. coli* and purified by affinity and gel filtration chromatography. Gel filtration analysis shows that Gp22 is a stable dimer in solution with an apparent molecular mass of ~130 kDa (expected mass of a dimer 136.4 kDa) (Figure 4a). Crystals of selenomethionine-Gp22 were grown at pH 7 in buffer containing Mg^{2+}. The phases were solved by selenomethionine single-wavelength anomalous dispersion, and data diffracted to a 2.5 Å resolution (Table 1). Four copies of Gp22 were observed in the asymmetric unit, all forming dimers through association with adjacent polypeptides about the crystallographic symmetry axes. Each polypeptide comprises an N-terminal L-shaped α-helical superhelix domain (529 residues) and a C-terminal β-sandwich domain. The rod like stems of the L-shaped super helical domain is formed from short parallel α-helices (each of nine residues) with three helices per turn. A hairpin-like structure formed from a coiled-coil of α-helices (Figure 4b) is located in the middle of the superhelix.

The C-terminal domain contains two metal-binding sites. The coordination geometry is compatible with Mg^{2+} (Figure 4 d), present in the crystallization buffer. Potential coordination ligands include the side chains of Asp540, Gln602, and Asp605, and the main-chain carbonyl groups of Gly606 and Asp540 for one Mg^{2+} and the side chains of Asp540 and Asp605 for the second, more peripheral Mg^{2+}. Caution is needed in assigning coordination ligands given that the data were only of medium resolution and only certain water molecules were observed in the density.

The total dimer interface covers 5523 A^2 of surface, 2760 A^2 from one polypeptide, and 2763 A^2 from its partner. It is formed interactions between the two α-helical hairpin-like structures that pack together to form a 4-helix bundle. Addition contacts are made between the N-terminal portions of each superhelix domain. These domains come into contact through rotation of the short arms of the L-shaped polypeptides (by ~60°) in opposite orientations relative to the α-helical hairpin, thus creating an asymmetrical interface (Figure 4 c).

The role of Mg^{2+} in CDHS-1 phage infection

Many phages require divalent cations ions to infect their bacterial hosts. To determine if divalent cations have a role in the CDHS-1 infection, phage were added on a lawn of CD105LC1 in the presence of Mg^{2+} and/or Ca^{2+}. The data show that CDHS-1 was only able to infect when the Mg^{2+} was present in the media (Figure S3).

The S-layer proteins of *C. difficile* contain the receptors for CDHS-1

To identify if S-layer proteins also act as phage receptors for CDHS-1, was added to lawns of the S-layer deficient mutant FM2.5. The CDHS-1 could not infect this strain (Figure 5 b), however, infection was observed when the wild type R2029 strain was used (Figure 5 a). Thus, the receptors for CDHS-1 do indeed, reside within the S-layer proteins.

Discussion

The recent identification of a plethora of novel and diverse phages has not been matched by an understanding of their biology or the way by which they interact with bacterial cells. The reason for this mismatch is that the time taken to obtain sequence data is significantly less than that needed for structural studies. Furthermore, this data is key as it will improve our understanding of phage bacterial interaction and more importantly is downstream exploitation of phage for therapeutic purposes (Dowah and Clokie, 2018). One of the challenges of the phage therapy applications is the development of various phage resistance mechanisms via the bacterial host. Among these mechanisms, prevention of phage adsorption and DNA injection, restriction enzymes, and CRISPR/Cas systems (Principi, Silvestri and Esposito, 2019). one of particular interest here is the prevention of the phage attachment. And to overcome this mechanism is by incorporating several phages with different host-range specificities within a cocktail (Yang *et al.*, 2020). Therefore, understanding phage-host interactions on a molecular level is crucial prior to phage therapy applications.

The phage proteins that mediate the bacterial phage interaction and determine phage specificity are the RBPs. They have been characterized for several phages that target Gram negative bacteria. This includes the *E. coli* T4, whose RBPs are well describe and understood (Mahony and van Sinderen, 2012) in addition to, lambda phage and T5 phage that have a monomeric RBP (Ppb5) that binds to the *E. coli* receptor FhuA. Pb5 exhibits an elongated shape in solution, and it is unlike any protein of known structure (Flayhan *et al.*, 2012; Goulet *et al.*, 2020).

Relatively little is known about RBP's from phages that infect Gram-positive bacteria and less still for those that target *C. difficile*. The other RBPs from Siphoviruses that target Gram-positive bacteria that have been determined are ORF18 of phage P2, ORF 49 of phage Tp901-1, ORF20 from phage1358, ORF20 of bIL170, and ORF53 from Tuc2009 all of

which infect *L. lactis* (Dunne *et al.*, 2018) and Gp 45 from phage φ11 that infects *S. aureus* (Koç *et al.*, 2016). all known structures here, form trimeric complexes, with each monomer having a modular organization consisting of head, neck, and stem/shoulder domains, in which the shoulder domain contains the binding site for the bacterial host (Sciara *et al.*, 2008; Koç *et al.*, 2016; Dunne *et al.*, 2018).

The amino acid sequence comparison of those structurally determined RBPs reveal that, the RBPs of bIL170 and P2 phages are 89% identical at the N-terminal level and less identity found at C-terminal level which explained the diversity of the host range for the two phages (Dowah and Clokie, 2018). On the other hand, RBPs of phage φ11infectng *S. aureus* organised in three parts too "stem", a "platform" and a "tower". Only the first 30 amino-acids of the stem resemble those of phages TP901-1 or Tuc2009 (Koç *et al.*, 2016).

Here we demonstrated that gp22 is the RBP for the Siphovirus CDHS-1 and that this protein is essential for a successful infection of *C. difficile.* This is the first description of a RBP of a phage that infects *C. difficile.* Notably, CDHS-1, infects the virulent strain 027 (Nale *et al.*, 2016) so could potentially be developed as a novel antibiotic-independent therapeutic. the structure of Gp22 reported here is unlike any previously structurally described RBPs, and thus represents a new class of receptor binding proteins.

Gp22 is a U-shaped homodimer stabilised via a central crossbar formed from a four-helix bundle. Each subunit comprises an N-terminal L-shaped α-helical superhelix domain, and a C-terminal β-sandwich domain. The protomers of each dimer are asymmetrical with the short arms rotated in opposite orientations to form the base of the U-shape. Such asymmetry is not uncommon in phage proteins. Gp22 was crystallised in the presence of Mg^{2+} two cations, compatible with Mg^{2+} are bound at the tip of the β-sandwich domain. This likely explains the observation that , infection by CDHS-1 is also Mg^{2+}-dependent.

We showed that CDHS-1 could not infect the S-layer mutant FM2.5 when CDHS-1 was spotted on the strain, indicating that the CDHS-1 receptors within the S-layer. The majority of phages infecting Gram-positive bacteria and have their receptors on the surface of the bacteria being identified have receptors within two main components on the surface of Gram positive bacteria; teichoic acid or peptidoglycan (Bertozzi Silva, Storms and Sauvageau, 2016). To date, no work has been done to indicate that the S-layer proteins may act as phage receptors on *C. difficile.*

This is the first study to identify the RBP from phages that infect *C. difficile.* The results obtained from this study establishes a solid basis for understanding how phage attaches to *C. difficile.* Furthermore, the development of sensitive affinity-based infection diagnostics and therapeutics for this organism. Furthermore, a similar study ongoing for myoviruses that infect *C. difficile.*

Importantly this work changes the paradigm of existing phage receptor binding proteins having a specific structure, one of the key messages from the study is that there are likely to be completely novel classes of receptor binding proteins in existence which can only be identified through this type of biochemical and structural biological approaches. Interestingly, performing such structural studies was not trivial, Gp22 was quick to form a crystal however to produce a high-quality crystal that are able to diffract at a high resolution, it took extended period of time (3 years) with various optimisation steps. The key steps that influence Gp22 crystals to diffract well were using, The BCS (Basic Chemical Space) Screen, Additive screen and seeding technique.

Methods:

Bacteria, phages, and culturing

The *C. difficile* strain used in this study for propagating phage CDHS-1 is CD105LC1 of ribotype 027, in addition to CDR20291 of ribotype 027, They both were from our Laboratory collections, and they have been previously described in detail (Nale *et al.*, 2016; Thanki *et al.*, 2018). Moreover, FM2.5 SlpA knock out (truncate SlpA at a site N-terminal to the post-translational cleavage site and, thereby, prevent the formation of an S-layer) derivative of R20291. *C. difficile* culturing and phage propagation were carried out according to a previously published protocol (Shan *et al.*, 2012). Briefly, *C. difficile* was grown under anaerobic conditions (10% H_2, 5% CO_2, and 85% N_2) on Brain-Heart Infusion (BHI, Oxoid, Basingstoke, UK) 1% agar plates, supplemented with 7% defibrinated horse blood (DHB, TCS Biosciences Ltd., Buckingham, UK) for 24 hours at 37°C. Liquid cultures were prepared by taking a single colony from the blood agar plate and then inoculated in a bijou tube containing 5 ml of Fastidious Anaerobic Broth (FAB; BioConnections, Kynpersley, UK). The liquid cultures were left to grow anaerobically overnight. Then, 500 µl of overnight FAB culture was inoculated into 50 ml of pre-reduced BHI and incubated until an OD_{550} of 0.2 was reached. Subsequently, 500 µl of phage CDHS-1was added to the culture and incubated for an additional 24 hours, followed by centrifuging at 3,400-x *g* for 10 min. The resulting supernatant was filtered using a 0.22 µm filter, and the phage titer in the filtrate was determined using a spot test assay according to a peer-reviewed protocol (Nale *et al.*, 2016).

Protein Expression and Purification of the baseplate proteins (gp21 & gp22) for CDHS-1 phage

DNA extraction was carried out as previously described (Nale *et al.*, 2016). Post DNA extraction, both *gp21 & gp22* were amplified using PCR. The primers used for the PCR are; 5- GTGATAAATTTGAGAGATAG-3 and 5-TTAACTCACCTCTTCTTTTATTTC-3 targeting *gp21*gene, and 5-TACTTCCAATCCATGAGTTGGGCGGAGACATACAAAG-3 and 5-TATCCACCTTTACTGTCATTAAATTGCTTGATACATTGCGTAA-3 to amplify *gp22* gene. Then, the amplified genes (*gp21* and *gp22*) were cloned into pET-based expression plasmids with the help of the cloning service (PROTEX) based at the University of Leicester. The resulting plasmids were used to transform *E. coli* BL21 (DE3) and an established protocol of protein expression in *E. coli* using isopropyl β-D-1-thiogalactopyranoside (IPTG) as inducer was followed (Campanacci *et al.*, 2010). The proteins were purified using Affinity chromatography purification on Glutathione Sepharose 4B beads affinity column (GE Healthcare). After that, the proteins were further purified by gel filtration on 200 16/60 columns (GE Healthcare). In 20 mM Tris pH 7.5, 20 mM NaCl and, then concentrated by filtration using a 10- kDa molecular mass cut off the membrane (Amicon) before further usage.

Phage Neutralisation

To determine which phage CDHS-1 tail protein was responsible for binding to *C. difficile*, the purified gp21 and gp22 proteins were sent to the Eurogenetec Company (Brussels, Belgium) for the generation of polyclonal antibodies. The resulting anti-serum antibodies were used in a 'phage neutralization test to determine which of these anti-sera would be able to neutralize

the phage infection. The assay was carried out according to a published protocol (Li *et al.*, 2016). Briefly, each antibody serum was diluted using SM buffer (10 mM NaCl, 8 mM, $MgSO_4.7H_2O$, and 50 mM Tris-HCL pH 7.5), into 1:10, 1:100, 1:1000 and 1:10000. Then 10^5 phage CDHS-1 was added to each of the dilutions mentioned above. The mixture was incubated for 20 minutes at 37 °C. Then the mixture was serially diluted using SM buffer and spotted on the lawn of the *C. difficile* strain used.

The interaction of gp22 protein with CD105LC1 strain

To determine if the recombinant gp22 interacts with CD105LC1, an adsorption inhibition assay was performed as in (Thanki *et al.*, 2018). Briefly, a strain grown in anaerobic conditions. Phage CDHS-1 (10^7 PFU/ml) and different concentration of gp22 proteins of 400µg, 200µg, 50µg, 0µg were added into CD105LC1 culture with an OD_{600} of around 0.2, respectively. After an incubation time of 30 minutes at 37 °C, a spot test was carried out to determine the phage titer. To visualize the direct interaction between gp22 and CD105LC1 Glutathione Sepharose 4B beads coated with either GST-tagged gp22 protein or only GST. Then they were mixed with CD105LC1 cultures and incubated at 4°C for 30 minutes before confocal microscope. This assay was done as described in (Stevens *et al.*, 2005).

Immune gold labeling

Immuno-gold labeling was carried out to localize the phage tail protein gp21 and gp22 on the phage particle. The assay was done as described previously with slight modification (Li *et al.*, 2016). In an Eppendorf tube, each antibody serum was diluted 1:100 using SM buffer and then incubated with 10^9 PFU/ml CDHS-1 phage for 20 minutes at room temperature. Then the mixture was applied to glow discharged, carbon-coated grids followed by grid washing with SM buffer for 10 minutes. Then a 1:30 diluted goat anti-rabbit IgG coupled with 12 nm gold colloids (Dianova, Hamburg) was added onto the grids and left for 20 minutes. Finally, the grids were negatively stained with 1% (w/v) uranyl acetate before examination under the TEM.

Gp22 crystallization

To determine gp22 structure, selenomethionine labeling was performed using the inhibition of the methionine pathway (Doublié, 2007).

Briefly, *E. coli* BL21 (DE3) was grown overnight in 5 ml of LB. Then the cells were centrifuged for 5 minutes at 1300 x *g*. After that, the pellet was resuspended in 1 ml of M9 (Molecular Dimensions, UK) (M9 prepared according to the manufacturer's guide) and added to 1 L of the same medium (M9). After this, the cells were left to grow in a 37 °C shaking incubator until they reached OD_{600} of 0.2, then amino acids were added; lysine, phenylalanine, and threonine at a final concentration of 100 mg/ml, and isoleucine, leucine, and valine at a final concentration of 50 mg/ml. Thereafter, the cells were grown until an optical density OD_{600} of 0.4 at 37°C, then IPTG was added at a concentration of 0.5 mM; finally, cells were incubated at 17°C overnight. Selenomethionine-gp22 protein purified as described above.

The purified Selenomethionine-gp22 protein was crystallized using the sitting-drop vapour diffusion technique. The crystals were grown in 0.1M HEPES pH 7, containing 0.05M ammonium acetate, 0.15M magnesium sulphate heptahydrate, 12% PEG smear medium, and 4% acetone. Then the crystals were transferred to a reservoir solution containing 30 % glycerol as a cryo-protectant before freezing in liquid nitrogen. The diffraction data were collected from the Diamond light source.

References

Bertozzi Silva, J., Storms, Z. and Sauvageau, D. (2016) 'Host receptors for bacteriophage adsorption', *FEMS Microbiology Letters*, 363(4), pp. 1–11. doi: 10.1093/femsle/fnw002.

Bielmann, R. *et al.* (2015) 'binding proteins of Listeria monocytogenes bacteriophages A118 and P35 recognize serovar-specReceptorific teichoic acids', *Virology*. Elsevier, 477, pp. 110–118. doi: 10.1016/j.virol.2014.12.035.

Burke, K. E. and Lamont, J. T. (2014) 'Clostridium difficile infection: A worldwide disease', *Gut and Liver*, 8(1), pp. 1–6. doi: 10.5009/gnl.2014.8.1.1.

Campanacci, V. *et al.* (2010) 'Solution and electron microscopy characterization of lactococcal phage baseplates expressed in Escherichia coli', *Journal of Structural Biology*, 172(1), pp. 75–84. doi: 10.1016/j.jsb.2010.02.007.

Doublié, S. (2007) 'Production of selenomethionyl proteins in prokaryotic and eukaryotic expression systems', *Methods in Molecular Biology*, 363, pp. 91–108. doi: 10.1385/1-59745-209-2:91.

Dowah, A. S. A. and Clokie, M. R. J. (2018) 'Review of the nature, diversity and structure of bacteriophage receptor binding proteins that target Gram-positive bacteria', *Biophysical Reviews*. Biophysical Reviews, 10(2), pp. 535–542. doi: 10.1007/s12551-017-0382-3.

Dunne, M. *et al.* (2018) 'Molecular basis of bacterial host interactions by gram-positive targeting bacteriophages', *Viruses*, 10(8). doi: 10.3390/v10080397.

Flayhan, A. *et al.* (2012) 'New insights into pb5, the receptor binding protein of bacteriophage T5, and its interaction with its Escherichia coli receptor FhuA', *Biochimie*, 94(9), pp. 1982–1989. doi: 10.1016/j.biochi.2012.05.021.

Goulet, A. *et al.* (2020) 'Conserved and diverse traits of adhesion devices from siphoviridae recognizing proteinaceous or saccharidic receptors', *Viruses*, 12(5), pp. 1–21. doi: 10.3390/v12050512.

Hargreaves, K. R. and Clokie, M. R. J. (2014) 'Clostridium difficile phages: Still difficult?', *Frontiers in Microbiology*, 5(APR), pp. 1–14. doi: 10.3389/fmicb.2014.00184.

Javed, M. A. *et al.* (2013) 'Bacteriophage Receptor Binding Protein Based Assays for the Simultaneous Detection of Campylobacter jejuni and Campylobacter coli', *PLoS ONE*, 8(7). doi: 10.1371/journal.pone.0069770.

Koç, C. *et al.* (2016) 'Structure of the host-recognition device of Staphylococcus aureus phage Φ11', *Scientific Reports*. Nature Publishing Group, 6(June), pp. 1–11. doi: 10.1038/srep27581.

Le, S. *et al.* (2013) 'Mapping the Tail Fiber as the Receptor Binding Protein Responsible for Differential Host Specificity of Pseudomonas aeruginosa Bacteriophages PaP1 and JG004', *PLoS ONE*, 8(7), pp. 1–8. doi: 10.1371/journal.pone.0068562.

Li, X. *et al.* (2016) 'An essential role for the baseplate protein Gp45 in phage adsorption to Staphylococcus aureus', *Scientific Reports*. Nature Publishing Group, 6(May), pp. 1–11. doi: 10.1038/srep26455.

388 Mahony, J. and van Sinderen, D. (2012) 'Structural aspects of the interaction of dairy phages with
389 their host bacteria', *Viruses*, 4(9), pp. 1410–1424. doi: 10.3390/v4091410.

390 Nale, J. Y. *et al.* (2016) 'Bacteriophage combinations significantly reduce Clostridium difficile growth
391 in vitro and proliferation in vivo', *Antimicrobial Agents and Chemotherapy*, 60(2), pp. 968–981. doi:
392 10.1128/AAC.01774-15.

393 Principi, N., Silvestri, E. and Esposito, S. (2019) 'Advantages and limitations of bacteriophages for the
394 treatment of bacterial infections', *Frontiers in Pharmacology*, 10(MAY), pp. 1–9. doi:
395 10.3389/fphar.2019.00513.

396 Rakhuba, D. V. *et al.* (2010) 'Bacteriophage receptors, mechanisms of phage adsorption and
397 penetration into host cell', *Polish Journal of Microbiology*, 59(3), pp. 145–155. doi: 10.33073/pjm-
398 2010-023.

399 Sciara, G. *et al.* (2008) 'A topological model of the baseplate of lactococcal phage Tuc2009', *Journal
400 of Biological Chemistry*, 283(5), pp. 2716–2723. doi: 10.1074/jbc.M707533200.

401 Shan, J. *et al.* (2012) 'Prophage carriage and diversity within clinically relevant strains of Clostridium
402 difficile', *Applied and Environmental Microbiology*, 78(17), pp. 6027–6034. doi: 10.1128/AEM.01311-
403 12.

404 Simpson, D. J., Sacher, J. C. and Szymanski, C. M. (2016) 'Development of an assay for the
405 identification of receptor binding proteins from bacteriophages', *Viruses*, 8(1). doi:
406 10.3390/v8010017.

407 Stevens, J. M. *et al.* (2005) 'Actin-binding proteins from Burkholderia mallei and Burkholderia
408 thailandensis can functionally compensate for the actin-based motility defect of a Burkholderia
409 pseudomallei bimA mutant', *Journal of Bacteriology*, 187(22), pp. 7857–7862. doi:
410 10.1128/JB.187.22.7857-7862.2005.

411 Thanki, A. M. *et al.* (2018) 'Unravelling the links between phage adsorption and successful infection
412 in Clostridium difficile', *Viruses*, 10(8), pp. 1–15. doi: 10.3390/v10080411.

413 Vardakas, K. Z. *et al.* (2012) 'Treatment failure and recurrence of Clostridium difficile infection
414 following treatment with vancomycin or metronidazole: A systematic review of the evidence',
415 *International Journal of Antimicrobial Agents*. Elsevier B.V., 40(1), pp. 1–8. doi:
416 10.1016/j.ijantimicag.2012.01.004.

417 Waseh, S. *et al.* (2010) 'Orally administered P22 phage tailspike protein reduces Salmonella
418 colonization in chickens: Prospects of a novel therapy against bacterial infections', *PLoS ONE*, 5(11).
419 doi: 10.1371/journal.pone.0013904.

420 Yang, Y. *et al.* (2020) 'Development of a Bacteriophage Cocktail to Constrain the Emergence of
421 Phage-Resistant Pseudomonas aeruginosa', *Frontiers in Microbiology*, 11(March), pp. 1–12. doi:
422 10.3389/fmicb.2020.00327.

423

424

425

426

427

428

429

430

431

432

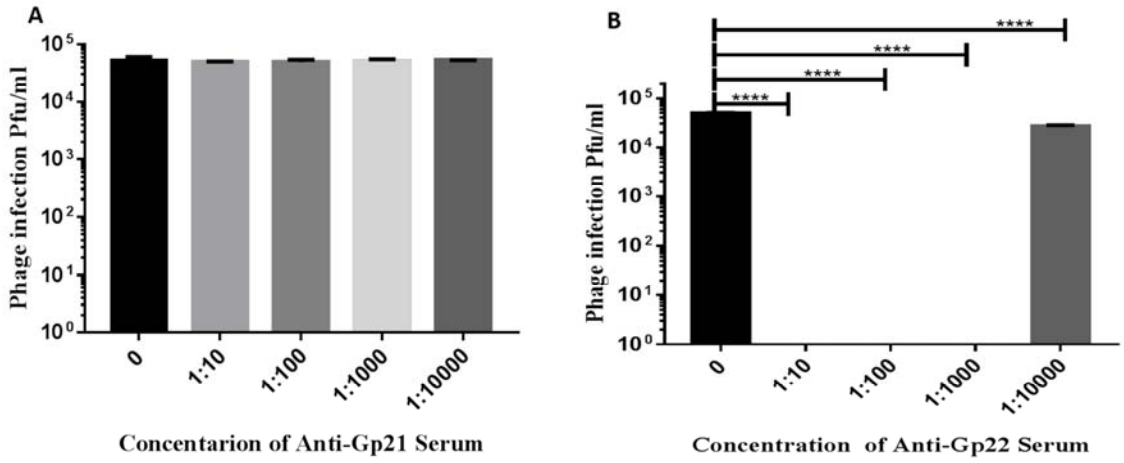

Figure 1: The role of gp21 and gp22 proteins in CDHS1 attachment:

CDHS1 was pre-incubated with different polyclonal antibodies against the tail proteins (gp21 and gp22). The phage was then tested for the ability to infect CD105LC1 strain using spot tests with serial 10-fold dilutions. (A) And (B) represent the plaque-forming units (Pfu/ml) of CDHS1 on CD105LC1 after incubation with different concentration of polyclonal antibody against gp21 and gp22 respectively. The result in (A) shows no significant differences between the positive control (gray column) and the negative control (black column) when anti- gp21 was used. (B) Represents the inhibition of phage CDHS1 infection on CD105LC1 strain, using different concentrations of anti-gp22 serum. The assay was conducted with three biological repeats each with three technical repeats. Error bar represent means ± standard deviations (SD, n = 3). Statistical differences calculated by two-way ANOVA.

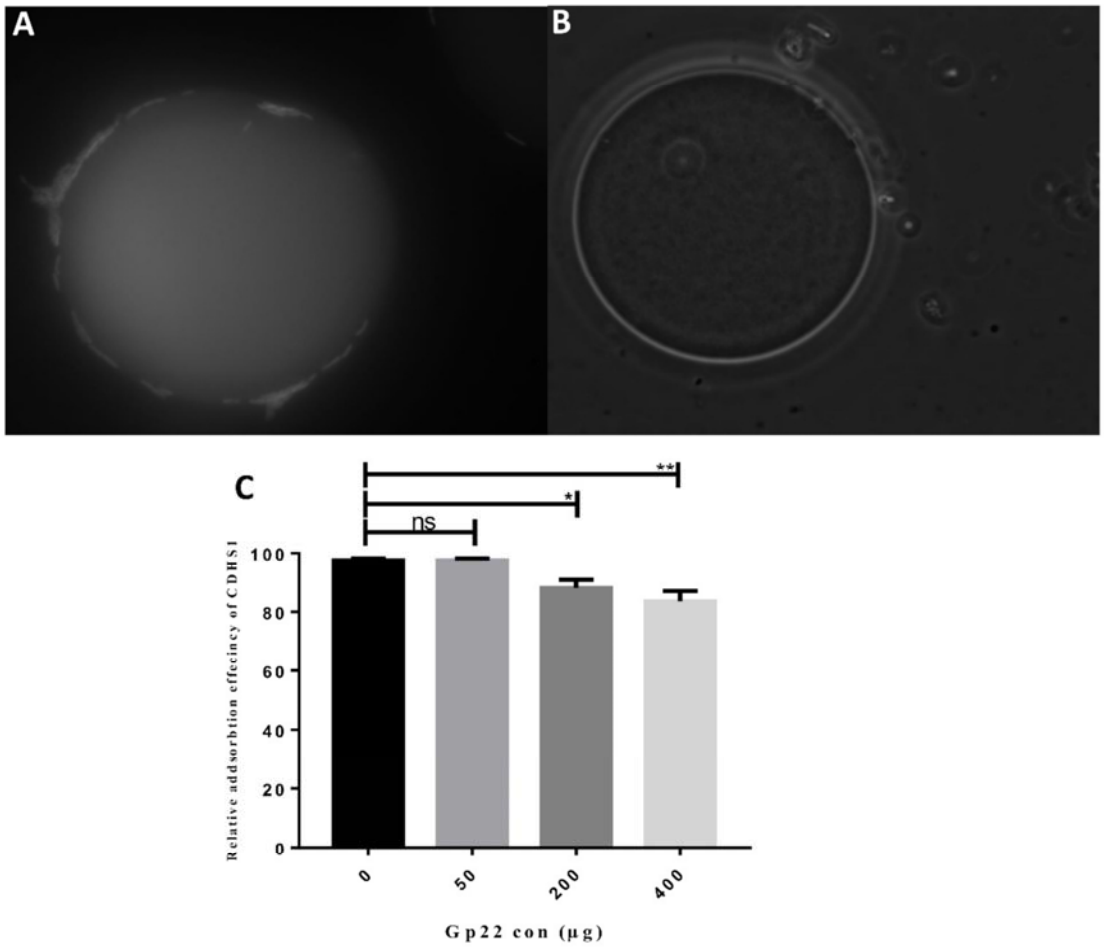

Figure 2: The attachment of gp22 protein with CD105LC1 strain.

(A) is a Confocal microscope image taken post-incubation of CD105LC1 strain with the Glutathione sepharose beads coated with gp22 protein tagged with GST. The result shows that the CD105LC1 cells attached to the beads coated with gp22 tagged with GST and indicates the gp22 protein has role in phage binding with *C. difficile*. As a negative control, the CD105LC1 cells pre-incubated with Glutathione sepharose beads coated with GST only and the result shows no attachment in image (B). (C) is a Dose-dependent inhibition of CDHS1 adsorption with recombinant gp22.

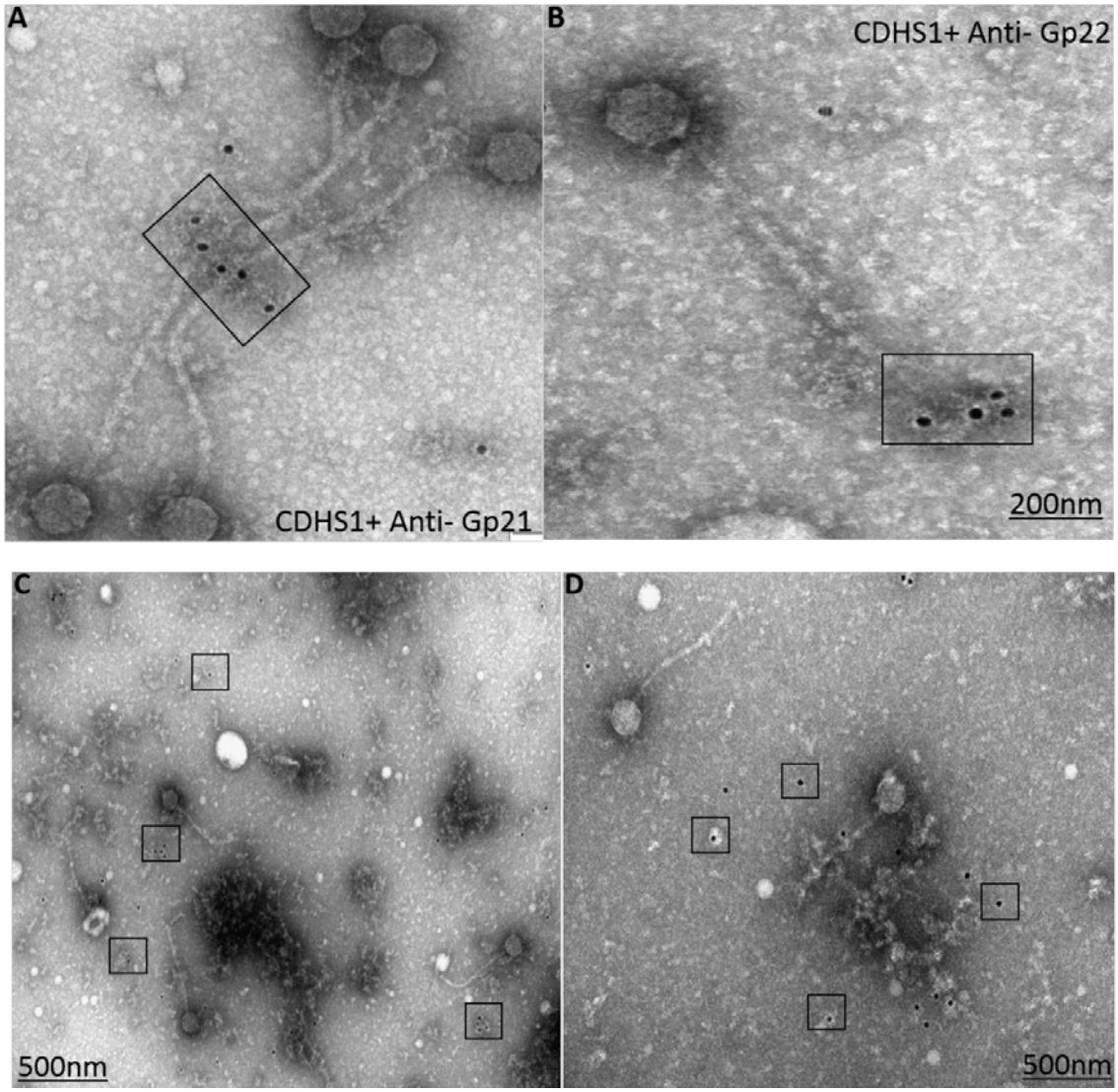

Figure 3: Immunogold labelling of tail proteins, gp21 and gp22.

Transmission electron microscopy (TEM) images of negatively stained phage CDHS1 after immunogold labelling with Anti- gp21 serum (A) and Anti -gp22 serum (B). (C) and (D) are the negative control (pre-immune sera for each antibody raised).

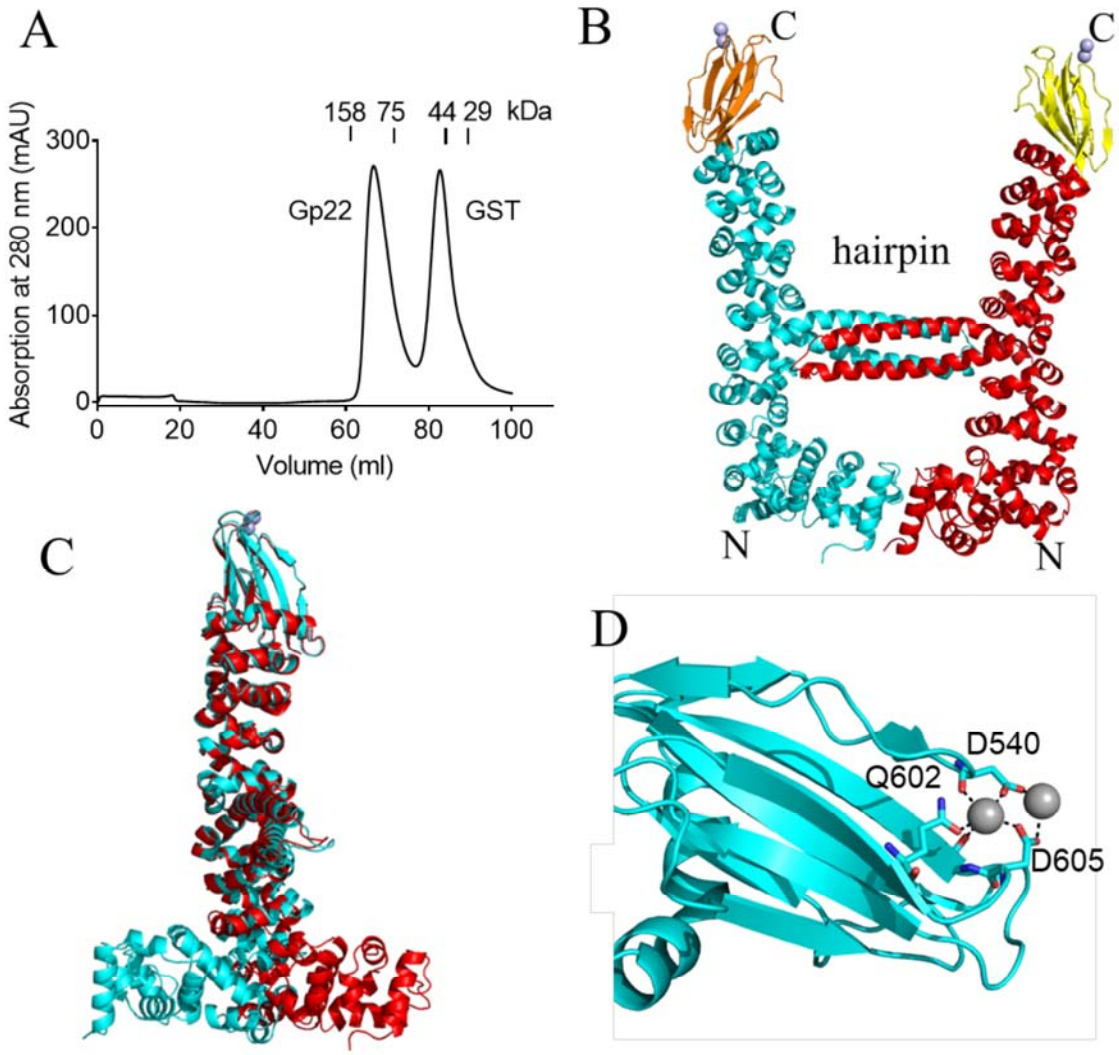

Figure 4 Gp22 is a stable homodimer. A, gel filtration of the gp22-GST fusion following cleavage by TEV protease. gp22 elutes with an apparent molecular mass of ~130 kDa from a Superdex 200 16/60 column based on the elution of molecular mass standards indicating that it is a dimer (monomer molecular mass = 68.4 kDa). GST is also dimeric as expected with a molecular mass of ~50 kDa B, the crystal structure of the gp22 dimer. The α-helical superhelix domains are in red, cyan, and the β-sandwich domains in yellow and orange. N- And C- termini are indicated C, overlay of the gp22 polypeptides of the dimer. The β-sandwich domains and the C-terminal portions of the superhelix domains superpose closely, however the N-terminal portions are rotated by ~60° in opposite directions. D, the β-sandwich domain showing bound Mg^{2+}. Putative coordination ligands are indicated.

Table 1. Data collection and refinement statistics.

Data collection	
Beamline	Diamond Light Source I03
Wavelength, *Å*	0.9793
Space group	P 2_1 2_1 2
a, b, c, Å	178.2, 227.5, 114.9
α, β, γ, °	90, 90, 90
Resolution, *Å*	140.3 - 2.52 (2.82 - 2.52)
No. reflections	104702 (5235)
R*sym*	0.167 (1.07)
CC(1/2)	1.0 (0.7)
I/δI	9.1 (1.6)
Completeness	95.8 (73.3)
Redundancy	6.6 (7.1)
Refinement	
Resolution, *Å*	140.3 - 2.52 (2.61 - 2.52)
No. reflections	104684 (291)
Multiplicity	6.6 (7.1)
R*work*/R*free*	18.8/22.8
No. atoms	18877
Protein	18738
Ligands	8
Water	133
B-factors, $Å^2$	46.4
Protein	45.7
Ligand	65.8
Water	40.5
RMS (bonds) Å	0.012
RMS (angles) °	1.45

Statistics for the highest-resolution shell are shown in parentheses.

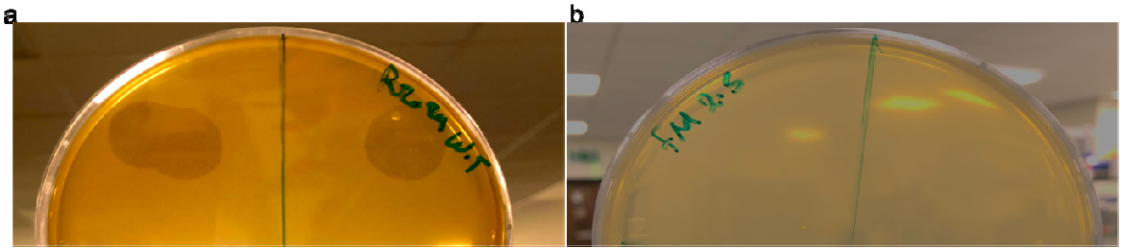

Figure 5: **CDHS1 phage infection to CDR20291 and FM2.5 SlpA knock out derivative of R20291:** CDHS1 Spotted on a loan of CDR20291(a) and FM2.5 (b)strains, CDHS1 was not able to infect the FM2.5 (b). As it did when spotted on the wild type CDR20291 (a). That is indicate that receptors that CDHS1 targeted are within the S-layer protein of CDR2029.