Title	Functional carbohydrate binding modules identified in evolved Dits from siphophages infecting various Gram-positive bacteria
Authors	Hayes, Stephen;Vincentelli, Renaud;Mahony, Jennifer;Nauta, Arjen;Ramond, Laurie;Lugli, Gabriele A.;Ventura, Marco;van Sinderen, Douwe;Cambillau, Christian
Publication date	2018-09-11
Original Citation	Hayes, S., Vincentelli, R., Mahony, J., Nauta, A., Ramond, L., Lugli, G. A., Ventura, M., van Sinderen, D. and Cambillau, C. (2018) 'Functional carbohydrate binding modules identified in evolved Dits from siphophages infecting various Gram-positive bacteria', Molecular Microbiology. doi:10.1111/mmi.14124
Type of publication	Article (peer-reviewed)
Link to publisher's version	10.1111/mmi.14124
Rights	© 2018, John Wiley & Sons Inc. This is the peer reviewed version of the following article: Hayes, S., Vincentelli, R., Mahony, J., Nauta, A., Ramond, L., Lugli, G. A., Ventura, M., van Sinderen, D. and Cambillau, C. (2018) 'Functional carbohydrate binding modules identified in evolved Dits from siphophages infecting various Gram-positive bacteria', Molecular Microbiology. doi:10.1111/mmi.14124, which has been published in final form at https://doi.org/10.1111/mmi.14124. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.
Download date	2025-01-02 20:51:25
Item downloaded from	https://hdl.handle.net/10468/6822

University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

DR. CHRISTIAN CAMBILLAU (Orcid ID : 0000-0001-5502-4729)

Article type : Research Article

Functional Carbohydrate Binding Modules Identified in evolved Dits from Siphophages Infecting various Gram-positive Bacteria

Stephen Hayes[1#], Renaud Vincentelli[2,3#], Jennifer Mahony[1], Arjen Nauta[4], Laurie Ramond[2,3], Gabriele A. Lugli[5], Marco Ventura[5], Douwe van Sinderen*[1,6] and Christian Cambillau*[1,2,3]

[1] School of Microbiology, University College Cork, Cork, Ireland.

[2] Architecture et Fonction des Macromolécules Biologiques, Aix-Marseille Université, Campus de Luminy, Marseille, France.

[3] Architecture et Fonction des Macromolécules Biologiques, Centre National de la Recherche Scientifique (CNRS), Campus de Luminy, Marseille, France.

[4] FrieslandCampina, Amersfoort, The Netherlands.

[5] Laboratory of Probiogenomics, Department of Life Sciences, University of Parma, Parma, Italy.

[6] APC Microbiome Ireland, University College Cork, Cork, Ireland.

#Contributed equally to the work.

*Corresponding authors: Douwe van Sinderen or Christian Cambillau

emails: d.vansinderen@ucc.ie or cambillau@afmb.univ-mrs.fr

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/mmi.14124

Keywords: Bacteriophage, *Lactococcus lactis*, receptor-binding protein, distal tail protein, carbohydrate binding module, phage-host interactions.

Abstract

With increasing numbers of 3D structures of bacteriophage components, combined with powerful *in silico* predictive tools, it has become possible to decipher the structural assembly and associated functionality of phage adhesion devices. Recently, decorations have been reported in the tail and neck passage structures of members of the so-called 936 group of lactococcal siphophages. In the current report, using bioinformatic analysis we identified a conserved carbohydrate binding module (CBM) among many of the virion baseplate Dit components, in addition to the CBM present in the 'classical' receptor binding proteins (RBPs). We observed that, within these so-called 'evolved' Dit proteins, the identified CBMs have structurally conserved folds, yet can be grouped into four distinct classes. We expressed such modules in fusion with GFP, and demonstrated their binding capability to their specific host using fluorescent binding assays with confocal microscopy. We detected evolved Dits in several phages infecting various Gram-positive bacterial species, including mycobacteria. The omnipresence of CBM domains in Siphophages indicates their auxiliary role in infection, as they can assist in the specific recognition of and attachment to their host, thus ensuring a highly efficient and specific phage-host adhesion process as a prelude to DNA injection.

IMPORTANCE

This work shows that many siphophages, in particular those infecting various Gram-positive hosts, employ auxiliary carbohydrate-binding modules (CBMs) that are present within their distal tail (Dit) protein to adhere to their host. In the case of lactococcal phages the specificity of the observed binding of such a CBM is the same as that of the corresponding receptor binding protein, suggesting that the Dit-associated CBM supports the phage in its efforts to specifically and exclusively bind to a host with a particular cell wall-associated polysaccharide. Overall, this work expands our insights into phage-host interactions of phages that exploit saccharidic receptors.

INTRODUCTION

One of the crucial steps in the infection process of a phage is the specific recognition of, and binding to, its host, followed by DNA injection. Tailed phages (*Caudovirales*) utilize a molecular device at the distal end of their tail for host recognition/binding. Members of the *Myoviridae*, of which phage T4 is the paradigm, possess a capsid and a contractile tail bearing a large molecular device at the tip, called the baseplate, which is responsible for host adhesion (Kostyuchenko *et al.*, 2003, Taylor *et al.*, 2016, Taylor *et al.*, 2018). Members of the most abundant group of the *Caudovirales*, the *Siphoviridae*, possess a capsid to which a long non-contractile tail is attached (Davidson *et al.*, 2012).

Until recently, structural knowledge on *Siphoviridae* was based primarily on the coliphages Lambda (Davidson *et al.*, 2012, Edmonds *et al.*, 2007, Edmonds *et al.*, 2005, Fraser *et al.*, 2006, Maxwell *et al.*, 2002, Pell *et al.*, 2009), T5 (Zivanovic *et al.*, 2014, Arnaud *et al.*, 2017) and HK97 (Cardarelli *et al.*, 2010, Duda *et al.*, 2013), as well as on SPP1 (Plisson *et al.*, 2007, Goulet *et al.*, 2011, Chaban *et al.*, 2015), which infects the Gram-positive bacterium *Bacillus subtilis*. It was established that the three coliphages recognize and attach to outer membrane proteins, and SPP1 to the ectodomain of a large membrane-associated protein (Sao-Jose *et al.*, 2006), using a tail tip component (Breyton *et al.*, 2013). Conversely, most lactococcal phages, among which are those of the so-called 936 group, lack a tail tip, but possess a baseplate that binds to a cell surface-associated polysaccharide (Sciara *et al.*, 2010, Veesler *et al.*, 2012). The X-ray structure of the baseplate from phage p2 (a representative of the 936 group) was determined, revealing that: i) the p2 baseplate is an assembly of three proteins, and ii) it changes conformation upon activation (Sciara *et al.*, 2010) (Fig. 1). This conformational change, brought about by the presence of Ca^{++} and other divalent cations (Mahony *et al.*, 2015), results in the downwards rotation of the Receptor Binding Proteins (RBPs) from an inactive "heads-up" conformation to a downwards facing "active" state. Subsequently, the X-ray structure of the baseplate of phage TP901-1, a member of the lactococcal P335 phage group, revealed that the activation mechanism is not shared by all carbohydrate-binding phages (Veesler *et al.*, 2012, Spinelli *et al.*, 2014). In both cases, however, a unique domain, the RBP head domain (Spinelli *et al.*, 2006b, Spinelli *et al.*, 2006a), was found to be involved in the recognition of the host.

The cell wall-associated polysaccharide (CWPS) of *Lactococcus lactis* (visually present as a so-called 'pellicle' layer around the cell wall (Chapot-Chartier *et al.*, 2010, McCabe *et al.*, 2015), is the host receptor of 936-type phages (and other groups), and its biosynthetic apparatus is encoded by a large gene cluster (called *cwps*) encompassing, when comparing different lactococcal genomes, a highly conserved region and a region of genetic diversity. Comparative analysis of these diversity regions allowed the classification of lactococcal strains to belong to different CWPS types based on the genetic composition of the *cwps* gene cluster: CWPS type A (representative *L. lactis* strains are UC509.9 and CV56), CWPS type B (present in *L. lactis* strains IL1403 and KF147), and CWPS type C (with *L. lactis* strains MG1363/SK11 as representatives) (Ainsworth *et al.*, 2014). Based on a PCR-typing method at least one more CWPS type appears to exist, while a correlation was found to exist between the CWPS type of a given lactococcal strain and the host range of 936-type phages (Ainsworth *et al.*, 2014). Importantly, this correlation was found to be due to a link between the CWPS type of a given strain and the RBP (sequence) of a phage capable of infecting that strain (Mahony *et al.*, 2013).

Recently, certain phages were shown to possess an additional carbohydrate binding module (CBM) in addition to the RBP (Legrand *et al.*, 2016, Dieterle *et al.*, 2016, Dieterle *et al.*, 2017). In phage Tuc2009, another P335 lactococcal phage that resembles TP901-1, such an additional CBM is present within the BppA accessory baseplate protein (Legrand *et al.*, 2016).

Phages of *L. lactis* are classified into ten groups based on morphology and DNA relatedness (Deveau *et al.*, 2006). Among these ten groups, the 936 group are the most frequently encountered in the dairy processing environment and represent a significant threat in this industrial context (Moineau, 2002, Deveau *et al.*, 2006). Here, we report a 'pan' 936 phage group study aimed at analyzing (where relevant) their evolved distal tail protein (Dit), identifying their CBM insertions and determining that they are involved, together with the RBP, in adhesion to the host. In a first step, we used a combination of HHpred analysis and multiple alignments of Dit sub-groups to identify different Dit-associated CBM classes (Zimmermann *et al.*, 2017, Hildebrand *et al.*, 2009). Briefly, HHpred builds a multiple sequence alignment of query-related sequences, even if they exhibit low amino-acid identities. From this alignment, a hidden Markov model profile (HMM) of the query is calculated. This query model is used to identify structural hits with known structures in the PDB database. HHpred provides as output a ranked list of matches (including E-values and

probabilities for a true relationship) and pairwise query-database sequence alignments. In a second step, we expressed the most interesting evolved Dits or domains in fusion with GFP, and tested the binding of these proteins to the phage's specific host(s) in order to validate the functionality of these modules. Finally, we identified the presence of CBMs for several phages infecting Gram-positive bacteria.

RESULTS

Bioinformatic analysis of the Distal Tail (Dit) proteins

The 936 phage group represents an industrially important group of phages that have been thoroughly studied in recent years (Mahony *et al.*, 2006, Castro-Nallar *et al.*, 2012, Rousseau & Moineau, 2009, Murphy *et al.*, 2016). The genome sequences of nearly ninety 936 phages are currently available in public databases and in this study an additional 26 distinctly different members of this phage group were isolated, characterized and sequenced. This provided a total of 115 lactococcal bacteriophages of the 936 group for which genomic sequences are now known and which were analyzed in this study (Table S1).

The availability of 115 genome sequences of a single phage group is rather unique, making the 936 group of phages an excellent model system for the analysis of adhesion domains and modules. The structure of the entire virion of the 936 phage p2 has been resolved by electron microscopy (EM) (Bebeacua *et al.*, 2013), while RBP structures of phages p2 and bIL170, and the complete phage p2 baseplate structure have been determined by X-ray crystallography (Spinelli *et al.*, 2006b, Ricagno *et al.*, 2006), providing a useful resource for the analysis of structural components of these phages. However, as more genome sequences became available, it has become clear that multiple host interactions may be in operation, warranting a close inspection of the adhesion devices of these phages. Ultimately, we wanted to define if adhesion mechanisms employed by 936 phages are similar to those of other genetically and structurally analyzed phages in order to assess the universality or uniqueness of adhesion systems utilized by tailed phages. From the available genome sequences of members of the 936 phage group only those ORFs were selected that encode products corresponding to the phage p2 baseplate Dit component (ORF15 in phage p2) (Sciara *et al.*, 2010).

The Dit sequences were aligned and two types of Dit proteins (or Dits) were discerned based on size differences, termed here as "classical" (~250-300 amino acids) and long (~500 amino acids) or "evolved" Dit types (Dieterle *et al.*, 2016, Dieterle *et al.*, 2017) (Fig. 2). Interestingly, the 135 N-terminal residues and the ~140 C-terminal residues of the long Dits align well with the full length of the classical Dit protein from phage p2. Dit proteins from phages SPP1 (Veesler *et al.*, 2010), TP901-1 (Veesler *et al.*, 2012) and p2 (Sciara *et al.*, 2010), for which the X-ray structures have been determined, are assembled as hexamers with each monomer composed of an N-terminal belt domain (residues 1-132), forming a hexameric ring, and a galectin-like domain protruding from the periphery of the ring (Fig. 1). Noteworthy, Dit_{T5} shares a similar structure in which the galectin domain is replaced by an OB-fold domain (Flayhan *et al.*, 2014). However, compared to Dit_{SPP1} and $Dit_{TP901-1}$, Dit_{p2} possesses a ~40 residue insertion, the "arm" and "hand", within the galectin loop, that serves to attach the RBP to the rest of the baseplate (Sciara *et al.*, 2010). The hand, at the extremity of the arm extension, inserts in a groove of the RBP's trimeric N-terminal or 'shoulder' domains in order to facilitate its attachment to the baseplate ((Sciara *et al.*, 2010); PDB ID 2WZP). Any variation in the Dit arm loop should therefore correlate with corresponding variations in the RBP shoulders. Several Dit residues in the "arm" and "hand" interact with RBP 'shoulders', and are, as might be expected, conserved within the phages analyzed in this study

Internal insertions of ~146-172 residues are observed that make up the length difference between the classical and evolved Dits (Fig. 2). These insertions were aligned and grouped into four distinct classes (Fig. 3). The first class, Class 1, comprises 18 members with an insertion length of 171/172 residues and 24 variable positions (Fig. 3a). Class 2, with 16 members, possesses an insertion of 166 residues and 18 variable residue positions (Fig. 3b). The two members of Class 3 have an insertion length of 151 and 163 residues and exhibit a larger number of variable positions (Fig. 3c). Class 4 (2 members) Dits have a short insertion of 146 residues and 6 variable positions (Fig. 3d).

HHpred analysis was performed on the CBM region of a selection of Dit proteins, one from each class (i.e. phages Phi17, Phi10.5, 712 and PhiL.6). HHpred analysis of the Dit insertions of these four phages reported PDB ID 5E7T as the first hit with probabilities of 98.5-99.3% (Fig. 4). The PDB entry PDB ID 5E7T corresponds in part to the baseplate structure of lactococcal phage Tuc2009, and encompasses the RBP, the C-terminal domain of BppU (a baseplate scaffolding module) and BppA, an extra CBM module attached to BppU

through a titin-like domain linker (Legrand *et al.*, 2016). More precisely, HHpred alignments employing the Dit insertions identified BppA, *i.e.* part of the linker and the complete CBM domain. Other CBM hits were reported with lower probabilities (65-45%) for phage Phi17 Dit insertion (Class 1), all covering the central part (residues 66-147) of the insertion (Fig. 4a). In contrast, BppA was the unique hit of Class 2 phage 10.5 insertion (Fig. 4b). For Class 3, several CBM hits of high probability (90-96%) followed that of BppA, covering residues 24-134 out of 151 (Fig. 4c). The same pattern was observed for Class 4, with CBM hit probabilities of 88-93% covering insertion residues 32-128 (Fig. 4d).

Alignments of phages Phi17, Phi10.5, 712 and PhiL.6 Dit insertions with BppA reveal low but significant identities of 24%, 21%, 21% and 25%, respectively (Fig. S1a). We noticed, however, that a stretch of residues corresponding to BppA residues ~180-210 is much more conserved (65-70%). In BppA, these residues, together with the N-terminal domain (1-15), form a small linker domain between the titin-like and the CBM domain (Fig. S1b). However, in the Dit insertions, the N-terminal part of this small domain was not identified at the N-terminus of the insertion (Fig. S2). We wondered whether this high identity stretch had biased the HHpred result, pushing BppA to the top of the hit list at the expense of other CBM domains. We therefore performed HHpred searches following deletion of residues from the search sequence, revealing that BppA remained the best hit and that our initial results were not biased (Fig. S2).

The borders of the Dit 'junction' that flank the BppA-like CBM module within the long Dits were precisely defined from the sequence alignments between evolved Dits and (the classical) Dit_{p2}. These alignments revealed that insertions are located within the Dit_{p2} hand domain, in the most peripheral part of Dit_{p2}. We therefore modeled a Dit insertion using the Dit_{p2} structure ((Sciara *et al.*, 2010); PDB ID 2WZP) and the Tuc2009 BppA structure ((Legrand *et al.*, 2016); PDB ID 5E7T) (Fig. 1b,c)). This analysis and modelling revealed that the six BppA-like modules (i.e. one for each Dit monomer of the six present in the baseplate) are located at the Dit hexamer periphery and protrude from it, a location that would be compatible with host binding (Fig. 1b,c).

Production of Dit or Dit CBM Insertions and Fluorescent Binding Assays

In order to validate the predictions of the bioinformatics study, we decided to assess the CBM/host association *in vivo* by incubating fluorescently tagged phage proteins or protein domains with host cells, followed by microscopy imaging to determine specific host-binding capabilities. For this purpose, we expressed 5 Dits or Dit domains, encompassing putative CBMs from 3 of the 4 "evolved" Dit classes (attempts to express a Dit protein from Class 3 failed due to insolubility of the protein) as translational N-terminal fusions to a GFP tag at the N-terminus to be visualized by fluorescent microscopy. We also selected and expressed two classical Dits as negative controls (Table 1).

Phi17 full-length Dit (Class 1), Phi10.5 Dit insertion and Phi2R14S full-length Dit (both Class 2), and PhiL.6 full-length Dit and Phi8R06S full length Dit (both Class 4), were all found to bind to their specific host when adding the proteins at a quantity of 5, 25,10, 15, and 15 µg (corresponding to 0.53, 4.8, 1.0, 2.83, and 2.82 µM), respectively (Table 1, Fig. 5). For each Dit, the quantity of protein added corresponds to the minimum required amount to consistently achieve and observe binding. The two Dit controls (*i.e.* with no CBM insertion present, corresponding to phages p2 and 4.2) were found, as expected, to be unable to bind their host (Fig. S3), even at quantities up to 100 µg (6.1 / 3.9 µM, respectively). Those Dits that did elicit host-specific attachment appeared to perform localized binding to the host cell surface, in areas where cell division and septum formation would be expected to occur (Fig. 6). This is highlighted by the similarity of a number of the fluorescent micrographs to images from previous studies on the localization of the bacterial septum (Kramer *et al.*, 2008, Le Bourgeois *et al.*, 2007, Weiss, 2004, Ghigo *et al.*, 1999).

Strain and CWPS specificity

Having verified the expected host-binding capacity of the evolved Dits, we then sought to investigate their specificity towards the host and its putative CWPS receptors. To this end, we examined the binding of phage Phi17 Dit (Dit_{Phi17}; belonging to Class 1) to five hosts that are infected by this phage, and five others that are not. The binding results show that Dit_{Phi17} binds to all examined lactococcal strains that are infected by Phi17, and not to any examined strain that is not infected by Phi17 (Table 2, Fig. 7). In terms of the Dit_{Phi17} receptor, it should be noted that the five infection-positive hosts belong to CWPS type C

(except for strain L, whose CWPS type is unknown), while the five strains that do not act as hosts for Phi17 belong to CWPS types A and B, as well as a non-host type C strain, thus indicating a link between Dit binding to the CWPS type of the host and infectivity.

We then tested the specificity of Dits from phages Phi10.5 and Phi2R14S (belonging to Class 2) towards four hosts and their CWPS receptors and four non-hosts. Here again, the binding results show that these Dits bind to all examined lactococcal strains that are infected by the corresponding phage, while they do not exhibit any binding activity towards strains that are non-hosts (Table 2, Fig. S4, S5). Noteworthy, they bind to hosts that possess a type B CWPS, while they do not elicit binding activity towards strains belonging to CWPS types A or C. The Dits of phages PhiL.6 and Phi8R06S (Class 4) were similarly assayed (Table 2, Fig. S6, S7), and likewise were shown to bind to the four hosts (three C type strains and one undetermined), and none of the four non-host strains (A, B, and C type CWPS). These results therefore suggest that the Dit-associated CBD specifically recognizes a saccharidic component of the CWPS.

From previous findings we know that the RBP of a given 936 group phage specifically recognizes a particular CWPS type (Mahony *et al.*, 2013). Based on this knowledge we decided to perform similar experiments with the RBP of phage Phi17 in order to compare its binding to that of the corresponding Dit, and confirm that the Dit also specifically recognizes a particular CWPS type. To this end, we expressed Phi17 RBP with a mCherry tag at its N-terminus and performed fluorescent binding assays with the same *L. lactis* strains as for the Dit protein (Table 2). We observed exactly the same binding pattern for Phi17 RBP and Dit_{Phi17}, confirming that both proteins specifically recognize a particular CWPS type (Table 2, Fig. 8). Worth noticing, for both the Phi17 RBP and Dit, cell binding appears to be localized, focused on areas where cell growth would be expected to take place (Fig. 6A and F).

Examples of Dit insertions beyond members of the lactococcal phage 936 group

We already mentioned that *Lactobacillus* phages Dit-associated CBM decorations were only recently shown to promote binding to the host (Dieterle *et al.*, 2014, Dieterle *et al.*, 2017). We wondered whether these Dit insertions found in *Lactobacillus* phages and now in members of the *Lactococcus lactis* 936 phage group are even more widespread. To this end,

we randomly examined various siphophages genomes, cherry-picking a few evolved Dits of interest.

We observed the presence of similarinsertions in other lactococcal phages besides those from the 936 group. Lactococcal phage 1706, from the eponymous group, possesses a 547 residue long Dit. HHpred analysis indicates the presence of a long insertion (PDB ID 5LY8) between residues 204 and 472 inside an SPP1-like Dit (Fig. 9). The structure closest to this insertion is the CBM2 module of *Lactobacillus casei* BL23 phage J-1 evolved Dit, the first example of an evolved Dit reported in literature (Dieterle *et al.*, 2017). Furthermore, we identified ~50 phages infecting *Streptococcus thermophilus* that possess a long Dit with a CBM insertion. HHpred analysis of one of these streptococcal phages (phage 0091) reveals a pattern close to that of phage 1706 with the same PDB ID 5LY8 insertion (Fig 9). The same pattern applies to several other phages: *Bacillus cereus* phage BDRD-cer4, *Bacillus thuringiensis* phage ATCC 10782, and *Listeria monocytogenes* phage A006, as identified examples. Phage OE33PA from *Oenococcus oeni* (Philippe *et al.*, 2017) exhibits an insertion with another type of CBM, PDB ID 1GUI, corresponding to the CBM4 from *Thermotoga maritima* (Boraston *et al.*, 2002) (Fig. 10). Several mycobacterial phages are also predicted to possess evolved Dits. Both *Mycobacterium smegmatis* phage Gage and *Mycobacterium avium* phage XTB exhibit insertions related to PDB ID 4B1L and PDB ID 4B1M, describing the structure of CBM66 from *Bacillus subtilis* (Cuskin *et al.*, 2012) (Fig. 9). Thus, it is apparent that the incorporation of these CBM insertions is repeated time and again in phages infecting Gram-positive hosts.

DISCUSSION

In addition to their RBP, many members of the so-called 936 group of bacteriophages harbor CBM-encompassing ‘decorations’ on the Dit baseplate component. Dit decorations have largely been ignored to date, while such decorations on *Lactobacillus* phages were only recently shown to promote binding to the host (Dieterle *et al.*, 2014, Dieterle *et al.*, 2017). These decorations are not specific to *Lactobacillus* or *Lactococcus* phages: our non-exhaustive survey indicates that they are present, and even abundant, in Gram-positive bacterial species, including *Mycobacterium*.

It is also highly gratifying to note that, with a rather limited number of known module structures in the siphophage sphere, our analysis retrieved so many hits against these modules. However, it should be mentioned that the hits retrieved by HHpred generally share low amino-acid identity with their query (15-25%). This indeed tells us that the folds are similar, but that the binding pockets are likely to be very different. This should not be taken as an indication of the CWPS recognized by the hit. Similarly, queries that share the same hit can differ widely, as this is only an indication of a common fold. This is illustrated by our 936 group analysis that distinguished 4 different classes, even though BppA is reported as a common HHpred hit.

Here we also show that the full-length Dits (or Dit CBM) from members of the lactococcal 936 phage group bind to host strains with the same selectivity as the complete virion. Experiments with different infective and non-infective strains have demonstrated that the Dit and RBP of Phi17 bind to the same host(s), and hence almost certainly to the same CWPS. The correlation between the binding abilities of the Dits from phages Phi10.5, 2R14S, PhiL.6, and 8R06S on the one hand and the host ranges exhibited by the full phage on the other indicates that this is the case for many if not all "evolved" Dit proteins. The RBPs of the 936 phages have previously been categorized into five groups (I-V) based on sequence similarity and host preference (Mahony *et al.*, 2013, Murphy *et al.*, 2016). In a further step, each RBP group has been associated with a CWPS type (Mahony *et al.*, 2013, Ainsworth *et al.*, 2014, Murphy *et al.*, 2016). Gathering the results concerning all the 936 groups used in this study, we can extend the RBP/CWPS correlation to a wider Dit/RBP/CWPS correlation (Table 3). For example, we observe that RBP groups IV and I, both associated with CWPS type C, correlate with Dit Classes 1 and 3, respectively, while RBP group III, associated with CWPS type B, is correlated with Dit Class 2 (Table 3). This evolved-Dit/RBP/CWPS correlation suggests that RBP and Dit proteins have evolved in a concerted manner to adapt to the same host-specific CWPS type.

It appears also that these modules bind most specifically to the cellular growth regions. This may be to ensure that a phage preferentially attacks an actively growing bacterium which is more likely to produce a high number of progeny phages in comparison to a non-growing, perhaps starved or dying, stationary bacterial cell. Furthermore, phages from the 936 group can infect their host only during the exponential phase, in contrast to, for example, phages from the P335 group that can infect their host both during the stationary and exponential phases. It has been suggested that these contrasted lifestyles reflect the fact that

phages from the P335 group possess cell wall hydrolysing enzymes (Kenny *et al.*, 2004), while those of the 936 group lack them. The peptidoglycan is cross-linked at stationary phase, and P335 phages need the peptidase activity for their DNA to penetrate the cell wall. TP901-1 mutants lacking the peptidase exhibit abolished infectivity (Stockdale *et al.*, 2013). Reversely, cell wall growth regions are not cross-linked and hence are more favourable to DNA passage.

RBPs are the main host adhesion moieties of phages, as they bring specificity and strong adhesion (McCabe *et al.*, 2015). It was reported that phage adhesion to hosts may occur by means of a two-tier mechanism: an initial CBM-mediated reversible and non-specific adhesion which is intended to scout for favorable hosts, followed by specific and irreversible adhesion involving RBPs. The finding that Dit-associated CBMs exhibit the same specificity as the virion (as elicited by RBPs), suggests that they are not involved in such a mechanism. Instead, we suggest that their position, projected far from the baseplate core, may help to grasp the host in any orientation of the virion, thus allowing the RBPs to deploy and orientate the virion properly for DNA injection, perpendicular to the host cell-wall (see also (Bebeacua *et al.*, 2013)).

We anticipate that this example of structural analysis of phage sequences will stimulate further explorations of lactococcal phages and beyond. The combination of sequence/structure analysis with HHpred prevents functional mis-assignments of (phage) proteins and improves existing annotations. This approach is ideal for the identification of adhesion domains of currently unknown function/structure in order to facilitate a next generation of structure-function studies to advance current knowledge of phage-host interactions of the ubiquitous *Siphoviridae* phages. Therefore, it may be more efficient to concentrate structural efforts on unknown parts of phage proteins, which may allow a cascade of new functional assignments.

METHODS

Hosts and bacteriophages strains

A total of 15 strains of the bacterial species *L. lactis* were used in this study: *L. lactis* strains 2, 3, 4, 5, 8, 9, 10, 13, 17, A, C, E, L, M, NZ9000 (Table S2). Bacterial strains were cultured overnight at 30 °C in M17 broth (Oxoid, Hampshire, UK) supplemented with either

0.5 % w/v lactose (LM17) in the case of strains A-T and 1-20, or 0.5 % w/v glucose (GM17) for all other strains.

Of the 115 phages from the 936 group included, 25 are reported for the first time in this study. Of these, 24 phages were isolated from whey samples collected in 2015 from a Dutch dairy plant (F3) as previously reported (Murphy *et al.*, 2013). An additional phage, phage MP1, was isolated from a waste sample in 2017. Whey sample screening, phage isolation, and phage DNA extraction were performed as described previously (Murphy *et al.*, 2013, Murphy *et al.*, 2016). Whey samples were screened against a selection of forty bacterial strains (designated 1-20 and A-T) previously isolated from two undefined mixed starter cultures via the double-layer plaque assay method (Lillehaug, 1997), with previously described modifications (Murphy *et al.*, 2013, Murphy *et al.*, 2016). Phage isolates were differentiated via multiplex PCR (del Rio *et al.*, 2007), as well as enzyme restriction digests performed on phage DNA as recommended by the manufacturer (Thermo Scientific, United States).

Sequencing of novel phages was performed using the Illumina MiSeq Sequencing System, and genome assembly using MIRA v4.0.2. Open reading frame prediction was performed with Prodigal v2.6, and the quality of the final contigs was improved with Burrows-Wheeler Aligner, SAMtools suite and VarScan v2.2.3. Annotation of the ORFs was performed using BLAST against NCBI databases and HMMER against the PFAM database. The genome sequences of the phages have been deposited in the Genbank database, with accession numbers listed in Table 1. The remaining 936 phage genomes were retrieved from the NCBI database (http://www.ncbi.nlm.nih.gov/) and accession numbers are listed in Table S1.

Bioinformatic analysis

The protein sequence and structure analysis was performed using HHpred (Homology detection & structure prediction by HMM-HMM comparison) using standard mode (Hildebrand *et al.*, 2009, Soding *et al.*, 2005). The sequence alignments were performed with Multalin (Corpet, 1988). Modular visualization and modeling were performed with Coot (Emsley & Cowtan, 2004, Emsley *et al.*, 2010), Chimera (Pettersen *et al.*, 2004) and Pymol (Pymol).

CBM-encoding gene Cloning

Ten DNA regions encoding 8 "evolved" Dits or domains from 6 phages, and two "classical" Dits (short) from two phages, were designed for expression in *Escherichia coli* (Fig. 3). Phage DNA was isolated using the Norgen Biotek Phage DNA Isolation Kit (Ontario, Canada), and the 10 genes or regions of interest were amplified by PCR (KOD Hot Start DNA Polymerase, Merck, Ireland). Purified PCR products were then cloned into a GFP vector using the NZYEasy Cloning Kit (NZYTech genes & enzymes, Portugal). Phi17 RBP gene was amplified by PCR, and the purified PCR product was then cloned into pHTP-mCherry. All recombinant proteins possessed a hexa-histidine tag for purification, as well as a TEV recognition sequence. Recombinant plasmids were transformed into NZY5α competent *E. coli* as previously described (Turchetto *et al.*, 2017). The list of protein-associated DNA sequences is presented in Dataset S2.

Protein production

All steps were carried out in 24-deep-well plates (DW24) with minor modifications of the lab standard protocol (Saez & Vincentelli, 2014), which is described briefly below. The recombinant plasmids were used to transform Rosetta™ (DE3) pLysS *E. coli* cells. For the production, in each DW24; 12 × 2 mL of auto-induction medium (NZYTech genes & enzymes, Portugal) supplemented with kanamycin (50 μg/mL) and chloramphenicol (34 μg/mL) was inoculated (1/40 v/v) for each protein. DW24 plates were then incubated for 24 h at 25 °C in a Microtron shaking incubator (INFORS-HT, Switzerland) (600 rpm). Cells were collected by centrifugation and each well re-suspended in 0.5 mL of lysis buffer (50 mM Tris, 300 mM NaCl, 10 mM imidazole, pH 8.0, 0.25 mg/mL Lysozyme) by 15 minutes shaking at 20 °C in a Microtron shaking incubator (INFORS-HT, Switzerland) (800 rpm). The DW24 plates were frozen at −20 °C. For purification, plates were thawed for 10 minutes at 37 °C in a water-bath followed by 15 minutes shaking at 20 °C in a Microtron shaking incubator. Protein purification was then performed as previously described (Saez & Vincentelli, 2014, Turchetto *et al.*, 2017).

We selected for large-scale production for further study proteins that demonstrated promising results during fluorescent binding assays. In this case, 100 mL of auto-induction medium containing kanamycin (50 μg/mL) and chloramphenicol (34 μg/mL) was inoculated

(1/100 v/v) for each protein. Cultures were then incubated for 24 h at 25 °C in a Multitron Standard shaking incubator (INFORS-HT, Switzerland) (300 rpm). Cells were collected by centrifugation and resuspended in a volume of lysis buffer dependent on the final culture OD of the individual expression, which was then frozen at −80 °C overnight. Cells were thawed, and incubated with 10 µg/mL DNAse and 20 mM $MgSO_4$ for 30 minutes shaking at 24 °C. To ensure complete cell lysis, cells were sonicated (Soniprep 150, MSE, UK) for 5 minutes (power 15, 30 s ON/OFF cycles), and subsequently centrifuged for 30 minutes at 20,000 × *g*. Proteins were then purified using Ni Sepharose 6 Fast Flow resin (GE Healthcare, Sweden) and eluted with 250 mM imidazole elution buffer (50 mM Tris, 300 mM NaCl, 250 mM imidazole, pH 8.0). Purified protein was dialyzed overnight against protein buffer (50 mM Tris-HCl, 300 mM NaCl, pH 8.0), before being frozen at −80 °C until further use.

Fluorescent Binding Assays

Cell binding assays using fluorescently labeled Dit proteins or domains were performed as described previously (Dieterle *et al.*, 2017), with a number of modifications. Briefly, 0.3 ml of the relevant host cells (Fig. 3) in the exponential growth phase were harvested and resuspended in 100 µl of SM buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM $MgSO_4$). Cells were incubated with between 5 µg and 100 µg (as appropriate) of fluorescently labeled protein for 12.5 minutes at 37 ºC. Cells were washed twice in SM buffer, and fluorescent binding was visualized by microscopy. Initial fluorescent binding assays using proteins produced via high-throughput expression were viewed via fluorescent microscopy at a magnification of 60X (Olympus AX70 Provis Upright Research Microscope, Olympus Corporation, Japan). Fluorescent binding assays performed on proteins of particular interest were viewed using confocal microscopy (Zeiss LSM 5 Exciter, Zeiss, Germany) to achieve high-resolution images, with a wavelength of 488 nm for GFP excitation, and 514 nm for mCherry. Images were analyzed and compiled using the Zen 2.3 Lite software package (Zeiss, Germany).

Nucleotide sequence accession numbers

Accession numbers are listed in Table S1.

Acknowledgments

We are grateful to Petr G. Leiman who alerted us to the potential of HHpred in module analysis. We gratefully acknowledge Suzanne Crotty (Department of Anatomy & Neuroscience, University College Cork), Sinéad O'Loughlin (School of Biochemistry and Cell Biology, University College Cork), and Eoghan Casey (School of Microbiology, University College Cork), for their assistance with the microscopy. SH is funded by an Industry Partnership Irish Research Council (IRC) studentship. JM is the recipient of a Starting Investigator Research Grant (Ref. No. 15/SIRG/3430) funded by Science Foundation Ireland (SFI). DvS is supported by a Principal Investigator award (Ref. No. 13/IA/1953) through SFI. This work was supported by the French Infrastructure for Integrated Structural Biology (FRISBI) ANR-10-INSB-05-01.

Competing interests: A.N. is an employee of FrieslandCampina, The Netherlands.

REFERENCES

-Ainsworth, S., I. Sadovskaya, E. Vinogradov, P. Courtin, Y. Guerardel, J. Mahony, T. Grard, C. Cambillau, M.P. Chapot-Chartier & D. van Sinderen, (2014) Differences in lactococcal cell wall polysaccharide structure are major determining factors in bacteriophage sensitivity. MBio 5: e00880-00814.

-Arnaud, C.A., G. Effantin, C. Vives, S. Engilberge, M. Bacia, P. Boulanger, E. Girard, G. Schoehn & C. Breyton, (2017) Bacteriophage T5 tail tube structure suggests a trigger mechanism for Siphoviridae DNA ejection. Nat Commun 8: 1953.

-Bebeacua, C., D. Tremblay, C. Farenc, M.P. Chapot-Chartier, I. Sadovskaya, M. van Heel, D. Veesler, S. Moineau & C. Cambillau, (2013) Structure, Adsorption to Host, and Infection Mechanism of Virulent Lactococcal Phage p2. J Virol 87: 12302-12312.

-Boraston, A.B., D. Nurizzo, V. Notenboom, V. Ducros, D.R. Rose, D.G. Kilburn & G.J. Davies, (2002) Differential oligosaccharide recognition by evolutionarily-related beta-1,4 and beta-1,3 glucan-binding modules. J Mol Biol 319: 1143-1156.

-Breyton, C., A. Flayhan, F. Gabel, M. Lethier, G. Durand, P. Boulanger, M. Chami & C. Ebel, (2013) Assessing the conformational changes of pb5, the receptor-binding protein of phage T5, upon binding to its Escherichia coli receptor FhuA. J Biol Chem 288: 30763-30772.

-Cardarelli, L., R. Lam, A. Tuite, L.A. Baker, P.D. Sadowski, D.R. Radford, J.L. Rubinstein, K.P. Battaile, N. Chirgadze, K.L. Maxwell & A.R. Davidson, (2010) The crystal structure of bacteriophage HK97 gp6: defining a large family of head-tail connector proteins. J Mol Biol 395: 754-768.

-Castro-Nallar, E., H. Chen, S. Gladman, S.C. Moore, T. Seemann, I.B. Powell, A. Hillier, K.A. Crandall & P.S. Chandry, (2012) Population genomics and phylogeography of an Australian dairy factory derived lytic bacteriophage. Genome Biol Evol 4: 382-393.

-Chaban, Y., R. Lurz, S. Brasiles, C. Cornilleau, M. Karreman, S. Zinn-Justin, P. Tavares & E.V. Orlova, (2015) Structural rearrangements in the phage head-to-tail interface during assembly and infection. Proc Natl Acad Sci U S A 112: 7009-7014.

-Chapot-Chartier, M.P., E. Vinogradov, I. Sadovskaya, G. Andre, M.Y. Mistou, P. Trieu-Cuot, S. Furlan, E. Bidnenko, P. Courtin, C. Pechoux, P. Hols, Y.F. Dufrene & S. Kulakauskas, (2010) Cell surface of Lactococcus lactis is covered by a protective polysaccharide pellicle. J Biol Chem 285: 10464-10471.

-Corpet, F., (1988) Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res 16: 10881-10890.

-Cuskin, F., J.E. Flint, T.M. Gloster, C. Morland, A. Basle, B. Henrissat, P.M. Coutinho, A. Strazzulli, A.S. Solovyova, G.J. Davies & H.J. Gilbert, (2012) How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity. Proc Natl Acad Sci U S A 109: 20889-20894.

-Davidson, A.R., L. Cardarelli, L.G. Pell, D.R. Radford & K.L. Maxwell, (2012) Long noncontractile tail machines of bacteriophages. Advances in experimental medicine and biology 726: 115-142.

-del Rio, B., A.G. Binetti, M.C. Martin, M. Fernandez, A.H. Magadan & M.A. Alvarez, (2007) Multiplex PCR for the detection and identification of dairy bacteriophages in milk. Food Microbiol 24: 75-81.

-Deveau, H., S.J. Labrie, M.C. Chopin & S. Moineau, (2006) Biodiversity and classification of lactococcal phages. Appl Environ Microbiol 72: 4338-4346.

-Dieterle, M.E., C. Bowman, C. Batthyany, E. Lanzarotti, A. Turjanski, G. Hatfull & M. Piuri, (2014) Exposing the secrets of two well-known Lactobacillus casei phages, J-1 and PL-1, by genomic and structural analysis. Appl Environ Microbiol 80: 7107-7121.

-Dieterle, M.E., J. Fina Martin, R. Duran, S.I. Nemirovsky, C. Sanchez Rivas, C. Bowman, D. Russell, G.F. Hatfull, C. Cambillau & M. Piuri, (2016) Characterization of prophages containing "evolved" Dit/Tal modules in the genome of Lactobacillus casei BL23. Appl Microbiol Biotechnol 100: 9201-9215.

-Dieterle, M.E., S. Spinelli, I. Sadovskaya, M. Piuri & C. Cambillau, (2017) Evolved distal tail carbohydrate binding modules of Lactobacillus phage J-1: a novel type of anti-receptor widespread among lactic acid bacteria phages. Mol Microbiol 104: 608-620.

-Duda, R.L., B. Oh & R.W. Hendrix, (2013) Functional domains of the HK97 capsid maturation protease and the mechanisms of protein encapsidation. J Mol Biol 425: 2765-2781.

-Edmonds, L., A. Liu, J.J. Kwan, A. Avanessy, M. Caracoglia, I. Yang, K.L. Maxwell, J. Rubenstein, A.R. Davidson & L.W. Donaldson, (2007) The NMR structure of the gpU tail-terminator protein from bacteriophage lambda: identification of sites contributing to Mg(II)-mediated oligomerization and biological function. J Mol Biol 365: 175-186.

-Edmonds, L., R. Thirumoorthy, A. Liu, A. Davidson & L. Donaldson, (2005) NMR assignment of the gpU tail protein from lambda bacteriophage. J Biomol NMR 32: 91-92.

-Emsley, P. & K. Cowtan, (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60: 2126-2132.

-Emsley, P., B. Lohkamp, W.G. Scott & K. Cowtan, (2010) Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66: 486-501.

-Flayhan, A., F.M. Vellieux, R. Lurz, O. Maury, C. Contreras-Martel, E. Girard, P. Boulanger & C. Breyton, (2014) Crystal Structure of pb9, the Distal Tail Protein of Bacteriophage T5: a Conserved Structural Motif among All Siphophages. J Virol 88: 820-828.

-Fraser, J.S., Z. Yu, K.L. Maxwell & A.R. Davidson, (2006) Ig-like domains on bacteriophages: a tale of promiscuity and deceit. J Mol Biol 359: 496-507.

-Ghigo, J.M., D.S. Weiss, J.C. Chen, J.C. Yarrow & J. Beckwith, (1999) Localization of FtsL to the Escherichia coli septal ring. Molecular microbiology 31: 725-737.

-Goulet, A., J. Lai-Kee-Him, D. Veesler, I. Auzat, G. Robin, D.A. Shepherd, A.E. Ashcroft, E. Richard, J. Lichiere, P. Tavares, C. Cambillau & P. Bron, (2011) The opening of the SPP1 bacteriophage tail, a prevalent mechanism in Gram-positive-infecting siphophages. J Biol Chem 286: 25397-25405.

-Hildebrand, A., M. Remmert, A. Biegert & J. Soding, (2009) Fast and accurate automatic structure prediction with HHpred. Proteins 77 Suppl 9: 128-132.

-Kenny, J.G., S. McGrath, G.F. Fitzgerald & D. van Sinderen, (2004) Bacteriophage Tuc2009 encodes a tail-associated cell wall-degrading activity. J Bacteriol 186: 3480-3491.

-Kostyuchenko, V.A., P.G. Leiman, P.R. Chipman, S. Kanamaru, M.J. van Raaij, F. Arisaka, V.V. Mesyanzhinov & M.G. Rossmann, (2003) Three-dimensional structure of bacteriophage T4 baseplate. Nat Struct Biol 10: 688-693.

-Kramer, N.E., H.E. Hasper, P.T. van den Bogaard, S. Morath, B. de Kruijff, T. Hartung, E.J. Smid, E. Breukink, J. Kok & O.P. Kuipers, (2008) Increased D-alanylation of lipoteichoic acid and a thickened septum are main determinants in the nisin resistance mechanism of Lactococcus lactis. Microbiology 154: 1755-1762.

-Le Bourgeois, P., M. Bugarel, N. Campo, M.-L. Daveran-Mingot, J. Labonté, D. Lanfranchi, T. Lautier, C. Pagès & P. Ritzenthaler, (2007) The unconventional Xer recombination machinery of Streptococci/Lactococci. PLoS genetics 3: e117.

-Legrand, P., B. Collins, S. Blangy, J. Murphy, S. Spinelli, C. Gutierrez, N. Richet, C. Kellenberger, A. Desmyter, J. Mahony, D. van Sinderen & C. Cambillau, (2016) The Atomic Structure of the Phage Tuc2009 Baseplate Tripod Suggests that Host Recognition Involves Two Different Carbohydrate Binding Modules. MBio 7.

-Lillehaug, D., (1997) An improved plaque assay for poor plaque-producing temperate lactococcal bacteriophages. J Appl Microbiol 83: 85-90.

-Mahony, J., H. Deveau, S. Mc Grath, M. Ventura, C. Canchaya, S. Moineau, G.F. Fitzgerald & D. van Sinderen, (2006) Sequence and comparative genomic analysis of lactococcal bacteriophages jj50, 712 and P008: evolutionary insights into the 936 phage species. FEMS Microbiol Lett 261: 253-261.

-Mahony, J., W. Kot, J. Murphy, S. Ainsworth, H. Neve, L.H. Hansen, K.J. Heller, S.J. Sorensen, K. Hammer, C. Cambillau, F.K. Vogensen & D. van Sinderen, (2013) Investigation of the relationship between lactococcal host cell wall polysaccharide genotype and 936 phage receptor binding protein phylogeny. Appl Environ Microbiol 79: 4385-4392.

-Mahony, J., D.M. Tremblay, S.J. Labrie, S. Moineau & D. van Sinderen, (2015) Investigating the requirement for calcium during lactococcal phage infection. International journal of food microbiology 201: 47-51.

-Maxwell, K.L., A.A. Yee, C.H. Arrowsmith, M. Gold & A.R. Davidson, (2002) The solution structure of the bacteriophage lambda head-tail joining protein, gpFII. J Mol Biol 318: 1395-1404.

-McCabe, O., S. Spinelli, C. Farenc, M. Labbé, D. Tremblay, S. Blangy, S. Oscarson, S. Moineau & C. Cambillau, (2015b) The targeted recognition of Lactococcus lactis phages to their polysaccharide receptors. Molecular microbiology 96: 875-886.

-Moineau, S., Tremblay, D. & Labrie.S., (2002) Phages of lactic acid bacteria: from genomics to industrial applications. In: ASM News. pp. 388-393.

-Murphy, J., F. Bottacini, J. Mahony, P. Kelleher, H. Neve, A. Zomer, A. Nauta & D. Van Sinderen, (2016b) Comparative genomics and functional analysis of the 936 group of lactococcal Siphoviridae phages. Scientific reports 6: 21345.

-Murphy, J., B. Royer, J. Mahony, L. Hoyles, K. Heller, H. Neve, M. Bonestroo, A. Nauta & D. van Sinderen, (2013) Biodiversity of lactococcal bacteriophages isolated from 3 Gouda-type cheese-producing plants. J Dairy Sci 96: 4945-4957.

-Pell, L.G., A. Liu, L. Edmonds, L.W. Donaldson, P.L. Howell & A.R. Davidson, (2009) The X-ray crystal structure of the phage lambda tail terminator protein reveals the biologically relevant hexameric ring structure and demonstrates a conserved mechanism of tail termination among diverse long-tailed phages. J Mol Biol 389: 938-951.

-Pettersen, E.F., T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng & T.E. Ferrin, (2004) UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25: 1605-1612.

-Philippe, C., F. Jaomanjaka, O. Claisse, R. Laforgue, J. Maupeu, M. Petrel & C. Le Marrec, (2017) A survey of oenophages during wine making reveals a novel group with unusual genomic characteristics. Int J Food Microbiol 257: 138-147.

-Plisson, C., H.E. White, I. Auzat, A. Zafarani, C. Sao-Jose, S. Lhuillier, P. Tavares & E.V. Orlova, (2007) Structure of bacteriophage SPP1 tail reveals trigger for DNA ejection. EMBO J 26: 3720-3728.

-Pymol, The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.

-Ricagno, S., V. Campanacci, S. Blangy, S. Spinelli, D. Tremblay, S. Moineau, M. Tegoni & C. Cambillau, (2006) Crystal structure of the receptor-binding protein head domain from Lactococcus lactis phage bIL170. J Virol 80: 9331-9335.

-Rousseau, G.M. & S. Moineau, (2009) Evolution of Lactococcus lactis phages within a cheese factory. Appl Environ Microbiol 75: 5336-5344.

-Saez, N.J. & R. Vincentelli, (2014) High-throughput expression screening and purification of recombinant proteins in E. coli. Methods Mol Biol 1091: 33-53.

-Sao-Jose, C., S. Lhuillier, R. Lurz, R. Melki, J. Lepault, M.A. Santos & P. Tavares, (2006) The ectodomain of the viral receptor YueB forms a fiber that triggers ejection of bacteriophage SPP1 DNA. J Biol Chem 281: 11464-11470.

-Sciara, G., C. Bebeacua, P. Bron, D. Tremblay, M. Ortiz-Lombardia, J. Lichiere, M. van Heel, V. Campanacci, S. Moineau & C. Cambillau, (2010) Structure of lactococcal phage p2 baseplate and its mechanism of activation. Proc Natl Acad Sci U S A 107: 6852-6857.

-Soding, J., A. Biegert & A.N. Lupas, (2005) The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33: W244-248.

-Spinelli, S., V. Campanacci, S. Blangy, S. Moineau, M. Tegoni & C. Cambillau, (2006a) Modular structure of the receptor binding proteins of Lactococcus lactis phages. The RBP structure of the temperate phage TP901-1. J Biol Chem 281: 14256-14262.

-Spinelli, S., A. Desmyter, C.T. Verrips, H.J. de Haard, S. Moineau & C. Cambillau, (2006b) Lactococcal bacteriophage p2 receptor-binding protein structure suggests a common ancestor gene with bacterial and mammalian viruses. Nat Struct Mol Biol 13: 85-89.

-Spinelli, S., D. Veesler, C. Bebeacua & C. Cambillau, (2014) Structures and host-adhesion mechanisms of lactococcal siphophages. Frontiers in microbiology 5: 3.

-Stockdale, S.R., J. Mahony, P. Courtin, M.P. Chapot-Chartier, J.P. van Pijkeren, R.A. Britton, H. Neve, K.J. Heller, B. Aideh, F.K. Vogensen & D. van Sinderen, (2013) The lactococcal phages Tuc2009 and TP901-1 incorporate two alternate forms of their tail fiber into their virions for infection specialization. J Biol Chem 288: 5581-5590.

-Taylor, N.M., N.S. Prokhorov, R.C. Guerrero-Ferreira, M.M. Shneider, C. Browning, K.N. Goldie, H. Stahlberg & P.G. Leiman, (2016) Structure of the T4 baseplate and its function in triggering sheath contraction. Nature 533: 346-352.

-Taylor, N.M., M.J. van Raaij & P.G. Leiman, (2018) Contractile injection systems of bacteriophages and related systems. Molecular microbiology 108: 6-15.

-Turchetto, J., A.F. Sequeira, L. Ramond, F. Peysson, J.L. Bras, N.J. Saez, Y. Duhoo, M. Blemont, C.I. Guerreiro, L. Quinton, E. De Pauw, N. Gilles, H. Darbon, C.M. Fontes & R. --Vincentelli, (2017) High-throughput expression of animal venom toxins in Escherichia coli to generate a large library of oxidized disulphide-reticulated peptides for drug discovery. Microb Cell Fact 16: 6.

-Veesler, D., G. Robin, J. Lichiere, I. Auzat, P. Tavares, P. Bron, V. Campanacci & C. Cambillau, (2010) Crystal Structure of Bacteriophage SPP1 Distal Tail Protein (gp19.1): A Baseplate Hub Paradigm In Gram-Positive Infecting Phages. J Biol Chem 285: 36666-36673.

-Veesler, D., S. Spinelli, J. Mahony, J. Lichiere, S. Blangy, G. Bricogne, P. Legrand, M. Ortiz-Lombardia, V. Campanacci, D. van Sinderen & C. Cambillau, (2012) Structure of the phage TP901-1 1.8 MDa baseplate suggests an alternative host adhesion mechanism. Proc Natl Acad Sci U S A 109: 8954-8958.

-Weiss, D.S., (2004) Bacterial cell division and the septal ring. Molecular microbiology 54: 588-597.

-Zimmermann, L., A. Stephens, S.Z. Nam, D. Rau, J. Kubler, M. Lozajic, F. Gabler, J. Soding, A.N. Lupas & V. Alva, (2017) A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. J Mol Biol. 430: 223-2243.

-Zivanovic, Y., F. Confalonieri, L. Ponchon, R. Lurz, M. Chami, A. Flayhan, M. Renouard, A. Huet, P. Decottignies, A.R. Davidson, C. Breyton & P. Boulanger, (2014) Insights into bacteriophage T5 structure from analysis of its morphogenesis genes and protein components. J Virol 88: 1162-1174.

FIGURE LEGENDS

Figure 1. The baseplate of the 936 phage group. a. Electron microscopy structure of the tail (green) and baseplate (orange) of phage p2, which possesses a “classical” Dit (Bebeacua *et al.*, 2013). **b**. Side-view ribbon structural model of a 936 phage non-activated baseplate depicting RBPs (blue), the Tals (green) and Dits (orange for the "classical" domain) and the BppA insertion (yellow) within the arm extension of Dit (Sciara *et al.*, 2010). **c**. 90° view (from bottom) with respect to **b**. **d**. Side-view ribbon structural model of a 936 phage Ca^{2+}-activated baseplate. Same colors as in **b** and **c**.

Figure 2. Sequence alignment of CBM insertions from Dit proteins listed in Fig. 1 and identified in currently available genome sequences of members of the lactococcal 936 phage group. Fully conserved amino acids are in red and partially conserved in blue. Otherwise in black.

Figure 3. Sequence alignment of the four classes of insertions in evolved Dit proteins. Classes 1 to 4 from top to bottom. Fully conserved amino-acids are in red and the partially conserved in blue, otherwise in black.

Figure 4. Full sequence HHpred analysis of the insertions of evolved Dit proteins.The HHpred output and statistics are followed by the alignment of each class leader with BppA, from residue 8 to 208. **a.** Analysis of the CBM insertion of the phage Phi17 Dit, the phage representative of Dit Class 1. **b.** Analysis of the CBM insertion of the Dit of phage Phi10.5, the phage representative of Dit Class 2. **c.** Analysis of the CBM insertion of the Dit of phage 712, the phage representative of Dit Class 3. **d.** Analysis of the CBM insertion of the Dit of phage PhiL.6, the phage representative of Dit Class 4.

Figure 5. Fluorescent binding assays of GFP-labelled Dits or Dit-CBMs to host cells. **a.** Binding of the Dit protein of phage Phi17 to host strain *L. lactis* E (CWPS Type C). Protein was added at a quantity of 5 µg. **b.** Binding of the CBM domain of the Dit of phage Phi10.5 to host *L. lactis* 10 (CWPS Type B). Protein was added at a quantity of 25 µg. **c.** Binding of the Dit of phage Phi2R14S to host strain *L. lactis* 2 (CWPS Type B). Protein was added at a quantity of 10 µg. **d.** Binding of the Dit of phage PhiL.6 to host strain *L. lactis* L (unknown CWPS Type). Protein was added at a quantity of 15 µg. **e.** Binding of the Dit of phage Phi8R06S to host strain *L. lactis* 8 (CWPS Type C). Protein was added at a quantity of 15 µg. Cells were visualized using differential interference contrast (DIC) microscopy (panels on the left), and fluorescent confocal microscopy (panels on the right) at GFP excitation wavelength of 488 nm. Scale bars correspond to 10 µm.

Figure 6. Fluorescent binding assays of GFP-labelled Dits or Dit-CBMs to host cells demonstrating the localized nature of binding. **a.** Binding of the Dit protein of phage Phi17 to host strain *L. lactis* E (CWPS Type C). Protein was added at a quantity of 5 µg. **b.** Binding of the CBM domain of the Dit of phage Phi10.5 to host *L. lactis* 10 (CWPS Type B). Protein was added at a quantity of 25 µg. **c.** Binding of the Dit of phage Phi2R14S to host strain *L. lactis* 2 (CWPS Type B). Protein was added at a quantity of 10 µg. **d.** Binding of the Dit of phage PhiL.6 to host strain *L. lactis* L (unknown CWPS Type). Protein was added at a

quantity of 15 µg. **e.** Binding of the Dit of phage Phi8R06S to host strain *L. lactis* 8 (CWPS Type C). Protein was added at a quantity of 15 µg. **f.** Binding of the mCherry labelled RBP of Phi17 to host strain *L. lactis* E (CWPS Type C). Protein was added at a quantity of 5 µg. Cells were visualized using differential interference contrast (DIC) microscopy (panels on the left), and fluorescent confocal microscopy (panels on the right) at the GFP excitation wavelength of 488 nm, or at the mCherry excitement wavelength of 514 nm. Scale bars correspond to 10 µm.

Figure 7. Host binding range assay of the GFP labelled Dit of Phi17 against 10 strains: five hosts sensitive to the indicated phage (a-e), and five non-host strains (f-j). **a.** Binding to host strain *L. lactis* 8 (CWPS Type C). **b**. Binding to host strain *L. lactis* 9 (CWPS Type C). **c**. Binding to host strain *L. lactis* E (CWPS Type C). **d**. Binding to host strain *L. lactis* L (CWPS Type Unknown). **e**/ Binding to host strain *L. lactis* M (CWPS Type C). **f**/ Binding to host strain *L. lactis* 2 (CWPS Type B). **g**/ Binding to host strain *L. lactis* 10 (CWPS Type B). **h**/ Binding to host strain *L. lactis* 17 (CWPS Type B). **i**/ Binding to host strain *L. lactis* 5 (CWPS Type C). **j**/ Binding to host strain *L. lactis* C (CWPS Type A). In all cases, protein was added at a quantity of 5 µg. Cells were visualized using differential interference contrast (DIC) microscopy (panels on the left), and fluorescent confocal microscopy (panels on the right) at the GFP excitement wavelength of 488 nm. Scale bars correspond to 10 µm.

Figure 8. **Host binding range assay of the mCherry labelled RBP of Phi17 against 10 strains:** five hosts which the mature phage is capable of infecting (a-e), and five non-host strains (f-j). **a**. Binding to host strain *L. lactis* 8 (CWPS Type C). **b**. Binding to host strain *L. lactis* 9 (CWPS Type C). **c**. Binding to host strain *L. lactis* E (CWPS type unkown). **d**. Binding to host strain *L. lactis* L (CWPS Type Unknown). **e**. Binding to host strain *L. lactis* M (CWPS Type C). **f**. Binding to host strain *L. lactis* 2 (CWPS Type B). **g**. Binding to host strain *L. lactis* 10 (CWPS Type B). **h**. Binding to host strain *L. lactis* 17 (CWPS Type B). **i**. Binding to host strain *L. lactis* 5 (CWPS Type C). **j**. Binding to host strain *L. lactis* C (CWPS Type A). In all cases, protein was added at a quantity of 5 µg (0.81 µM). Cells were visualized using differential interference contrast (DIC) microscopy (panels on the left), and fluorescent confocal microscopy (panels on the right) at the mCherry excitement wavelength of 514 nm. Scale bars correspond to 10 µm.

Figure 9. Full sequence HHpred analysis of evolved Dit proteins from various bacterial species. The name of the bacterial species/strain is indicated above the HHpred output.

Table 1. Phage Dit/Dit-insertions selected for production and binding assays.

Table 1. Phage Dit/Dit-insertions selected for production and binding assays.

Phage name	Host	Protein	Dit Class	Protein Yield (µg/µL)	Protein added (µg ; µMM)	Binding
Phi17	*L. lactis* E	Dit	1	0.63	5 ; 0.8989	**YES**
Phi10.5	*L. lactis* 10	Dit	2	Insoluble	-	-
Phi10.5	*L. lactis* 10	Dit-CBM	2	2.24	25 ; 4.8	**YES**
Phi2R14S	*L. lactis* 2	Dit	2	0.85	10 ; 1.8	**YES**
712	NZ9000	Dit	3	Insoluble	-	-
712	NZ9000	Dit-ΔCBM	Classical	Insoluble	-	-
PhiL.6	*L. lactis* L	Dit	4	0.91	15 ; 2.83	**YES**
Phi8R06A	*L. lactis* 8	Dit	4	0.82	15 ; 2.82	**YES**
Controls						
p2	NZ9000	Dit	Classical	0.65	100; 6.1	**NO**
Phi4.2	*L. lactis* 4	Dit	Classical	0.34	100; 3.9	**NO**

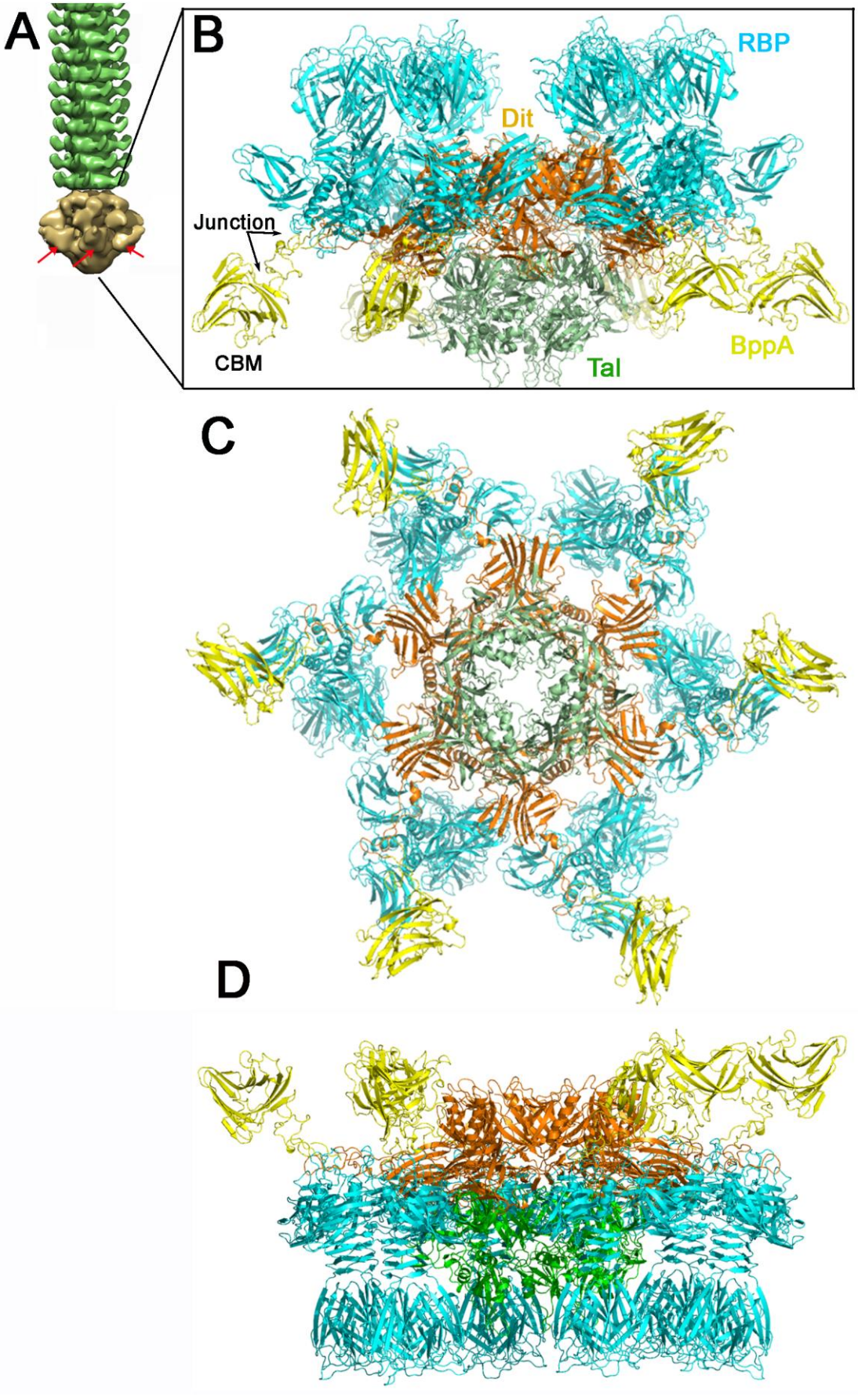
Table 2. Binding of the RBP and Dit of phagePhi17,the CBM region of the Dit of Phi10.5, and the Dits of Phi2R14S, PhiL.6, and 8R06S Dit to a selection of host and non-host strains. (nd: non determined; CWPS: cell wall polysaccharide). Types are according to (Mahony *et al.*, 2013, Ainsworth *et al.*, 2014). + denotes binding was observed, while – indicates no binding.

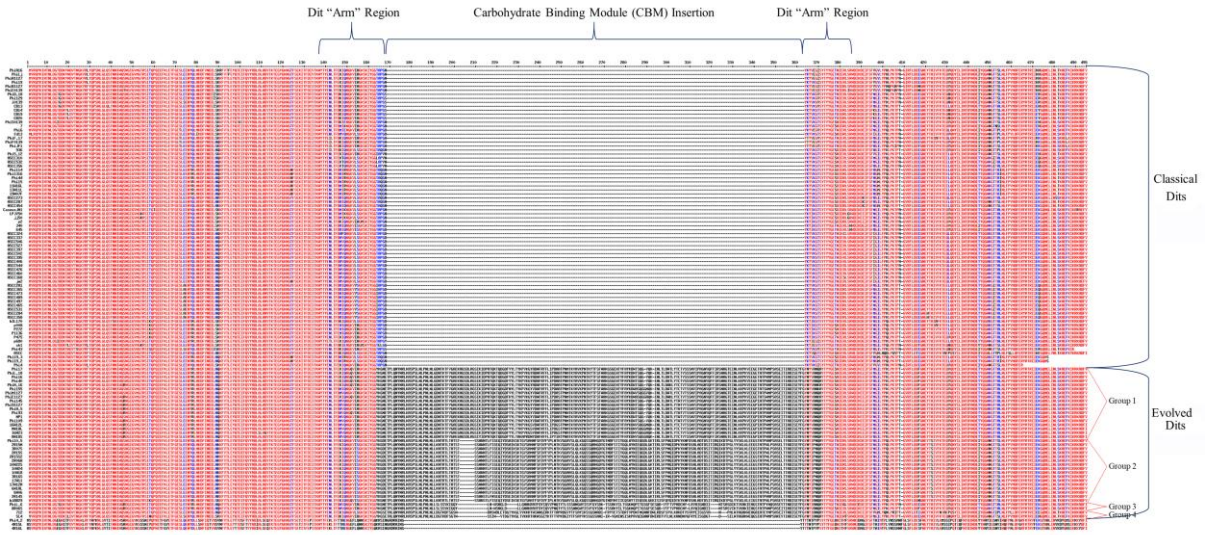
	Host Strains					**Non-Host Strains**				
Class 1 Dits										
Strain (CWPS Type)	**8(C)**	**9(C)**	**E(C)**	**L(nd)**	**M(C)**	**2(B)**	**10(B)**	**17(B)**	**5(C)**	**C(A)**
Phi17 Dit	+	+	+	+	+	-	-	-	-	-
Phi17 RBP	+	+	+	+	+	-	-	-	-	-
Class 2 Dits										
Strain (CWPS Type)	**2(B)**	**10(B)**	**17(B)**	**3(B)**		**C(A)**	**4(C)**	**13(C)**	**A(C)**	
Phi2R14S Dit	+	+	+	+		-	-	-	-	
Phi10.5 Dit-CBM	+	+	+	+		-	-	-	-	
Class 4 Dits										
Strain (CWPS Type)	**8(C)**	**9(C)**	**L(nd)**	**M(C)**		**C(A)**	**2(B)**	**10(B)**	**13(C)**	
PhiL.6 Dit	+	+	+	+		-	-	-	-	
Phi8R06S Dit	+	+	+	+		-	-	-	-	

Table 3: Correlation between Dit classes, RBP groups(Murphy *et al.*, 2016) and CWPS types (Mahony *et al.*, 2013, Ainsworth *et al.*, 2014).

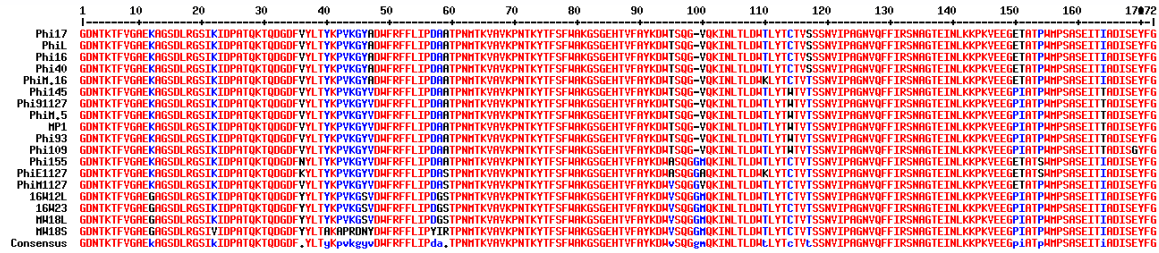
Phage	Dit class	RBP Group	Host CWPS Type
Phi17	1	IV	C
PhiL.18	1	IV	C
Phi16	1	IV	C
Phi40	1	IV	C
PhiM.16	1	IV	C
Phi145	1	IV	C
Phi91127	1	IV	C
PhiM.5	1	IV	C
MP1	1	IV	C
Phi93	1	IV	C
Phi109	1	IV	C
Phi155	1	IV	C
PhiE1127	1	IV	C
PhiM1127	1	IV	C
16W12L	1	IV	C
16W23	1	IV	C
MW18L	1	IV	C
MW18S	1	IV	C
Phi10.5	2	III	B
2R15M	2	III	B
2R15S	2	III	B
2R15S2	2	III	B
3R07S	2	III	B
2R06A	2	III	B
10W24	2	III	B
10W22S	2	III	B
10W18	2	III	B
3R16S	2	III	B
6W06	2	III	B
6W18L	2	III	B
17W11	2	III	B
17W12M	2	III	B
2R14S	2	III	B
biBB29	2	III	B
712	3	I	C
SL4	3	I	nd
PhiL.6	4	IV	C
8R06S	4	IV	C

nd: CWPS type is not known

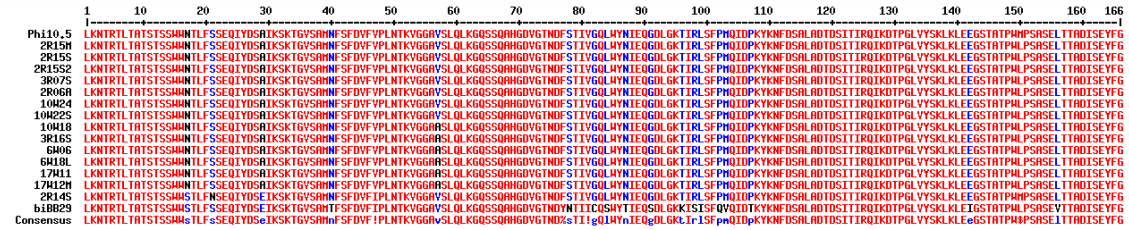




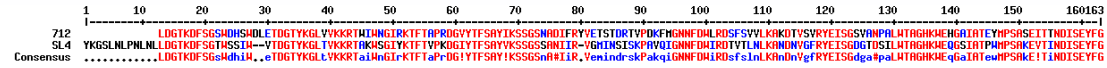
Class 1



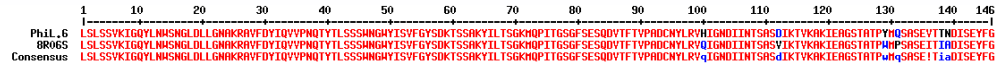
Class 2



Class 3

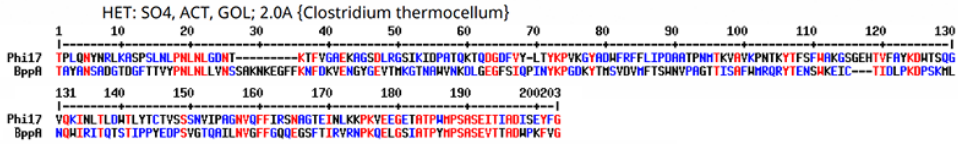


Class 4



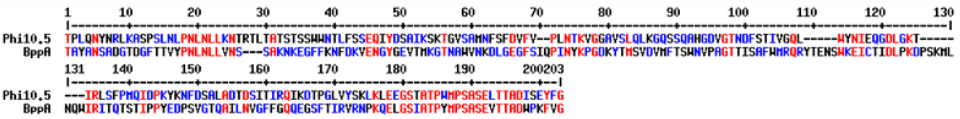
a. Dit Class 1

Hit	Name	Probability	E-value	SS	Cols	Target Length
5E7T_B	Minor structural protein 4, Minor; bacteriophages, Lactococcus lactis, Siphoviridae, nanobody; HET: SO4; 2.9A {Lactococcus phage Tuc2009}	99.74	1.5e-20	11	148	286
4QPW_A	Putative glycosyl hydrolase family 10; Carbohydrate-Binding Module (CBM), hydrolase; HET: XYP; 1.14A {Bacteroides intestinalis DSM 17393}	90.88	3.4	10	79	142
3P6B_A	Cellulose 1,4-beta-cellobiosidase (E.C.3.2.1.91); beta-sandwich, carbohydrate binding, SUGAR BINDING; HET: SO4, ACT, GOL; 2.0A {Clostridium thermocellum}	89.15	7.7	11.2	88	205



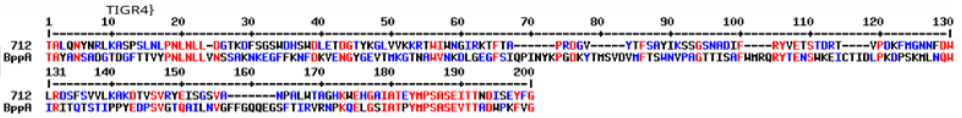
b. Dit Class 2

Hit	Name	Probability	E-value	SS	Cols	Target Length
5E7T_B	Minor structural protein 4, Minor; bacteriophages, Lactococcus lactis, Siphoviridae, nanobody; HET: SO4; 2.9A {Lactococcus phage Tuc2009}	99.72	3e-20	10.3	166	286

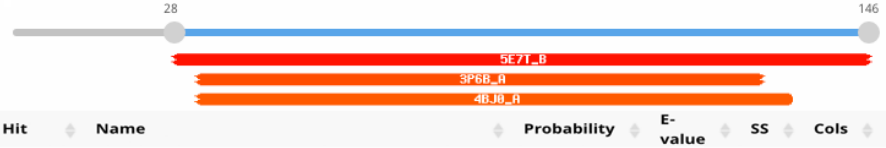


c. Dit Class 3

Hit	Name	Probability	E-value	SS	Cols	Target Length
5E7T_B	Minor structural protein 4, Minor; bacteriophages, Lactococcus lactis, Siphoviridae, nanobody; HET: SO4; 2.9A {Lactococcus phage Tuc2009}	99.43	4e-15	12	159	286
4MGQ_A	Glycosyl hydrolase family 10; Endo-xylanase, HYDROLASE; 1.68A {Prevotella bryantii}	96.33	0.012	12	106	168
4D0Q_A	HYALURONATE LYASE (E.C.4.2.2.1); LYASE, HYALURONAN BINDING CARBOHYDRATE BINDING; HET: EDO; 1.2A {STREPTOCOCCUS PNEUMONIAE TIGR4}	96.13	0.054	14	111	169



d. Dit Class 4



Hit	Name	Probability	E-value	SS	Cols	Target Length
5E7T_B	Minor structural protein 4, Minor; bacteriophages, Lactococcus lactis, Siphoviridae, nanobody; HET: SO4; 2.9A {Lactococcus phage Tuc2009}	98.12	7.7e-8	8.1	114	286
3P6B_A	Cellulose 1,4-beta-cellobiosidase (E.C.3.2.1.91); beta-sandwich, carbohydrate binding, SUGAR BINDING; HET: SO4, ACT, GOL; 2.0A {Clostridium thermocellum}	92.24	2.4	10.8	95	205
4D0Q_A	HYALURONATE LYASE (E.C.4.2.2.1); LYASE, HYALURONAN BINDING CARBOHYDRATE BINDING; HET: EDO; 1.2A {STREPTOCOCCUS PNEUMONIAE TIGR4}	91.01	2.9	11	92	169