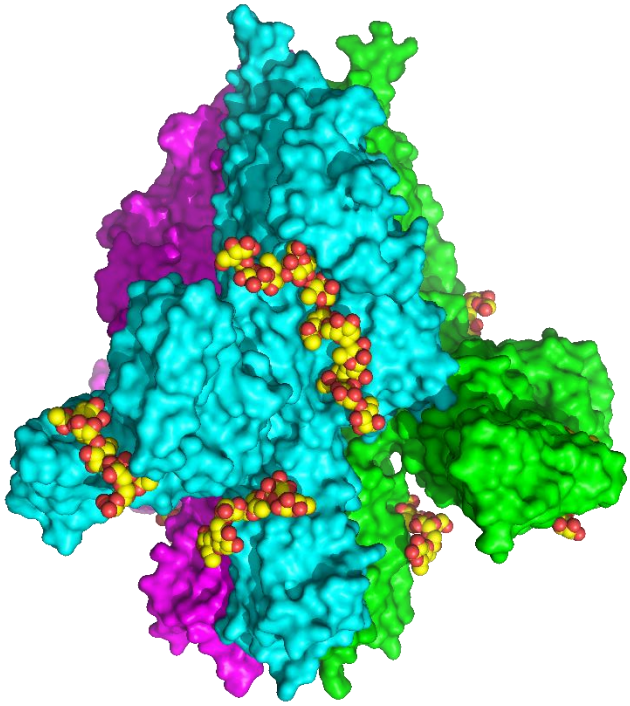
Universidad Autónoma de Madrid

Facultad de Ciencias

Departamento de Biología Molecular

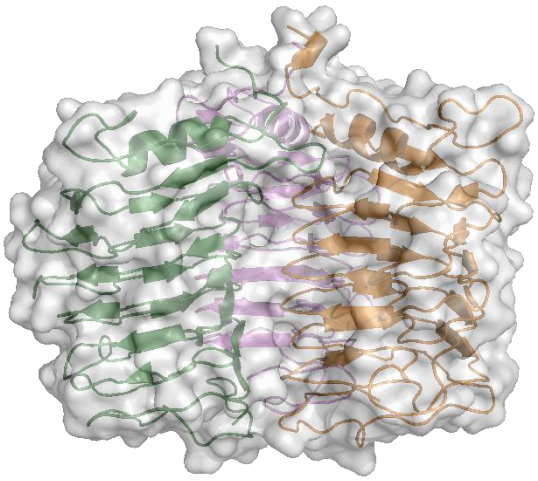
STRUCTURE AND FUNCTION OF *SALMONELLA VIRUS EPSILON15* AND *CAMPYLOBACTER VIRUS F358* TAILSPIKES



Mateo Seoane Blanco

Madrid, 2021

Memoria presentada por Mateo Seoane Blanco, graduado en Biología, para optar al título de Doctor en Bioquímica, Biología Molecular, Biomedicina y Biotecnología (Biociencias Moleculares)



DIRECTOR DE LA TESIS:

DR. MARK JOHAN VAN RAAIJ

CENTRO NACIONAL DE BIOTECNOLOGIA

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

El trabajo recogido en esta memoria ha sido realizado en el Centro Nacional de Biotecnología (CNB-CSIC) bajo la dirección del Dr. Mark Johan van Raaij, y actuando como tutor académico el Dr. Mario Mencía Caballero. Su financiación corrió a cargo de un contrato FPI y los proyectos BFU2014-53425-P y BFU2017-82207-P, financiados por el Ministerio de Economía y Competitividad y cofinanciados por el fondo FEDER de la Unión Europea.

Index

Abbreviations

ASA: Accessible Solvent Area

AU: Asymmetric Unit

BSA: Buried Surface Area

CAZy: Carbohydrates-Active enZymes

CC: Correlation Coefficient

CE: Carbohydrate Esterase

CLD: Correlation with Local Deviation

CPS: Capsular PolySaccharide

CV: Column Volumes

DALI: Distance-matrix ALIgnment

ΔG^{diss}: predicted standard free energy of dissociation

DM: Density Modification

EMDB: Electron Microscopy Data Bank

FOM: Figure Of Merit

FT: Flow Through

GDSL: Glycine-Aspartic acid-Serine-Leucine

GH90: Glycoside Hydrolases 90

Gp208ΔN: N-terminally truncated gp208 protein

Gp20ΔN: N-terminally truncated gp20 protein

Gp20ΔC: C-terminally truncated gp20 protein

His-tag: six-histidine tag

ICTV: International Committee on Taxonomy of Viruses

IEC: Ion Exchange Chromatography

IM: Inner Membrane

IMAC: Immobilised Metal ion Affinity Chromatography

IPTG: isopropyl-β-D-1-thiogalactopyranoside

Kdo: Ketodeoxyoctonic acid

LacI: Lactose operon Inhibitor gene

LB: Luria-Bertani medium

Lpp: murein Lipoprotein

LPS: LipoPolySaccharide

LTA: LipoTeichoic Acid

MBTH: 3-Methyl-2-BenzoThiazolinone Hydrazone hydrochloride

MeOPN: O-methyl phosphoramidate

MWCO: Molecular-Weight Cut-Off

mRBP3: monomeric RBP3 protein

Ni-NTA: Nickel-NitriloTriacetic Acid

OM: Outer Membrane

OmpA: Outer membrane protein A

OmpC: Outer membrane protein C

OmpF: Outer membrane protein F

PB: Parallel Beta-sheet

PBD: Phage-Binding Domain

PDB: Protein Data Bank

PLS4: Point Loma Sewage 4 mutant

PS: PolySaccharide

RBP: Receptor Binding Protein

RES: ReticuloEndothelial System

RMSD: Root-Mean-Square deviation

RNA: RiboNucleic Acid

RSE: Reducing Sugar Ends

RU: Repeating Units

SAD: Single-wavelength Anomalous Diffraction

SeMet: Seleno-Methionine

SGNH: Serine-Glycine-Asparagine-Histidine

SOC: Super Optimal broth with Catabolite repression

ssDNA: singles-stranded DNA

ssRNA: single-stranded RNA

TAE: Tris base-Acetic acid-EDTA

tRBP3: trimeric RBP3 protein

TSS: Transformation and Storage Solution

T6SS: Type VI Secretion System

UV-C: UltraViolet C

W: Wash

WTA: Wall Teichoic Acid

Abstract

Bacteriophages are the most numerous organisms on Earth. They infect bacteria, including pathogenic ones, and have a huge influence on ecological processes. In the future, they can help to treat diseases caused by antibiotic-resistant bacteria. Phages use their receptor-binding proteins to adsorb to their receptors in bacteria. Consequently, receptor-binding proteins control the phage host range.

In my thesis, I focused on the *Salmonella virus epsilon15* receptor-binding proteins, gp20, and on one of the four *Campylobacter virus F358* receptor-binding proteins, RBP3. These phages infect *Salmonella enterica* serovar Anatum and *Campylobacter jejuni*, respectively. They are the two bacteria causing most of the food-borne diseases to humans in Europe. The epsilon15 receptor is the O-antigen of the lipopolysaccharide (LPS). It consists of repetitions of the trisaccharide D-Galp[6Ac]-α-1 – 6-D-Manp-β-1 – 4-L-Rhap-α-1 – ROH linked by α-1 – 3 bonds.

During my thesis, I determined the structure of gp20, which is a trimeric tailspike, with and without its receptor. Gp20ΔN is composed of three domains. From the N- to the C-termini, the β-helix domain, the β-sandwich domain and the petal domain. The β-helix domain is a right-handed 12-rung helix made of β-strands. Each rung has three β-strands linked by turns. Some of the turns have loops that create a groove facing outwards. The β-sandwich domain is composed of a six-stranded and a five-stranded antiparallel β-sheets. This is the most distal domain in gp20. A backward-running linker connects the β-sandwich with the petal domain. The petal domain has an α/β hydrolase fold with a β-barrel inserted in one of its loops. These two subdomains form a groove parallel to the one of the β-helix domain.

Salmonella enterica serovar Anatum O-antigen oligosaccharides bind to four regions of gp20; one to the petal domain, inside the groove; another to the β-sandwich domain and two to the β-helix domain. The two β-helix domain oligosaccharides are separated by the endorhamnosidase site. It is located in a negatively charged area, inside the β-helix groove. It is formed by two aspartic acids that probably act following the retaining mechanism, common to other glycosidases. This activity clears space to allow the virus to approach the bacterial membrane. The petal domain contains a putative esterase site, although it is unknown if this putative esterase site has activity. Removing the acetyl group bound to the galactose might reduce the interactions between uncut lipopolysaccharide chains and also help the virus approach the membrane.

I also determined the structure of the *Campylobacter virus F358* receptor-binding protein RBP3. It crystallised both as a monomer and as a trimer. Each RBP3 monomer is composed of a β-helix domain. It is shorter than homologous trimeric tailspikes and monomeric β-helical bacterial enzymes. This feature might be related to the existence of monomeric and trimeric states. The trimeric RBP3 has a negatively charged area in its interchain grooves. It resembles the catalytic sites of other tailspikes.

In summary, I present structural and functional findings of *Salmonella* and *Campylobacter* bacteriophage receptor-binding proteins. This knowledge will be useful to design engineered bacteriophages for the application of tailor-made phage therapy and other phage-based applications.

Resumen

Los bacteriófagos son los organismos más numerosos de la Tierra. Producen infecciones en bacterias, incluyendo las patógenas, y tienen una gran influencia en los procesos ecológicos. En el futuro, podrán ayudar a tratar enfermedades causadas por bacterias resistentes a los antibióticos. Los bacteriófagos usan sus proteínas de unión a receptor (RBP) para unirse a sus receptores en las bacterias. Por lo tanto, estas proteínas controlan el rango de hospedador de los bacteriófagos.

Durante mi tesis, estudié la RBP del *Salmonella virus epsilon15*, gp20, y una de las cuatro RBPs de *Campylobacter virus F358*, RBP3. Estos fagos infectan a las especies *Salmonella enterica* serovar Anatum y *Campylobacter jejuni*, respectivamente. Estas son las dos principales bacterias causantes de enfermedades humanas transmitidas por comida en Europa. El receptor de la RBP de epsilon15 es el antígeno O del lipopolisacárido (LPS). Este consiste en la repetición del trisacárido D-Galp[6Ac]-α-1 – 6-D-Manp-β-1 – 4-L-Rhap-α-1 – ROH unido por enlaces α-1 – 3.

Durante mi tesis, determiné la estructura de una forma truncada de gp20, gp20ΔN, que es una espícula de la cola trimérica, con y sin su receptor. Gp20ΔN se compone de tres dominios del extremo N al C, el dominio hélice β, el dominio sándwich β y el dominio pétalo. El dominio hélice β es una hélice dextrógira de 12 vueltas hecha de hebras β. Cada vuelta tiene tres hebras β unidas por giros. Algunos de estos giros tienen lazos que crean un surco orientado hacia el exterior. El dominio sándwich β se compone de dos láminas β antiparalelas hechas de cinco y seis hebras β. Esta es la parte más distal de gp20. Los dominios sándwich β y pétalo se conectan mediante una sección corta de aminoácidos orientados hacia atrás. El dominio pétalo tiene un plegamiento de hidrolasa α/β con un barril β insertado en uno de sus lazos. Estos dos subdominios forman un surco paralelo al del dominio hélice β.

Los oligosacáridos de antígeno O de *Salmonella enterica* serovar Anatum se unen a cuatro zonas de gp20: uno al dominio pétalo, dentro del surco; otro, al dominio sándwich β y dos, al dominio hélice β. Los dos oligosacáridos del dominio hélice β están separados por el sitio endoramnosidasa. Este se localiza en un área con carga negativa, dentro del surco. Este sitio está formado por dos ácidos aspárticos que, probablemente, actúan siguiendo el mecanismo de retención, común a otras glucosidasas. Esta actividad despeja el camino para la aproximación del virus a la membrana bacteriana. El domino pétalo tiene un posible sitio esterasa, aunque se desconoce si es activo. Eliminar el grupo acetil unido a la galactosa podría reducir las interacciones entre cadenas de LPS no cortadas y facilitar la aproximación del virus a la membrana.

También determiné la estructura de RBP3. Esta cristalizó como monómero y como trímero. Cada monómero de RBP3 está compuesta simplemente de un dominio hélice β. Es una proteína más corta que sus homólogas: las espículas de la cola triméricas y las enzimas bacterianas monoméricas con un plegamiento en hélice β. Esta característica podría estar relacionada con la existencia de estados monoméricos y triméricos. La RBP3 trimérica tiene una zona cargada negativamente en sus surcos intercatenarios. Esta zona recuerda a los sitios catalíticos de otras espículas de la cola.

En resumen, aquí muestro descubrimientos estructurales y funcionales de las RBPs de los bacteriófagos de *Salmonella* y *Campylobacter*. Este conocimiento podrá ser útil para diseñar bacteriófagos modificados para la aplicación de terapia fágica personalizada y otras aplicaciones basadas en fagos.

1. Introduction

1.1. Bacteriophages

Bacteriophages, also called phages, are viruses that infect bacteria. Frederik Twort (Twort, 1915) and Felix d'Hérelle (D'Herelle, 2007), who coined the name, independently discovered them more than 100 years ago. In those early times, phages were evaluated as a treatment against infectious diseases. However, the efficacy of phage therapy was ambiguous, probably due to insufficient quality control of many phage preparations. Alexander Fleming's discovery of antibiotics led to the decay of phage therapy (Salmond & Fineran, 2015). Phage therapy only continues in some Eastern Bloc countries, such as Georgia and Poland (Międzybrodzki et al., 2018). The first images of bacteriophages only appear about 30 years after their discovery. In the second half of the 20th century, bacteriophages were important in establishing the basic tools and principles of molecular biology, such as the central dogma of molecular biology, restriction enzymes and the first sequenced genome. Nowadays, the increase of infections caused by antibiotic-resistant bacteria has put phages back in the spotlight as tools to fight bacteria (Rohwer & Segall, 2015; Salmond & Fineran, 2015).

Bacteriophages are found almost everywhere their host bacteria are present. Therefore, they can be found virtually anywhere on Earth, from human gut to the ocean water. Since bacteria have an important role in the cycles of chemical elements, phages have a huge impact on ecosystems (Clokie et al., 2011; Salmond & Fineran, 2015). In our bodies, they may control bacterial growth. However, some phages codify for virulence factors, for example in *Vibrio cholerae*, *Salmonella* and *Shigella* species. In the dairy industry, bacteriophages may also be a source of problems, killing the bacteria needed for milk fermentation.

Other bacteriophages encode proteins involved in cyanobacteria photosynthesis. Therefore, they affect the carbon cycle in the ocean either increasing or reducing carbon capture (Rohwer & Segall, 2015; Salmond & Fineran, 2015).

1.2. Bacteriophage life cycle

The bacteriophage life cycle starts with phage adsorption to the host bacterium (Fig. 1). It is a poorly known process. Most information comes from the adsorption process of a few bacteriophages of Gram-negative bacteria. Adsorption is generally accepted to have two steps.

T4 phage, a myovirus, reversibly binds the outer membrane protein C (OmpC) or the outer core oligosaccharide; first, with a few long tail fibres and then, with all. This binding causes a conformational change of its baseplate that causes the extension of its short tail fibres. This extension enables them to reach the secondary receptor, ketodeoxyoctonic acid (Kdo) residues in the inner core oligosaccharide. After the second binding, the contraction of the tail sheath pierces the *E. coli* cell envelope and ejects the viral genome (Hu et al., 2015; Taylor et al., 2016).

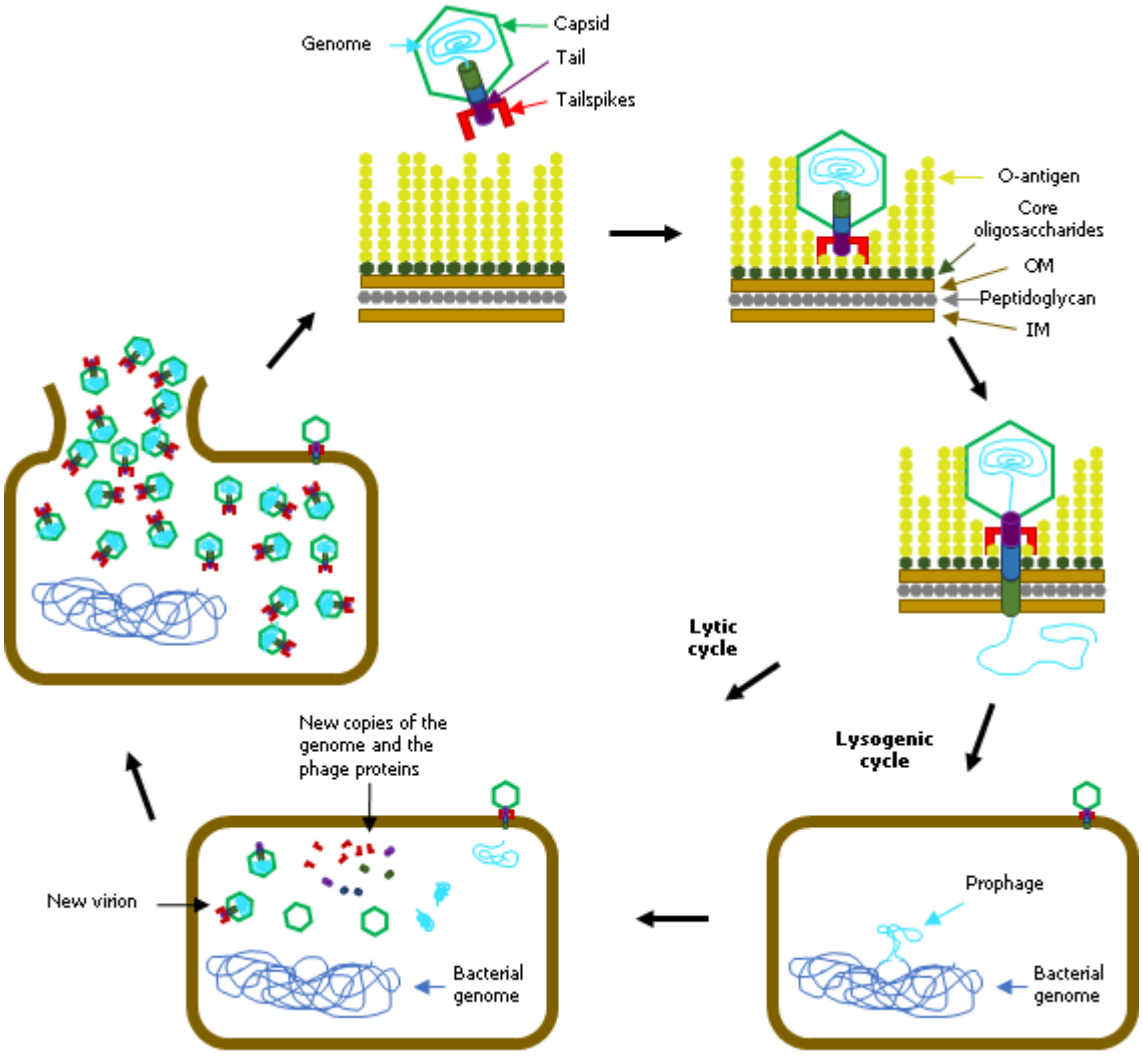


Figure 1. **Life cycle of a podovirus infecting a Gram-negative bacterium:** After adsorption, phage tailspikes degrade the O-antigen. This triggers the formation of the ejectosome, the core proteins that form a tube through the cell envelope. In the lysogenic cycle, genome ejection is followed by its integration in the bacterial genome as a prophage. In the lytic cycle, it is followed by its replication and the production of protein that will form the new virions. In the last step, holins provoke the lysis of the bacteria and the new progeny of phages can start a new cycle.

However, many bacteriophages have only one side RBP or one per bacterial host. Hence, they must have other mechanisms to introduce their genomes into the host bacteria. Podoviruses like P22 and T7 only need pure LPS to eject their genome (Andres et al., 2010;

González-García et al., 2015). T7 phage also interacts with OmpA and OmpF, which might work as primary receptors. Then, the T7 tail fibre interacts with the core oligosaccharides of LPS, the secondary receptor (González-García et al., 2015). This interaction triggers a conformational change that opens the tail channel through which core proteins and the genome go out of the phage.

After adsorption, the genome is translocated through the host membrane using protein tubes. Siphoviruses probably form the tube with the end of their tail and bacterial membrane proteins (Fig. 2A-B). In myoviruses, binding of the secondary receptor leads to conformational changes that end up in the contraction of their tail sheath, drilling the tail tube into the bacterial membrane (Fig. 2C-D). Podoviruses build a tube through the membrane and periplasmic space with core proteins from inside the capsid. Nevertheless, it is unknown how these proteins pass through the tail channel (Fig. 2E-F) (Sanz-Gaitero et al., 2019).

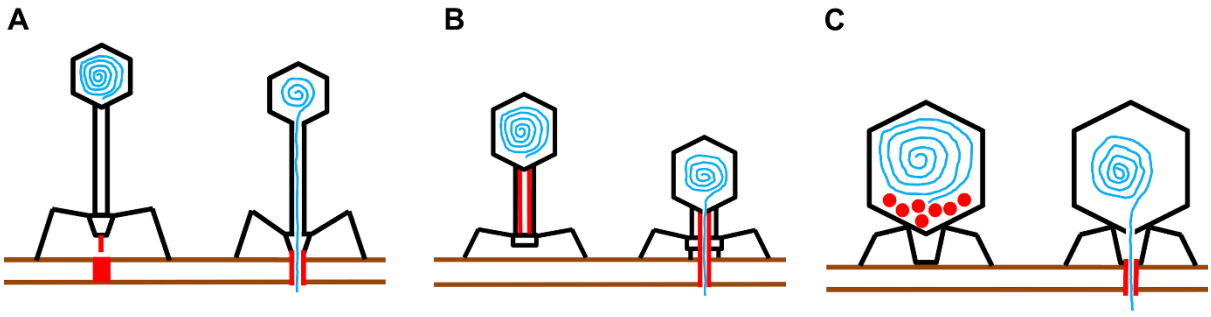


Figure 2: Structure and infection mechanism of the three families of the order Caudovirales. Pre- and post-ejection schematic drawings of a siphovirus (A and B), a myovirus (C and D) and a podovirus (E and F). Taken and modified from Sanz-Gaitero et al., 2019.

Once inside the cell, phages may use two different strategies: lytic or lysogenic (Fig. 6). A lytic phage uses the cell machinery to replicate their DNA and produce their proteins. Then, the holins open channels in the cytoplasmic membrane. Endolysins use these channels to reach the peptidoglycan layer and hydrolysing it, causing bacterial lysis. In this way, new virions get dispersed from the cell and are able to infect new hosts.

In the lysogenic cycle, the phage genome replicates along the bacterial genome as a prophage, either internalised in the genome or outside as plasmid. Hence, it is multiplied alongside the bacteria as a prophage. Usually, some prophage genes are expressed and protect the host bacteria against infection from other phages of the same or similar strain. A switch to the lytic cycle may be caused by cellular stress and/or the presence of a high number of uninfected bacteria, i.e. better opportunities of multiplying in other bacteria than in the current

host (Gandon, 2016). It can be induced experimentally by UV-C, mitomycin C, hydrogen peroxide, or other methods that cause stress (Weinbauer, 2004).

Competition between the phage pathogen and the bacterial host has led to an evolutionary arms race. Phage pressure selects for bacterial species with lower expression of membrane proteins used as receptors or the ones that changed the O-antigen composition. In this case, adapted bacteriophages may bind to another receptor or infect another bacterial strain. As an example, phage λ can be evolved to bind to membrane proteins OmpF and LamB with its protein J when *E. coli* B cells reduce the expression of LamB (Samson et al., 2013). In phages, this arms race has led to the generation of highly variable regions in the distal domains of RBPs (Islam et al., 2019; Trojet et al., 2011), similar to the hypervariable regions of antibodies. *Bordetella* phage BPP-1 even has a special mechanism to speed up RBP evolution. It uses a site-specific, error-prone reverse transcription to produce an offspring with modified RBPs (Salmond & Fineran, 2015). The arms race between bacteria and phages also caused the development of bacterial defence systems, such as restriction enzymes and CRISPR-Cas systems. Both are extensively used in laboratories and in the biotechnology businesses to modify DNA (Sorek et al., 2013; Wang et al., 2016).

Phage-bacterium interactions are not always a competitive process. Sometimes, phages are hijacked to transfer regions of the bacterial genome or specific genes. Therefore, bacteriophages are responsible for and an important part of horizontal gene transfer. This mechanism drives bacterial evolution by spreading antibiotic resistance or virulence genes. This is the case for the prophage CTX-Φ. It carries the virulence genes of *Vibrio cholerae*. Hence, it is partially responsible of cholera disease (Salmond & Fineran, 2015).

Besides transporting their genes, some bacteria have adapted phage genes as their own virulence factors. Myovirus tail-like genes were found in bacterial genomes, encoding "tailocins" (Salmond & Fineran, 2015). Bacteria use them to attack both prokaryotic and eukaryotic cells. Tailocins target specific strains with their tail fibres. Then, they contract like myovirus tails making a hole in the bacterial envelope. Type VI secretion systems (T6SS) are tail-like complexes fixed to the bacterial membrane. They have a different type of RBP to detect its victim. T6SS contraction pierces the membrane and delivers effector proteins into both prokaryotic and eukaryotic cells (Taylor et al., 2018).

1.3. Phage therapy and other applications

Interest in phage therapy has increased due to fears of a "post-antibiotic era". By now, phage only has been applied for compassionate use and no clinical trials have been successful (McCallin et al., 2019). However, there are 33 ongoing clinical trials involved phage or phage-related therapeutics (Theuretzbacher et al., 2020). Two main approaches in phage therapy may be identified: those using phage cocktails and those using a phage bank. Phage cocktails, mainly used in Eliava Institute (Tbilisi, Georgia), consist of a mixture of phages targeting a cohort of pathogenic bacteria. Phage banks have been used in the Hirszfeld Institute (Wroclaw, Poland). This approach uses one or more phages active against a pathogenic bacterium isolated from the patient (Kutter et al., 2009). Confirming that the phage is active against the relevant bacterium is a clear advantage, but culturing the bacterium for testing takes precious time some patients may not have. Treatment with phages may offer the following advantages over antibiotics (Bragg et al., 2014; Kakasis & Panitsa, 2019; Wittebole et al., 2014):

- Bacteriophages co-evolve with bacteria. They can adapt when bacteria become resistant to them.
- The huge diversity of phages makes it possible to find another bacteriophage able to infect a bacteria resistant to a previous phage.
- Bacteriophage numbers increase in the presence of their host bacterium, i.e. the dose may adapt to the seriousness of the infection.
- Due to their narrow host range, bacteriophages only attack targeted species or strains, eliminating collateral damage to commensal microbiota.
- Phages do not interfere with antibiotic activity and even can have synergetic functions.
- Some bacteriophages are also able to enzymatically disrupt biofilms.
- Bacteriophages spread throughout almost the whole body when administered intravenously.
- Bacteriophages typically show low toxicity and low occurrences of allergic reactions.

However, these characteristic are still preliminary, since no clinical trials of phase III have been performed. In addition, there are some drawbacks that must be addressed (Jault et al., 2019; Kakasis & Panitsa, 2019; Wittebole et al., 2014):

- Purified Gram-negative bacteriophages may contain traces of LPS, an endotoxin.
- The narrow host range means that phage banks might not target all pathogenic strains.
- Many bacterial diseases are acute infections that must be treated quickly. In contrast, selecting the right phages or phages takes several days, and this time may not be available.
- Not all bacteriophages have a stable shelf life.
- The immune system might detect phages and eliminate them through the reticuloendothelial system, reducing the efficacy after repeated applications.
- Clinical trials still have to define the optimal dose and route of administration.
- The regulatory framework is currently not adapted to the peculiarities of phage therapy.
- Some bacteriophages may carry bacterial virulence genes or induce horizontal gene transfer.

Some of these drawbacks might be solved by engineering phages or using phage-related antimicrobials instead of whole phages (Pires et al., 2016). For example, synthetic phages can express biofilm-degrading enzymes or quorum-sensing interfering enzymes to enhance their activity against structured microbial communities. Serial passage allows the generation of mutant phages in animal models that are retained less by the reticuloendothelial system. Endolysins can lyse Gram-positive bacteria, or even Gram-negative bacteria, when carrying a membrane permeabilising agent or outer membrane-destabilising peptide. Purified tailspikes may also help to dissolve bacterial biofilms.

Natural or engineered phages as well as other phage-derived proteins (tailocins or endolysins) can be used not only to treat bacterial diseases in humans. Some of these products have been approved to be used to kill bacteria in the food industry and veterinary medicine. Phage-derived experimental designs start to be applied to detect bacteria in patients, food or water (Pires et al., 2016).

Phage tailspikes alone can be used due to their depolymerase activity. Due to this activity, they serve as adjuvants of antibiotics (to ease antibiotic penetration), to increase phagocytic killing or to detect and identify bacterial strains. Besides antibacterial treatments, depolymerases can also be used to hydrolyse complex saccharides in smaller pieces in to improve taste of food as well as in the bioenergy or medical industry (Pires et al., 2016).

1.4. Phage structure

By 2012, just over 6000 phage species had been described (Ackermann & Prangishvili, 2012). Known bacteriophages are divided in eight families: *Leviviridae*, *Tectiviridae*, *Corticoviridae*, *Cystoviridae*, *Microviridae*, *Inoviridae*, *Siphoviridae*, *Myoviridae* and *Podoviridae* (Sanz-Gaitero et al., 2019). The family *Leviviridae* members have a small icosahedral capsid, containing a ssRNA genome, encoding four proteins. The capsids of the families *Tectiviridae* and *Corticoviridae* are both icosahedral and contain an internal lipid membrane. Both families have dsDNA genomes. Tectiviruses eject their genome by forming a tube with the lipid membrane that goes through the bacterial membrane. The *Corticoviridae* family is less known than the *Tectiviridae*. The *Cystoviridae* family members have a lipid membrane outside their icosahedral capsid. Inside, their genome is made of three dsRNA segments. The *Microviridae* family phages have icosahedral heads that encapsulate ssDNA genomes. The *Inoviridae* family members are the only rod-shaped filamentous bacteriophages. Their capsid protein binds to the circular ssDNA forming a helical capsid. The most famous member is the M13 bacteriophage. It has been used in several applications, such as phage display or directed evolution.

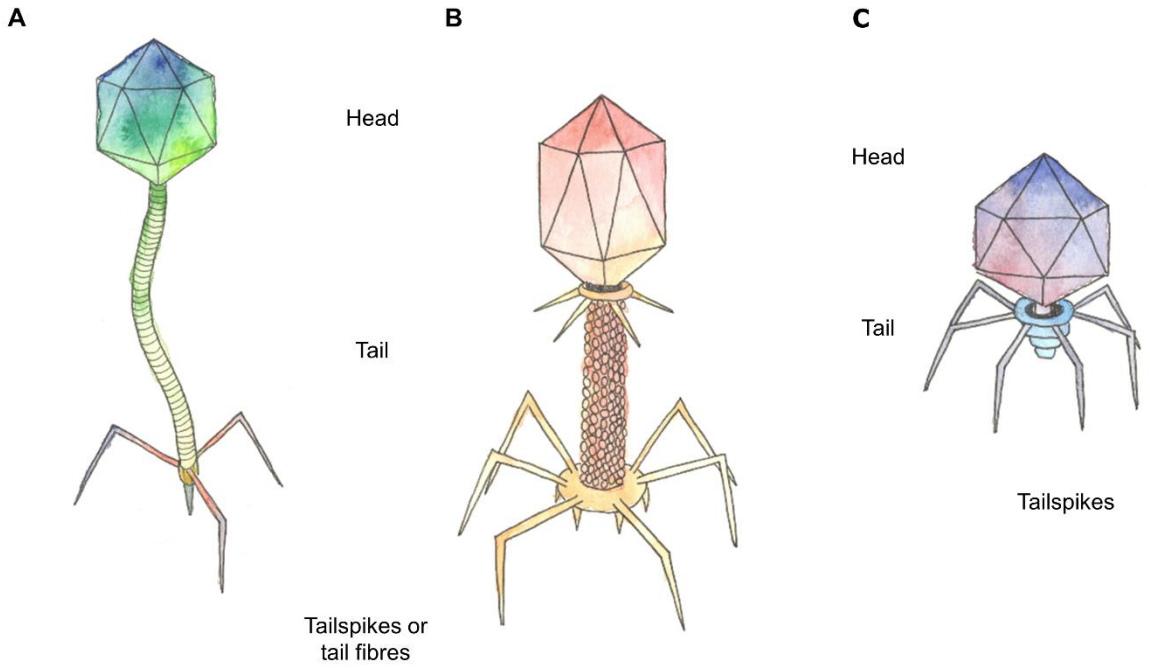


Figure 3. Caudovirales morphologies: Artistic representation of a prototypical siphovirus (A), myovirus (B) and podovirus (C). Siphovirus and podovirus have icosahedral heads while myovirus head is prolate, like T4 phage head. Taken and modified from Sanz-Gaitero et al., 2019.

Archaeal viruses are less well-studied. They display a large diversity of morphologies, ranging from families similar to tailed bacteriophages to others with shapes not observed in

other viruses such as *Ampullaviridae* (bottle-shaped) and *Spiraviridae* (coil-shaped). Besides, only a small part of their genomes show homology with previously known genomes.

The bacteriophages studied in this thesis belong to the *Caudovirales* order of tailed bacteriophages. Of known bacteriophage species, tailed bacteriophages make up 95% of them, i.e. almost 5900 are tailed bacteriophages (Ackermann & Prangishvili, 2012). However, some authors estimate that 100 million tailed-phage species may exist (Rohwer, 2003). The order *Caudovirales* is subdivided in three families according to their morphology: *Siphoviridae*, *Myoviridae* and *Podoviridae* (accounting for approximately 60, 25 and 15% of all known tailed phages, respectively). Siphoviruses have a long, flexible and non-contractile tail; myoviruses have a long, rigid and contractile tail and podoviruses have a short, rigid and non-contractile tail (Fig. 3).

Caudoviruses are dsDNA viruses with a wide range of genome sizes. Siphovirus *Leuconostoc phage L5* has only 2435 b (Almpanis et al., 2018; Rohwer & Edwards, 2002), while the *Bacillus virus G* has nearly 498 kb and hundreds of genes (Ageno et al., 1973). The genome is lodged in phage heads, or capsids, of between 43 and 160 nm in diameter. Most tailed phages have icosahedral heads, but other heads are prolate (icosahedral heads elongated along one axis) (Fig. 4 A-B). Capsids consist of at least 60 copies of the same protein, but in more complex tailed phages a separate protein makes up the five-fold vertices. Furthermore, in some viruses, there are proteins cementing or decorating the capsid.

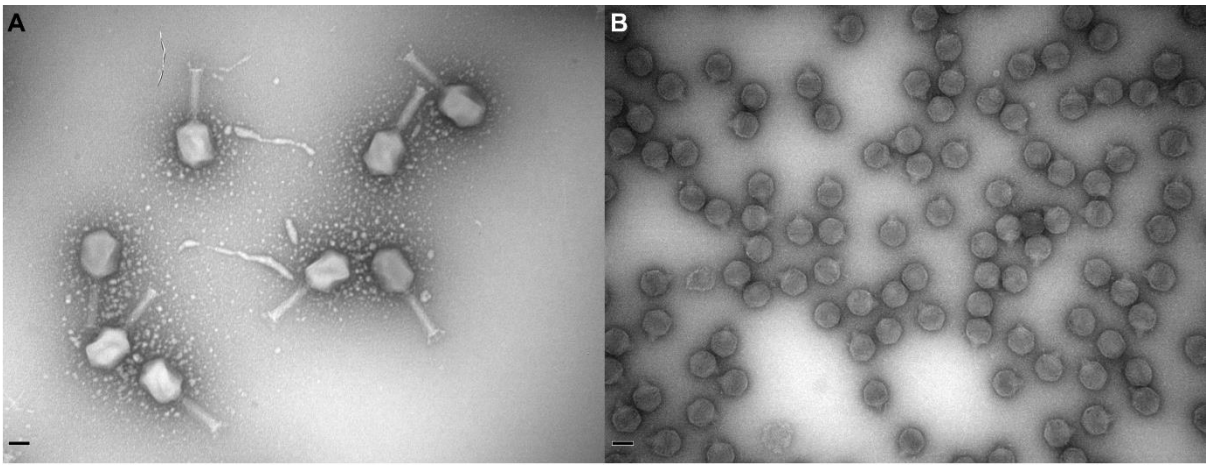


Figure 4: Electron microscopy of tailed bacteriophages. (A) Electron microscopy of the myovirus T4, with a prolate head and a long contractile tail. (B) Electron microscopy of the podovirus T7, with an icosahedral head and a short, barely visible, tail. Scale bar, in black, represents 50 nm.

As their names implies, phages belonging to the order *Caudovirales* have a tail connected to the heads through the portal protein. Podoviruses have a short, stubby tail. Siphoviruses have a tail tube formed by multiples copies of a single type of protein, ending with a pointed or plate-

shaped receptor-binding complex. In myoviruses, several copies of two different proteins form a double-walled tail, consisting of a tail tube covered by a tail sheath. Families of the order Caudovirales are classified according to the morphology of their tails (Sanz-Gaitero et al., 2019). However, this classification is subject to change based on genetic analysis (Lefkowitz et al., 2018). At the end of the tail, myoviruses and some siphoviruses have complexes called baseplates.

1.5. Tailspikes and tail fibres

Caudoviruses use their receptor binding proteins (RBPs) to select for and adsorb to their bacterial prey. They are fibrous homotrimers mainly located around the phage tail tube, but also at its tip. The RBPs on the side may be classified into tail fibres and tailspikes. Tail fibres are long, thin homotrimers, while tailspikes are shorter and stubbier complexes. Tailspikes often trimerise in three parallel beta-helices, but may also contain other folds (Seoane-Blanco et al., 2020). In addition, tailspikes generally have enzymatic activity. They are mostly endoglycosidases (endorhamnosidases or endosialidases), acetyl esterases and enzymes with anti-biofilm activity (Pires, Oliveira, et al., 2016; Prokhorov et al., 2017). Both tailspikes and tail fibres are highly stable homotrimers (Mitraki et al., 2006), with a stoichiometry of multiples of three complexes per phage particle (Nobrega et al., 2018). They emerge from the tail tube (in siphoviruses and podoviruses) or from the baseplate (in myoviruses and siphoviruses). They bind to lipopolysaccharide (LPS) of Gram-negative bacteria, teichoic acids (in Gram-positive bacteria), membrane proteins, flagella, pili or capsular polysaccharides (CPS). As phage adsorption is mediated by RBPs, they determine the host specificity and host range of their bacteriophages. Some phages have only one type of side RBP (P22; Fig. 5E-F), while others have two (T4; Fig. 5A-D) or more, up to eleven in the case of myovirus phiK64. Viruses without tail fibres or tailspikes also exist, for example SPP1 (Pan et al., 2017; Sanz-Gaitero et al., 2019).

One of the strategies to widen the host range is to bind to receptors common to different strains or species. For example, the long tail fibres of T4 phages can bind to the *E. coli* strains B and K12 core oligosaccharide and to the outer membrane protein C (OmpC) (Islam et al., 2019). The short tail fibres irreversibly bind to the secondary receptor only after the reversible binding of the long tail fibre. Another option to widen the host range is to use more than one type of RBP per phage particle, each for different host bacteria. K1-5 viruses bind to K1 and K5 *E. coli* CPS with the endosialidase K1 tailspikes and with the K5 lyase tailspikes (Scholl et

al., 2001). Det7 has four different tailspikes, each of them binding to different *Salmonella* serovars (Casjens, 2005). The lysogenic phage Mu is an interesting case, as its genome encodes for two sets of tail fibre and chaperone genes with the same N-terminal region, one set in forward sense (G+) and another in reverse sense (G-). When integrated into the host genome, mu can invert a gene segment, expressing the alternative carboxy-terminal part of the fibre and the alternate chaperone protein. This alternative carboxy-terminus changes the host range. G(+) phages infect *Escherichia* coli type K, while G(-) phages can use *E. coli* type C, *Citrobacter freundii* and *Shigella sonnei* as host (Sakai et al., 2019; van de Putte et al., 1980).

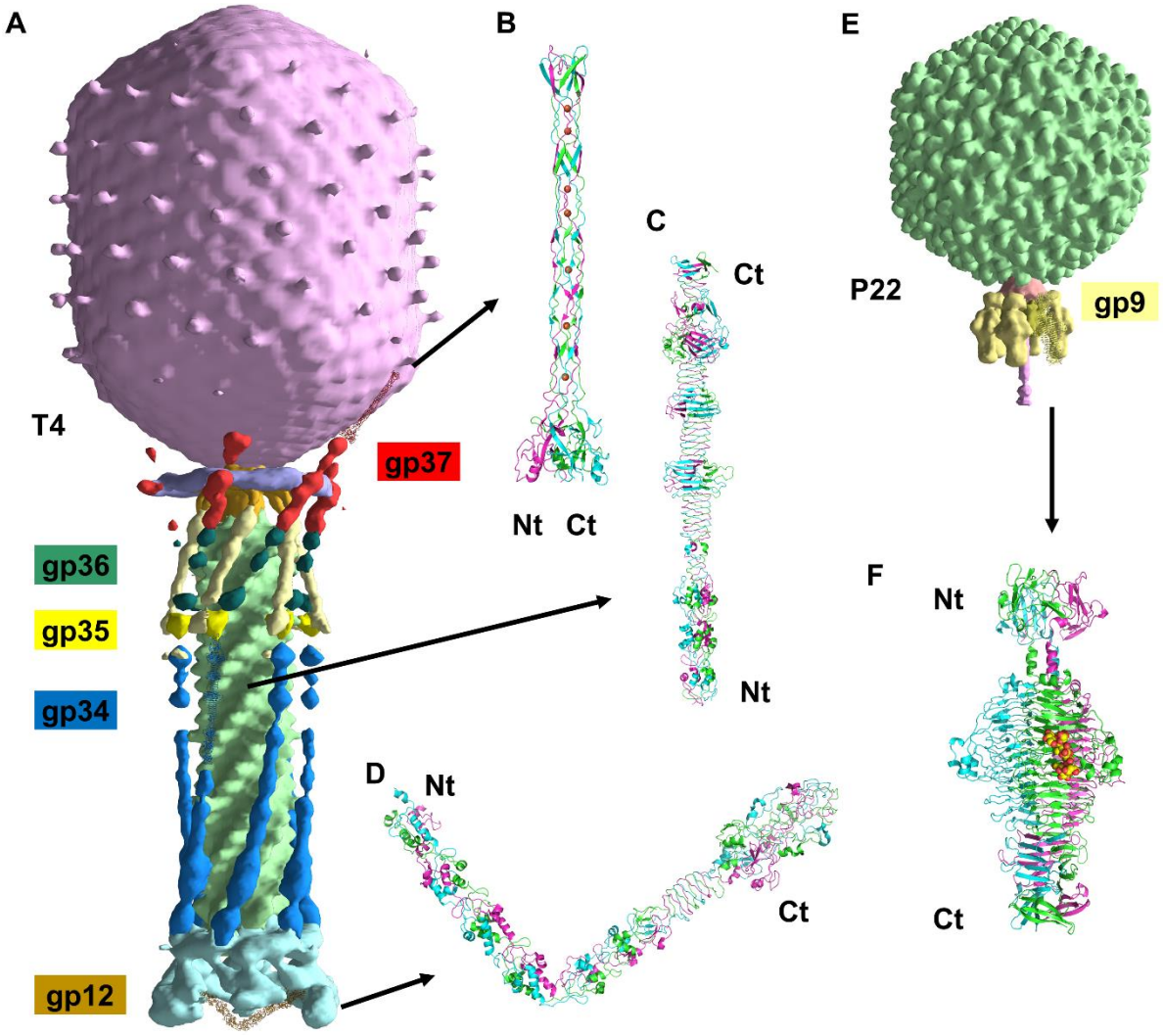


Figure 5: Tail fibres and tailspikes. (A) Bacteriophage T4 (EMDB entry EMD-2774). Tail fibre proteins are shown in blue (proximal tail fibre protein gp34), yellow ("knee" protein gp35), dark green (distal tail fibre protein gp36) and red (distal tail fibre protein gp37). The structure of the C-terminal region of gp34 and gp37 (dark ribbon diagram) are shown in their approximate positions. The baseplate (light blue) contains the six short tail fibres, which are made up of gp12 (one is shown as a brown ribbon diagram). The carboxy-terminal part of gp37 (B; PDB entry 2XGF), the carboxy-terminal part of gp34 (C; PDB entry 5NXH) and full-length gp12 (d; from PDB entry 5IV5) are shown as ribbon diagrams. (E) Bacteriophage P22 (EMDB entry EMD-1220). P22 tailspikes (gp9) are in yellow. A close-up of a gp9 trimer with each chain coloured differently is also shown (F; PDB entry 2XC1). The *Salmonella* O-antigen receptor (ball representation in red and yellow; from PDB entry 1TYX) is shown bound

to one of the chains. Amino- and carboxy-termini are indicated for each protein (Nt and Ct, respectively). Taken and modified from Seoane-Blanco et al., 2020.

1.6. The bacterial cell envelope

The bacterial cell envelope establishes the limit between the cytoplasm and the outer space. It shapes the bacterium, gives it rigidity, and controls the intake of nutrients and the exit of waste. In addition, it also protects the cell from desiccation, antibiotics, predators and pathogens. The bacterial cell envelope is composed of at least a lipid bilayer, the cytoplasmic membrane. It is made of phospholipids and membrane proteins. Gram-negative bacteria also have a thin cell wall, made of peptidoglycan, and another lipid bilayer (Outer Membrane, OM). In contrast, Gram-positive bacteria only have a thick cell wall outside the cytoplasmic membrane. In addition, some bacteria also have other structures such as capsule, flagella, pili and transport systems (Fig. 6A-B) (Silhavy et al., 2010).

A less common cell envelope is found in the family *Mycobacteriaceae* such as *Mycobacterium tuberculosis*. Its cell wall consists of two covalently bound layers: the inner one is made of peptidoglycan bound to arabinogalactan bound to mycolic acids molecules. Mycolic acids form the inner leaflet of the mycomembrane and the outer leaflets is composed of lipids and glycolipids. Pathogenic species also have a capsule mainly composed of polysaccharides and proteins (Daffé & Marrakchi, 2019). In addition, the peptidoglycan cell wall is absent in the family *Mycoplasmataceae* (Rottem, 2003) and substituted by a layer of cross-linked proteins via disulphide bonds in the family *Chlamydiaceae* (Christensen et al., 2019).

1.6.1. The Gram-negative cell envelope

The inner layer of the Gram-negative cell envelope is called the cytoplasmic membrane or inner membrane (IM). In *E. coli*, this phospholipid bilayer is mainly made of phosphatidyl ethanolamine and phosphatidyl glycerol. Then, periplasm is located between IM and OM. It is densely packed with proteins, including nucleases such as RNase and alkaline phosphatase. It also contains the peptidoglycan chains, which are oriented perpendicular to the bacterial long axis. The peptidoglycan forms a rigid exoskeleton that shapes the bacterium. The peptidoglycan consists of repetitions of the disaccharide N-acetyl glucosamine – N-acetyl muramic acid. The polysaccharide chains of the peptidoglycan are cross-linked by pentapeptides. Due to its key function, many natural and synthetic antibiotics target the proteins involved in peptidoglycan biogenesis. Furthermore, to leave the bacterium, phages produce

endolysins that degrade peptidoglycan causing bacterial lysis from within. The Gram-negative outer membrane and peptidoglycan are linked through the murein lipoprotein (Lpp), the most abundant protein in *E. coli* (about 500000 molecules per cell) (Fig. 6A) (Silhavy et al., 2010).

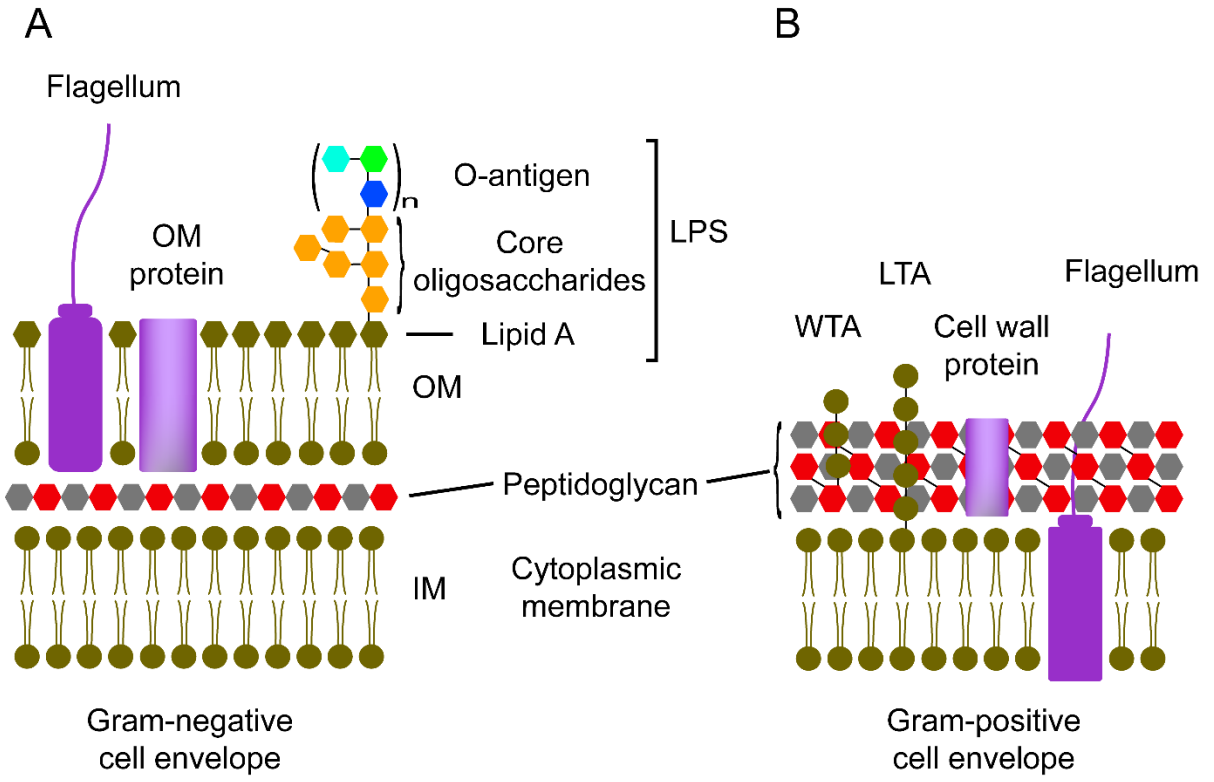


Figure 6: The bacterial cell envelope. Schematic representation of a Gram-negative (A) and a Gram-positive cell envelope (B). IM, inner membrane; LTA, lipoteichoic acid; OM, outer membrane; WTA, wall teichoic acid.

The outer membrane is important for bacteriophage host recognition. Bacterial receptors for phages are located here. It has membrane proteins, which can be used as primary or secondary receptors by bacteriophages (Marti et al., 2013). The inner leaflet of the OM is made of phospholipids (like the IM). However, the outer leaflet is composed of lipopolysaccharides (LPS). LPS consists of lipid A, core oligosaccharide and O-antigen. Lipid A is a glucosamine disaccharide with four or seven saturated acyl chains and two phosphate groups. Its structure is conserved at species level, but can be modified depending on the environmental conditions. Lipid A causes the endotoxic shock during septicaemias, as it is recognised by the immune system as an indicator of infection by Gram-negative bacteria. Core oligosaccharides are bound to the glucosamine disaccharide. Its modified composition varies among species and, sometimes, even among strains. In rough strains (no O-antigen), they function as a receptor for phages (Bertani & Ruiz, 2018; Silhavy et al., 2010).

O-antigen polysaccharide is bound to the core oligosaccharides. It is only present in smooth bacteria. O-antigen consists of a number of repetitions of two to eight saccharides. Its composition is quite variable and determines the serovar of the bacterium. It is also used as a receptor for some phages. The difference between using the core oligosaccharides or the O-antigen as receptors is important. Phages targeting core oligosaccharides of rough strains have wider host range than the ones with tropism for the O-antigen of smooth strains. Hence, O-antigen polysaccharides, which cover the core oligosaccharides, protect bacteria against most phages (Knirel et al., 2015; McConnell & Schoelz, 1983). Consequently, smooth strains would have less possible predators. For the same reason, rough strains are mainly found in laboratories or as mutants of smooth strains. O-antigens may also protect bacteria from attack by host enzymes and the immune system (Bertani & Ruiz, 2018; Kulikov et al., 2017; Osawa et al., 2013).

1.6.2. The Gram-positive cell envelope

The main difference with the Gram-negative cell envelope is the lack of an OM. Consequently, the peptidoglycan layer is thicker to protect the cell, 30-100 nm in Gram-positive bacteria and only few nm in Gram-negative. The composition and structure of the peptidoglycan is similar to that present in Gram-negative cells.

Chains of anionic polymers cross peptidoglycan layer and protrude out of it. They are called teichoic acids and are grouped in lipoteichoic acids (LTA), when bound to the cytoplasmic membrane, and wall teichoic acids (WTA), when bound to the peptidoglycan itself. Both are made of glycerol phosphate, glucosyl phosphate or ribitol phosphate repeats (Fig. 6B). Teichoic acids account for 60% of the cell envelope mass. For this reason, many phages use them as receptor, other use proteins of the cell envelope.

1.6.3. The capsule

Some bacterial strains have yet another layer outside the cell envelope called capsule (Fig. 7). It is mainly made up of polysaccharides and, in some cases, of polypeptides. Capsular polysaccharides (CPS) consist of long chains made of repetitions of negatively charged oligosaccharides. CPSs are covalently attached to phospholipids or Lipids A. They are negatively charged and hydrated molecules that form a thick layer around the cell envelope. They are present in Gram-negative and Gram-positive bacteria and have high intraspecific diversity. The *Escherichia coli* genus has 80 known CPSs (K antigens); *Streptococcus pneumonia,* 93 CPSs (serovars) and 47 CPSs (serotypes) have been found in *Campylobacter*

jejuni up to now (Wen & Zhang, 2015). The capsule protects the bacterium from desiccation, phagocytosis by neutrophils and macrophages, and complement-mediated killing (Limoli et al., 2015; Rohde, 2019). In *C. jejuni*, it is also important for chicken gut colonisation and invasion of human epithelial cells (Bachtiar et al., 2007; Bacon et al., 2001; Grant et al., 2005; Jones et al., 2004). Hence, it increases the pathogenicity of bacteria. In addition, some bacteria have receptors similar to the host saccharides. This is called molecular mimicry. This way of evading the host immune system could lead to the onset of autoimmune diseases. Guillain-Barré disease appears in 0.1% of people infected with *Campylobacter spp.* bacteria. They contain a lipooligosaccharide of the capsule that mimics the gangliosides present in neurons. After attacking bacteria, the immune system attacks the neurons (Burnham & Hendrixson, 2018).

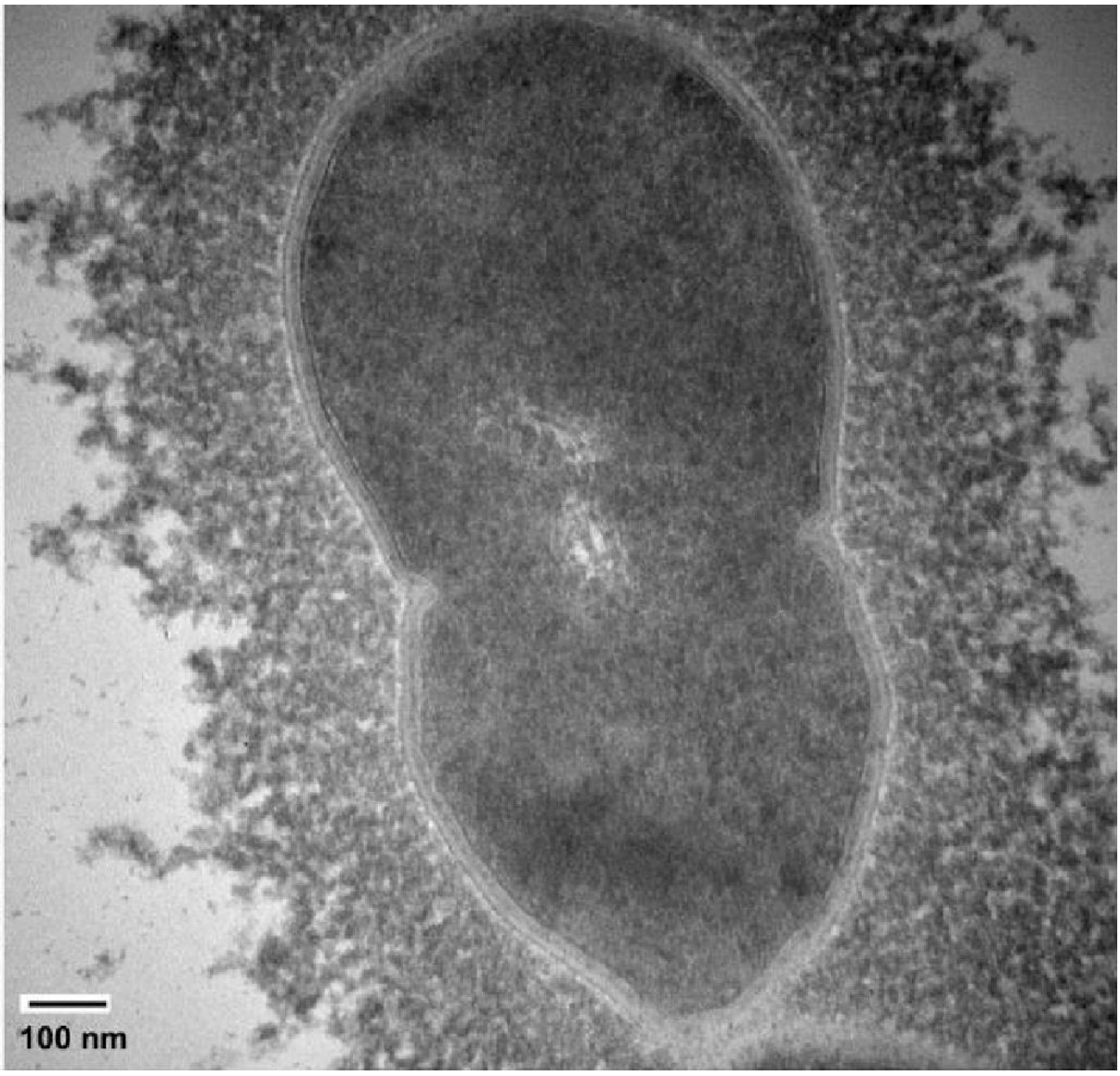


Figure 7: Electron micrograph of bacterial capsules from *S. pneumoniae* strain ST556. Figure from Wen & Zhang, 2015.

One of the main concerns for the future of health is the rise of diseases caused by antibiotic-resistant bacteria. Resistance to penicillin was reported shortly after its discovery and now, multidrug-resistant bacteria are common in hospitals (Arias & Murray, 2009; Bragg et al., 2014). Although developing resistance to toxic agents is a natural process driven by natural selection, resistance against antibiotics have been enhanced by human mishandling. Misuse of antibiotics happens in human and veterinary medicine, meat and fish production and agriculture (Holmes et al., 2016). In addition, due to a lack of financial incentives, the last novel classes of antibiotics were developed in the 1980s (Durand et al., 2019).

Though not the most concerning diseases related to antibiotic resistance, campylobacteriosis and salmonellosis are the first and the second most prevalent gastrointestinal diseases in Europe in 2018, with 246571 and 91857 cases, respectively. They account for 95% of detected zoonotic cases, mainly food-borne; 90% of hospitalisations and 30% of deaths ('The European Union One Health 2018 Zoonoses Report', 2019). As *Salmonella spp.* and *Campylobacter spp.* are an important source of morbidity and mortality and antibiotic resistance is spreading, new ways to fight and detect these bacteria are needed.

1.7. Enzymatic activity in tailspikes

To avoid phage infection, some bacteria hide their secondary receptors with thick layers of polysaccharides (PS). In order to approach the lipid membrane, adapted bacteriophages possess tailspikes that processively degrade them. After hydrolysing one bond, the released fragment diffuses away and the next saccharides of the chain can move to the empty position. In this manner, tailspikes clear the way and place the phage perpendicular to the cell envelope (Broeker & Barbirz, 2017).

Most bacteriophages have endoglycosidase activity. For example, P22 tailspikes hydrolyse the α1-3 glycosidic link between galactose and rhamnose. Hence, they have endorhamnosidase activity (Iwashita & Kanegasaki, 1976; Steinbacher et al., 1994) and belong to the family GH90 in the CAZy classification (http://www.cazy.org/ (Lombard et al., 2014)). Glycosidases show negatively charged catalytic sites, and the catalytic residues usually are aspartic or glutamic acids. Glycosidases have two mechanisms of action: retaining and inverting. These families are differentiated by the distance between the general acid and nucleophile: 5.5 Å and 10 Å, respectively (Davies & Henrissat, 1995; Zechel & Withers, 2000). The retaining mechanism takes places in two steps. First, the general acid protonates the

glycosidic oxygen, which cleaves the glycosidic bond. The nucleophile binds the anomeric carbon and form a glycosyl-enzyme complex. Eventually, the general acid deprotonates an incoming water molecule, which attacks the anomeric centre and releases the saccharide (Fig. 8A). In the inverting mechanism, the catalytic residues are more separated, because a water molecule has to fit between the glycosidic bond and the nucleophile. The inverting mechanism only has one step. The general acid protonates the glycosidic oxygen, which is released, and a deprotonated water molecule substitutes it (Fig. 8B).

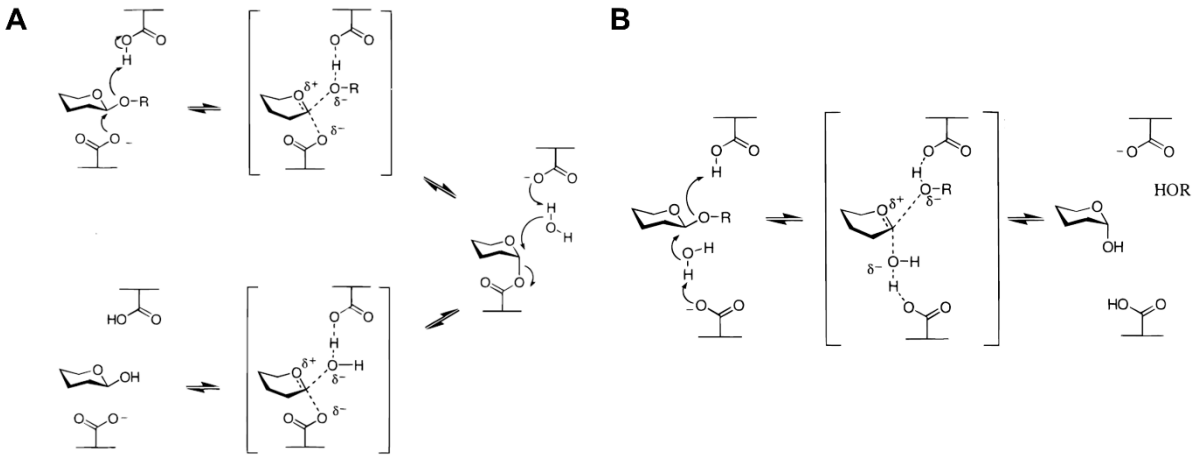


Figure 8: Schemes of the enzymatic hydrolysis of the glycosidic bond by the retaining (A) and the inverting glycosidases (B). Taken and modified from Zechel & Withers et al., 2000.

A group of phages targeting the *S. enterica* serovar Typhi Vi capsule have a conserved acetyl esterase domain in their tailspikes. The esterase cleaves the acetyl moieties that decorate the capsular polysaccharide without hydrolysing the backbone chain. The biological function might be to reduce the number of hydrogens bonds between polysaccharide chains and to allow the virus to approach the membrane (Pickard et al., 2010). Phage G7C tailspike gp63.1 and phage PP99 tailspike gp55 also deacetylate acetyl-decorated lipopolysaccharides (LPS). Gp63.1 cleaves the ester bonds of one acetyl group each six saccharide residues, while PP99 tailspike completely cleaves the acetyl groups present in the deoxytalose, but leaves the mannose-bound acetyl intact (Lukianova et al., 2020; Prokhorov et al., 2017). Despite containing a beta-helical fold, Gp63.1 tailspike does not have the glycosidase activity other tailspikes have (Steinbacher et al., 1994). Therefore, while the O-antigen backbone stays intact, some of the acetyl groups are removed. This activity is essential for adsorption, as it cannot bind to deacetylated LPS. In this case, acetyl groups are not a barrier, but essential for bacterial infection (Broeker & Barbirz, 2017).

1.8. *Salmonella virus epsilon15*

Bacteriophage epsilon15 is a podovirus that infects Salmonella species belonging to the Salmonella serogroup E1, of which *Salmonella enterica subsp. enterica* serovar Anatum A1 is a member (Guichard et al., 2013; Sechter & Sechter-Mooreville, 1990). Epsilon15 has a 40-kb genome. The genome is packed inside an icosahedral capsid with a diameter of 70 nm (Fig. 9A). The capsid is composed of 60 hexamers and 11 pentamers of the major capsid protein, in total 415 copies of gp7 (Jiang et al., 2006). The same number of gp10 proteins is present in the capsid. They form dimers that work as staples between two neighbouring hexamers (Baker et al., 2013; Jiang et al., 2008). A so-called "special" vertex contains a dodecameric portal complex instead of a pentamer of the major capsid protein. A core complex is bound to the portal inside the capsid. Outside the capsid, a hexameric tail hub is bound to the portal protein. The tail hub is a 17 nm long tube to which six tailspikes were thought to be bound (Jiang et al., 2006). However, more recent reports suggest that the tailspikes, made of gp20 proteins, bind to directly to the portal complex (Guichard et al., 2013). Tailspikes are divided in two regions separated by a kink (Jiang et al., 2006): a thin phage-binding arm and a three-fold flower-like region (Murata et al., 2010). This suggests that the epsilon15 tailspike is also trimeric, as most tail fibres and tailspikes are (Fig. 9B) (Seoane-Blanco et al., 2020).

Epsilon15 tailspikes reversibly bind the *Salmonella enterica subsp. enterica* serovar Anatum A1 (O:3,10) O-antigen (Kanegasaki & Wright, 1973; Sechter & Sechter-Mooreville, 1990). This O-antigen is made up of about 40 repetitions of the acetylated trisaccharide D-Galp[6Ac]-α-1 – 6-D-Manp-β-1 – 4-L-Rhap-α-1 – ROH linked by α-1 – 3 bonds (Fig. 9C-D) (Gajdus et al., 2008; Mcconnell & Schoelz, 1983). Polysaccharide acetylation prevents hydrolysis by rhamnogalacturonases (Kofod et al., 1994; Schols et al., 1990; Searle-van Leeuwen et al., 1992), lysozyme penetration in Gram-negative bacteria (Kulikov et al., 2017) and acts as a non-specific shield against phage infection (Golomidova et al., 2016; Knirel et al., 2015). Some bacteriophages have acquired the capacity of degrade ester bonds to overcome this protection or even depend on the acetyl group presence to infect their host bacteria, such as G7C (Knirel et al., 2015). Epsilon15 can infect bacteria with or without the acetyl group (Wright, 1971) and acetyl groups are not required for phage adsorption (Hagiwara et al., 1966).

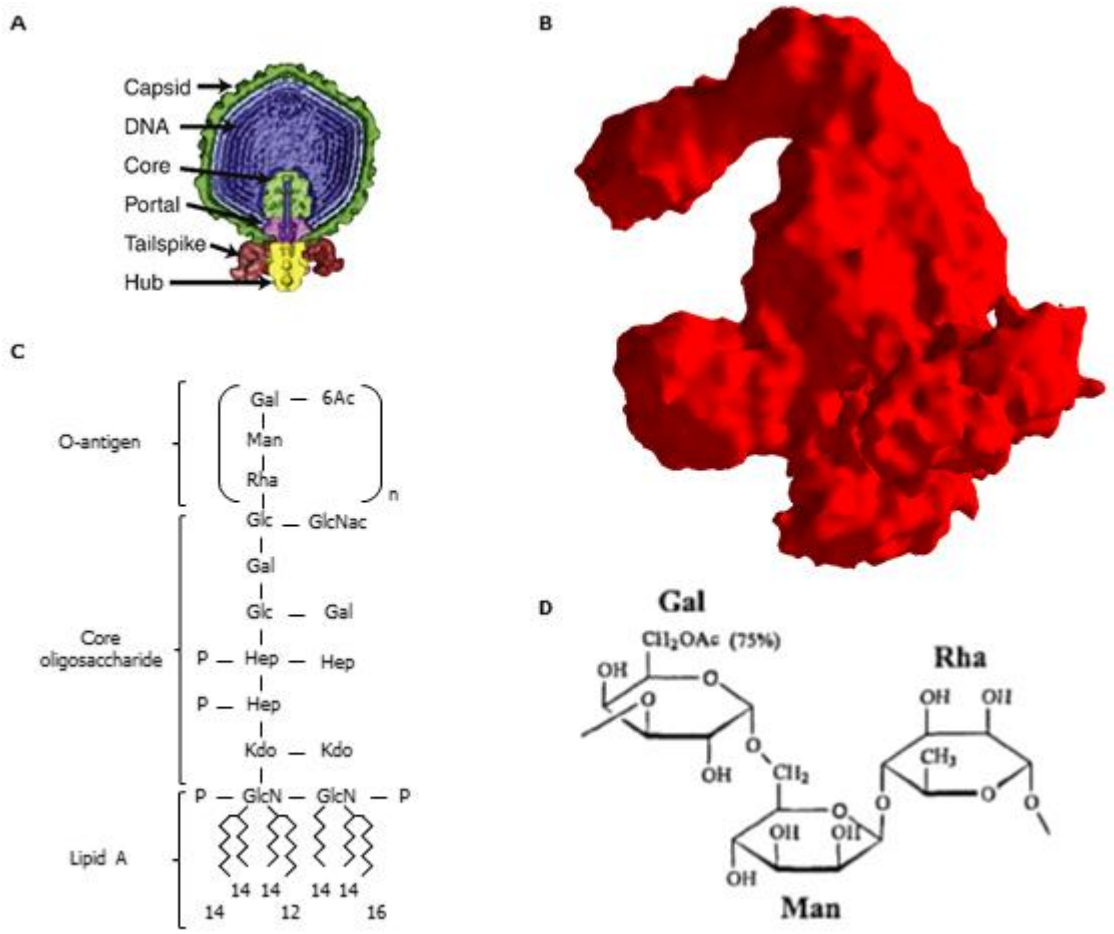


Figure 9. Schematic representation of Gram-negative and Gram-positive cell envelopes: (A) Cut-view of a epsilon15 phage single particle map (accession number: EMD-5216). Taken from Chang et al., 2010. (B) Epsilon15 tailspike taken from a whole-virus epsilon15 map (accession number: EMD-5207). (C) Salmonella LPS structure. Lipid A and core oligosaccharide correspond to *S. enterica* serovar Typhimurium LPS (Bertani & Ruiz, 2018 and Di Lorenzo et al., 2019). O-antigen corresponds to S. enterica serovar Anatum (Gadjus et al., 2008). (D) Haworth projection of the *S. enterica* serovar Anatum O-antigen repeating unit shown in (C).

Gp20 hydrolyses the α1-3 link between the L-rhamnose and the D-galactose, generating hexasaccharides and other oligosaccharides multiples of three (Kanegasaki & Wright, 1973). Epsilon15 uses its endorhamnosidase activity to penetrate the O-antigen layer. Eventually, it approaches the lipid A and the core oligosaccharide, where a putative secondary receptor may exist (Chang et al., 2010). Then, genome is ejected into the bacterium.

Epsilon15 may adopt the lytic or to a lysogenic pathway. In the lysogenic cycle, the epsilon15 prophage induces the bacterial ORF17.4 gene, which is normally dormant (McConnell et al., 2001). Expression of the orf17.4 protein leads *Salmonella enterica subsp. enterica* serovar Anatum to produce type 15 O-antigen instead of type 10. In type 15 O-antigen, the three-saccharide repeating units are joined by a β-glycosidic bond (between rhamnose and

galactose) instead of the α-glycosidic bond of type 10. Bacteria with the epsilon15 lysogen are known as Anatum A15. Other epsilon15 phages cannot bind to its LPS molecules; hence, the prophage protects the bacterium by the same phage.

1.9. *Campylobacter virus F358*

Most of the found campylobacter bacteriophages are myoviruses belonging to the subfamily *Eucampyvirinae* (Javed et al., 2014). They are grouped in two genera, *Firehammervirus* (formerly *Campylobacter virus CP220*) and *Fletchervirus* (formerly *Campylobacter virus CP81*). They mainly differentiate in the genome length (180 vs 140 kb) and tail length (150 vs 115 nm); both have a baseplate at the end of the tail. The capsid diameter is about 95 nm in both groups (Javed et al., 2014). Receptor-binding proteins are 100-nm long in firehammerviruses with small globular domains in the distal region. In fletcherviruses, RBPs are short, thin and folded back towards the tail.

Campylobacter phages use flagella or capsular polysaccharides as bacterial receptors. Firehammerviruses usually, but not universally, bind to bacterial flagella and fletcherviruses usually, but not always, bind to capsular polysaccharides (CPS) (Sørensen et al., 2015). Both phage genera and bacterial receptor correlates with isolation strain. Phages isolated from RM1221 *C. jejuni* strain belong to the genus *Firehammervirus* and bind to flagella. On the contrary, phages isolated from the NCTC12662 and the NCTC12658 strains belong to the genus *Fletchervirus* and mostly bind to CPS (Baldvinsson et al., 2014; Sørensen et al., 2015).

Capsular polysaccharides are formed by repetitions of sugars. Sometimes, moieties such as O-methyl phosphoramidate (MeOPN) and O-methyl groups are bound to the backbone chain. Many CPS-dependent phages need MeOPN to adsorb to a number of *C. jejuni* strains (Sørensen et al., 2015). In NCTC11168 strain, phages produce plaques when galactose-bound MeOPN moieties are present, but heptose-bound MeOPN groups inhibit plaque formation. Besides, phages require O-methyl groups bound to the carbon 3 of heptoses to form plaques, but located in carbon 6, they reduce or even inhibit plaque formation (Sørensen et al., 2012).

About 70% of *C. jejuni* strains have MeOPN in their CPSs (McNally et al., 2007). It protects the bacterium against the attack of the immune system and increases its pathogenicity (Pequegnat et al., 2017; van Alphen et al., 2014). Its presence is partially controlled by phase-variable genes (Gencay et al., 2018). Phase-variation occurs by variations in homopolymeric G/C tract lengths turning gene expression on or off. It allows *C. jejuni* strains to produce >700 CPS structural variants or phasotypes, which can randomly change (Aidley et al., 2017;

Sørensen et al., 2012). Such a diverse population is an advantage against dangers such as the immune system or bacteriophages.

Campylobacter virus F358 is a Fletchervirus that infects the campylobacter strain NCTC12662 (Gencay et al., 2018). This strain is able to gain phage resistance with phase-variable expression of MeOPN-producing proteins. To compensate for this evasion mechanism, F358 has four RBP genes (RBP1, RBP2, RBP3 and RBP4), and two of them (RBP2 and RBP3) are phase-variable (Martine Sørensen, personal communication). Next to the RBP genes, genomic analysis revealed another gene corresponding to a putative chaperone, which may help to fold the RBPs. The phase-variable RBPs and the multiplicity of RBPs allows F358 to adsorb to different receptors and avoid the phage resistance.

1.10. Conclusion

Phage research has led to many molecular biology applications. Recently, phage therapy has gained new impulses. However, relatively little is known about a key step in phage cycle: adsorption. In this thesis, the structures of two phage receptor binding tailspikes have been determined: that of *Salmonella enterica* serovar Anatum phage epsilon15 and of the *Campylobacter jejuni* strain NCTC12662 phage F538. For the epsilon15 tailspike, structures with the O-antigen receptor have also been determined and the endorhamnosidase activity has been characterised.

Detailed structural and functional knowledge of these receptor-binding proteins will inform us on their receptor-binding properties and allow us to better understand phage infection. It may also allow site-directed mutation to direct phages to different receptors, or the generation of chimeric receptor-binding proteins for the same purpose. Possible applications are detection and elimination of bacteria and incorporation into protein-based biotechnological devices.

2. Objectives

Tailspikes are key to phage-bacterium adsorption. The general aim of this thesis is to structurally characterise this bacteriophage organelle and unravel its interactions with bacterial receptors at atomic level. In the future, this information might benefit phage therapy, designing of bacterial detectors and other applications.

In particular, the following objectives were pursued:

1- Cloning, expression and purification of the *Salmonella enterica* serovar Anatum O3,10 phage epsilon15 gp20 tailspike and determination of its structure without and with the receptor.

2- Characterisation of the endorhamnosidase activity of the gp20 tailspike.

3- Expression and purification of the *Campylobacter jejuni* strain NCTC12662 phage F358 RBP3 tailspike and determination of its structure.

3. Materials and Methods

3.1. Materials

3.1.1. Vectors

Construct gp20 (434-1070) was ordered (GenScript) as a construct already inserted in the plasmid pET28a(+). Codons were optimised using the OPTIMIZER web server (Puigbò et al., 2007) with the codon list and frequencies for *Escherichia coli* B obtained from the web site http://www.kazusa.or.jp/codon (Nakamura et al., 2000).

An expression plasmid for the RBP3 tailspike was kindly provided to us by Dr. Martine Holst-Sørensen from the laboratory of Dr. Lone Brønsted (University of Copenhagen, Denmark). The RBP gene and its chaperone are cloned in pET28a(+).

3.1.2. Primers

Table 1: Primers used to clone gp20 constructs.

Name	Sequence	Restriction site	Constructs
E15fib812for	GGTCGGATCCGGGGCGGGGCTAGTGATGTTGAAG	BamHI	812-1070
E15fib734for	GGTCGGATCCTCGGTGATAGCGCGGACCAATTGC	BamHI	734-1070
E15fib520for	GGTCGGATCCTTTCTGAAAATATAATCACCGACAAC	BamHI	520-1070
E15fib287for	GGTCGGATCCATGCAGTTTTATGCATCAATAAAAG	BamHI	287-1070
E15fib200for	GGTCGGATCCCTGCTTTTGATAGTAATGGAAATG	BamHI	200-1070
E15fib2for	GGTCGGATCCTGACGGTTTCAACCGAAGTTGACC	BamHI	2-1070
E15fibrev	ATAAGAATGCGGCCGCTTACCCACCCCACCCCTTCGACTT	NotI	2-1070 200-1070
E15fibrevHindIII	CCGGAAGCTTTTACCCACCCCACCCCTTCGACTT	HindIII	248-1070 287-1070 520-1070 734-1070 812-1070
E5Fw	TCAGCAAGGGCTGAGGGAAGTTGACCACAATGACTACAC	[a]	5-768
E248Fw	TCAGCAAGGGCTGAGGGAAAAGCTTCAAAACGTTGTTTATCC	[a]	248-768
E264Fw	TCAGCAAGGGCTGAGGCCAACCGATGGGACTAGC	[a]	264-768
E15a768Rv	TCAGCGGAAGCTGAGGTTAGCCAACCATACGTAGATCTG	[a]	5-768 249-768 264-768
E15del248fw	GGACAGCAAATGGGTCGGATCCTGGAAAAGCTTCAAAACGTTGTTTATCC	[b]	248-1070

E15del248rf	GGATAAACAACGTTTTGAAGCTTTTCCAGGATCCGAC CCATTTGCTGTCC	[b]	

[a] Constructs without restriction site were cloned in the plasmid pHTP1 system (NZYTech) based on a single ligase-independent reaction.

[b] Generated by deletion from the gp20(2-1070) construct.

Figure 10: Gp20 and gp208 sequence alignment. Sequence alignment of epsilon15 tailspike and one of the Det7 tailspikes, gp208. Starting redidues of the gp20(248-1070) construct and gp208ΔN structure are marked by blue and magenta arrows respectively.

Table 2: Primers used to generate gp20(248-1070) mutants.

Name	Sequence
D449N_Fw	GGGACGATACCAGTAACTACTGGAATGGTATTGTTGGCGGCGGTGAATG G
D449N_Rv	CCATTCACCGCCGCCAACAATACCATTCCAGTAGTTACTGGTATCGTCCC
S477A_Fw	GAGTTTGAGGATAATGGTCAGGCCGGTATCTATGCTGGTGGCAACGGGG G
S477A_Rv	CCCCCGTTGCCACCAGCATAGATACCGGCCTGACCATTATCCTCAAACTC
N503D_Fw	ACCACATACATGGAAACTGGGATCGCGGTATAGATTTTGGGGTTG
N503D_Rv	CAACCCCAAAATCTATACCGCGATCCCAGTTTCCATGTATGTGGT
D507N_Fw	GGAAACTGGAATCGCGGTATAAATTTTGGGGTTGTACAGCGTCTTGC
D507N_Rv	GCAAGACGCTGTACAACCCCAAAATTTATACCGCGATTCCAGTTTCC
S787A_Fw	GCCGTTCAGTTACCGTGGTTGGGCATGCCTTCTGTACCAGTGATGTTATG AGC

S787A_Rv	GCTCATAACATCACTGGTACAGAAGGCATGCCCAACCACGGTAACTGAA CGGC
H1038A_Fw	CGCTCTCTGCGAGAAGATAACCTGGCCCCATCTGAAACACTACAGCC
H1038A_Rv	GGCTGTAGTGTTTCAGATGGGGCCAGGTTATCTTCTCGCAGAGAGCG
W441R_Fw	GGATGTCTTCTGAAGCCAAGCCGGCGGACGATACCAGTAACTACTGG
W441R_Rv	CCAGTAGTTACTGGTATCGTCCGCCGGCTTGGCTTCAGAAGACATCC

Primers with restriction sites were designed by Mark J. van Raaij based on secondary structure prediction. I designed construct 434-1070 based on secondary structure prediction and the other primers based on knowledge from secondary structure prediction and the (by then known) petal domain structure. Gp20(248-1070) was designed based on sequence homology with the gp208ΔN protein, one of the Det7 tailspikes (Table 1, Fig. 10), for which the structure became available during my thesis period. Specifically, amino acid 248 was chosen to include a restriction site for HindIII in the amino-acidic position 249 and 250 (bp 745-750). Primers used for mutagenesis were designed based on structural knowledge from gp20ΔN (Table 2).

3.1.3. Bacterial strains

For cloning and expression experiments, the following bacterial strains were used:

Table 3: *E. coli* strains used in this work.

E. coli strain	Genotype	Purpose	Supplier
Top10	F- *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*araleu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G	Cloning and plasmid propagation	Invitrogen
DH5α	F- φ80*lac*ZΔM15 Δ (*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17(rk^-, mk^+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1λ-	Cloning and plasmid propagation	Invitrogen
XL10-Gold	Tet^R Δ(*mcr*A)183 Δ(*mcr*CB-*hsd*SMR-*mrr*)173 *end*A1 *sup*E44 *thi*-1 *rec*A1 *gyr*A96 *rel*A1 lac Hte [F' proAB $lacI^q$ ZΔM15 Tn10(Tet^R) Amy Cam^R]	Cloning	Stratagene
BL21 (DE3)	*E. coli* strain B F^- *ompT hsdS*$_B$ (r_B^-, m_B^-) *gal dcm* (DE3)	Protein expression	Novagen
B834 (DE3)	*E. coli* strain B F^- *ompT hsdS*$_B$ (r_B^-, m_B^-) *gal dcm met* (DE3)	Protein expression; methionine auxotroph	Novagen
BL21-CodonPlus (DE3) RIL	*E. coli* strain B F^- *ompT hsdS*(r_B^- m_B^-) dcm^+ Tet^R *galλ* (DE3) *end*A Hte [*argU ileY leuW* Cam^R]	Protein expression	Agilent

E. coli strains Top10 and DH5α were used for cloning and plasmid propagation, because they carry the recA1 gene (a mutant of recA) (Table 3). This increases insert stability and prevents unwanted recombination. In addition to recA1, the XL10-Gold strain exhibits the Hte phenotype. It increases the transformation efficiency of ligated and large DNA molecules.

For protein expression, strains carrying the λ (DE3) lysogen were used. It contains the T7 RNA polymerase gene under the control of *lac*UV5 promotor. The lactose operon inhibitor gene (*lac*I), presents in the plasmids, represses lacUV5 in non-induced conditions. Protein expression is induced by adding isopropyl-β-D-1-thiogalactopyranoside (IPTG). IPTG blocks the repressor; hence, *lac*UV5 can start the expression of T7 RNA polymerase and the protein of interest.

Strain BL21-CodonPlus (DE3)-RIL has additional features. It codifies for four tRNAs translating rare codons (AGA, AGG, AUA and CUA).

3.1.4. Plasmids

Plasmids pET28a(+), pET28c(+) (Novagen) and pHTP1 (NZYTech) were used as vectors. Plasmids pET28a(+) and pET28c(+) encode six-histidine tags (His-tags) just before and after the multi-cloning site, for inclusion at the N- and C-termini of the resultant expressed peptides. However, the sequence encoding the C-terminal his-tag was avoided by inserting a stop codon in front of it. On the other hand, pHTP1 only encodes a his-tag for inclusion at the N-terminal end. All plasmids include a kanamycin resistance gene to select for bacteria containing the plasmid. Besides, they have a T7 promotor and terminator, so they can use the T7 RNA polymerase provided by the host cells to generate the messenger RNA necessary for expressing the target protein.

3.1.5. Anatum O-antigen polysaccharide and oligosaccharides

Purified samples of *Salmonella enterica subsp. enterica* Anatum polysaccharide, nonasaccharide and hexasaccharide were a kind gift from Dr Stefanie Barbirz laboratory (University of Potsdam, Germany). They purified these samples from *Salmonella enterica subsp. enterica* serovar Anatum O3,10 (1300188), a *Salmonella* strain from the serogroup E. One repeating unit (RU) of Anatum O-antigen consists of the trisaccharide D-Galp[6Ac]-α-1 – 6-D-Manp-β-1 – 4-L-Rhap-α-1 – ROH linked by α-1 – 3 bonds. The acetyl group is labile and can be degraded upon freeze/thaw cycles.

3.1.6. Media for bacteria

a) LB medium (Luria-Bertani): 10 g/L tryptone, 5 g/L yeast extract, 10 g/L sodium chloride. Used for growing *E. coli* strains. Supplemented with 50 µg/mL of kanamycin when cells contain some of the plasmids mentioned above.

b) LB-agar: 10 g/L tryptone, 5 g/L yeast extract, 10 g/L sodium chloride, 15 g/L bacto-agar. Used for growing *E. coli* strains on a solid medium to obtain single colonies. Kanamycin was added as above when necessary.

c) SOC (Super Optimal broth with Catabolite repression): 20 g/L tryptone, 5 g/L yeast extract, 10 mM sodium chloride, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 20 mM glucose. Recovery medium used in plasmid transformation to incubate *E. coli* cells before plating.

d) TSS (Transformation and Storage Solution): LB supplemented with 12% (w/v) PEG 8000, 5% (v/v) DMSO, 25 mM $MgCl_2$ and 1% (v/v) glycerol. Used to store chemically competent cells.

3.1.7. Buffers and reagents

a) TAE buffer (1X): 40 mM Tris-HCl pH 8.3, 20 mM acetic acid and 1 mM EDTA.

b) DNA loading buffer (6X) 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol FF and 30% (v/v) glycerol.

c) SDS-PAGE electrophoresis buffer: 25 mM Tris-HCl pH8.3, 192 mM glycine and 0.1 (w/v) SDS.

d) Protein loading buffer (5X): 0.25% (w/v) bromophenol blue, 0.5 M DTT, 50% (w/v) glycerol, 10% (w/v) SDS and 0.25 M Tris-HCl pH 6.8.

e) Coomassie brilliant blue staining buffer: 0.005 (w/v) Coomassie brilliant blue R-250, 20% (w/v) ethanol and 5% (w/v) acetic acid.

f) Lysis buffer: 20 mM Tris-HCl pH 7.5, 0.5 M NaCl, 5% (w/v) glycerol and 20 mM imidazole.

3.2. Methods

3.2.1. DNA methods

3.2.1.1. DNA amplification by PCR

DNA amplification was performed using the Taq DNA Polymerase kit (Thermo Scientific) following kit instructions. Primers (Sigma-Aldrich) and deoxynucleotides (Sigma Aldrich) were supplied from external sources and epsilon15 was kindly provided by Dr Michael McConnell (Point Loma Nazarene University, San Diego, USA) to be used as a template. Polymerase Chain Reactions (PCR) took place in a MJ Mini^{TM} Personal Thermal Cycler (Bio-Rad) and consisted of the following steps:

Initial denaturation:		94 ºC, 4 min
30 cycles of	Denaturation:	94 ºC, 1 min
	Annealing:	T_m ºC, 1 min
	Extension:	72 ºC, 1 min/kb
Final extension:		72 ºC, 10 min

Amplified gene fragments from a PCR were purified with the GeneJET PCR Purification kit (Thermo Fisher Scientific).

For cloning experiments using pHTP1 plasmids, PCR products were digested with DpnI enzyme after the amplification according to manufacturer instructions (NZYTech).

3.2.1.2. DNA agarose electrophoresis

Amplified gene fragments and constructions were checked by DNA agarose electrophoresis. Agarose gels were prepared by mixing 1% agarose (w/v) in TAE buffer with 0.5 μg/mL of ethidium bromide. Samples were mixed with DNA loading buffer (6X) before being loaded.

3.2.1.3. Restriction enzyme digestion

Amplified gene fragments and circular plasmids were digested at 37 ºC with the necessary restriction enzymes (Table 1) according to the manufacturer guidelines (Thermo Fisher Scientific). Digested gene fragments and plasmids were resolved in an agarose gel and their bands cut with a scalpel and subsequently purified with the GeneJET Gel Extraction kit (Thermo Fisher Scientific).

3.2.1.4. Ligation

In ligase-dependent cloning, digested gene fragments and digested pET28c(+) plasmids were ligated at 16 ºC overnight using T4 DNA ligase (New England Biolabs), following provider instructions. 50 ng of vector were used and the proportion of vector to insert was 1:3. In ligase-independent cloning, DpnI digested products and linear pTHP1 were ligated using the NZYEasy Cloning and Expression kit I (NZYTech). Reactions consisted of the following steps: a 60-min incubation at 30 ºC, then another 10-min incubation at 80ºC and finally, 10 min at 30 ºC.

After ligations, resulting products were transformed into *E. coli* TOP10 or DH5alpha strains to amplify the constructs and purified. The amplified plasmids were digested with the appropriate restriction enzymes and analysed by agarose gel electrophoresis to identify successful ligations.

3.2.1.5. Plasmid purification

Plasmids were purified from 10-mL cultures of TOP10 or DH5alpha grown overnight at 37 ºC. The GeneJET Plasmids Miniprep kit (Thermo Fisher Scientific) or NZYMiniprep kit (NZYTech) were used as indicated in their protocols.

3.2.1.6. Deletion and Mutagenesis

PCR was also used for construction of the gp20(248-1070) gene fragment using the gp20(2-1070) construct as a template. In this case, QuikChange II XL Site-Directed Mutagenesis Kit reagents and guidelines were used. Then, amplified DNA fragments were digested with DpnI and, eventually, transformed in Top10 cells. Mutagenesis of the gp20(248-1070) construct were performed with the NZYMutagenesis kit (NZYTech) according to supplier instructions. The same steps as in deletion were performed here.

3.2.2. Bacterial methods

3.2.2.1. Chemical transformation of bacteria

Competent cells were thawed on ice and then, 50 µL of competent bacteria were incubated with 50-100 ng of plasmid on ice for 5-30 min. Afterwards, a heat shock at 42 ºC for 45 s was performed. Subsequently, the tube was left on ice for 2 min. Later, 250 µL of pre-warmed SOC or LB medium were added to the tube and it was incubated at 37 ºC for 1 h at

200 rpm. The mixture was subsequently plated onto pre-warmed LB-agar plates supplemented with kanamycin and incubated overnight at 37 ºC.

3.2.2.2. Preparation of chemically competent cells

One colony of the required *E. coli* strain was incubated in LB medium, with an antibiotic if necessary, at 37 ºC overnight at 120 rpm. One mL of this culture was mixed with 99 mL of LB medium and then incubated as before until the culture reached an $O.D._{600}$ = 0.3. Then, cells were centrifuged at 3000 x *g* for 10 min at 4 ºC and resuspended in 10 mL of ice-cold TSS. Aliquots of 150 µL were flash-frozen in liquid nitrogen and stored at -80 ºC.

3.2.2.3. Cell culture, protein expression and cell disruption

One colony from a bacterial strain transformed with the target construct was incubated in LB with 100 mg/mL of kanamycin at 37 ºC overnight at 100 rpm. Then, the overgrown culture was diluted 1:100 with fresh medium. When the $O.D._{600}$=0.6-0.8, 1 mL of IPTG 1 M was added to induce protein expression at 15-25 ºC overnight at 120 rpm. Cells were pelleted at 6000 x *g* for 10 min at 4 ºC and resuspended in Lysis buffer. Then, cells were disrupted with a Digital Sonifier 250 (Branson) and centrifuged at 15000 x *g* for 45 min at 4 ºC. Gp20 constructs were expressed in *E. coli* BL21(DE3) or in *E. coli* B834 for selenomethionine tagged proteins. RBP3 was expressed in *E. coli* BL21-CodonPlus (DE3)-RIL.

3.2.2.4. Polysaccharide purification

Salmonella enterica serovar Anatum polysaccharide was purified from 10 L of a bacterial culture incubated at 37 ºC in LB medium overnight. Cells were harvested by centrifuging at 5000 x *g* for 10 min at 4 ºC. They were resuspended in a solution of 10% (v/v) acetic acid and boiled for 1.5 h under reflux. Afterwards, the resuspension was cooled on ice for 30 min and centrifuge at 9200 x *g* for 20 min. The supernatant contains the polysaccharide. The pellet was resuspended in acetic acid 10%, boiled and centrifuged again. The two supernatants were mixed and dialysed for three days against flowing water in dialysis tubing cellulose membranes of a molecular-weight cut-off (MWCO) of 3.5 kDa (Merck-Millipore). The dialysate was centrifuged at 16200 x *g* for 15 min. The polysaccharide was separated from bacterial DNA and proteins using an ion exchange chromatography performed in an ÄKTApurifier 10 FPLC system (Cytiva) with a RESOURCE^{TM} Q 6 mL column. The column was equilibrated with 10 mM ammonium bicarbonate. Polysaccharide eluted in the flow through while proteins and DNA bound to the column. Polysaccharide-containing fractions were pooled and lyophilised.

It was later dissolved in ddH_2O at 1 mg/mL and centrifuged at 7650 x *g* for 20 min at 4 ºC. The supernatant was adjusted to 80% (v/v) ethanol and incubated for 1 h at -40 ºC to precipitate the polysaccharide. Then, it was centrifuged at 11700 x *g* for 30 min at 4 ºC. The pellet was dissolved in 1 mL of ddH_2O. The last two centrifugations and the precipitation were repeated until no more pellet was formed in the first centrifugation. The aqueous fraction was lyophilised and 80 mg of pure polysaccharide were obtained. Polysaccharide purification was done together with Dr Nina K. Bröeker at University of Potsdam.

3.2.3. Protein methods

3.2.3.1. SDS-PAGE

Protein samples were analysed by SDS-PAGE performed according to Laemmli (Laemmli, 1970). The resolving gel had a polyacrylamide concentration of 12-15% (w/v), depending on the expected size of the proteins to be analysed, and 5% (w/v) for the stacking gel. Samples were incubated with Protein loading buffer at 95 ºC for 5 min. In some cases, to assess whether proteins form SDS-stable oligomers, heated and non-heated samples were loaded in denaturing gels. PageRuler™ Unstained Protein Ladder (Thermo Fisher Scientific) was loaded together next to the samples to estimate the molecular weight of the proteins. After electrophoresis, gels were stained with Coomassie brilliant blue staining buffer.

3.2.3.2. Immobilised metal ion affinity chromatography (IMAC)

Supernatants from the centrifugation of the disrupted cells were incubated with a nickel-nitrilotriacetic acid (Ni-NTA) agarose (Jena Bioscience) on ice for 30 min. The mixture was poured into an Econo-Pac® Chromatography Column (Bio-Rad) and eluted by gravity. His-tagged proteins bound to Ni ions while other bacterial proteins pass through the column. Then, the column was washed with two column volumes (CV) of lysis buffer. The target protein was later eluted by passing lysis buffer volumes with increasing concentrations of imidazole (50-1000 mM). The imidazole group is the functional group of the histidine side chain; hence, it binds to nickel ions and displaces the his-tag histidines. Then, eluted fractions were analysed by denaturing gel electrophoresis to see where the target protein eluted. Often, some other proteins and other impurities like nucleic acids elute with the target protein because they non-specifically bind Ni-NTA resin. Therefore, a second step of purification is needed.

3.2.3.3. Dialysis and ion exchange chromatography (IEC)

Fractions containing the target protein were dialysed in dialysis tubing cellulose membranes of 14 kDa MWCO (Merck-Millipore), first, against 20 mM Tris pH 7.5 and 0.2 M NaCl for 3-4 h at 4 ºC and then, against 20 mM Tris pH 7.5 overnight at 4 ºC. Dialysis was carried out in two steps to decrease aggregation by changing buffer more slowly. Prior to IEC, the dialysed sample was centrifuged at 15000 x *g* for 10 min at 4 ºC to remove aggregates. IEC was performed using an ÄKTApurifier 10 FPLC system (Cytiva) with a RESOURCE™ Q 6 mL column. Proteins were eluted with a gradient of 0 to 1 M NaCl buffer with 20 mM Tris pH 7.5. In some cases, since proteins eluted in the first half of the gradient, this part had 10 CV, so the peak of the target protein and its contaminants could be more separated, and the second, 5 CV, as target proteins does not normally elute in this part. Thus, the peaks in the first half were more separated and contaminant proteins were avoided. Fractions with significant absorbance at 280 nm were analysed by denaturing gel electrophoresis.

3.2.3.4. Diafiltration and concentration

Fractions with negligible amounts or without contaminant proteins were selected to be desalted and concentrated with Amicon® Ultra centrifugal filters of 3 kDa or 10 kDa MWCT (Millipore). Concentrated samples were centrifuged at 15000 x *g* for 10 min prior to crystallisation trials; the pellet was discarded.

3.2.3.5. MBTH assay

This assay measures the amount of sugar reducing ends present in a solution. As tailspikes hydrolyse the bond where a reducing end is present, the MBTH assay quantifies the amount of O-antigen repeating units produced by a tailspike from the polysaccharide. The reagent 3-methyl-2benzothiazolinone hydrazone hydrochloride (MBTH) (Sigma Aldrich Scientific) reacts with the reducing sugar ends (RSE) generating a chromogenic compound, which has an absorption peak at 655 nm. The protocol (Anthon & Barrett, 2002; Zhang et al., 2011) was followed with modifications. Briefly, 50 µL of protein (1 µM) and Anatum polysaccharide (1 mg/mL) were incubated for 2 h at 37 ºC. The reaction was stopped with 50 µL of NaOH 0.25 M. Then, 50 µL of a freshly prepared mixture of MBTH (1.5 mg/mL) and DTT (0.5 mg/mL, dithiothreitol) were added to the tube. At neutral pH, the saccharide condenses with a MBTH molecule to form and adduct. Later, the tube was heated at 80 ºC for 15 min. Immediately after the incubation, 100 µL of an oxidising solution (0.5% (w/v) $FeNH_4(SO_4)_2$, 0.5% (w/v) sulfamic acid and 0.5% M HCl) were added. In acidic conditions,

the adduct reacts with another MBTH molecule to form a highly coloured final product. Spectrophometric measurements were performed in a Spectra Max ID3 reader (Molecular Devices). A standard curve was done with glucose at 25 to 1000 mM.

3.2.4. Crystallographic methods

3.2.4.1. Overview of X-ray crystallography

X-ray crystallography is the most-used technique to solve protein structures. It relies on crystals formed by proteins or protein complexes. In order to increase crystallisation probability, protein samples must be highly concentrated (more than 5 mg/mL), pure, monodisperse and conformationally homogenous. Crystallisation is a special type of aggregation. However, crystallisation conditions are different for each protein and are not generally known *a priori*. For this reason, tens, hundreds or, in some cases, thousands of different crystallisation conditions are mixed with concentrated protein, in attempts to obtain crystals.

The structure of a protein with its ligand gives information about the detailed interaction of the protein with its ligand and, in favourable cases, the possible mechanism of action of the protein. To achieve such information, protein and ligand must be crystallised together by soaking or co-crystallisation. Soaking consists in the addition of a ligand to a drop with protein crystals. In co-crystallisation, trials are set up with a solution of protein and ligand. Hence, co-crystallisation requires more amount of ligand. On the other hand, ligand addition in soaking might disrupt the order of the crystal, hindering diffraction. Besides, ligand binding sites may be blocked by a neighbouring protein.

For diffraction experiments, crystals are harvested and flash-cooled in liquid nitrogen to preserve the crystals before and during the diffraction. A gas flow of liquid nitrogen keeps the crystal at about 100 K and protect the crystal from X-ray radiation damage. However, the flash-cooling process may also damage the internal order of the crystal, which would lead to a loss of diffraction quality. Therefore, compounds like glycerol, small polyethylene glycols or lithium salts are used as cryo-protectants. Their presence leads to the formation of vitreous ice – a glass-like, amorphous type of ice – upon flash-freezing, instead of crystalline ice, which would diffract X-rays.

Crystals are formed by millions of translational repetitions of a unit cell. A unit cell can have one asymmetric unit or more. If it has more than one, asymmetric units are related by

crystallographic symmetry operations (rotation axes and/or screw axes in the case of proteins). Protein crystals diffract X-ray beams producing diffraction patterns according to the periodical order of its crystalline matter (protein crystals also contain unordered molecules, like water molecules). Synchrotron radiation facilities produce highly intense, monochromatic X-ray beams with adjustable wavelengths. The length of the sides and the angles of the unit cell can be obtained from the diffraction patterns (spot locations). Diffraction occurs if the wavelength of the radiation is similar to the spacing between the units of a lattice, i.e. the atoms in this case. Hence, the wavelengths used in these experiments are around 1 Å, similar to the bond length between atoms in proteins (2013-Llamas-Sainz).

The position of every atom in the unit cell (real space) can be determined by a vector. The reflections scattered during the diffraction experiment, define a new lattice and a new vector space (reciprocal space). As both spaces are related, the Fourier transform (Fig. 11) can be used to obtain the real space vectors from the diffraction patterns, the records of the reciprocal space. However, while the amplitude (F_{hkl}) of the diffracted waves can be obtained from the intensity of the reflections in diffraction patterns, the phase (α_{hkl}) is missing. Both parameters compose the structure factor of the wave and are necessary in the Fourier transform. In crystallography, this is called the phase problem.

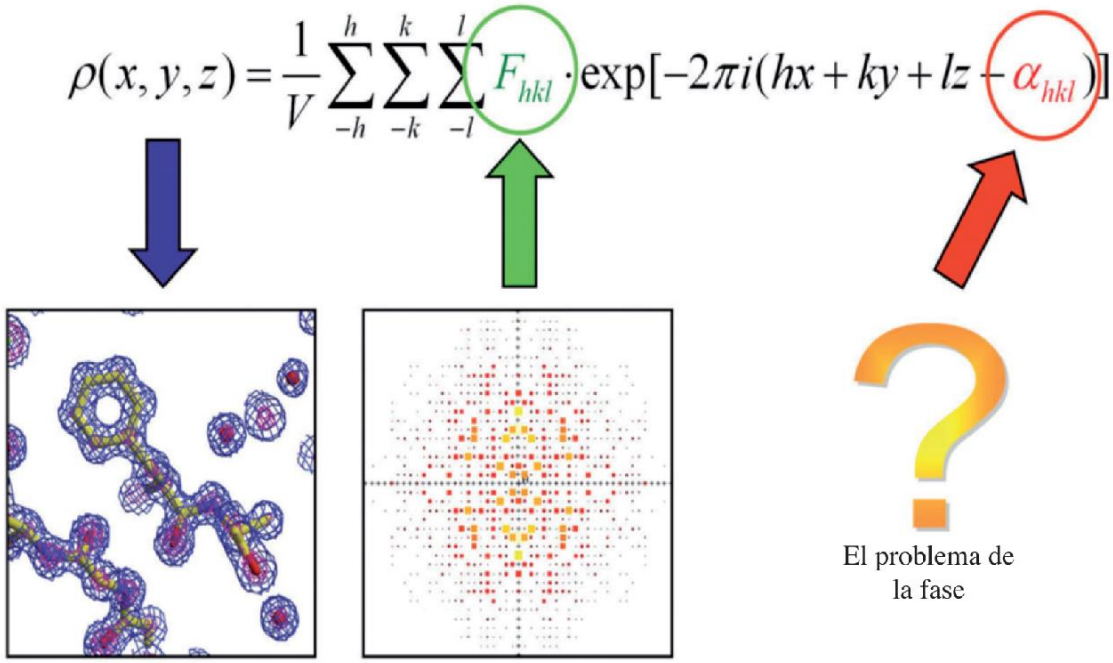


Figure 11: Equation of the Fourier transform. Structure factors, ρ(x, y, z), the real space, are waves that form the electron density. As all waves, they are composed of amplitudes, F_{hkl}, and phases, α_{hkl}.

To obtain the wave phase of diffraction patterns from protein crystals, I used molecular replacement and single-wavelength anomalous diffraction (SAD). In molecular replacement, a structurally similar protein is needed. As the structure of the protein of interest is unknown, protein structures with more than 25% of sequence identity are employed for this task, because they are likely to have a sufficiently similar structure to allow structure solution by molecular replacement. First, the homologous protein structure is rotated in the asymmetric unit to maximise the correlation between the calculated and measured Patterson function (or a similar correlation). Then, for the top rotation solutions, the structure is translated in the asymmetric unit and different positions are scored for optimal Patterson correlation or fit to the diffraction data. The phases of the rotated and translated protein model are then used to determine structure factors for map calculation.

The SAD method relies on "anomalous signal", i.e. significant differences in diffraction intensities between Bijvoet pairs. For this, each asymmetric unit needs to contain a small number of well-ordered atoms that give rise to these differences at a wavelength attainable at the synchrotron beamline. Mercury compounds, which bind covalently to cysteine residues, or selenium atoms incorporated as selenomethionine fulfil these criteria. Given high-quality diffraction datasets with significant anomalous differences and a number of "heavy" atoms that is not too high, the positions of these atoms can be determined. From these positions, initial phases for the electron density map can be calculated.

The intensity of the reflections and the initial phase estimations produce the first electron density map. This allows building if a first, partial, protein model, which improve the initial phases and, subsequently, the electron density map. This electron density map is optimised by several cycles of model building and refinement.

3.2.4.2. Protein crystallisation

Crystallisation trials were performed in MRS crystallisation plates (SWISSCI) using the sitting-drop vapour-diffusion technique. The following commercial screens were used:

- Crystallization kit for Protein-Protein Complexes (Sigma Aldrich)
- Crystallization Cryo kit for Proteins (Sigma Aldrich)
- Crystallization Extension kit for Proteins (Sigma Aldrich)
- JBScreen Classic 1-10 (Jena Bioscience)

- Pi-minimal Screen (Jena Bioscience)
- JBScreen PEG/Salt 1-2 (Jena Bioscience)
- Morpheus I, II and III (Molecular Dimensions)

Reservoirs were filled with 50 µL of crystallisation condition and drops with 1.5 µL (1 µL of protein and 0.5 µL of condition). Due to ligand scarcity, protein drops with crystals were soaked with highly concentrated samples of Anatum O-antigen oligosaccharides (Table 4).

Table 4: Protein and ligand concentrations

Protein-ligand	Initial protein concentration (mg/mL - mM)	Initial ligand concentration (mM)	mM Ratio
Petal-nonasaccharide	16 - 0.40	20	50
Gp20ΔN-hexasaccharide	10 - 0.11	10	92

3.2.4.3. Soaking experiments

In some cases, crystals were soaked with cryo-protectant agents, heavy metals or ligands. Similar procedures were followed for the three materials. Crystals of the native and selenomethionine-derivative gp20(734-1070) crystals were harvested and soaked in their respective crystallisation solution supplemented with increasing concentration of glycerol (5, 10, 15 and 20% (v/v)). Immediately, they were harvested and flash-frozen. Other gp20(734-1070) crystals were also soaked with ddH_2O-dissolved Anatum O-antigen nonasaccharides at 20 mM for 17 h. These crystals were not cryo-protected.

Gp20(248-1070) crystals were soaked with ddH_2O Anatum O-antigen hexasaccharides at 10 mM for 2 min. These crystal were cryo-protected with the addition of its crystallisation condition supplemented with 20% (v/v) glycerol.

Methylmercury chloride powder was added to the reservoirs of RBP3 hits. After an overnight equilibration with the drop, crystals were incubated with an equal volume of methylmercury chloride-saturated crystallisation condition for 5 h. RBP3 crystals were not cryo-protected.

3.2.4.4. Crystal harvesting

Crystals were harvested with MicroMounts (MiTeGen) and Litholoops (Molecular Dimensions). If necessary, 20% (v/v) glycerol was added to the harvesting buffer. Sometimes,

crystals dissolved upon cryo-protectant addition. Hence, the first crystals were harvested without added glycerol to avoid solubilisation of the crystal. Then, 20% (v/v) glycerol was added to the crystal drop to harvest cryo-protected crystals.

3.2.4.5. Data collection, structure determination and analysis

Crystals were diffracted at the XALOC beamline of the ALBA-CELLS synchrotron (Cerdanyola del Vallès, Spain) (Juanhuix et al., 2014). Diffraction image data sets were processed (indexed, integrated, merged and scaled) with autoPROC (Vonrhein et al., 2011) using the XALOC automated pipeline or with CCP4 (Winn et al., 2011), (Table 5). Different programs were used to process the same data. The results with higher quality (CC1/2, resolution, completeness, multiplicity, R-work, R-free, etc.) were used.

Table 5A: Data processing

	Indexing	Integration	Merging/scaling
gp20(734-1070)-SeMet	iMOSFLM		AIMLESS & POINTLESS
gp20(734-1070)	iMOSFLM		
gp20(734-1070)-nonasaccharide	XDS[a]		
gp20(248-1070)	XIA2/DIALS		
gp20(248-1070)-hexasaccharide	XDS[a]		
RBP3 C2-CH_3HgCl	XDS[a]		
RBP3 C2	XIA2/DIALS		
RBP3 $P6_522$	XDS		

[a] Data was automatically processed inside autoPROC, using the programs XDS, AIMLESS and POINTLESS.

Table 5B: Structure solution and refinement

	Automated package	Substructure determination	Hand determination	Model building
gp20(734-1070)-SeMet	CRANK2	SHELXC/D	Solomon & MULTICOMB	Parrot, REFMAC5 & Buccaneer
RBP3 C2-CH_3HgCl				

Table 5C: Molecular replacement

	gp20(734-1070)	gp20(734-1070) 3 RU	gp20(248-1070)	gp20(248-1070) 2 RU	RBP3 C2	RBP3 $P6_522$
Molecular replacement	MOLREP	PHASER	PHASER	MOLREP	MOLREP	PHASER

References of the used programs: iMOSFLM (Battye et al., 2011), AIMLESS (P. R. Evans & Murshudov, 2013), POINTLESS (P. Evans, 2006), MOLREP (Vagin & Teplyakov, 2010), PHASER (McCoy et al., 2007), XDS (Kabsch, 2010), xia2 (Winter, 2010), DIALS (Winter et al., 2018), CRANK2 (Skubák & Pannu, 2013), SHELX (Sheldrick, 2010), Solomon (Abrahams & Leslie, 1996), MULTICOMB (Skubák et al., 2010), Parrot (Kevin Cowtan, 2010), REFMAC5 (Murshudov et al., 2011) and Bucaneer (K. Cowtan, 2006).

Protein models were further built and refined with Buccaneer, REFMAC5 and Coot (Emsley et al., 2010). Models were validated with MolProbity (Williams et al., 2018), either using their web server (http://molprobity.biochem.duke.edu/) or integrated in REFMAC5.

PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC) and UCSF Chimera (Pettersen et al., 2004) were used for analysis and visualisation of the models. The representations of the surface electrostatic potential were created with the PyMOL plugin APBS Electrostatics.

3.2.5. Electron microscopy of gp20(2-1070)

3.2.5.1. Negative staining sample preparation for electron microscopy

Gp20(2-1070) was analysed by transmission electron microscopy using the negative staining technique. A 5-µL drop at 0.01 mg/mL was adsorbed for 5 min onto a glow-discharged carbon/collodion coated copper grids (Gilder Grids). The excess liquid was removed by quick blotting with Whatman paper (Whatman). Then, the grid was washed with 2% (w/v) uranyl acetate and blotted again. This step was repeated once. Later, the grid was incubated with 2% (w/v) uranyl acetate for 2 min. Eventually, the grid was blotted and air-dried on a filter paper.

3.2.5.2. Electron microscopy data acquisition

The grid was initially checked in a 100 kV JEOL JEM 1011 transmission electron microscope (JEOL) with a Gatan ES1000Ww camera (Gatan). Later, images were acquired with a FEI Tecnai FEG200 electron microscope (FEI) operated at 200 kV using a 4K x 4K Eagle CCD camera (FEI) at a 62000x magnification and a defocus range of 0.25-4 µm. The camera has a pixel size of 15 µm, giving a nominal sampling rate of pixel rate of 3.614 Å/pixel.

3.2.5.3. Image processing

Images were processed using the Scipion Software Framework (de la Rosa-Trevín et al., 2016). The contrast transfer function was determined by CTFfind4 (Rohou & Grigorieff, 2015) for 441 images. Then, 29777 particles were manually picked with Xmipp3 – manual-picking software (Abrishami et al., 2013; de la Rosa-Trevín et al., 2013; Sorzano et al., 2013) in 110 x 110-pixel boxes. Bad particles were removed after being classified with the Xmipp3 – cl2d software (de la Rosa-Trevín et al., 2013; Sorzano et al., 2010, 2013). Then, the best averages were classified with the Relion – 2D classification software (Scheres, 2012; Zivanov et al., 2018) in one class and aligned with Xmipp3 – apply alignment 2d (de la Rosa-Trevín et al., 2013; Sorzano et al., 2013).

4. Results

4.1. The gp20 tailspike of *Salmonella virus* Epsilon15

4.1.1. Expression and purification of gp20 constructs

Based on primary sequence and secondary structure prediction, eleven gp20 constructs were designed to be cloned in pET28c(+). The beginning and the end of the constructs were aimed to be set at disordered regions between domains, although before solving the structure, the exact domain boundaries are of course unknown. This would increase the possibility of producing soluble protein. C-terminal affinity tags were also avoided. The C-terminal end has a key folding role in most trimeric fibre and tailspike proteins, such as demonstrated for T4 fibritin, P22 tailspike and adenovirus type 2 fibre (Gage & Robinson, 2003; Mitraki et al., 1999; Tao et al., 1997). Besides, tail fibres and tailspikes often bind to their bacterial receptor through their C-terminal domain. Therefore, a foreign sequence could hinder receptor-binding. Instead, an N-terminal affinity tag containing six consecutive histidine residues was included to allow first-step purification by immobilised metal ion affinity chromatography.

Only some of the cloned constructs yielded soluble protein (Table 6). These were the full-length protein (2-1070) and the constructs gp20(248-1070) and gp20(734-1070). As most of the primers were designed with no prior structural knowledge, for many of the other constructs, the start point may fall inside folded gp20 domains.

Table 6: Cloned constructs solubility.

Constructs of gp20	Solubility and yield (mg/L of culture)	Crystal/solved structure
2-1070	2.5	Cryo-EM map (Murata et al., 2010)
200-1070	No expression	No
287-1070	Insoluble	No
434-1070	Insoluble	No
520-1070	Low expression	No
734-1070	4.5	Yes (778-1070)
812-1070	No expression	No
5-768	Insoluble	No
248-768	Insoluble	No
264-768	Insoluble	No
248-1070	8.5	Yes (248-1070)*

Constructs designed based on secondary structure prediction are shadowed orange and constructs designed based on existing structures which appeared during my thesis are shadowed green.
*Part of the affinity tag are also resolved in the crystal structure.

Proteins were purified following the procedures explained in the Materials and Methods section. In the first step, immobilised metal affinity chromatography removed most of the

soluble contaminants. Then, the purest fractions were dialysed against 20 mM Tris pH 7.5 to remove glycerol, imidazole and sodium chloride. Anionic exchange chromatography was performed in Tris-HCl buffer at pH 7.5. Usually, only one peak was observed. The protein eluted at a concentration of about 0.25 M sodium chloride for gp20(2-1070) construct and 0.19 M sodium chloride for gp20(248-1070) and for gp20(734-1070) (Fig 12, 13 & 14). The fractions corresponding to the peak were desalted and concentrated to about 10 mg/mL. The three constructs gave enough protein per litre of culture to carry out crystallisation trials. The gp20(248-1070) construct yielded about two and four times more protein than gp20(734-1070) and gp20(2-1070), respectively (Table 6).

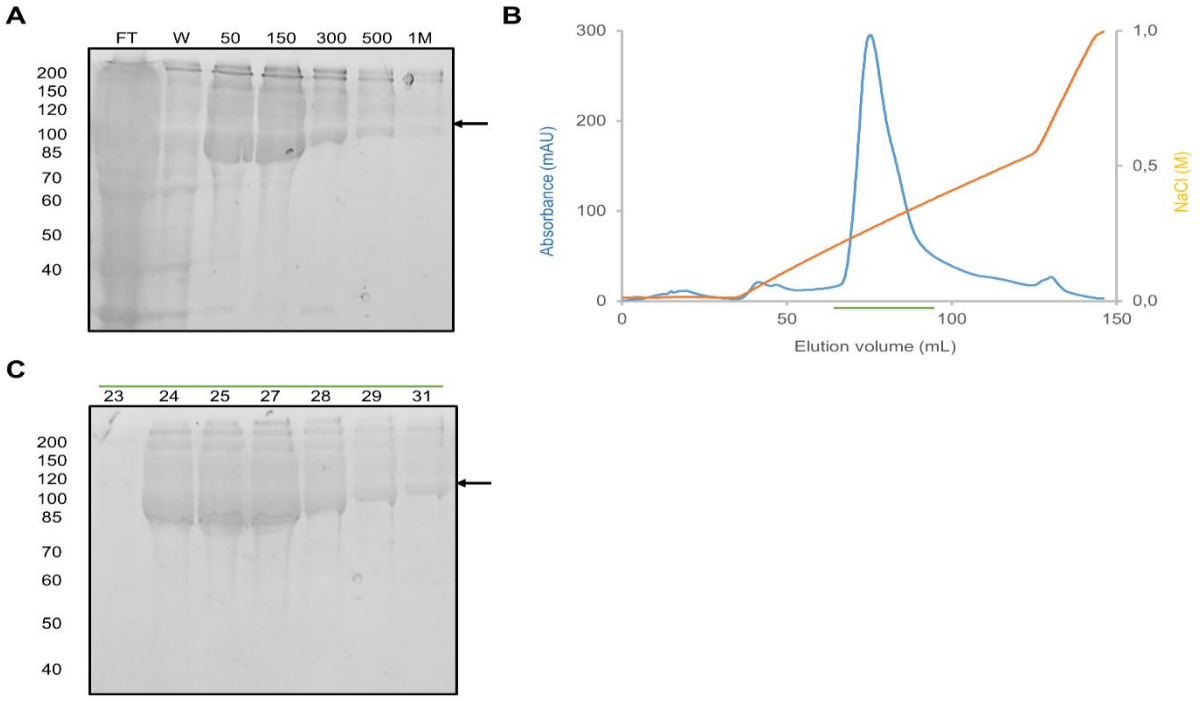


Figure 12: Gp20(2-1070) purification. (A) Denaturing gel electrophoresis of samples from the IMAC (FT, flow through; W, wash; imidazole concentration in mM). (B) Anion exchange chromatography profile. Fractions analysed by denaturing gel electrophoresis (C) are indicated with a green bar (Numbers above the gel indicate the fractions from the chromatography). Black arrows in (A) and (C) indicate the expected molecular weight of gp20(2-1070), 119 kDa.

In denaturing gels, constructs gp20(2-1070) and gp20(248-1070) produced a smear above the molecular weight of its monomer. However, the smear was not present for gp20(734-1070) gels. The constructs gp20(2-1070) and gp20(248-1070) also show a band of trimeric size when samples are loaded without being heated at 95 ºC before electrophoresis. This behaviour is common among tail fibres and tailspikes (Barbirz et al., 2009). However, unheated gp20(734-1070) samples remain as monomers in denaturing gels (data not shown). In conclusion, longer constructs show an electrophoretic behaviour consistent with trimeric structures, while gp20(734-1070) only shows monomeric bands.

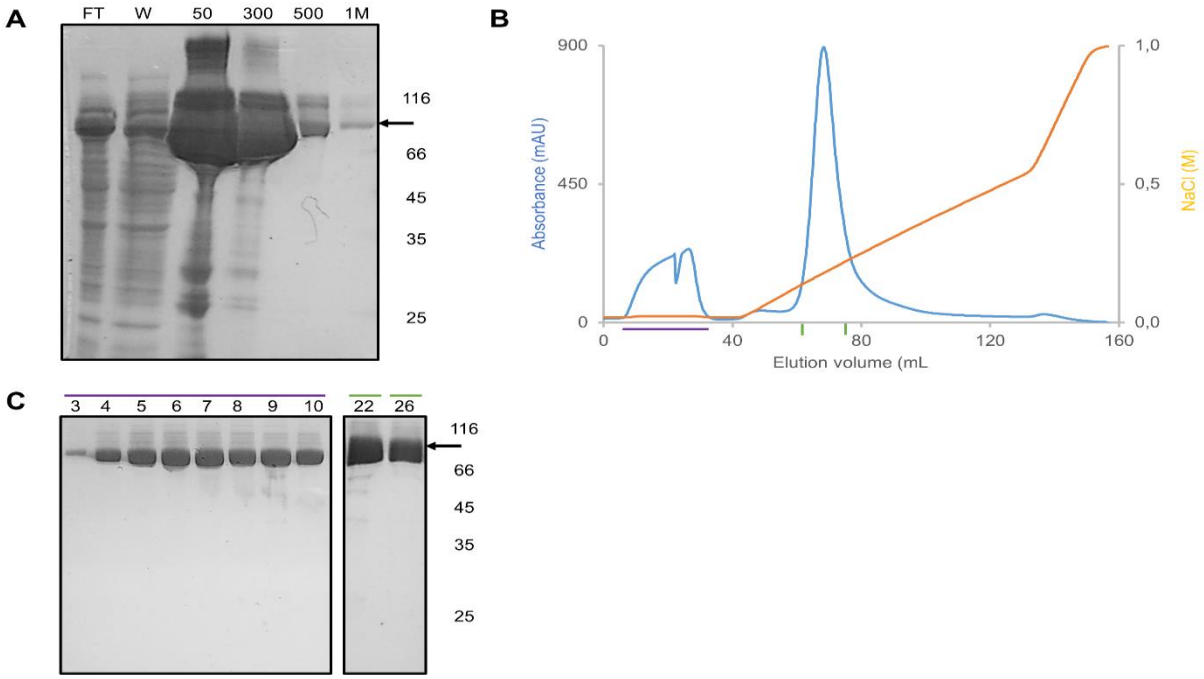


Figure 13: Gp20(248-1070) purification. (A) Denaturing gel electrophoresis of samples from IMAC (FT, flow through; W, wash; imidazole concentration in mM). (B) Anion exchange chromatography profile. Fractions analysed by denaturing gel electrophoresis (C) are indicated with the purple and green bars (Numbers above the gel indicate the fractions from the chromatography). Black arrows in (A) and (C) indicate the expected molecular weight of gp20(248-1070), 92 kDa.

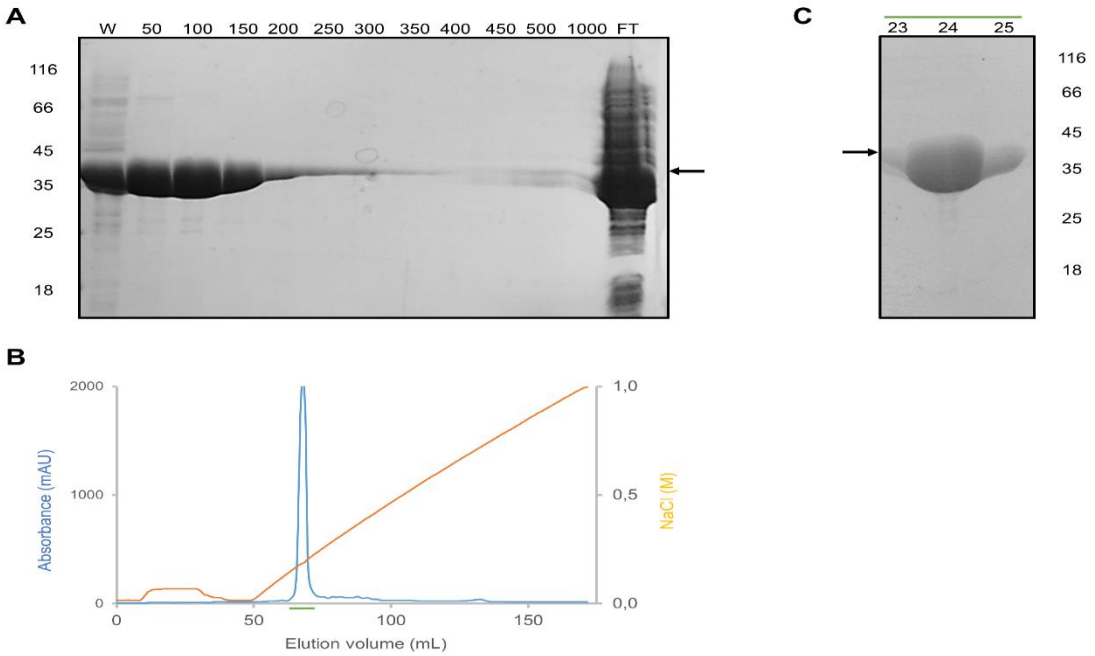


Figure 14: Gp20 (734-1070) purification. (A) Denaturing gel electrophoresis of samples from the IMAC (FT, flow through; W, wash; imidazole concentration in mM). (B) Anion exchange chromatography profile. Fractions analysed by SDS-PAGE (C) are indicated with green bars (Numbers above the gel indicate the fractions from the chromatography). Black arrows in (A) and (C) indicates the expected molecular weight of gp20(734-1070), 40 kDa.

4.1.2. Crystallisation and structure determination of gp20(734-1070) and gp20(248-1070)

Crystallisation trials performed with constructs gp20(734-1070) and gp20(248-1070) produced enough crystals to resolve their atomic structures without and with ligands. The best gp20(734-1070) crystals grew in a solution consisting of 0.1 M Tris pH 8.0, 12% (w/v) PEG 8000. They grew as clusters of plates that were separated to harvest single crystals (Fig 15A). The best tested native crystal diffracted to 1.17 Å resolution (Fig. 15B) and belonged to space group C2 with cell parameters of 154.2, 41.4 and 49.4 Å for the cell edges a, b and c, respectively, and a beta angle of 102º (Table 7).

As no homologous protein was found in the PDB, experimental phasing was necessary to solve its structure. Crystals grown in the same crystallisation plate were soaked with methylmercury chloride to collect SAD data (Fig. 15E). At the same time, a batch of selenomethionine-derivatised gp20(734-1070) was purified and crystallised. The best selenomethionine crystal grew in 0.1 M sodium citrate pH 5.5, 12% (w/v) PEG 4000 (Fig. 15C). Good data was collected up to 2.5 Å resolution (Fig. 15D), with significant anomalous signal (Table 7A). Its space group was $P2_12_12_1$ with cell edges of 41.7, 49.7 and 144.7 Å. In methylmercury chloride-soaked crystals, the space group and lattice were similar to native crystal, also with sufficient anomalous signal to allow structure solution (not shown). As the quality of data from the selenomethionine dataset was higher, only its structure solution was reported.

Crystals obtained in 0.1 M KCl, 0.1 M HEPES pH 7.0, 12% (w/v) PEG 8000 were soaked with a nonasaccharide of three repeating units (RU) of the Anatum O-antigen. A crystal harvested after seventeen hours of incubation gave a twinned diffraction pattern (Fig. 15F). It diffracted at 2.1 Å and belonged to the same space group as the selenomethionine derivative. A difference electron density map showed ring-like densities corresponding to carbohydrate molecules. A pentasaccharide corresponding to a two-repeat unit saccharide lacking the reducing-end rhamnose was modelled into these densities,

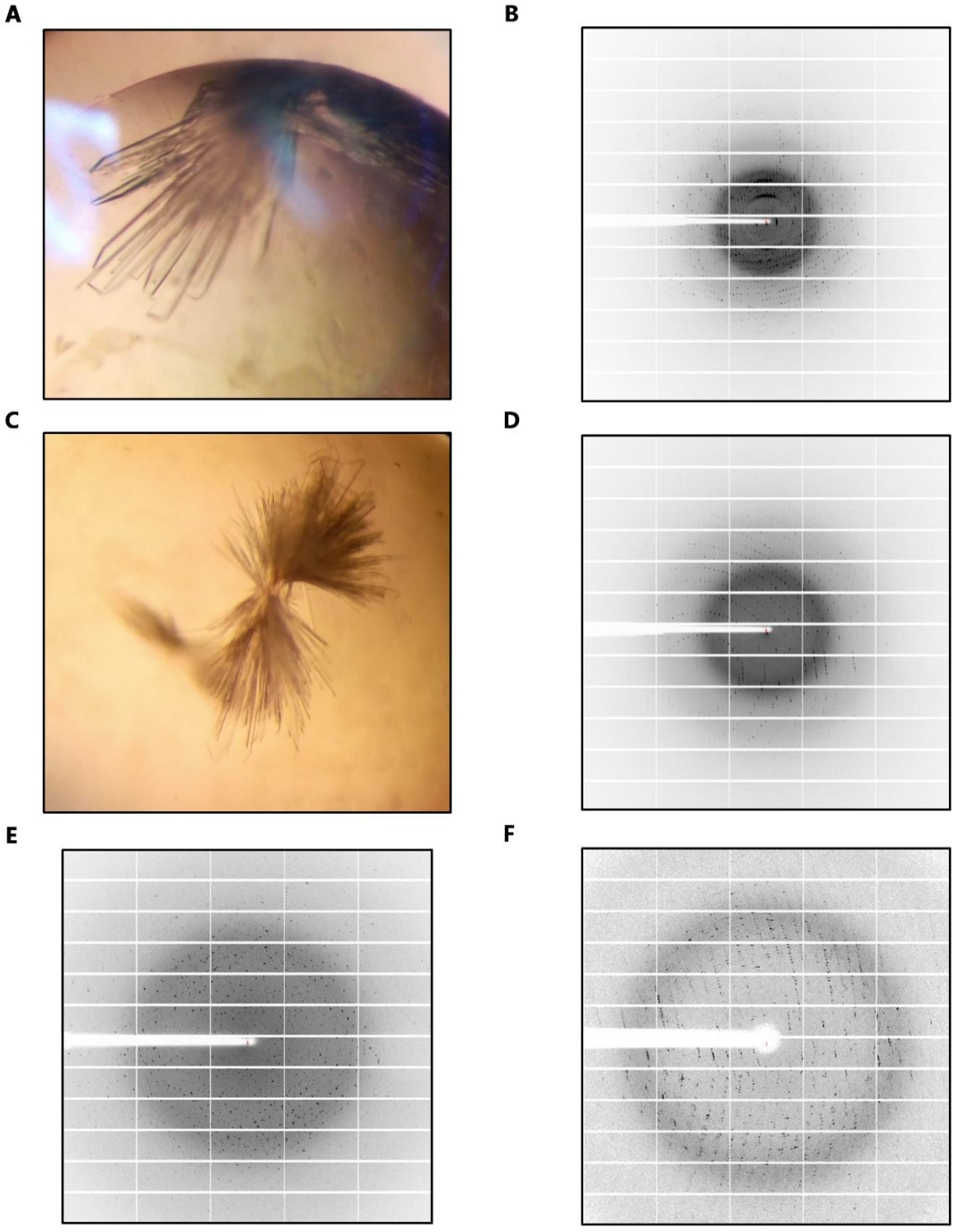


Figure 15: Crystals and diffraction patterns of gp20(734-1070). Cluster of native plate-shaped (A) and of selenomethionine-derivatised crystals (C). Diffraction patterns of gp20(734-1070) (B), selenomethionine-derivatised gp20(734-1070) (D), methylmercury-derivatised gp20(734-1070) (E) and gp20(734-1070) soaked with 3 RU of the Anatum O-antigen (F).

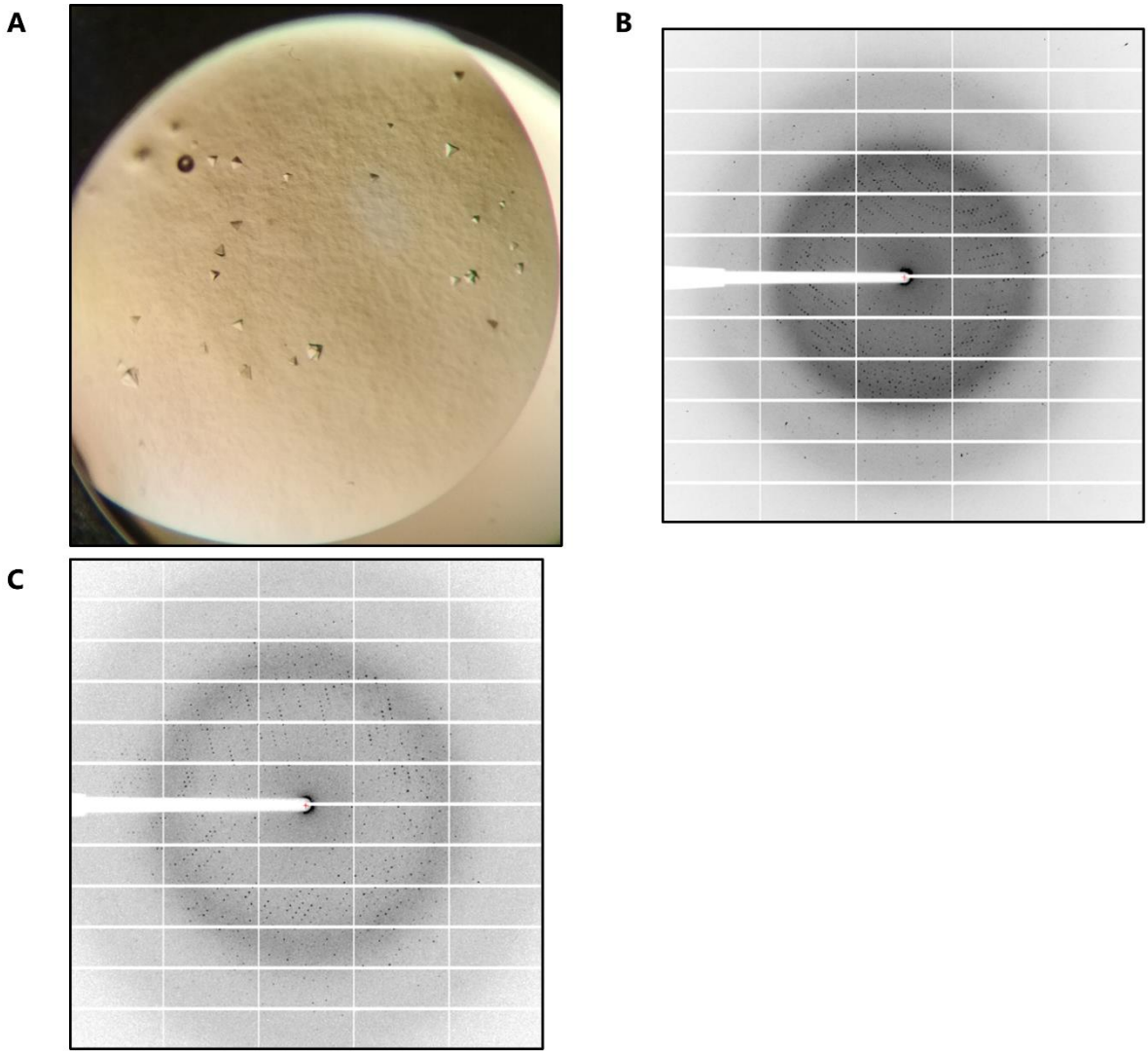


Figure 16: Crystals and diffraction patterns of gp20(248-1070) (A) Tetrahedron-shaped crystals. Diffraction patterns of gp20 (248-1070) (B) and gp20 (248-1070) soaked with 2 RU of the Anatum O-antigen (C).

Tetrahedron-shaped gp20(248-1070) crystals grew in 3 M sodium formate and similar crystallisation conditions after more than four months (Fig. 16A). The best crystal yielded data to 1.8 Å resolution (Fig. 16B) and belonged to space group R3. In the hexagonal setting (H3), cell parameters were 117.4, 117.4 and 251.2 Å (Table 7). The structure was solved by molecular replacement with the gp20(734-1070) structure and the *Azotobacter vinelandii* Mannuronan C-5 epimerase AlgE6 A-module (PDB entry 5LW3). This was the most structurally homologous protein found when the dataset was obtained. Crystals were also soaked with a hexasaccharide (two repeating units) of the Anatum O-antigen. The best of these crystals had the same shape and diffracted giving a dataset with similar statistics and quality to the apoprotein dataset (Fig. 16C, Table 7).

Table 7A: Crystallographic statistics of gp20(734-1070) datasets

	gp20(734-1070) SeMet	gp20(734-1070)	gp20(734-1070) 3 RU
Data collection			
Radiation source	BL13-XALOC (ALBA-CELLS)	BL13-XALOC (ALBA-CELLS)	BL13-XALOC (ALBA-CELLS)
Wavelength (Å)	0.97854	0.97854	0.97980
Detector	PILATUS 6M - DECTRIS	PILATUS 6M – DECTRIS	PILATUS 6M - DECTRIS
Crystal-to-detector distance (mm)	242.5	188.5	502.3
Oscillation per image (º)	0.2	0.25	0.1
Number of images	1800	524	3600
Data processing			
Space group	P 21 21 21	C2	P 21 21 21
Cell edges (a, b, c; Å)	41.7, 49.7, 144.7	154.2, 41.4, 49.4	41.5, 49.7, 144.4
Cell angles (α, β, γ, º)	90.0, 90.0, 90.0	90.0, 102.0, 90.0	90.0, 90.0, 90.0
Resolution range (Å)	144.69–1.64 (1.67-1.64)	48.31-1.17 (1.19-1.17)	144.42-2.08 (2.14-2.08)
Total number of reflections	717685 (26902)	218679 (9097)	185308 (5215)
Number of unique reflections	64570 (2896)	94416 (4263)	17754 (1043)
Completeness (%)	100.0 (100.0)	92.7 (85.8)	95.6 (75.7)
Anomalous completeness (%)	99.9 (100)	-	-
Multiplicity	1.9 (1.9)	6.0 (1.0)	10.4 (5.0)
Anomalous multiplicity	6.4 (6.0)	-	-
CC 1/2[a]	0.998 (0.960)	0.997 (0.450)	0.993 (0.554)
Resolution (Å) at $CC_{anomfit}$ = 0.15	2.53	-	-
Rmeas (all I+ & I-)[b]	0.055 (0.480)	0.082 (0.974)	0.186 (0.917)
$\langle I/\sigma(I) \rangle$	10.4 (2.2)	6.0 (1.0)	10.0 (1.6)
Phasing			
Heavy atoms sites	5 Se	-	-
Correlation coeff. (all/weak)[c]	48.2/30.9	-	-
Combined DM FOM + phasing CLD score (chosen/rejected hand)[d]	19.5/0.0	-	-
Refinement			
Resolution range (Å)	-	48.29-1.30	72.31-2.08
Reflections used	-	75179	17700
Reflections used for R-free	-	3741	878

R-factor/R-free[e]	-	0.13/0.16	0.18/0.24
Model statistics			
Amino acid coverage	-	778-1069	777-1070
Atoms (protein/ions/ligands/water)	-	2278/0/29/271	2251/0/61/175
Ramachandran[f] (%favoured/allowed/outliers)	-	96.55/2.64/0.81	95.89/3.08/1.03
RMSD (bonds, Å / angles, º)[g]	-	0.028/2.3	0.008/1.543
B-factor (protein/ions/ligands/water)[g]	-	13.52/-/ 29.84/26.48	25.3/-/46.44/29.24
MolProbity score/percentile[h]	-	1.33/90th	1.24/100th
PDB code	-		

Table notes are below Table 7B.

Table 7B: Crystallographic statistics of gp20(248-1070) datasets

	gp20 (248-1070)	gp20 (248-1070) 2 RU
Data collection		
Radiation source	BL13-XALOC (ALBA-CELLS)	BL13-XALOC (ALBA-CELLS)
Wavelength (Å)	0.97926	0.97926
Detector	PILATUS 6M - DECTRIS	PILATUS 6M - DECTRIS
Crystal-to-detector distance (mm)	355.3	394.1
Oscillation per image (º)	0.2	0.1
Number of images	900	1220
Data processing		
Space group	R3	R3
Cell edges (a, b, c; Å)	107.7, 107.7, 107.7	107.6, 107.6, 107.6
Cell angles (α = , β, γ, º)	66.0, 66.0, 66.0	66.2, 66.2, 66.2
Resolution range (Å)	79.01-1.81 (1.84-1.81)	48.07-1.88 (1.91-1.88)
Nº of reflections (Total)	595315 (44330)	195550 (7830)
Nº of reflections (Unique)	117496 (8782)	103257 (4458)
Completeness (%)	99.9 (100.0)	98.3 (86.4)
Multiplicity	5.1 (5.0)	1.9 (1.8)
CC 1/2[a]	0.994 (0.356)	0.999 (0.311)
Rmeas[b] (all I+ and I-)	0.178 (2.426)	0.041 (0.990)
<I/σ(I)>	6.9 (1.2)	7.5 (0.5)
Refinement		
Resolution range (Å)	58.75-1.81	48.07-1.88
Reflections used	117442	103257
Reflections used for R-free	5877	5227
R-factor/R-free[e]	0.16/0.18	0.16/0.19
Model statistics		
Amino acid coverage	239-1070	239-1070
Atoms (protein/ions/ligands/water)	6387/1/30/640	6357/1/267/495

Ramachandran (%favoured/allowed/outliers)[f]	95.90/4.10/0.0	95.54/4.46/0.00
RMSD (bonds, Å / angles, º)[g]	0.007/1.42	0.008/1.48
B-factor (protein/ions/ ligands/water)[g]	25.28/18.16/39.74/33.19	39.44/31.78/57.31/42.70
Molprobity score/percentile[h]	1.23/99th	1.30/99th
PDB code		

*The values in parentheses correspond to the high-resolution shell

[a] CC1/2 correlation coefficient between intensity estimates from half data sets (Karplus & Diederichs, 2015).

[b] $Rmeas=(\sum_{hkl}[n/n-1)]^{1/2}\sum_i|I_{i\,hkl}\sum_i I_i(hkl)$

[c] Found by SHELXD (Sheldrick, 2010)

[d] According to Solomon (Abrahams and Leslie, 1996) and Multicomb (Skubák et al., 2010)

[e] $R = (\sum hkl\ |Fobs(hkl) - Fcalc(hkl)| / \sum hkl\ |Fobs\ (hkl)|$

[f] Calculated by MolProbity (Williams et al., 2018)

[g] Calculated by Refmac5 (Evans and Murshudov, 2013)

[h] MolProbity score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalised to be on the same scale as X-ray resolution. 100th percentile is the best among the structures of comparable resolution; 0th percentile is the worst. The comparative set of structures for MolProbity score was selected in 2006 (http://molprobity.biochem.duke.edu/; Williams et al., 2018)

4.1.3. Structure of gp20(734-1070) without and with its receptor

In the asymmetric unit, one monomer containing amino acids 778-1069 is resolved. The N-terminal affinity tag, residues 734-777 and 1070 are disordered. The resolved amino acids form a globular protein composed of two subdomains (Fig. 17A). The largest subdomain consists of seven α-helices and two 3_{10}-helices surrounding a parallel four-stranded β-sheet (residues 778-825 and 907-1070). The smallest domain is a six-stranded antiparallel β-barrel spanning residues 826-906. The loops between strands 4 and 5 and the strands 6 and 7 are extensive (Fig. 17B). Together, they form a groove with the large α/β subdomain.

The largest packing interface area between two molecules in the crystal measures 0.5 x 10^3 Å^2, two orders of magnitude less than the total accessible surface area, 11.9 x 10^3 Å^2. The small interface area is in line with the absence of trimeric proteins in non-heated samples loaded in denaturing gels. This suggests that this part of gp20 does not take part in the intra-trimer contacts.

Residues 778-1069 fit into domains of the tailspike in a cryo-EM map of epsilon15 (EMDB entry: EMD-5209). Monomers adjust well to each of the three lateral blobs, called petals (Jiang et al., 2006) (Fig 18A-B). In the best fitting position, petal domain grooves face outwards and the N- and C-termini inwards. This position makes biological sense, as the petal N-terminal end could join to the rest of the gp20 chain through the inner part of the petal. In addition, since many binding or hydrolytic activities take place in open clefts (Sanz-Gaitero and van Raaij, 2020; Seoane-Blanco et al., 2021), the petal groove could only carry out a role

in the protein function facing outwards. Hereinafter, the gp20(778-1069) structure will be referred to as the "petal domain".

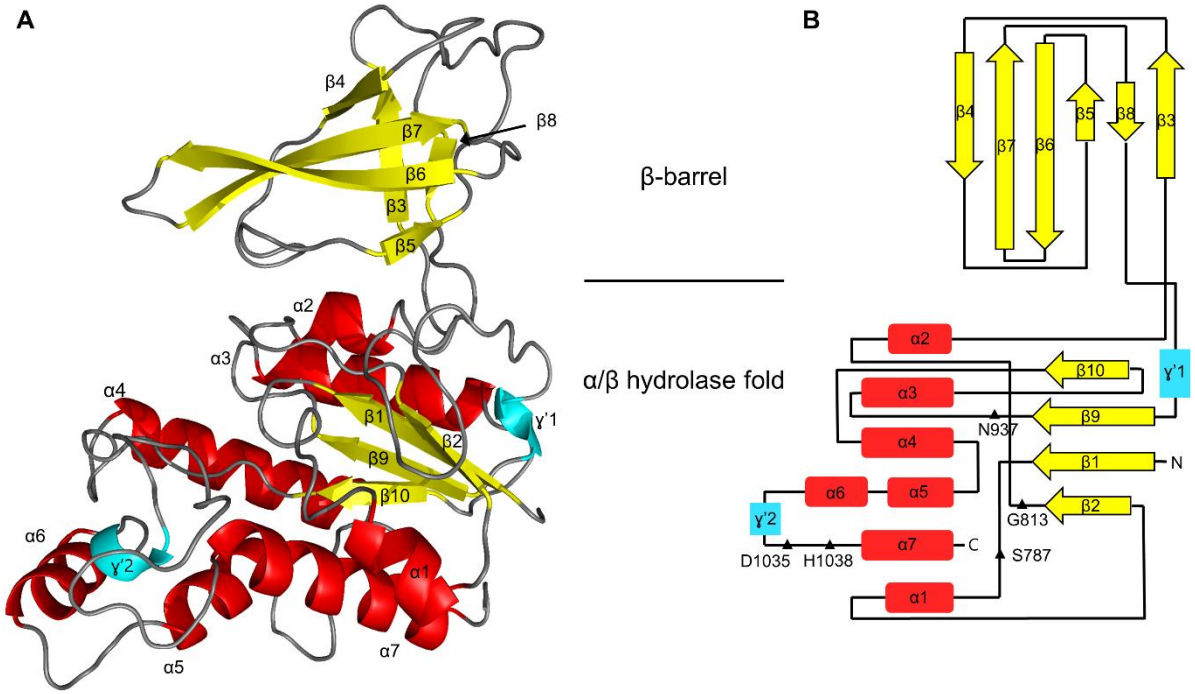


Figure 17: Gp20 petal structure. Ribbon representation (A) and topology (B) of the gp20 C-terminal petal unit with α-helices coloured in red (α), β-strands in yellow (β), 3_{10}-helices in cyan (ɣ' is the Greek numeral for our Arabic numeral 3) and loops in grey. This fragment of gp20 is divided in two subdomains, the β-barrel and the α/β hydrolase fold. Together, they form a structure of about 52 x 38 x 28 Å (height x width x depth). In (B), residues responsible for the esterase activity and the oxyanion hole are represented as black triangles in their approximate position.

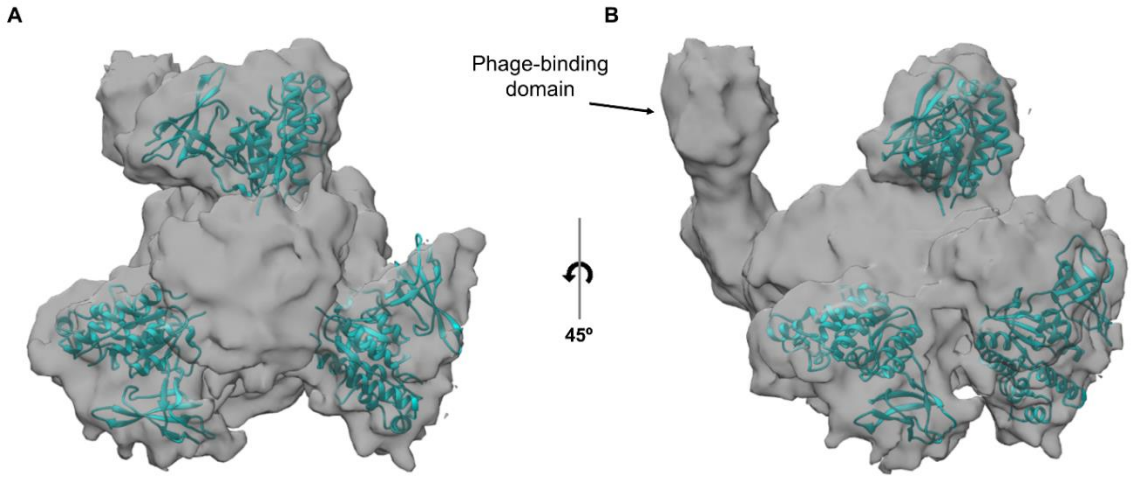


Figure 18: Gp20 petal domain crystal structure fitting in gp20 cryo-EM map. Bottom view (A) and lateral view (B) of a tailspike from an epsilon15 cryo-EM map (13 Å-resolution; EMDB entry: EMD-5209)) with three petal domain crystal structures.

A structural homology search found no homologues for the full petal domain. However, the α/β subdomain is structurally homologous to members of the SGNH-hydrolase superfamily. Despite the low sequence identity (10-20%), the RMSD values among the first

fifty hits are below 3 Å (PDB code 4JGG: 2.4 Å, 3MIL: 2.4 Å or 6WN9: 2.4 Å) (Fig. 19B). Members of this family have a five-stranded β-sheet sandwiched by two layers of α-helices. In the petal domain, the last β-strand is missing. Instead, there is a long loop between α4 and α5 covering the β-sheet in the β10 end. Another loop between β8 and γ'1 covers the other end of the β-sheet.

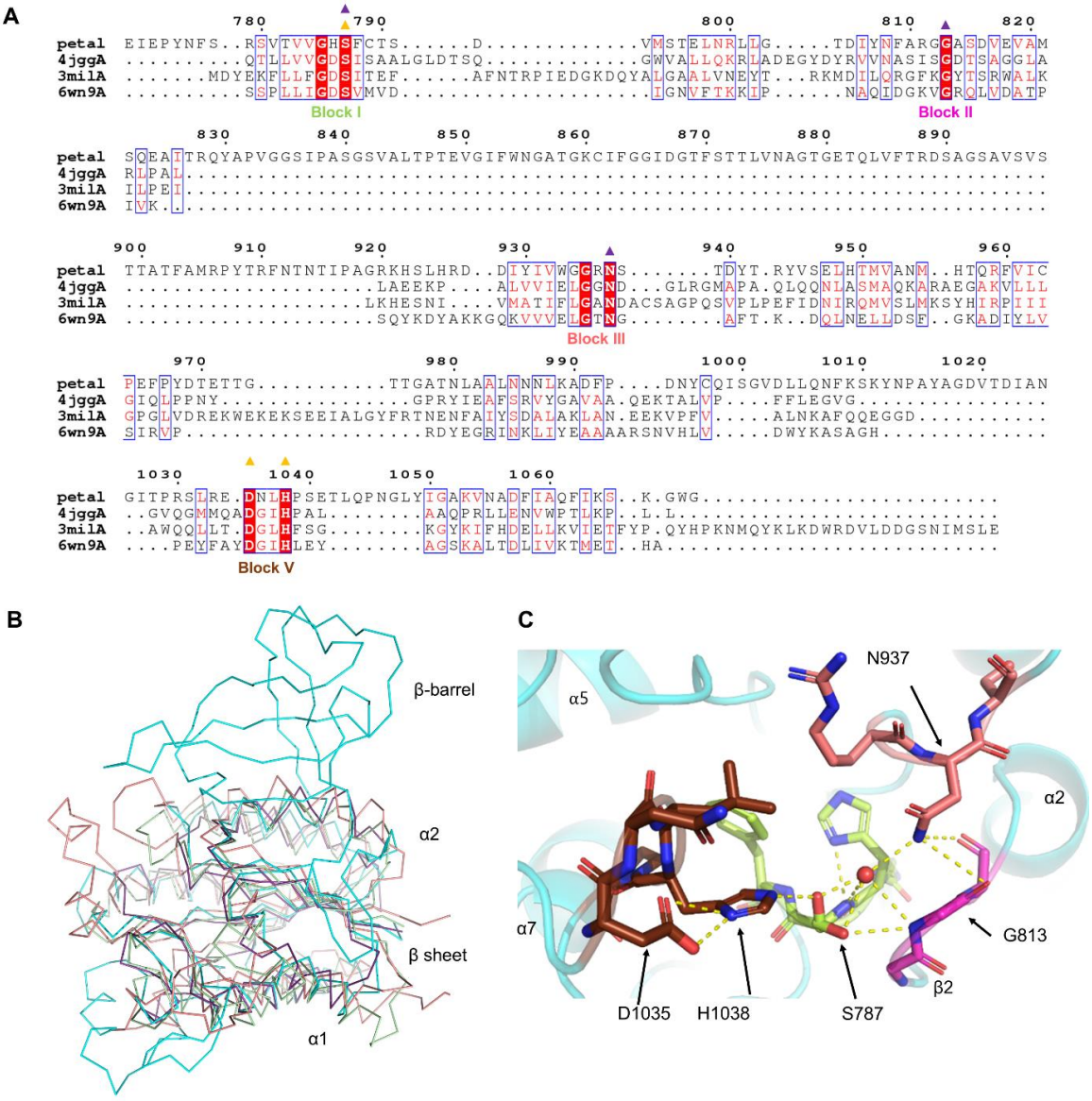


Figure 19: Structural homologues and putative active site of the gp20 petal domain. (A) Structure-based sequence alignment of the petal domain, TesA (PDB entry 4JGG), IAH1 (PDB entry 3MIL) and OatA (PDB entry 6WN9). Structural alignment was performed in the DALI server (Holm, 2020) and represented by ESPript - http://espript.ibcp.fr (Robert & Gouet, 2014). Gly785 and Ser787 are fully conserved in Block I; in Block II, only Gly813; in Block III, Gly935 and Asn937; in Block V, Asp1035 and His1038. Residues putatively involved in the oxyanion hole have a purple triangle above them and those belonging to the putative catalytic triad another in orange. (B) Ribbon view of the structural alignment shown in (A). (C) Close-up view of the putative catalytic centre. Block I (pale green) contains Ser787, Block II (magenta) contains Gly813, Block III (salmon) contains Asn937 and Block V (brown) contains Asp1035 and His1038. Hydrogen bonds around among members of the catalytic triad and the oxyanion hole are shown as yellow dashed lines.

The main characteristic of this superfamily is the presence of four conserved residues (Ser, Gly, Asn and His) in four conserved blocks I, II, III and V, respectively (Akoh et al., 2004) (Fig. 19A). Block IV was firstly proposed when this superfamily was described (Upton et al., 1995; Dalrymple et al. 1997), but later studies discarded it since they no longer found convincing homologies in this region when more protein sequences were included (Lo et al., 2003). In the petal domain, conserved residues Ser787, Asp1035 and His1038 form the putative catalytic triad. These three amino acids are bound through hydrogen bonds. Two among Asp1035 carboxylic oxygens and His1038 Nδ1 and another hydrogen bond between His1038 Nδ2 and Ser787 hydroxyl oxygen. The Ser787 residue has a double conformation in the crystal structure, so it also interacts with the Gly813 amide (Fig. 19C).

The SGNH-superfamily was previously called "GDSL family" after the motif where the serine residue was found. However, the leucine residue is often substituted by other hydrophobic residue such as isoleucine or valine. Besides, aspartic acid can also be substituted by asparagine (PDB entry: 7BR2). The petal domain has the motif GHSF. Phe879 is an aromatic residue larger than leucine, though it is also found in other SGNH-superfamily members. However, no histidine residue has been found at this position in another member of this superfamily. Fig. 19C shows an hydrogen bond between the Nδ1 of the imidazole group of His1038 and the Ser787 amide. This amide would form the putative oxyanion hole together with the Gly813 amide and the Asn937 side chain Nδ2. In the crystal structure, these three nitrogen atoms and the Oγ of Ser787 have hydrogen bonds with a water molecule (Fig. 19B). This water molecule is also present in other proteins (such as the IAH1 protein), together with a glycerol molecule. This suggest it may be the position where the substrate is cleaved.

Regarding the β-barrel subdomain, the structural homology search showed structural homologies with two type of proteins. The best hit (Z-score = 6.4) was a Macrolide-efflux protein (PDB entry 3OP1). In these secretion systems, the β-barrel domain would be in charge of the interaction with another proteins of the secretion system (Yu 2009). The β-barrel subdomain is also homologous to riboflavin binding domains in bacterial flavin adenine dinucleotide synthetase proteins, such as PDB entries 1S4M and 5A88 (Z-score: 6.2 and 6.1 respectively). All hits found have identity percentages of 10 % or less, but the RMSD values are still below 3 Å.

The crystal structure of gp20(734-1070) soaked with the *Salmonella enterica* serovar Anatum nonasaccharide shows the petal domain from residue Phe777 to Gly1070 and five

consecutive carbohydrate residues in the open cleft between the β-barrel and the SGNH subdomain. The mannose in the position 5 only establishes one water-mediated hydrogen bond with the protein. Considering its position, the next saccharide (rhamnose) would not be close to any protein residue (Fig.20A). The lack of links would make the rest of the nonasaccharide too flexible to be resolved in the crystal structure. The acetyl groups in the galactoses are also partially unresolved. The Coot electron density map contoured at 1.0σ covers the ester oxygen and the acetyl central carbon of Gal1 (Fig. 20B) and the all acetyl atoms but the carbonyl oxygen of Gal4 (Fig 20C). However, the PyMOL electron density map contoured at 1.0σ covers the ester oxygen of Gal1 and all the acetyl atoms of Gal4 (Fig. 20D). When the map was contoured at 0.5σ in PyMOL, the electron density map reaches the acetyl central carbon and the carbonyl oxygen of Gal1 and all the acetyl atoms of Gal4 (Fig. 20E). This differences are probably caused by different algorithms to calculate 2Fo-Fc electron density maps. In addition, the lack of density in the acetyl groups might be caused by the flexibility or the lack of acetyl groups in part of the nonasaccharides forming the crystals. The electron density of the acetyl groups shows the sum of the acetyl group signals from each asymmetric unit. However, if the signal is missing in part of asymmetric units due to the lack of acetyl group, the final electron density will be incomplete. In this case, the lack of the acetyl group may be caused by its degradation – the acetyl group is very labile – or because it was hydrolysed by the protein. Remarkably, the Gal1 acetyl signal is smaller than the Gal4 one. As Gal1 is closer to the putative esterase site, this may suggest that the active site is functional.

The Cα backbones of the apo-protein and the complex are almost identical (RMSD = 0.3 Å). Despite the groove between the two subdomains, they appear to be very rigid even in absence of a ligand. The sequence of the five ordered residues of the bound oligosaccharide is D-Galp[6Ac]-α-1 – 6-D-Manp-β-1 – 4-L-Rhap-α-1 – 3-D-Galp[6Ac]-α-1 – 6-D-Manp-β-1 – ROH. The oligosaccharide binds to the petal through a network of hydrogen bonds, most of them through a water molecule. Only four amino acids out of sixteen establish direct interactions with the oligosaccharide. The Oδ1 atom of Asn937 interacts with the Gal1, the amide of Tyr941 and Nδ2 of Asn980 interact with the O-acetyl of Gal4 and the carboxylic oxygen atom of Thr939 interacts with Man5. Seven amino acids in the loop from β9 to α3 (Arg936-Thr942) intervene in this hydrogen bond network. The putative oxyanion hole residue Asn937 also belongs to this loop (Fig. 20A). Surprisingly, the ester group of Gal1 is located 12 Å away of the Ser787 in the crystal structure. Unlike Ser787 in the apo-protein, here it only has one conformation. The Oγ atom of Ser787 interacts with O3 of Gal1 through a water

molecule. The amide of the oxyanion Gly813 also has a hydrogen bond with this water. Asn937 is the only conserved residue that interacts directly with the Gal1 (Fig. 20A).

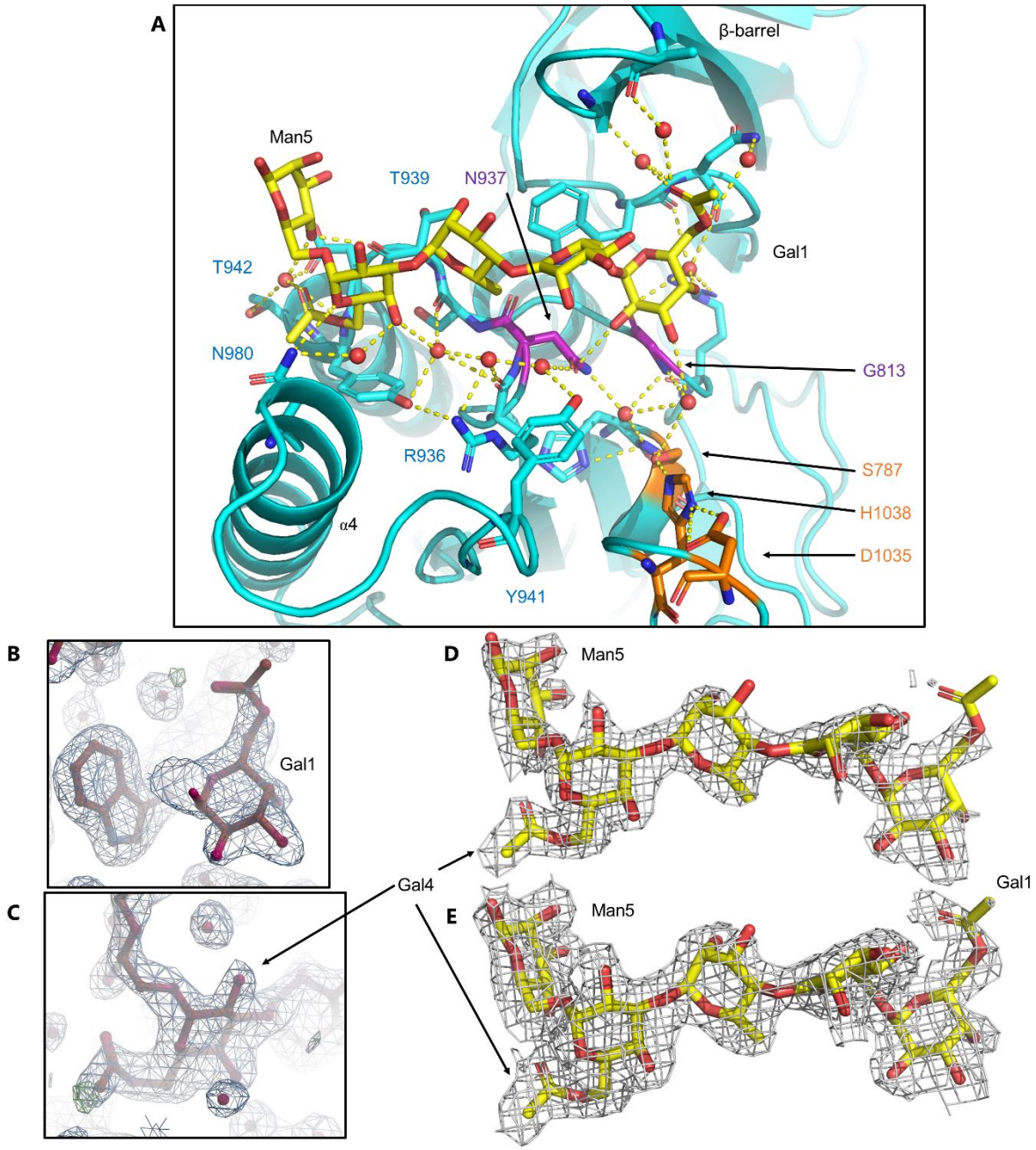


Figure 20: Petal-oligosaccharide complex. (A) Close-up view of the pentasaccharide and the petal biding site. Petal domain is shown in carton view (cyan). Main chain and side chain of residues interacting with the pentasaccharide are shown as sticks. The putative catalytic triad is shown in orange and the oxyanion residues in purple (Ser787 belongs to both groups). The water molecules of the interaction network are shown as small red balls and the hydrogen bonds as yellow dashed lines. Coot screenshot of the modelled Gal4 (B) and Gal1 (C) (brown) with the 2Fo-Fc electron density map contoured at 1σ (grey). Saccharides shown in panels B-D are rotated 90º counter-clockwise with regard to the pentasaccharide in (A). (D) Stick representation of the pentasaccharide (yellow) with the 2Fo-Fc electron density map contoured at 1σ (grey). (E) Same as in (D) with the 2Fo-Fc electron density map contoured at 0.5σ.

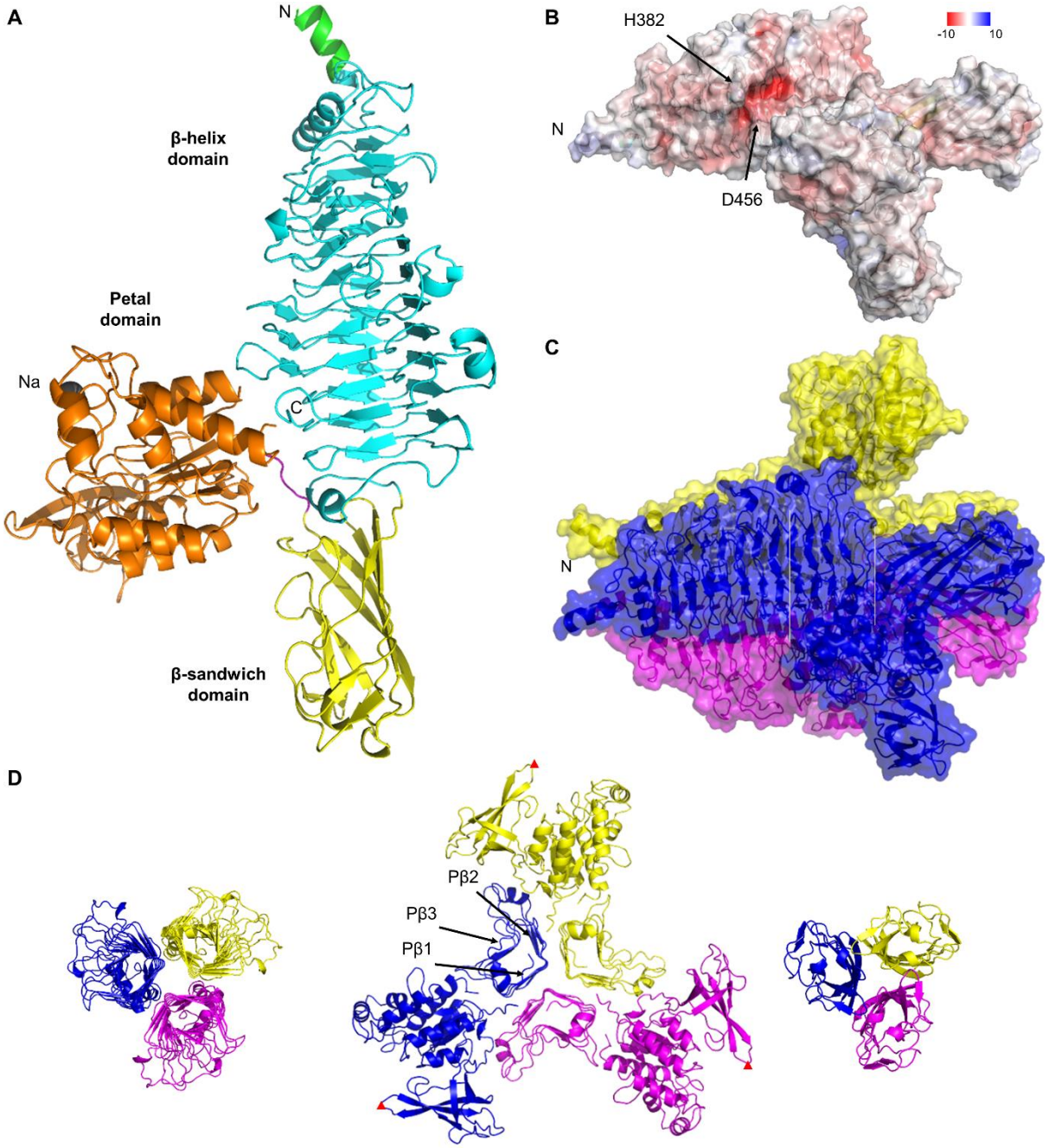


Figure 21: The structure of gp20ΔN. (A) Cartoon representation of the gp20ΔN (residues 248-1070) monomer with affinity tag residues coloured in green, the β-helix domain in cyan, the β-sandwich domain in yellow, a loop in magenta and the petal domain in orange. (B) Cartoon and surface representation at 20% of transparency of a gp20ΔN monomer coloured according to the charge, red for acid, blue for basic and white for neutral charge. (C) Cartoon and surface representation at 40% of transparency of a gp20 ΔN trimer with one colour per chain. Protein is divided from the proximal end (N-terminal, at the left) to the distal end (right) as follows β-helix domain (rungs 1-8), β-helix (rungs 9-12) and petal domain, and β-sandwich domain. (D) Transversal views of the portions from the distal end of (C). Parallel β-sheets 1, 2 and 3 are indicated in the β-helix. Gly878 position is represented with a red triangle.

4.1.4. Structure of gp20ΔN without and with its receptors

The crystal structure of gp20ΔN (residues 248-1070) shows the last nine residues of the N-terminal affinity tag and residues 248-1070 of gp20. The nine his-tag amino acids and the first two gp20 amino acids form an α-helix. Then, a β-β-α fold comprising residues 250 to 280 caps one end of a right-handed parallel β-helix (amino acids 281 to 633). This capping fold is common to other β-helices (Fig. 21A). Each of the twelve rungs of the β-helix is formed by three β-strands. Some turns between strands have longer loops, mainly between the second and the third strands and between the third and the first strands of each rung. Rungs 6, 7 and 8 have another β-strand in the loops after their third strand. These longer loops form a long and open groove interrupted by an arch built by an hydrogen bond between His382 and Glu456, located in the loop between the second and the third strands of the fourth rung and the loop between the third and fourth strands of the sixth rung, respectively. A negatively charged patch is located exactly next to this position (Fig. 21B), something common in the active sites of other tailspikes (Plattner et al., 2019). The β-helix domain ends with an α-β fold (634-644).

The next domain is a β-sandwich (645-783). It is composed of a six-stranded and a five-stranded antiparallel β-sheet. The distal end has insertions with two 3_{10} helices, one α-helix and some long loops. A six-residue loop (774-779) links the petal domain (780-1070) to the rest of the sequence. In the petal domain, a sodium ion is bound to the loop in the α/β subdomain, between the β6 and β7, outside the binding and the putative catalytic site. The amide group of a Tris molecule binds in the same pocket in gp20(734-1070) crystal structure. Both the N- and the C-termini of the petal domains are in the interface with the last two rungs of the β-helical domain. The petal domain is the same as in the gp20(734-1070) structure (RMSD = 0.4 Å).

Most of the tailspikes are trimers. The asymmetric unit (AU) of this crystal has only one monomer. Nevertheless, the crystallographic three-fold rotation axis is parallel to the long axis of the monomer, so adding two symmetry mates generates a trimer. This trimer was fitted in a 13-Å-resolution Cryo-EM map. All domains fit inside the map and only 974 atoms out of 21360 were outside the map (Fig. 22A-B). The missing 247 amino acids in the N-terminal end probably would fill the void seen on the other side of the kink in the cryo-EM map. Then, this N-terminal domain would connect the body of gp20 with the phage tail. The overall dimensions of gp20ΔN trimer are 131 x 114 Å (height x triangle side from petal Gly878 to petal Gly878; the furthest residues from the centre (60 Å away)) (Fig. 21C-D). In addition, an analysis of the interfaces by QtPISA (Krissinel, 2015) shows thirty hydrogen bonds and nine salt bridges as

inter-monomeric contacts. Trimer formation buries 17.3 x 10^3 Å^2 of surface area out of 96.5 x 10^3 Å^2 of total surface area (18%). This buried surface is in the low range of other structurally homologous tailspikes; possibly, due to the highly solvent-accessible petal domain. The predicted standard free energy of dissociation (ΔGdiss) is clearly positive, 48.7 kcal/mol, which is on the low side when compared to similar tailspikes (Table 8), but still clearly indicative of stable trimer formation.

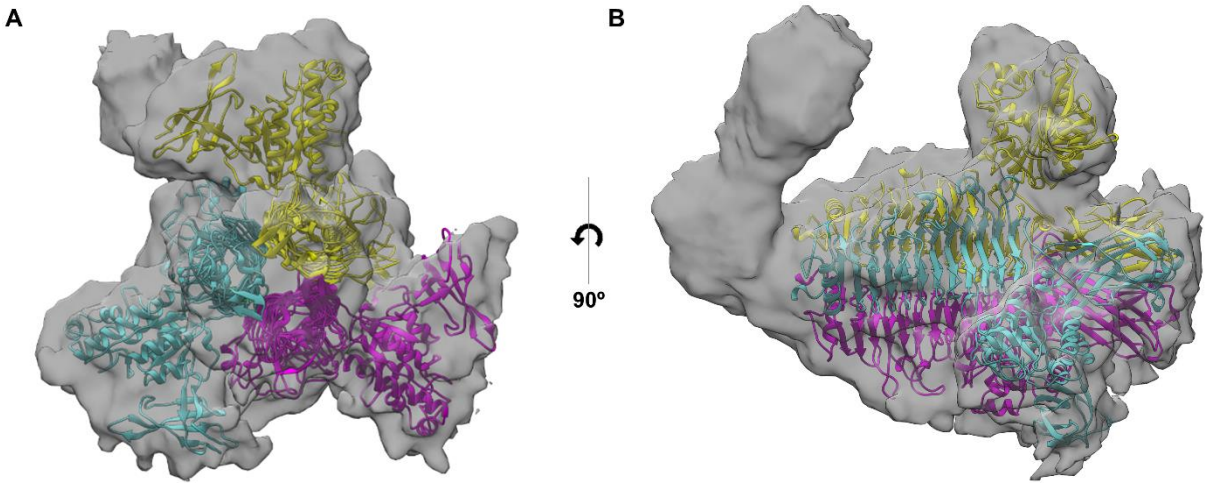


Figure 22: Gp20ΔN crystal structure fitting in gp20 cryo-EM map. Bottom view (A) and lateral view (B) of a tailspike segment of an epsilon15 cryo-EM map (13 Å resolution, EMDB entry: EMD-5209) with a trimer of gp20ΔN fitted into the map. Each chain is coloured in a different colour.

Table 8: Buried surface and contact analysis of gp20ΔN and its structural homologues.

Protein	ASA (x 10^3 Å^2)	BSA (x 10^3 Å^2)	BSA* (%)	ΔG^{diss} (kcal/mol)
gp20ΔN	79.2	17.3	18	48.7
Det7 gp208 (6F7D)	47.0	17.1	27	89.3
TSP2ΔN (5W6P)	56.8	19.2	25	121.4
K5 Lyase (2X3H)	43.2	33.2	43	256.1
Sf6 TSP (2VBE)	45.8	20.3	31	128.0

ASA: Accessible Solvent Area
BSA: Buried Surface Area
* BSA with respect to the total area
All values were calculated with QtPISA and refer to the monomeric protein.

A structural homology search found that gp20ΔN is structurally homologous to other trimeric β-helical proteins with endoglycosidase activity. The most similar structures are one of the Det7 tailspikes, gp208ΔN, (PDB entry 6F7K; Z-score of 46.6, RMSD value of 1.8 with 521 Cα in alignment and identity of 40%) and the phage CBA120 tailspikes, TSP2 (PDB entry 5W6P; Z-score of 31.8, RMSD value of 1.8 with 327 Cα in alignment and identity of 16%). Gp208ΔN is an epsilon15-like tailspike, which binds one hexasaccharide consisting in D-Galp-α-1 – 6-D-Manp-β-1 – 4-L-Rhap-α-1 – 3-D-Galp-α-1 – 6-D-Manp-β-1 – 4-L-Rhap-α-1 – ROH.

However, it misses the petal domain present in gp20 and some other long loops present around the β-helix groove and in the β-sandwich domain (Fig. 23).

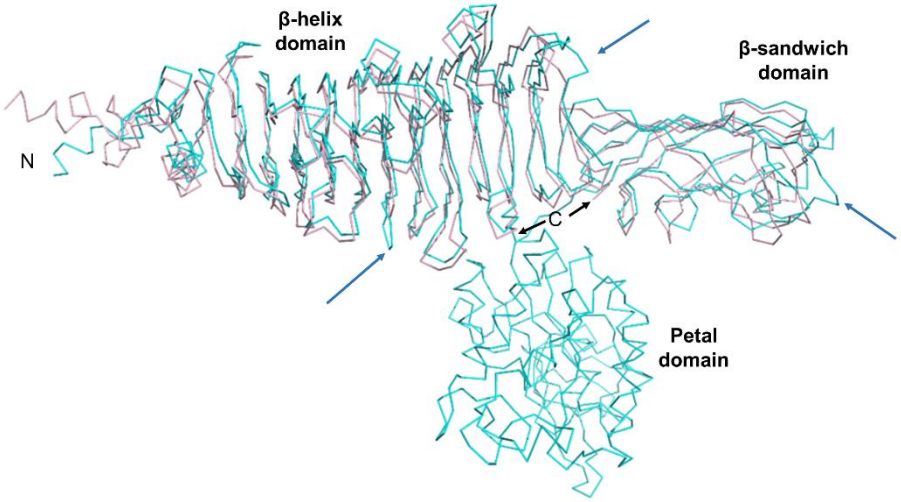


Figure 23: Gp20ΔN domain structural homologies. Wire representation of gp20ΔN (cyan) and gp208ΔN (light pink). N- and C-termini are indicated. Blue arrows indicate the points where backbones diverge.

The crystal structure of gp20(248-1070) with the ligand is identical to the apo-protein structure (RMSD of the Cα = 0.4 Å) with the addition of three hexasaccharide bound per monomer (Fig. 24A). The sequence of the added hexasaccharide is D-Galp[6Ac]-α-1 – 6-D-Manp-β-1 – 4-L-Rhap-α-1 – 3-D-Galp[6Ac]-α-1 – 6-D-Manp-β-1 – 4-L-Rhap-α-1 – ROH. The only difference with the gp208 ligand is the presence of acetyl groups bound to the galactoses. Hereinafter, hexasaccharides will be referred as Hex1, Hex2 and Hex3 in N- to C-termini order (Fig. 24A-D). Hex2 is located in gp20 almost in an identical position to the hexasaccharide in gp208 (See Fig. 37 in the Discussion section).

Hex1 binds to the negatively charged patch and near the His382-Asp456 arch in the β-helix. It is bound by amino acids located in PB3s of rungs 4, 5, 7 and 8 and their surrounding loops. The links are mainly direct or water-mediated hydrogen bonds connecting with all saccharides, but some aromatic stacking interaction are also present. Rha6 has hydrogen bonds with four amino acids (Asp449, Gln476, Gln503 and Asp507), the Hex2 Gal1 and an aromatic stacking with Tyr480 (Fig. 24E). Hex2 also binds to the β-helix domain from the distal end of the negatively charge patch up to the β-helix domain end. It shows an almost identical position to the hexasaccharide in the Dettilon TSP (Det7 gp208; PDB entry 6F7D). Binding amino acids belong to the PB3s of rungs 7-12 and some of their previous loops. Most of them are hydrogen bonds and there is also an aromatic stacking interaction between Tyr630 and Man5. This aromatic stacking interaction is also present in the Dettilon TSP structure. The Gal1 of Hex2,

in the negatively charged patch, is the saccharide with most hydrogen bonds to the protein (Fig. 24F). The β-helix cleft where these hexasaccharides are bound is 7.5 to 17.5 Å wide, measured at Hex1 Man5 and at Hex2 Man2 respectively, and 34 Å long. Hex3 binds to the loops 1, 5 and 7 of the β-sandwich domain proximal region that face the petal domain (Fig. 24G). Although the soaked hexasaccharide had an acetyl group bound to the C6 of the galactoses, this group is only resolved in Hex2 galactoses (Fig. 24B-D).

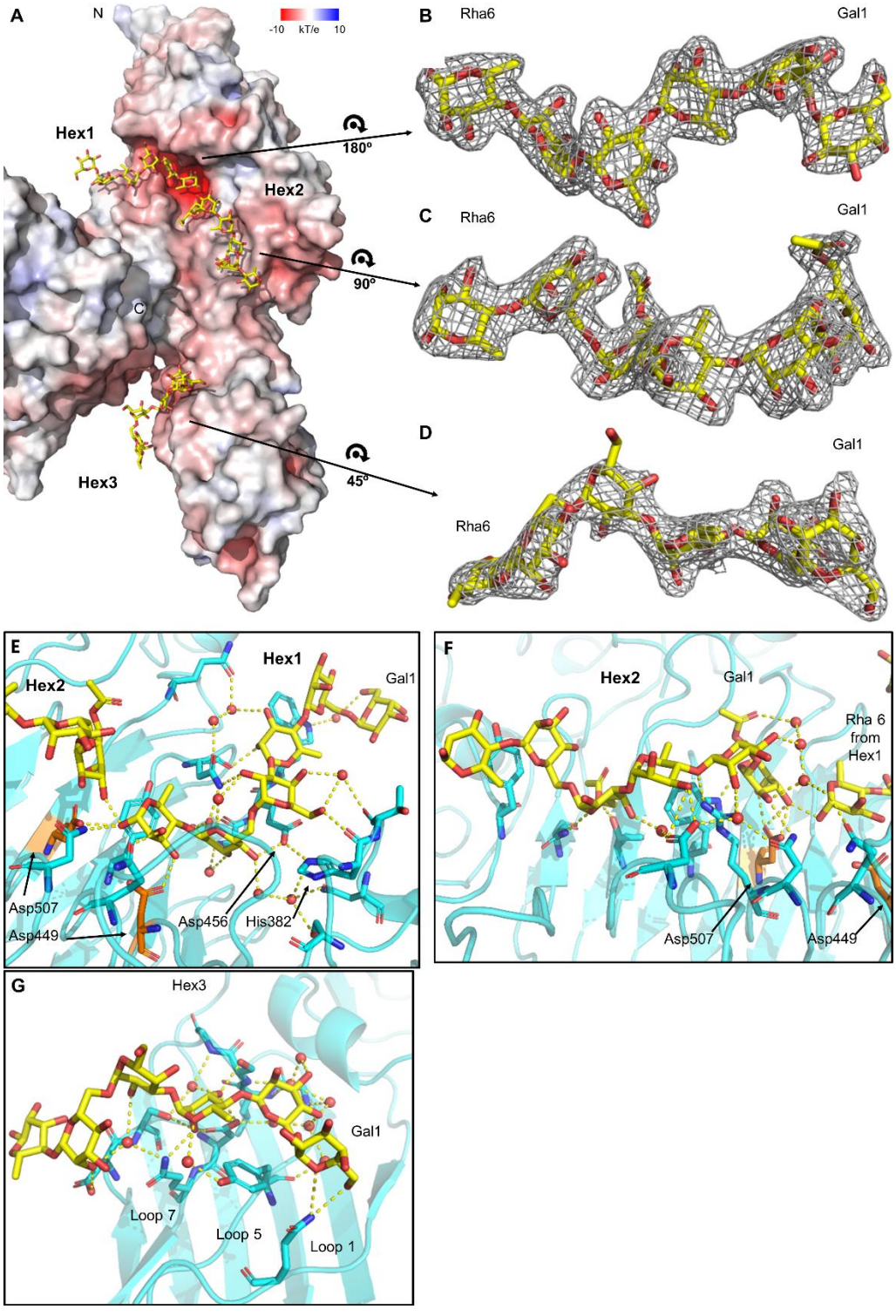


Figure 24: Gp20ΔN with three bound hexasaccharides. (A) Surface representation of gp20ΔN structure with three hexasaccharides according to its surface charge; red for acid, blue for basic and white for neutral charge. Stick representation of Hex1 (B), Hex2 (C) and Hex3 (D) (yellow) with their 2Fo-Fc electron density maps contoured at 1σ (grey). Close-up views of Hex1 (E), Hex2 (F) and Hex3 (G) binding sites showing the hydrogen bonds (yellow dashed lines). Hexasaccharides are shown in as sticks (yellow), gp20ΔN as ribbon (cyan), residues as sticks (cyan, Asp507 and Asp449 in orange) and water molecules as balls (red).

The first galactose of Hex2 (Gal1) and the reducing rhamnose of Hex1 (Rham6) are connected by a hydrogen bond (Fig. 24E). This interface is located above the negatively charged patch. Here, Asp449 and Asp507 closer carboxylic oxygens are at a distance of 5.3 Å from each other, constituting a putative catalytic site. This suggests gp20ΔN could use the retaining mechanism to hydrolyse *Salmonella enterica* serovar Anatum O-antigen polysaccharides.

4.1.5. Microscopy of the full-length protein

After unsuccessful crystallisation trials with full-length gp20 protein, I tried to characterise its structure by negative-staining electron microscopy. 2D classification sorted the gp20 particles in classes with up to 9 different bending angles between the N-terminal region and the main body (Fig. 25). This flexible region more or less corresponds to the void space in the epsilon15 tailspike depicted in Fig. 18 and 22. The most common angle is 60º (31% of the particles), similar to the angles found in the tailspikes of available epsilon15 maps. However, as shown here, gp20 angles can be as acute as 40º and as obtuse as 170º. Tailspikes bound to the phage might be less flexible due to the interactions of gp20 with epsilon15 tail. In addition, particle processing was probably focussed on the capsid, a much larger structure, and the less common alternative conformations of the tailspikes could be missed. The flexibility shown by the full-length trimer probably hindered crystallisation efforts.

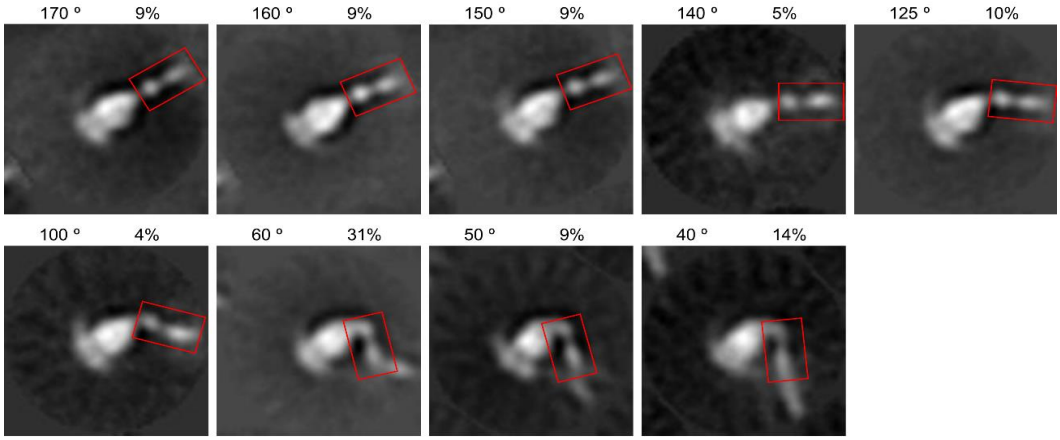


Figure 25. Full-length gp20 EM reconstruction. Nine class averages showing the gp20 N-terminal region (inside the red rectangle) bent at different degrees. In total, they come from almost 15000 particles. The approximate angle with respect the main body and the percentage of particles in that class are indicated above each image.

4.1.6. MBTH assays with gp20ΔN mutants

To unravel the importance of amino acids present around the interface of Hex1 Rha6 and Hex2 Gal1 and to assess the role of the petal triad on the endorhamnosidase activity, I performed a number of single-point mutations in gp20ΔN (W441R, D449N, S477A N503D, D507N, S787A, H1038A) (Fig. 26A-B). Then, I evaluated the hydrolytic activity of the mutants on the Rha6-Gal1 α 1 – 3 link by detecting the number of reducing ends produced after cleavage (Fig. 26C). The activity of the petal domain alone was not statistically different from the control with polysaccharide only, indicating that its putative active site is not involved in O-antigen cleavage. Hence, it can be used as a control. Besides, full-length gp20 and gp20ΔN results were not statistically different, so I decided to use gp20ΔN as the reference for the gp20ΔN mutants. On one hand, D449N, N503D and D507N reducing ends productions were the lowest, being statistically similar to the petal domains results. This means that these residues may be involved in the O-antigen cleavage reaction. On the other hand, esterase mutants S787A and H1038A were statistically similar to gp20ΔN. W441R and S477A showed intermediate activities, suggesting these residues are important for the binding of the oligosaccharide, but are probably not directly involved in the enzymatic activity.

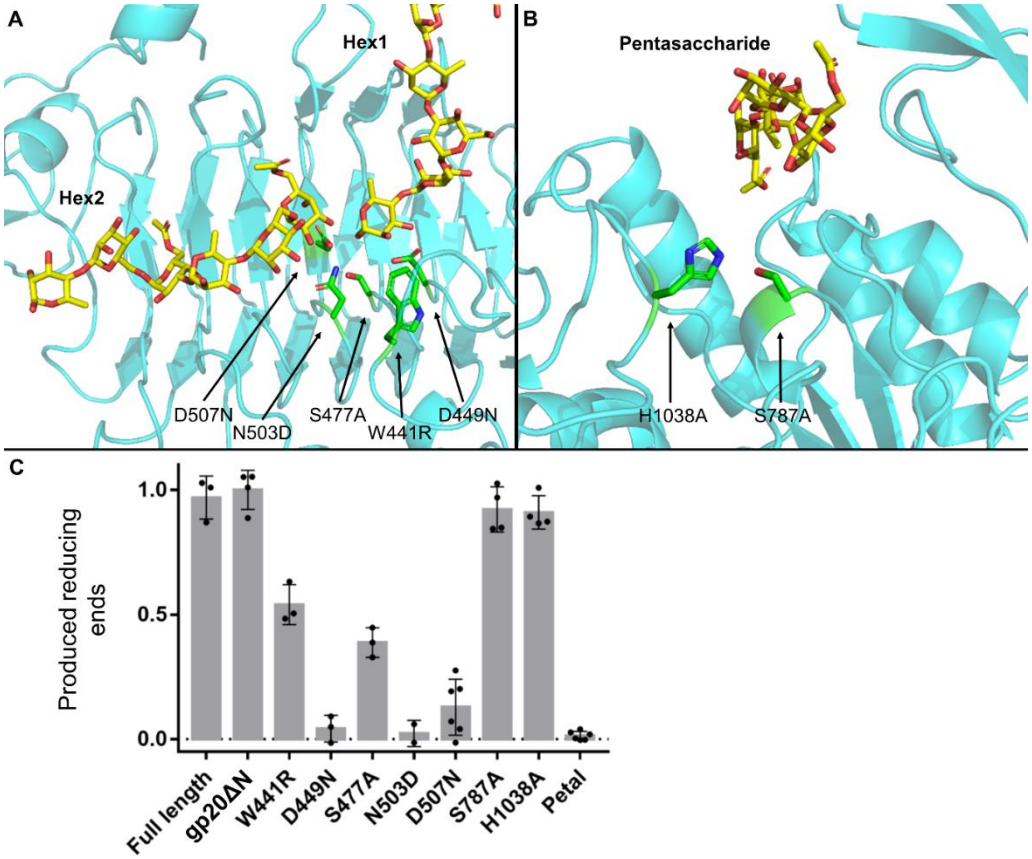


Figure 26: Endorhamnosidase activity. Putative endorhamnosidase (A) and esterase sites (B). Gp20 is coloured in cyan and mutated amino acids are coloured in green. (C) Relative amount of reducing ends produced by gp20

constructs and single-point mutants. The unhydrolysed polysaccharide signal was set as 0.0 and the signal from gp20ΔN-hydrolysed polysaccharide was set as 1.0.

4.2. The RBP3 tailspike of *Campylobacter virus F358*

4.2.1. Expression and purification of RBP3

The RBP3 tailspike of *Campylobacter jejuni* bacteriophage F358 was expressed in *E. coli* strain BL21-CodonPlus (DE3)-RIL. For purification, an immobilised metal affinity chromatography (IMAC) removed most of the soluble contaminant proteins (Fig 27A). Then, the purest fractions were dialysed against 20 mM Tris pH 7.5 to remove glycerol, imidazole and sodium chloride. Subsequently, RBP3 was purified by anionic exchange chromatography at pH 7.5. After increasing the concentration of sodium chloride, RBP3 eluted at a sodium chloride concentration of about 0.14 M in a narrow peak (Fig 27B). It was collected in a single tube and checked in a denaturing gel. The band of approximately 45 kDa may correspond to the secondary peak next to the main one (Fig 27C). However, the sample was regarded to be pure enough to perform crystallisation trials. Therefore, the sample was desalted and concentrated to about 13 mg/mL to set up crystallisation trials. The yield of purified RBP3 was about 3.7 mg from 3.4 litres of culture, enough to try almost 800 different crystallisation conditions.

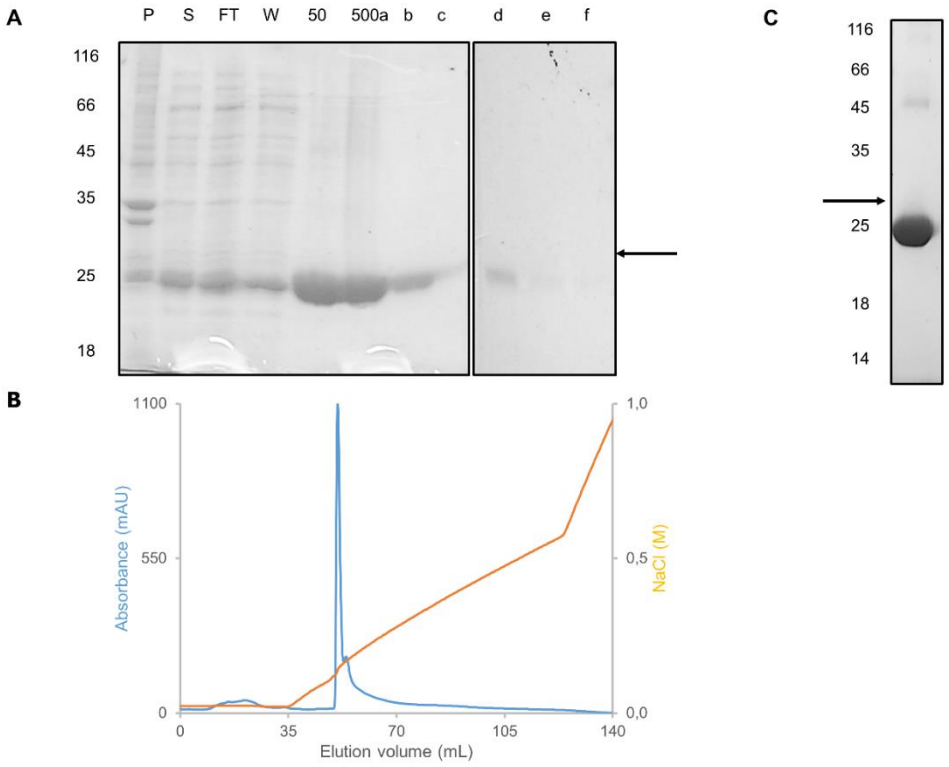


Figure 27: Purification of *Campylobacter virus F538* tailspike RBP3. (A) Denaturing gel with samples from the IMAC (P, pellet; S, supernatant; FT, flow through; W, wash; imidazole concentration in mM, 6 tubes of the

500 mM imidazole fraction were recovered). (B) Anion exchange chromatography profile. The peak was recovered in a single fraction. (C) Denaturing gel electrophoresis of purified RBP3. Black arrows in (A) and (C) indicate the expected molecular weight of RBP3, 28 kDa.

4.2.2. Crystallisation and structure determination of RBP3

Crystallisation trials performed with RBP3 produced crystals belonging to two different space groups. The best crystal belonging to space group C2 grew in 22% (w/v) PEG 4000, 0.1 M sodium acetate and 0.2 M ammonium sulphate (Fig. 28A), diffracted to 1.20 Å (Fig. 28C) and had a unit cell with cell axes of 98.8, 58.9 and 102.2 Å and a beta axis of 92.7º. Crystals belonging to the $P6_522$ space group grew in 1.5 M ammonium sulphate, 75 mM Tris-HCl pH 8.5, 25% (v/v) glycerol, diffracted to 1.52 Å (Fig. 28D) and had a unit cell of 93.2 x 93.2 x 288.1 Å (Fig. 28B). Some of the C2 crystals were soaked with methylmercury chloride to perform experimental phasing by single anomalous diffraction. The best of these crystals grew in 20% (w/v) PEG 8000, 0.1 M HEPES pH 7.5 and 0.2 M magnesium acetate and showed significant anomalous signal.

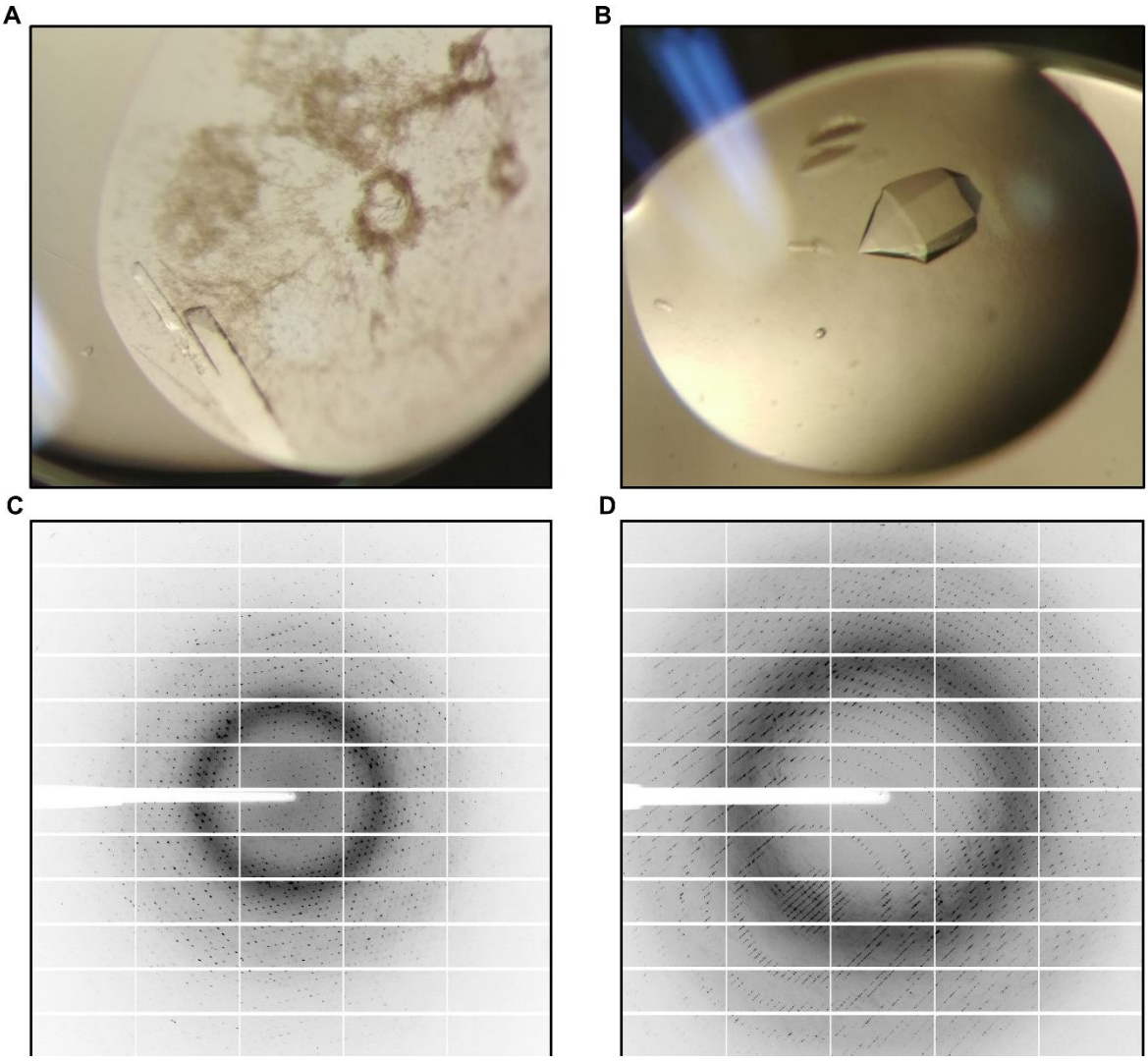


Figure 28: RBP3 crystals and diffraction patterns. Crystals (A) and diffraction pattern (C) of the C2 RBP3. Crystals (B) and diffraction pattern (D) of the $P6_522$ RBP3.

Table 9: Crystallographic data of RBP3 proteins.

	C2 RBP3 + Hg	C2 RBP3	$P6_522$ RBP3
Data collection			
Radiation source	BL13-XALOC (ALBA-CELLS)	BL13-XALOC (ALBA-CELLS)	BL13-XALOC (ALBA-CELLS)
Wavelength (Å)	1.0077998	1.005548	1.005551
Detector	PILATUS 6M – DECTRIS	PILATUS 6M – DECTRIS	PILATUS 6M – DECTRIS
Crystal-to-detector distance (mm)	464.44	179.67	295.46
Oscillation per image (º)	0.2	0.2	0.2
Number of images	1800	900	900
Data processing			
Space group	C2	C2	$P6_522$
Cell edges (a, b, c; Å)	98.7, 58.8, 101.1	98.8, 58.9, 102.2	93.2, 93.2, 288.1
Cell angles (α, β, γ; º)	90, 91.24, 90	90.0, 92.7, 90.0	90.0, 90.0, 120.0
Resolution range (Å)	50.52-2.36 (2.42-2.36)	102.10-1.20 (1.22-1.20)	48.02-1.52 (1.55-1.52)
Nº of reflections (Total)	137552 (13751)	566923 (26116)	2065069 (60974)
Nº of reflexions (Unique)	24401 (2368)	181869 (8883)	114387 (5561)
Completeness (%)	99.3 (99.4)	99.5 (98.1)	100.0 (100.0)
Anomalous completeness (%)	98.3 (98.1)	-	-
Multiplicity	5.6 (5.8)	3.1 (2.9	18.1 (11.0)
Anomalous multiplicity	2.9 (3.0)	-	-
CC ½[a]	0.996 (0.848)	0.983 (0.651)	1.000 (0.317)
Resolution (Å) at CC_{anom} = 0.15	3.61	-	-
Rmeas (all I+ and I-)[b]	0.134 (0.603)	0.149 (0.711)	0.067 (4.170)
$\langle I/\sigma(I)\rangle$	9.9 (3.2)	7.1 (4.0)	20.1 (0.5)
Phasing			
Heavy-atom sites	9 Hg	-	-
Correlation coeff. (all/weak)[c]	33.1/19.5	-	-
Phasing CLD for the (chosen/rejected hand)[d]	19.5/0.0	-	-
Refinement			
Resolution range (Å)	-	102.09	48.06-1.52
Reflections used	-	181875	114233
Reflections used for R-free	-	8871	5862
R-factor/R-free[e]	-	0.12/0.15	0.18/0.19
Model statistics			
Amino acid coverage	-	12-239	20-239
Atoms (protein/ions/ ligands/water)	-	5507/0/23/913	3641/0/118/403
Ramachandran[d] (% favoured/allowed/outliers)	-	95.85/3.26/0.89	96.09/3.68/0,23
Rmsd[e] (bonds, Å / angles, º)	-	0.010/1.48	0.008/1.37
B-factor (protein/ions/ ligands/water)	-	7.83/-/42.28/21.37	29.61/-/76.58/44.18

MolProbity score/ percentile[h]	-	1.40/83rd	1.52/84th
PDB code	-		

*The values in parentheses correspond to the high-resolution shell

[a] CC1/2 correlation coefficient between intensity estimates from half data sets (Karplus & Diederichs, 2015).

[b] $Rmeas=(\sum_{hkl}[n/n-1)]^{1/2}\sum_i|I_{i\,hkl}\sum_i I_i(hkl)$

[c] Found by SHELXD (Sheldrick, 2010)

[d] According to Solomon (Abrahams and Leslie, 1996) and Multicomb (Skubák et al., 2010)

[e] $R = (\sum hkl\ |Fobs(hkl) - Fcalc(hkl)| / \sum hkl\ |Fobs\ (hkl)|$

[f] Calculated by MolProbity (Williams et al., 2018)

[g] Calculated by Refmac5 (Evans and Murshudov, 2013)

[h] MolProbity score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalised to be on the same scale as X-ray resolution. 100th percentile is the best among the structures of comparable resolution; 0th percentile is the worst. The comparative set of structures for MolProbity score was selected in 2006 (http://molprobity.biochem.duke.edu/; Williams et al., 2018)

4.2.3. Structure of RBP3

In both crystal forms, RBP3 forms a beta-helix with seven rungs. However, interestingly, the quaternary organisation differs between them. On one hand, the asymmetric unit of the C2 crystals consist of three copies of RBP3 with non-crystallographic symmetry (Fig. 29A and B). According to the QtPISA analysis, each monomer has 27 hydrogen bonds with the other two. This suggests three copies of RBP3 form a stable quaternary structure i.e. a homotrimer, as all known tailspikes (Sanz-Gaitero et al., 2019). On the other hand, the asymmetric unit of the $P6_522$ crystals have two copies of RBP3 (Fig. 29E). QtPISA analysis of this structure shows only eight hydrogen bonds between the two monomers. Therefore, QtPISA concludes it likely does not form a complex in solution. Hereinafter, trimeric RBP3 from the C2 crystal form will be referred to as tRBP3 and the monomeric RBP3 from the $P6_522$ as mRBP3. The differences in quaternary structure and the conformation of the N-terminal region are likely caused by the presence of the high salt concentration (1.5 M ammonium sulphate) in the crystallisation mixture used to obtain the $P6_522$ crystal form.

One monomer of tRBP3 and another of mRBP3 are almost identical, having a RMSD value of 1.1 Å when Cα atoms are superposed (Fig. 29G-H). To compare, the RMSD values of the structural alignments among the chain A and the other two chains of tRBP3 are 0.2 and 0.3 Å (Fig. 29C) and the RMSD value of the structural alignment between the two mRBP3 monomers is 0.3 (Fig. 29F). One monomer of tRBP3 is composed of only one domain, spanning from amino acid 12 to 239 (Fig. 29D). It has a right-handed parallel β-helical fold, as other tailspikes. tRBP3 has an initial linker (residues 12-22) followed by the β-α cap (23-47) similar to other β-helices (Fig. 30E). A protruding loop (48-55) interacts with the cap and first rung of a neighbouring chain (Fig. 29 A-B). The β-helix fold has seven complete rungs starting in the amino acid 56. Loops in T2 and the PB3 of one chain and the PB1 of the next chain form

a V-shaped groove (Fig. 29A and 30G). RBP3 lacks the C-terminal cap common to most beta-helix folds (Bryan et al., 2011); instead, the entrance to the interior of RBP3 is blocked by the last T3 loop (226-233) and the side chains of Asn 212, Phe217 and Lys224 (Fig. 29I).

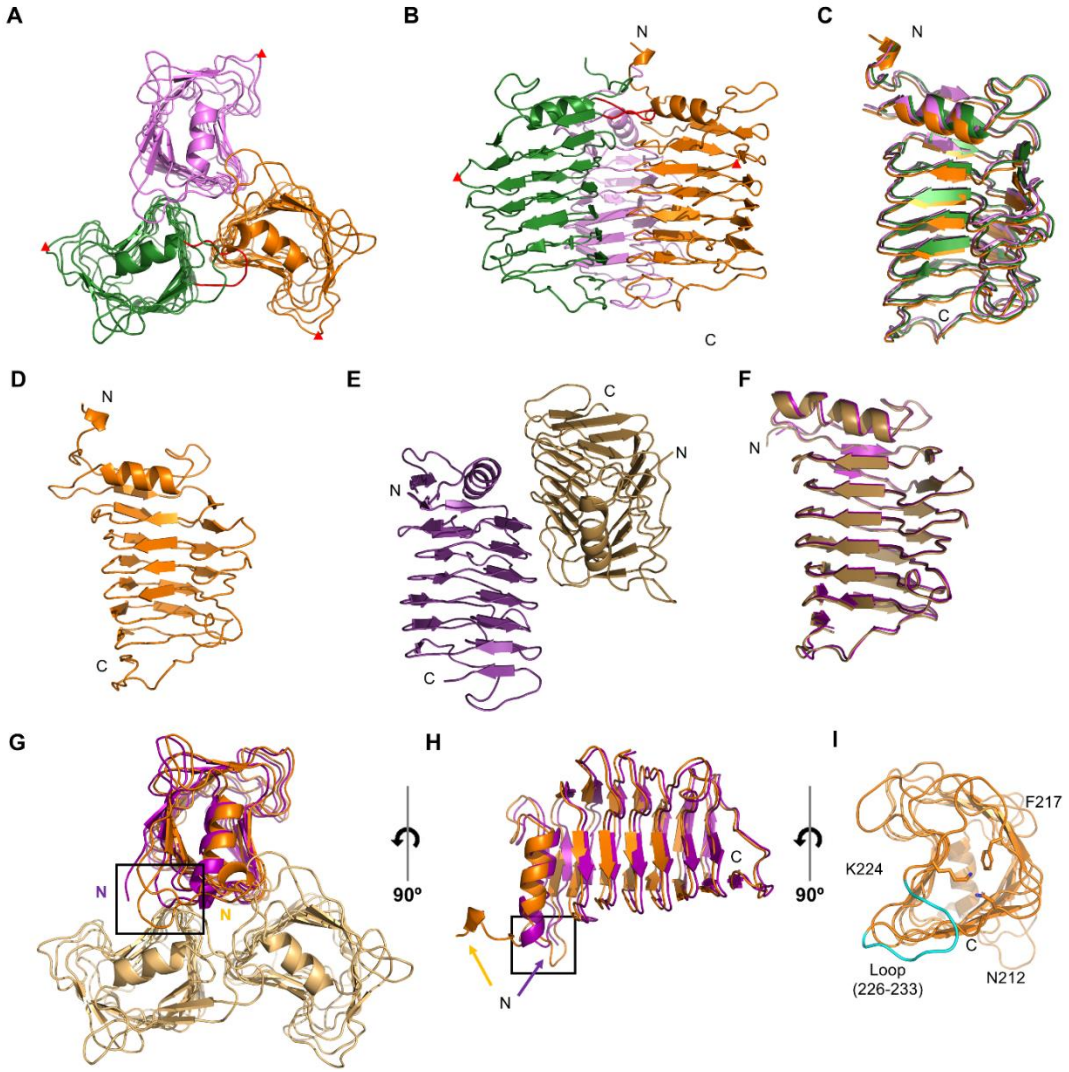


Figure 29: Structure of RBP3. (A) Top view of the tRBP3 structure in ribbon representation. Each chain is coloured orange, green and light violet. The red triangles indicate the position of Ala94. The protruding loop is coloured red. (B) Side view of the tRBP3 structure coloured as in (A). The red triangles indicate the same as in (A). (C) Structural alignment of the three tRBP3 chains coloured as in (A). (D) Side view of a monomer of tRBP3. (E) The two mRBP3 present in the $P6_522$ asymmetric unit. (F) Structural alignment of the two mRBP3 chains coloured as in (E). (G) Top view of tRBP3 with one chain in orange and two chains in light orange. One mRBP3 chain is aligned to the orange chain of tRBP3. (H) Side view of the tRBP3 and mRBP3 monomers aligned in (G). The protruding loop is framed inside a black rectangle in the panels G and H. (I) Bottom view of a tRBP3. The last T3 loop (226-233) is shown in cyan, and residues Asn212, Phe217 and Lys224 are shown in stick representation.

The only significant differences between the structures of tRBP3 and mRBP3 are located in the N-terminal region, before the beginning of the β-helix. Residues 12-22 are not ordered in mRBP3. In addition, the protruding loop is shorter in mRBP3, which makes the previous α-helix longer (Fig. 29E). The lack of interaction with the missing initial linker on one side and

the nearby cap and first rung in the other might cause the aforementioned shortening of the protruding loop (Fig. 29D).

The height of the tRBP3 is 54 Å and the triangular side is 60 Å from Ala94 to Ala94, in the T2 loop of the second rung (the longest distance between the same residue in different chains) (Fig. 29A and B). In the monomer, the widest distance is 34Å between the carbon α of Ala94 and Phe79.

In the trimer, RBP3 chains are rotated about 90º when compared with other tailspikes. Figure 30A shows their PB3 is almost perpendicular to the centre. Instead, the PB3 of LKA tailspike, K5 lyase, P22 tailspike and epsilon15 tailspike is approximately tangential to the centre (Fig. 21D & 30B-D). In tRBP3, the PB2 sheet and T1 turns completely face the outer space. Its chains contact each other with the T3 turns and the PB1 sheets only. In other tailspikes (Fig. 30B-D), the PB1 sheets, the T1 turns and the PB2 sheets intervene in the trimeric contacts. Despite this atypical conformation, the RBP3 trimer shows a similar buried surface area as the β-helix domains of other tailspikes and even, in some cases, a higher ΔG^{diss}; probably, to compensate for the lack of other domains that could help stabilise the trimer (Table 10).

Table 10: Buried surface and contact analysis of tRBP3 and other tailspikes.

β-helix domain	ASA (x 10^3 Å^2)	BSA (x 10^3 Å^2)	BSA* (%)	ΔG^{diss} (kcal/mol)
Trimeric RBP3	22.4	7.2	24.3	25.7
Gp49(221-680)	37.8	10.1	26.8	5.9
K5 lyase(7-374)	34.6	13.6	39.2	48.0
Gp20(248-645)	38.3	9.0	23.6	16.9
Gp9(111-628)	39.6	10.9	21.6	18.4

List of the tailspike-possessing phages: F358, Pseudomonas phage LKA1 (PDB entry 4RU4), K5 (PDB entry 2X3H), epsilon15 and P22 (PDB entry 2XC1).
ASA: Accessible Solvent Area
BSA: Buried Surface Area
* BSA with respect to the total area.
All values were obtained with QtPISA and refer to the monomeric protein

The RBP3 beta-helix fold is structurally homologous to trimeric phage tailspikes and other monomeric glycoside hydrolases; such as epimerases, alginate lyases or pectinases. The most similar structures are *Azotobacter vinelandii* mannuronan C-5 epimerase AlgE6 (PDB entry 5LW3, Z-score of 17.5, RMSD value of 2.3 with 188 Cα in alignment and identity of 13%), the tailspike gp49 from *Pseudomonas* phage LKA (PDB entry 4RU4, Z-score of 17.3, RMSD value of 2.4 with 184 Cα in the alignment and identity of 17%) and the tailspike K5

lyase A (PDB entry 2X3H, Z-score of 16.8, RMSD value of 2.6 with 183 Cα in alignment and identity of 10%). However, none of their catalytic or putatively catalytic amino acids are present in RBP3. Instead, RBP3 has a negatively charged area in the proximal region of the β-helix (Fig. 30F-G). This patch is located in the deepest part of an interchain groove. Six aspartic or glutamic acids surround this area with distances to each other between 7.6 and 11.6 Å (Fig. 30H). Catalytic amino acids in inverting enzymes are usually 10.5 Å apart (Davies & Henrissat, 1995).

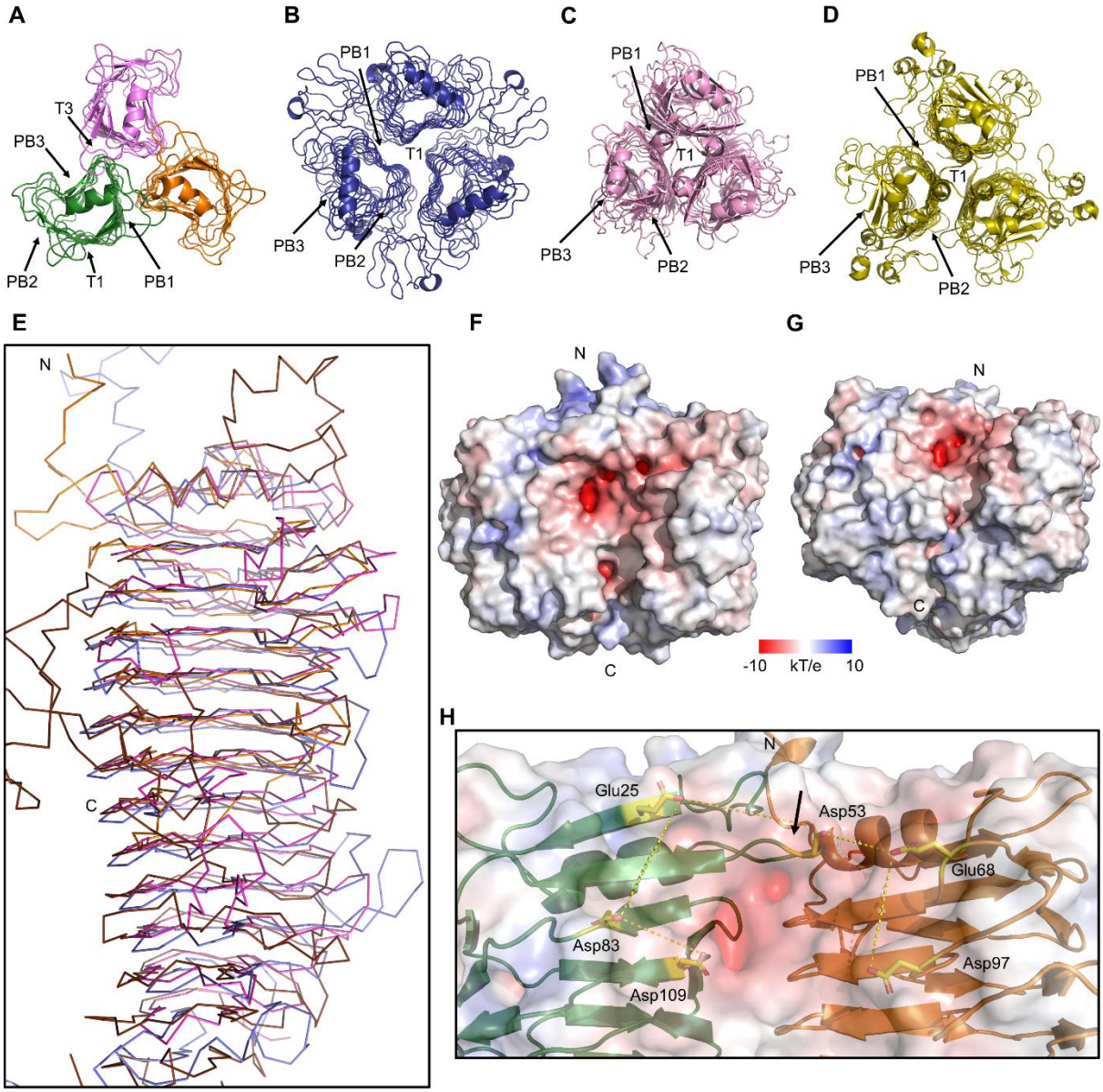


Figure 30: RBP3 homologues and function. Top view of the tRBP3 represented as in Fig. 29A (A), LKA tailspike gp49(221-680) (B), K5 lyase A (C) and P22 tailspike gp9 (D). Parallel β-sheets and turns mentioned in the text are indicated. (E) Side view of one monomer of tRBP3 aligned with its structurally homologous proteins: the C-5 epimerase AlgE6 (magenta), the LKA tailspike gp49 (brown) and the tailspike K5 lyase A (light blue) in wire representation. (F) Side view of the tRBP3 in surface representation coloured according to the surface charge. (G) Side-bottom view of the tRBP3 represented as in (H). N- and C-termini are indicated as N and C, respectively. (H) Detailed view of the tRBP3 interchain negatively charged area in ribbon and transparent surface

representation. The surface is coloured according to the predicted charge and chains are coloured as in (A). Side chains of six acidic amino acids are represented as sticks in yellow and red. Glu68 and Asp97 belong to the orange chain and Glu25, Asp53, Asp83 and Asp109 belong to the green chain. A black arrow points to the protruding loop of the green chain.

4.2.4. RBP3 oligomeric state

To further clarify the oligomeric state of the RBP3 protein in solution, size exclusion chromatography was performed with protein at 5 mg/mL as specified in the Materials and methods section.

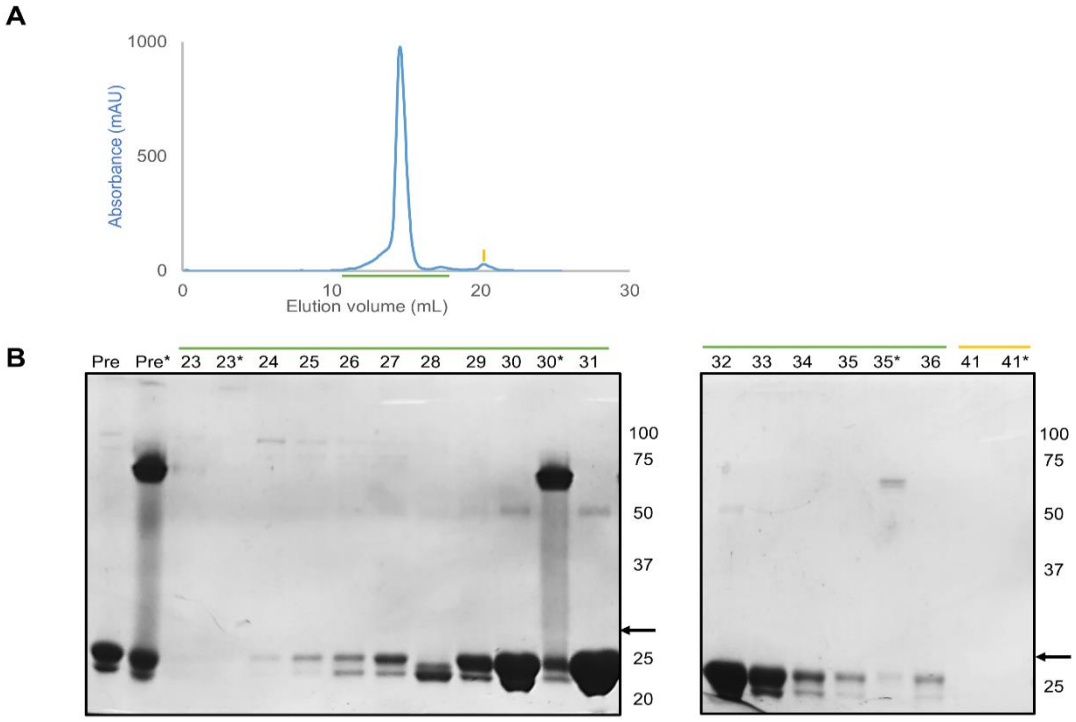


Figure 31: RBP3 SEC. (A) Size exclusion chromatography profile. Fractions loaded in the denaturing gel (B) are located on the green bar and below the orange bar. Numbers above the gel indicate the fractions from the chromatography, Pre stands for pre-loaded sample and the lane with an asterisk were not boiled at 95 ºC for 5 min. Black arrows in (B) indicate the weight where a monomer of RBP3 should be found, 28 kDa.

RBP3 mainly eluted in one peak after 14 mL (Fig. 31 A). This peak was distributed in fractions 24 to 33 with its absorbance maximum at fraction 30. The denaturing gel shows a minor band around 25 kDa, which probably correspond to a truncated version of a RBP3 monomer (Fig. 31A). In fact, the molecular weight of each resolved tRBP3 chain is 24.6 kDa. The presence of two bands in most of the lanes may indicate two truncated forms coexist in solution. Two samples of fraction 30 were loaded, one was boiled as the other samples and another was not. The non-boiled sample (lane 30*) shows a major band just below 75 kDa, a smear and a smaller band around 25 kDa. The upper band probably corresponds to the trimeric form of RBP3, while the lower contains denatured RBP3 monomers. As observed for other

tailspikes (Barbirz et al., 2009), RBP3 appears to remain trimeric in non-boiled samples. The smear in between probably corresponds to RBP3 trimers that monomerised and denatured during the electrophoresis at different times. Fractions 30, 31 and 32 also show a band around 50 kDa, which probably correspond to the contaminant present in Fig. 27C.

A 60-fold smaller peak after the main peak (Fig. 31A) also contains trimeric protein, as shown in the non-boiled lane corresponding to the fraction 35 (Fig. 31B). Another peak, 30 times smaller than the main one, eluted around fraction 41 does not appear to contain protein. This suggests that the main peak may contain dimers of trimers of RBP3 and the second peak trimers. The third peak probably corresponds to a minor impurity present in the sample.

5. Discussion

5.1. Structure of gp20 and RBP3 tailspikes

Tailspikes determine the host range of bacteriophages. They detect bacterial strains suitable for infection and adsorb to their LPSs. Most tailspikes also hydrolyse the O-antigen to allow the phage to approach the bacterial membrane. The β-helical fold is present in many tailspikes. It often contains an active site which can perform the endorhamnosidase reaction. Both the epsilon15 and the F358 tailspikes have a β-helical fold. They both have a groove containing acidic residues. The area surrounding these acidic residues is negatively charge, something characteristic of proteins with this activity. Nevertheless, gp20 and RBP3 β-helical domains are not exactly the same. Tailspikes are stable homotrimers. Upon trimerisation, they bury about a quarter or more of their surface; normally, by placing two of the three parallel β-sheets (PB) of a chain towards the other two chains. However, each monomer of a tRBP3 only has one PB facing the other chains.

5.2. Gp20 structure

A 13 Å-resolution cryo-EM map suggested an N-terminal arm working as a phage-binding domain and a main body joined by a kink. The crystallographic model of gp20ΔN fills the map after the kink. Processing of microscope images of the full-length tailspike revealed this kink is flexible (Fig. 25). Previous studies of epsilon15 structure (Chang et al., 2010; Jiang et al., 2006) show two out of six gp20 tailspikes interact with the edges of two two-fold symmetry sites. This interaction displaces the tailspikes about 20 Å with regard to the other four tailspikes. It is caused by a symmetry mismatch between the five-fold capsid and the six-fold, or more accurately, quasi six-fold, tail. My EM characterisation on free tailspikes shows the angle between the phage-biding domain (PBD) and the main body is variable. The flexibility of the kink goes from 40 to 170º. Just over half (54 %) of particles have an acute angle (40, 50 or 60º), 32% have an obtuse angle (140, 150, 160 and 170º) and two classes of particles remain in the middle with 10% (125º) and 4% (100º) of particles. A similar situation occurs in the bacteriophage phi29 (Farley et al., 2017; Xiang et al., 2006). Its appendages exist in two different conformations: extended and folded, obtuse and acute angle respectively. All appendages can switch between both conformations, but, due to steric hindrance, the folded one is preferred when the appendage is below a head fibre. In epsilon15, despite the lack of steric hindrance, tailspikes remain in the folded conformation. However, in free state, one third

of them are in extended conformation, angles 140º to 170º. This difference might come from the contacts of the tailspike with the phage particle. Besides the PBD contact with the tail, five out of six tailspikes bind to the capsid next to the kink (Fig. 32A-B). In addition, one petal domain of two tailspikes binds to the interface between the PBD and the tail (Fig. 32B). The addition of these interactions could reduce the flexibility seen in the free protein.

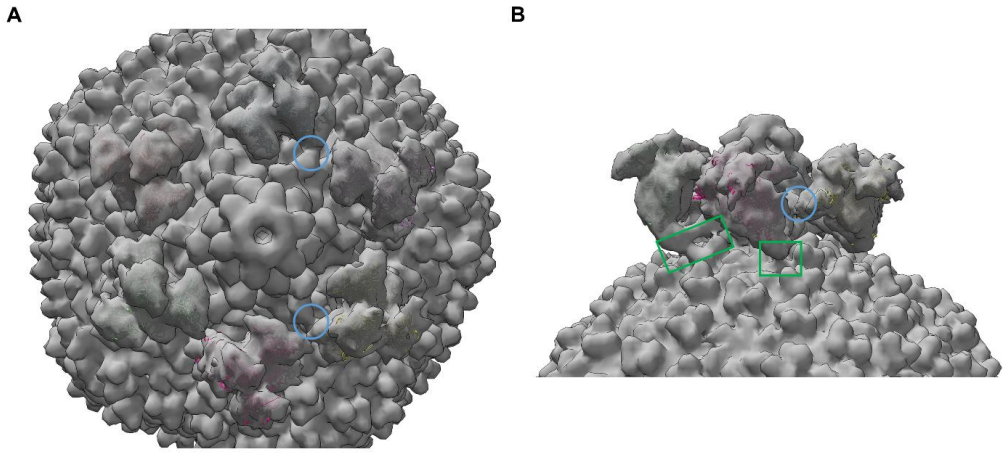


Figure 32: Gp20 contacts. (A) Bottom view of an epsilon15 phage map (EMDB entry: EMD-5209). (B) Side view of the map shown in (A). Blue circles mark the contacts between petal domains and PBD-tail interface. Green rectangles mark the contacts between PBDs and the capsid.

The construct gp20(287-1070) contains the complete parallel β-sheets (PB) of the β-helical domain (Fig. 2B). Compared to the crystallised gp20(248-1070) (Fig. 2A), it lacks the α-helix cap. Without it, a hydrophobic core (Leu289, Ile291, Phe295, Val297, Leu301, Ile303, Leu307 and Val 309) is exposed. This probably causes gp20(287-1070) constructs to aggregate and remain insoluble. This mechanism is similar to the formation of prion aggregates (Bryan et al., 2011; Spagnolli et al., 2019). Some of the initially designed constructs start in the middle of the β-helical domain. This is the case for gp20(434-1070) (Fig. 33C), which starts in the turn 2 of the sixth rung, and for gp20(520-1070) (Fig. 33D), starting in the T3 of the eighth rung. As in the previous case, they are probably insoluble because of exposing their hydrophobic core.

Two constructs are mainly focused on the petal domain, gp20(734-1070) and gp20(812-1070) (Fig. 33E-F). Of them, only the former could be expressed (Table 6). Gp20(812-1070) lacks the α-helix 1, β-strand 1 and β-strands 2 of the petal domain. The β1 and β2 strands form part of the core of the petal domain. The lack of these secondary structure elements seems to be enough to disrupt the production of the protein. In contrast, gp20(734-1070) comprises more than the petal domain. It also contains the joining linker, the last 3 β-strands and a 3_{10} helix-containing long loop of the β-sandwich domain. However, none of them appears in the crystal

structure. Both results together suggest that, in order to get soluble domains, it is better to leave residues than lack.

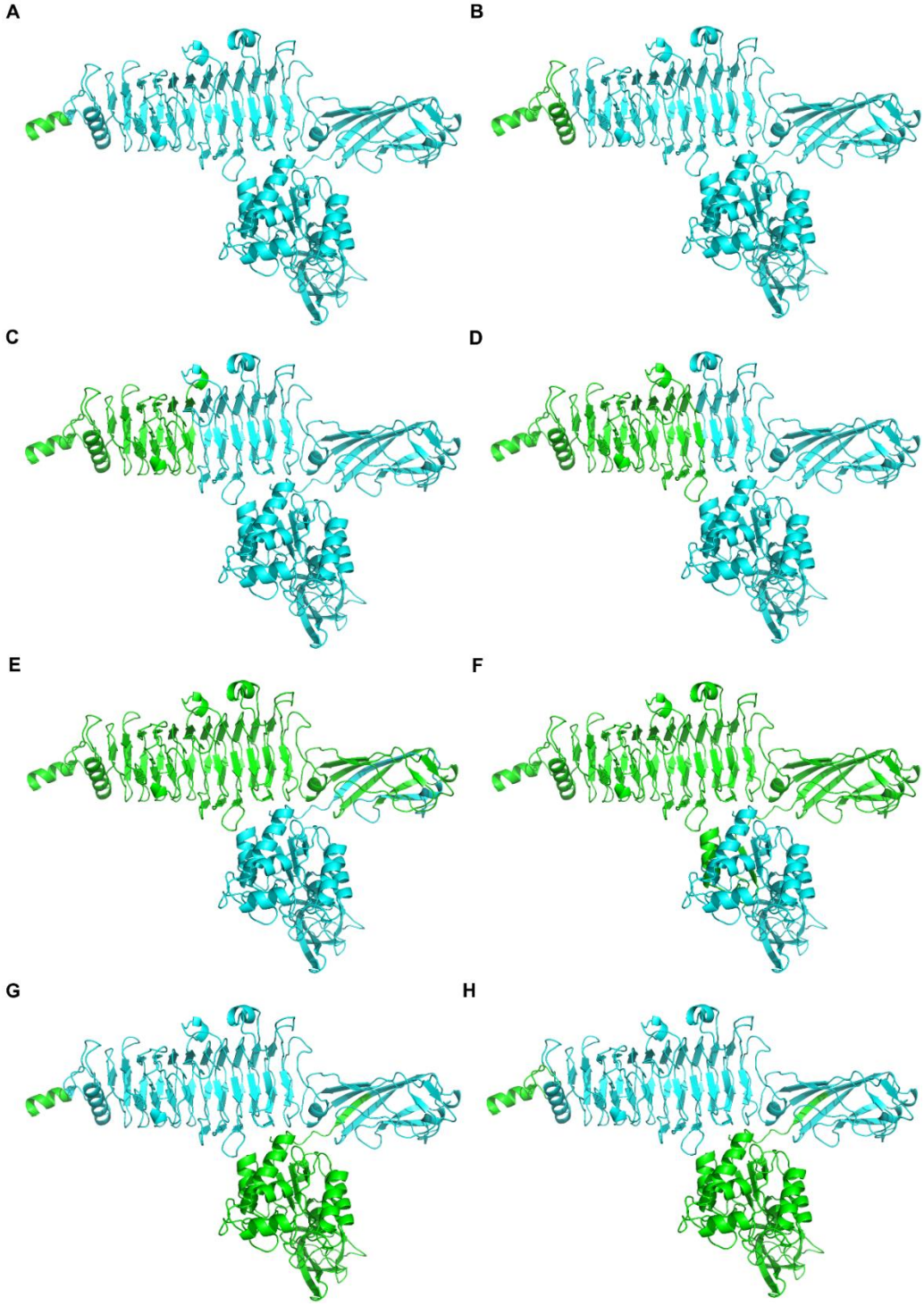


Figure 33: Extension of gp20 constructs. (A) Gp20(248-1070), (B) gp20(287-1070), (C) gp20(434-1070), (D) gp20(520-1070), (E) gp20(734-1070), (F) gp20(834-1070), (G) gp20(248-768) and (H) gp20(264-1070). Amino acids corresponding to the constructs are coloured cyan and the rest green. The N-terminal end is located in the left of the structures.

The constructs gp20(5-768), gp20(248-768) and gp20(264-768), called gp20ΔC, were insoluble too (Fig. 33G-H and Table 6). Gp20(264-768) lacks the β-strands presents in the cap,

which could reduce the solubility of the construct. In addition, the constructs miss the petal domain and the four residues of the last β-strands in the β-sandwich domain. In this case, I am not sure if the lack of the whole petal domain or the lack of those four residues of the last β-sandwich β-strand are causing the insolubility of gp20(5-768) and gp20(248-768), which start in a similar position to the soluble gp20(2-1070) and in the same amino acid as the crystallised construct, gp20(248-1070), respectively. The C-terminal end is essential for tailspikes and tail fibre trimerisation (Gage & Robinson, 2003; Mitraki et al., 1999; Tao et al., 1997). However, each C-terminal end of gp20 does not contact with other parts of the trimer. The insolubility of the gp20ΔC constructs triggers the doubt of if the petal domain starts folding gp20 or if it just covers hydrophobic patches in the β-helical domain that would cause aggregation when exposed. The highly homologous gp208ΔN from Det7 was successfully expressed and crystallised. Both proteins share a similar sequence. In the gp20 β-helix region facing the petal domain, hydrophobic amino acids not present in gp208 are only partially buried by petal residues. In addition, gp208 has two other hydrophobic amino acids facing outwards not present in epsilon15 gp20 (Fig. 34). This leads me to think that exposed hydrophobic patches are not causing aggregation in gp20ΔC constructs. Instead, the petal domain might work as an intra-molecular chaperone for gp20. Gp20ΔC constructs lack the last six residues of gp208ΔN – four from the β-strand of the β-sandwich domain and two additional flexible residues. The lack of these residues, which might disrupt the folding of the β-sandwich and, consequently, of the whole protein; may have been the difference between solubility and insolubility.

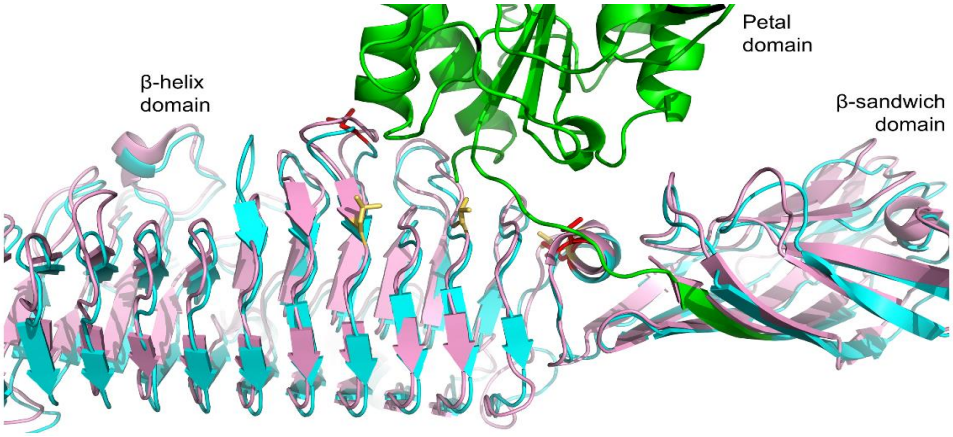


Figure 34: The interface between the petal domain and the β-helix domain. Detailed view of the β-helix and petal domains interface in g20ΔN (cyan and, from residue 768 onward, green) aligned with gp208ΔN (pink). Residues shown as sticks are coloured red for gp20 and yellow for gp208.

Despite the structural similarity between epsilon15 gp20 and Det7 gp208, there are some key differences. The largest one is the petal domain. It has a unique fold not previously seen. It is composed of an α/β hydrolase fold common to the SGNH superfamily and a β-barrel fold.

This extra domain brings a new carbohydrate-binding site (Fig. 20A). In addition, it also has the conserved SGNH catalytic triad formed by Ser787, Asp1035 and His1038. This triad together with the oxyanion residues Gly813 and Asn937 form the putative SGNH esterase site. Both the binding site and the esterase site are lodged in a groove between the α/β subdomain and the β-barrel. As the closest ester group is 12 Å away from the catalytic serine (Fig. 20A), the acetyl-containing galactose should rotate to approach the catalytic serine. Although is unknown if petal domain has esterase activity, a large polysaccharide could access its groove, where acetyl groups would be removed. G7C phage and Vi phages tailspikes possess esterase activity but do not hydrolyse the O-antigen chain. They probably use that activity to reduce the interactions between chains and to ease the approach of the phage particle to the membrane. Previous studies (Hagiwara et al., 1966; Wright, 1971) show epsilon15 phage could adsorb and infect bacteria without the galactose-bound acetyl group. This makes it clear that the presence of the acetyl group and its hydrolysis is not necessary for phage infection, as it is in phage G7C. However, the acetyl group could still be useful for the bacterium to help impede penetration of the bacteriophage through the O-antigen chains.

None of the two characterised Det7 tailspikes, gp207 and gp208 (Broeker et al., 2019; Walter et al., 2008) have been reported to possess esterase activity. A search in the Conserved Domain Database - https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi (Lu et al., 2020) and HHpred - https://toolkit.tuebingen.mpg.de/tools/hhpred (Zimmermann et al., 2018) does not show homologies with esterase conserved sites for the other two tailspikes. Nevertheless, phage Det7 is able to infect the strain *S. enterica subsp. enterica* serovar Anatum.

Another possibility is that petal domain has lost its esterase activity and only works as a binding site to ensure a better hold of the O-antigen chains while the β-helix domain hydrolyses them. Other α/β hydrolases, such as the CtCE2 of *Clostridium thermocellum* (PDB entry 2WAO), also bind carbohydrates near the conserved catalytic triad (Fig. 35A-D). The petal domain has a deeper groove than CtCE2 and the bound carbohydrate orientation is different. Both characteristics might be conditioned by its putative binding function. The bound O-antigen oligosaccharide binds to the petal domain with its non-reducing end pointing to the distal terminus of gp20 (Fig. 35E). The bacterium and the rest of the O-antigen are in this direction. It reinforces the idea that petal domain groove has, at least, a binding activity.

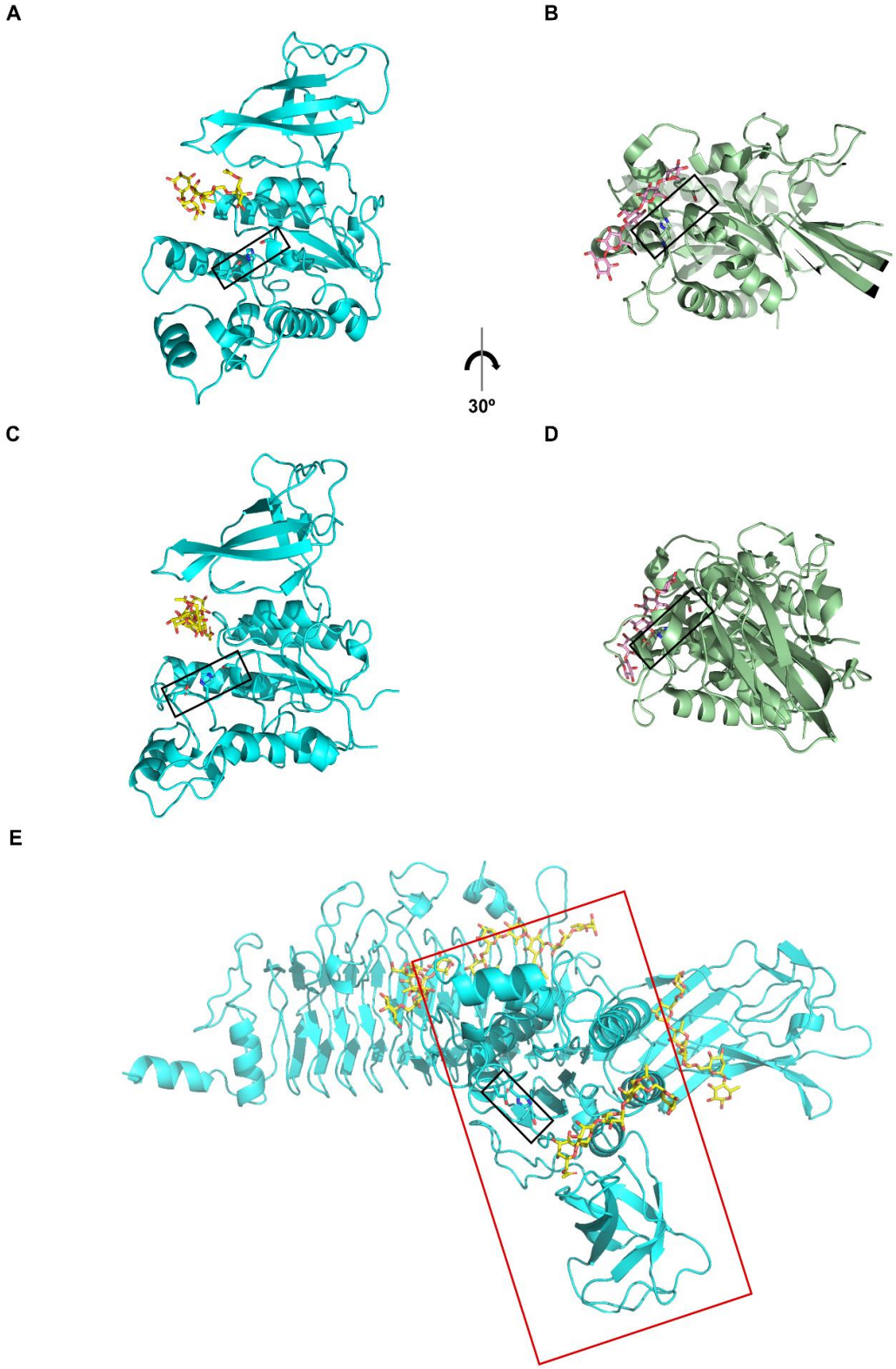


Figure 35: Anatum O-antigen binding by the petal domain. (A and C) Ribbon view of the petal domain (cyan) with Anatum oligosaccharide (yellow and red). The reducing end is closest to the viewer. (B and D) Ribbon view of the CtCE2 of *Clostridium thermocellum* (green) bound to a hexacellulose saccharide (pink and red). Each CtCE2 is aligned to the petal domain on its left. (E) Ribbon view of a gp20 monomer (cyan) with four fragments of the O-antigen (yellow and red). Distal end is on the right. In all panels, black rectangles frame the catalytic triad. The red rectangle frames the petal domain.

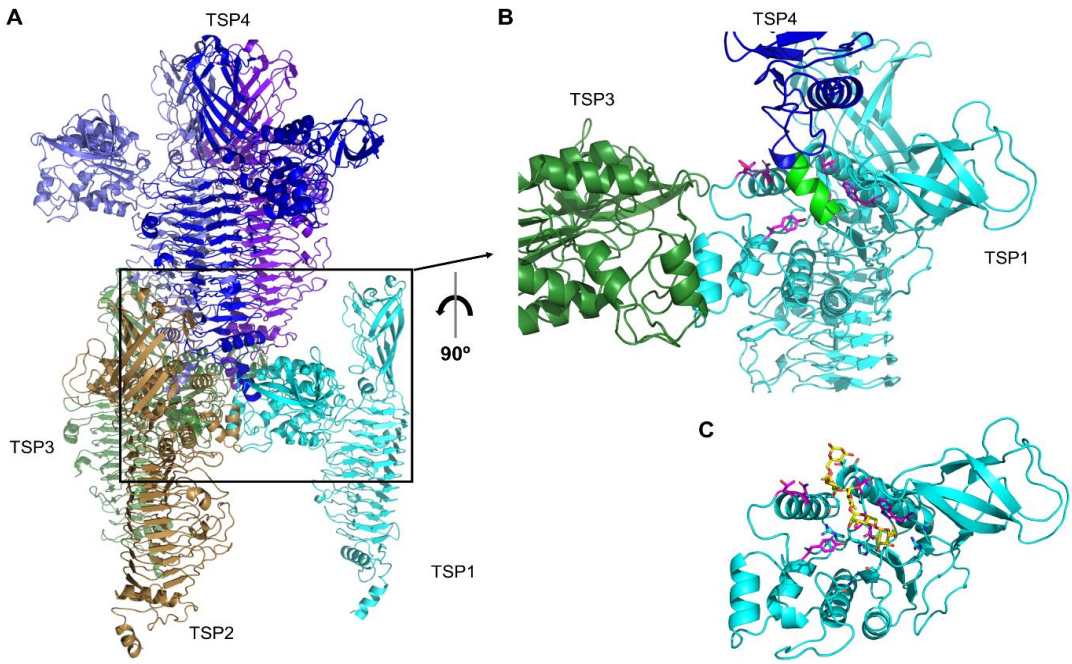


Figure 36: Gp20ΔN crystal packing. (A) Ribbon representation of four gp20ΔN tailspikes in their crystal packing. For the lower TSPs, only the chain contacting TSP4 is shown. Each tailspike is shown in one colour and, for the TSP4, each chain has a different tone of blue. (B) A close-up view of the region indicated by a rectangle in (A). For clarity, TSP2 was removed. It would be between the TSP1 and the viewer. The his-tag of TSP4 chain is coloured green. (C) Petal domain structure with its bound oligosaccharide. Amino acids contacting the TSP4 N-terminal end in (B) and the oligosaccharide in (C) are depicted in magenta.

The Anatum O-antigen is present in the petal domain crystal structure, but not in the petal domain of the gp20ΔN crystal structure soaked in hexasaccharide. The reason for this difference is the way gp20ΔN tailspikes are packed inside the crystal. One tailspike is held by other three tailspikes forming a pyramidal tetrahedron (Fig. 36A), similar to the crystal shape (Fig. 16A). Each tailspike makes contacts with other tailspikes in four regions: the N-terminal end and each of the petal grooves. The N-terminal α-helix is made by the first amino acids of the truncated gene and the last amino acids of the his-tag. Each of the three N-terminal helices interacts with the groove of one petal domain from three different tailspikes (Fig. 36B). The oligosaccharide bound to the petal domain structure binds exactly in the same position (Fig. 36C). Consequently, the hexasaccharides added during the soaking experiments could not reach their binding site.

The comparison between epsilon15 gp20 and Det7 gp208 suggests how new tailspikes may be created through the addition of new domains with new functions. If gp20 were to be adapted to a new host with a capsule, the petal domain could be exchanged for a capsule-degrading domain with similar interactions with the β-helix domain. The gp20 gene could have appeared from gp208 gene in a similar way in nature. The full petal domain sequence, i.e. both

subdomains, is homologous to the C-terminal region of the *Sodalis paecaptivus* prophage tail fibre N-terminal domain-containing protein sequence (NCBI Reference Sequence: WP_025423026.1). *Sodalis paecaptivus* is an opportunistic pathogen found in a human wound. The mentioned protein contains the GHSM motif, similar to the GHSF in gp20, instead of the canonical GDSL motif of the SGNH family. No other protein belonging to this family was found to have a histidine instead of an acid residue in this motif. However, the significance of this amino acid is unknown.

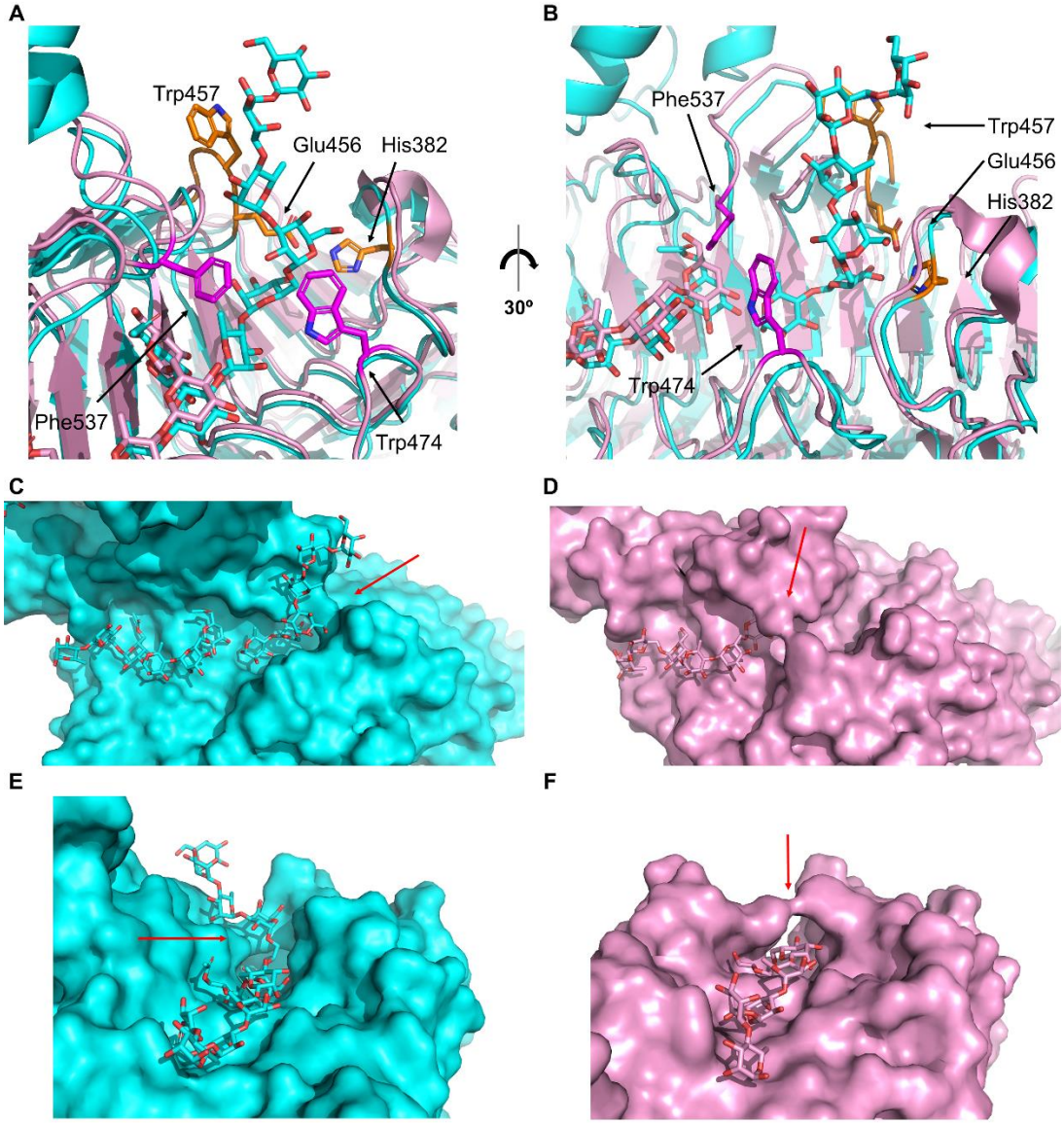


Fig 37: Gp20 and gp208 arches and the blocking loop. (A and B) Close-up views of gp20 (cyan) and gp208 (PDB entry 6F7K) (pink) in ribbon view and its respective ligands in stick view. Gp20 arch and loop amino acids are coloured orange and gp208 arch amino acids are coloured magenta. (C and E) Surface view of gp20 with its ligands in stick conformation; all coloured as in (A). (D and F) Surface view of gp208 with its ligands in stick conformation; all coloured as in (A). Red arrows indicate de position of the arches in each protein.

Another difference between gp20 and gp208 is the presence of di-amino acid "arches" in different zones near the endorhamnosidase site. In gp20, an arch is formed by a His382-Glu456 salt bridge at a distance of 13 Å from the catalytic site (Fig. 37 C and E). Glu456 is located in a loop in the T3 between rungs five and six. Next to it, Trp455 interacts with Man2 of Hex1 by π stacking. This arch blocks the groove, because the tunnel under it is only 6.3 by 8.0 Å large. A deacetylated galactose is up to 6.5 Å wide, so would not be able to pass through. In gp208, the arch is made of Trp474 and Phe537, which might be forming a π-stacked bond. Its dimensions are 10.9 x 12.4 Å, much wider than in gp20. In addition, the endorhamnosidase site is located just below it. Hence, the hydrolysed hexasaccharide probably goes below the arch right before being cut. Then, it can freely leave the catalytic site through the groove. Instead of the blocking loop, gp208 has a small turn between rungs five and six in the position T3. The loop in gp20 might increase the binding surface with Hex1 and be responsible for the presence of Hex1 in the crystal structure.

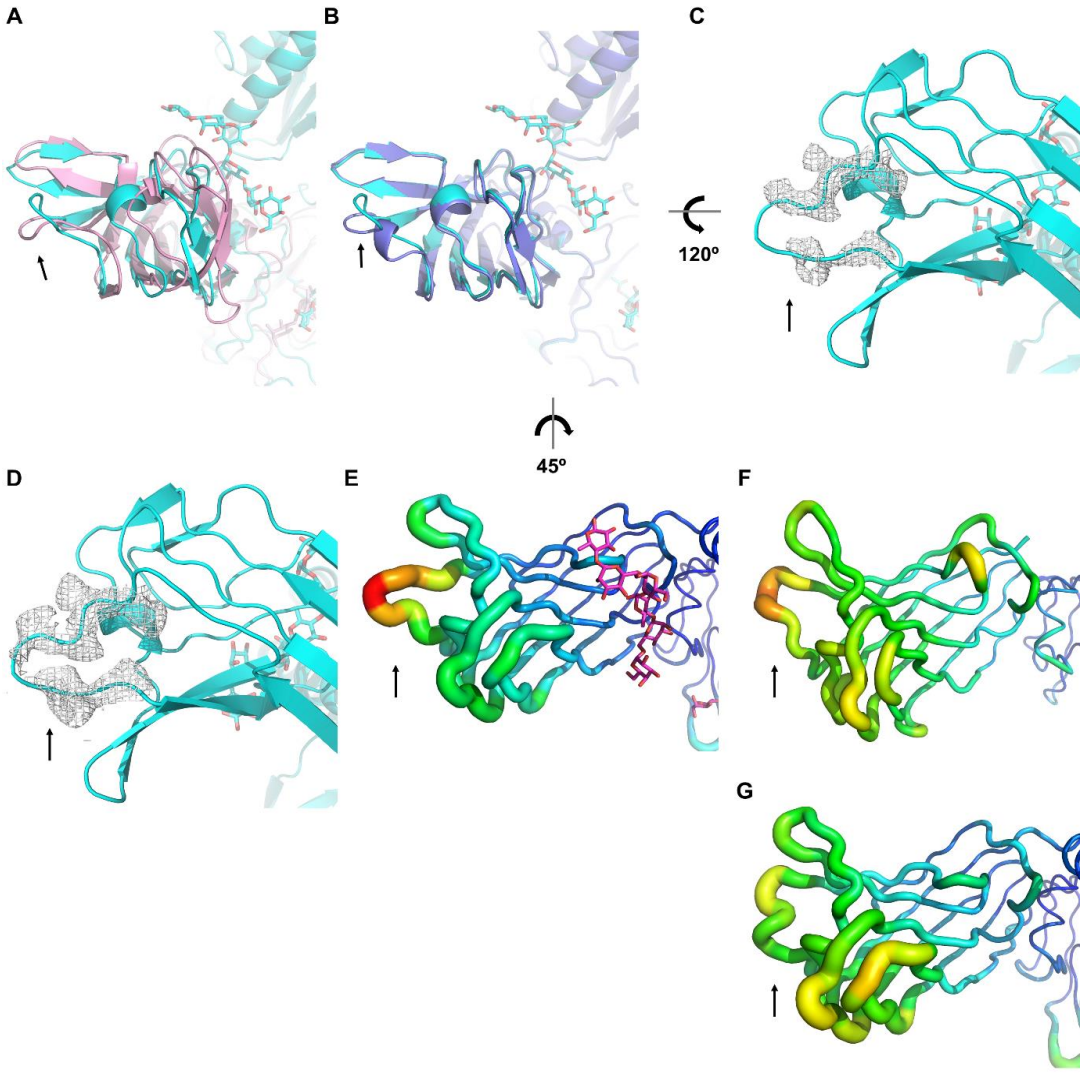


Figure 38: β-sandwich domain. (A) Distal end view of a gp20 complexed with hexasaccharide (cyan) and gp208 (PDB entry 6F7K) (pink) superposition. (B) Superposition of gp20 complexed with hexasaccharide (cyan) and apo-gp20 (blue). (C and D) Close up view of the distal β-sandwich. The 2Fo-Fc electron density map of loop 720-727 contoured at 1 σ (C) and 0.5 σ (D) is shown as a grey mesh. Shown structures are rotated 120 º as indicated. (E-G) Representations highlighting the temperature factors view of gp20 complexed with hexasaccharides, gp208 and apo-gp20 respectively. High values are indicated as red and low values as blue. A black arrow points to the loop formed by residues 720-727 in all panels.

The binding site in the β-sandwich domains of both proteins have similar structures and even share most of the binding amino acids. The presence of an O-antigen hexasaccharide in gp20 might come from the higher concentration used in gp20 crystals, 10 mM vs 5 mM in gp208. Nevertheless, this difference is not that large if compared with the range used in soaking experiments of other β-helical structures: from 2 mM of a *Salmonella sp.* octasaccharide (PDB entry 1TYX) to 570 mM of nigerose (PDB entry 6K0N). The differences between β-sandwich domains of gp20 and gp208 appear in the distal end. Distal loops and some β-strands of gp208 are slightly moved with regard to the gp20 ones. The most different region is the loop formed by residues 720-727 (Fig. 38A). Surprisingly, this loop also has a different conformation in the gp20 apo-protein (Fig. 38B). In the ligand structure, the Met724 2Fo-Fc electron density is missing at 1 σ (Fig. 38C), but it is partially visible at 0.5 σ (Fig. 38D). The temperature factor representation of the three β-sandwich domains shows higher values for that loop with regard to the rest of the domain (Fig. 38E-D). Higher flexibility and different conformations at the distal end of gp20 might be related to putative secondary receptor binding in the host membrane.

Apart from the three differences mentioned above, gp20 and gp208 are highly homologous. Both protein share the amino acids located around the endorhamnosidase site (Fig. 39); among them, the mutated amino acids already mentioned in the section 4.1.6 (Fig. 26B). This degree of conservation suggests that they are key residues for tailspike function. The mutation of Asp449 or Asp507 to Asn causes loss of endorhamnosidase activity in gp20ΔN (Fig. 26C). These aspartates are 5.3 Å away from each other, the typical distance for the general acid and the nucleophile of retaining glycosidases (Davies & Henrissat, 1995; Zechel & Withers, 2000). This is similar to what happens with the endorhamnosidase active sites of the P22 tailspike (Andres et al., 2010; Baxa et al., 1996), Sf6 tailspike (Müller et al., 2008) and CBA120 TSP2 (Plattner et al., 2019). However, while Sf6 and CBA120 tailspikes have their binding and catalytic sites in an interchain groove; the P22, epsilon15 and Det7 gp208 tailspikes catalytic sites are intrachain.

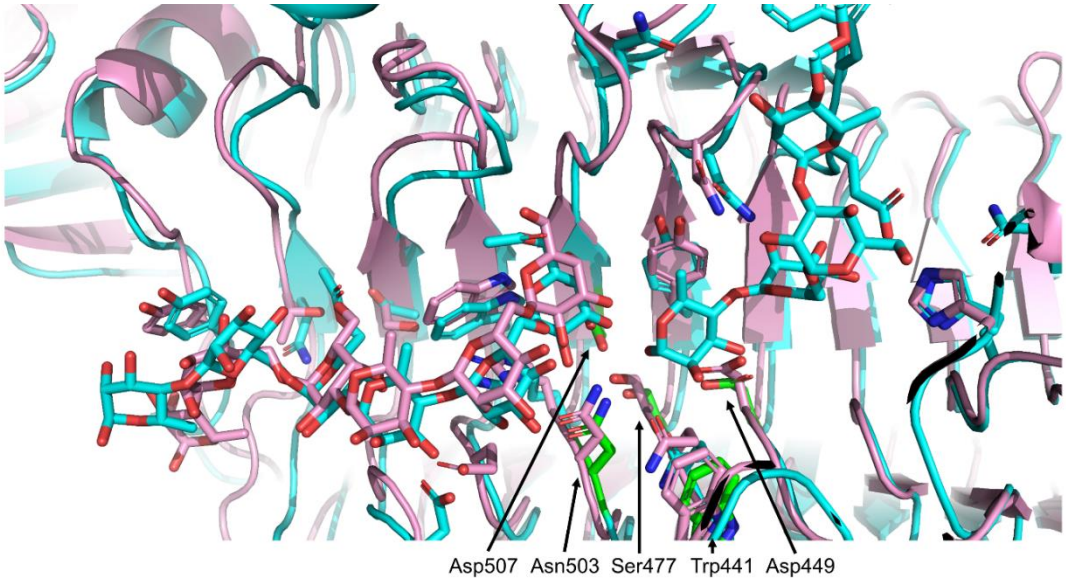


Figure 39: catalytic site. (A) Close-up view of the aligned catalytic sites of gp20 (cyan) and gp208 (pink). Residues contacting the ligands are depicted in stick conformation. Gp20 mutated residues are coloured green and pointed out with arrows.

Mutation of Asn503 to aspartic acid also supresses catalytic activity. The Nδ2 atom of Asn503 faces the catalytic site and has hydrogen bonds with the carboxylic oxygen of Asp507, the hydroxyl of Ser477, the Hex1 Rha6 O1 and the Hex2 Gal1 O3. Its positive charge could stabilise this acidic environment. Consequently, its exchange for an acidic residue could modify the hydrogen bond network, disrupting the catalytic site and/or repel the substrate. The Ser477 to Ala mutant also reduces the endorhamnosidase activity to 39%. Ser477 establishes hydrogen bonds with the Asp507 and Asn503. The lack of this residue may also modify the hydrogen bond network as the mutant Asn503 to Asp.

The last mutant in the catalytic area, W441R, reduces the endorhamnosidase activity by about one half. This mutant was present in the epsilon15 variant called PLS4 (Point Loma Sewage #4). This mutant is able to infect both *Salmonella* serovar Anatum and the epsilon15 lysogenic bacterium (Michael McConnell, personal communication). The former has an α-O-glycosidic bond between trisaccharides and the latter a β-O-glycosidic bond. This single mutation allows the phage to widen its host range. However, it also reduces the endorhamnosidase activity. Trp441 does not directly interact with any of the saccharides (Fig. 39). The more flexible Arg residue, when compared to Trp, probably allows the accommodation of both α- and β-linked oligosaccharides. At the same time, the increased local flexibility apparently affects the efficiency of the endorhamnosidase activity.

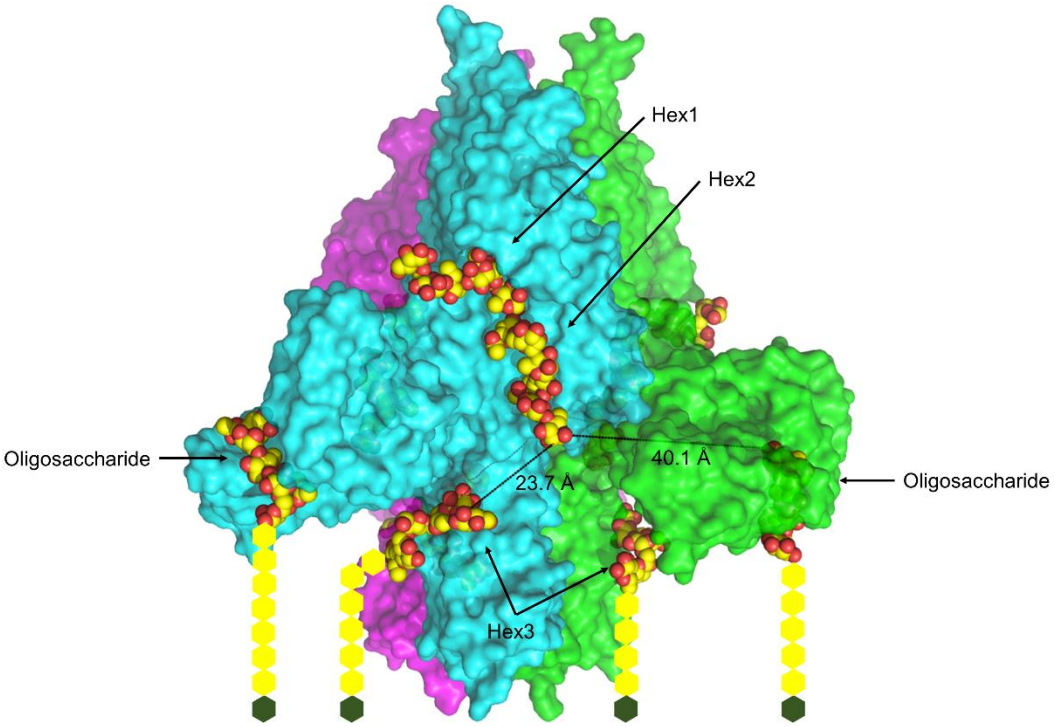


Figure 40: trimeric gp20. Surface representation of the gp20ΔN tailspike with each chain in one colour. The Anatum O-antigen saccharides are depicted as yellow (carbon atoms) and red (oxygen atoms) spheres. Black lines represent the distances from Hex3 and the oligosaccharide to the Hex2 Rha6. Yellow hexagons represent the rest of the O-antigen chain until the core-oligosaccharide, represent as a dark green hexagon.

As in the petal domain, the hexasaccharide reducing ends, Rha6, in the β-helix domain and the β-sandwich point towards the distal end of the protein. Therefore, they point to the LPS origin in the bacterial membrane. This is the case also for the hexasaccharide in gp208ΔN. The common direction of all oligosaccharides suggests they are not artefacts of crystallisation. This, together with the interface of Hex1 and Hex2 located on the catalytic site, suggests their positions in the crystal structure coincide with the positions during the adsorption step. The distance between Gal1 O3 of Hex3 to Rha6 O1 of Hex2 is 23.7 Å (Fig. 40). It is comparable to the distance of the modelled hexasaccharides, which ranges from 22.1 to 24.5 Å Å. This means another hexasaccharide could fit in between Hex 2 and Hex3. In contrast, the distance between Hex2 Rha6 O1 of one chain and the oligosaccharide Gal1 O3 of the next chain is 40.1 Å (Fig. 40). This straight line goes through the petal domain. A dodecasaccharide could surround the petal domain and connect those ends, although with notable turns. The connection between Hex2 and Hex3 through another hexasaccharide seems more straightforward; otherwise, the polysaccharide chain going to Hex2 would cross with the one continuing from Hex3.

To my knowledge, this is the only tailspike with four ligands bound to different binding sites per chain (Fig. 40). The closest example is G7C gp36.1. One O-antigen binding site was

located in the crystal structure and other two carbohydrate-binding sites were predicted by structural homology. In addition, if my model depicted in Fig. 40 is right, it would be the only known tailspike which binds more than one LPS chain per monomer. The implications of a higher binding capacity are unknown, but it may well allow a stronger binding of the phage particle to the host bacterium.

5.3. RBP3 structure

Expressed RBP3 protein crystallises in two oligomeric states, as a trimer and as a monomer. Known structures of proteins mainly composed of β-helical folds can be trimers or monomers, but not both. All known phage tailspikes are homotrimers while bacterial carbohydrate modifying enzymes are monomers (for examples, see PDB entries 2PYG and 6KFN). Tailspikes often have C-terminal domains that help in the trimerisation process. However, RBP3 counts only with the β-helical domain, as do monomeric bacterial enzymes. The N-terminal protruding loop increases the contacts between neighbouring chains and might stabilise the trimer. As this is also the putative region that links RBP3 to the rest of the phage, the contacts with the rest of the phage may help to impose and maintain the trimer state. The gp49 tailspike from *Acinetobacter baumannii* bacteriophage Fri1 (PDB entry 6C72) also has a protruding loop that increases the contacts between subunits (Fig. 41).

Figure 41: Protruding loop. Tailspike protein gp49 from Acinetobacter baumannii bacteriophage Fri1 (PDB entry: 6C72). Each chain is coloured differently, the loop 336-364 from the green chain is coloured orange.

Size exclusion chromatography (SEC) shows the existence of one main oligomeric state in solution (Fig. 31A). The non-boiled samples run in a denaturing gel showed this oligomeric state is probably a trimer, as monomeric bands are probably only caused by heating and denaturing conditions (Fig. 31B). Therefore, crystallisation of RBP3 protein as monomers might be an artefact. The high concentration of ammonium sulphate in the crystallisation

condition used to obtain monomeric RBP3 may have disrupted interchain bonds in the trimer core.

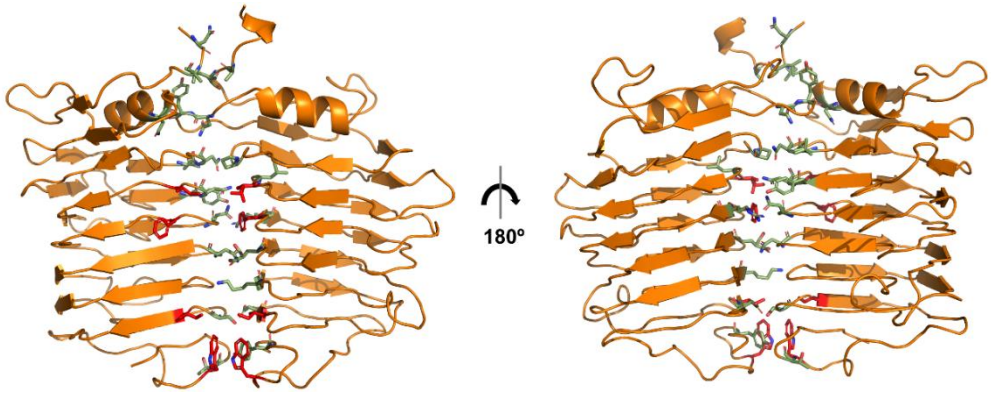


Figure 42: Tailspikes core. Ribbon representation of the two RBP3 monomers coloured orange. The frontal chain was removed in both panels for clarity. The amino acids establishing hydrogen bonds between two tRBP3 chains (green) and the hydrophobic core residues (red) are shown in stick representation.

In the tRBP3, the β-helix of one monomer makes fourteen hydrogen bonds with a neighbouring chain in its seven rungs. In total, they involved the side chains or the main chains of 21 residues (Asn15, Ala17, Leu19, Asn46, Asn50, Tyr51, Val73, Asn74, Pro78, Tyr103, Asn104, Leu107, Asn130, Arg135, Asn155, Lys185, Ser207, Asp211, Thr232 and Tyr236) (Fig 42). The core region of each tRBP3 monomer has four hydrophobic residues interacting with the other chains (Ile106, Phe133, Ile209 and Trp231) (Fig. 42). Together, they keep tRBP3 trimeric in solution. However, the ammonium sulphate present in the crystallisation solution might disrupt the hydrophilic interactions in the tRBP3 core. The four hydrophobic contact residues might not be enough to counteract this effect in the crystallisation condition.

The literature about *Campylobacter* phage RBPs is scarce. The members of the family *Fletchervirus* commonly need the presence of MeOPN to adsorb to and infect their host. Probably, MeOPN would form part of their receptor in *C. jejuni* bacteria. F358 phage belongs to the *Fletchervirus* family. It infects the *C. jejuni* NCTC12662 strain, whose capsular polysaccharides (CPS) contain MeOPN moieties. Apart from the fact that RBP2 allows phage F358 to infect a MeOPN deletion mutant strain (Martine Sørensen, personal communication), not much is known about the binding specificities of RBP1, 2, 3 and 4. Hence, some of the other RBPs may be involved with MeOPN binding. Solving the structure of the RBPs together with their receptors will shed light on the function of each of the four tailspikes.

Given the thickness of the capsule, bacteriophages would need a mechanism to go through it and reach the bacterial membrane. As for phages infecting bacteria with smooth LPSs, F358 will probably need to digest the capsule wall to approach the membrane. RBP3

might have catalytic activity, but the specific residues are unknown. F358 can still infect its host strain when antibodies individually block its RBPs. Hence, either the phage performs the infection without hydrolysing the CPS or more than one RBP has catalytic activity.

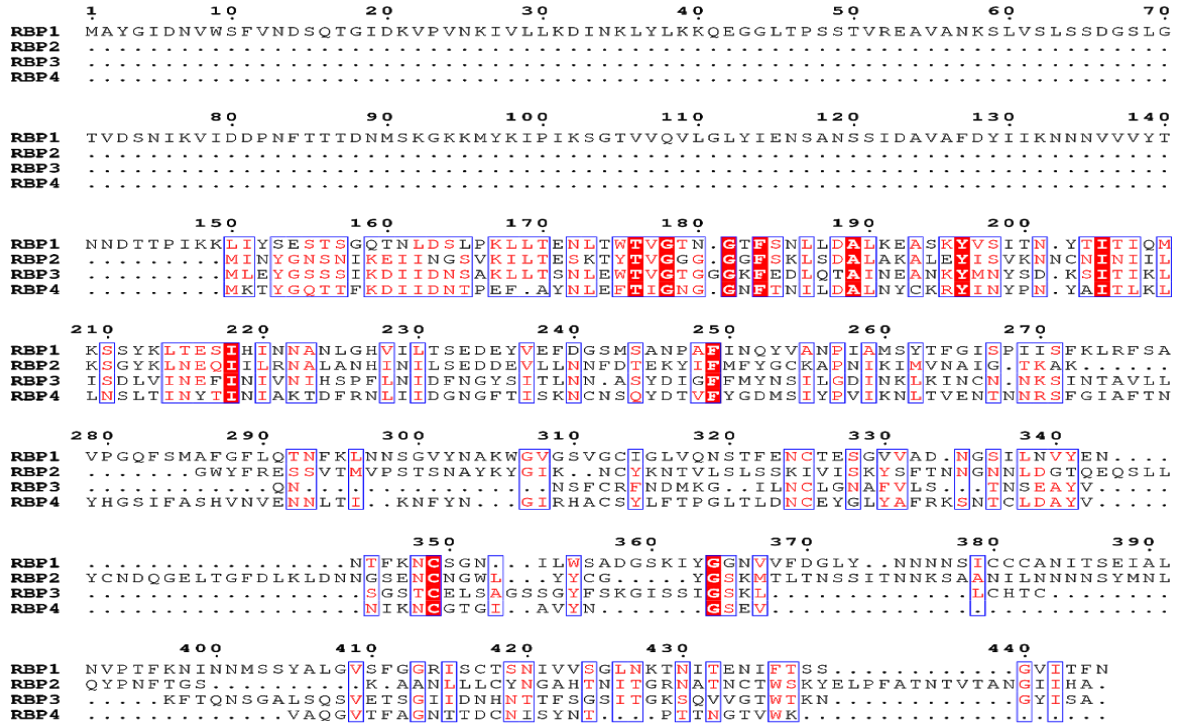


Figure 43: Sequence comparison of the four F358 RBPs. Amino acids shadowed in red are conserved in the four proteins. When conserved in three proteins, amino acids letters are coloured red. Blocks of conserved amino acids are inside a blue square. RBP1 sequence numbering is followed.

A sequence comparison shows all four RBPs have a similar region, which corresponds to the first 70 amino acids in RBP3 (Fig 43). This region is the β-α cap, the protruding loop and the first rung of the β-helix. In addition, a structure prediction search (HHpred; Zimmermann et al., 2018) with the RBPs revealed homologies to β-helix-containing proteins in all cases. These results suggest the four RBPs of F358 have a β-helix domain, which is present in many tailspikes and often present catalytic activity (Seoane-Blanco et al., 2021).

Sequence comparisons also reveal the existence of an N-terminal domain in RBP1, the longest of the four RBP. In phage CBA120, the N-terminal region of the longest of its tailspikes, TSP4, binds to the baseplate and serves as anchoring site for TSP1 and TSP2 (Plattner et al., 2019). In F358, RBP1 could perform the same function. However, it is still unknown if all RBPs are present in the same phage particle, as they are in phages CBA120 and K1-5 (Scholl et al., 2001). Structure determination by cryo-EM of the baseplate could help to unravel this question.

6. Conclusions

1) The *Salmonella virus epsilon15* gp20 protein forms a trimeric tailspike arranged in four domains: a stalk-like phage-binding domain, a β-helix domain, a β-sandwich domain and a petal domain.

2) The bend between phage-binding domain and the β-helix domain is flexible and hinders crystallisation assays.

3) The β-helix and the β-sandwich domains of gp20 are highly structurally homologous to the Dettilon tailspike.

4) The petal domain of gp20 is an α/β module made of two subdomains: an α/β hydrolase and a β-barrel and is, so far, unique to epsilon15.

5) The aspartic acids in positions 449 and 507 form the gp20 endorhamnosidase catalytic site in a negatively charged area. They hydrolyse the *Salmonella enterica subsp. enterica* serovar Anatum O-antigen following the retaining mechanism.

6) The *Salmonella virus epsilon15* gp20 protein has four O-antigen binding sites. One in the petal domain, another in the β-sandwich domain and two in the β-helix domain. The latter two are located just before and after the endorhamnosidase site.

7) The *Campylobacter virus F358* tailspike RBP3 protein is a short β-helix and crystallises both in monomeric and trimeric state. The monomeric state is probably an artefact caused by the crystallisation condition. The trimeric structure resembles to that of other phage tailspikes and the monomeric one looks like carbohydrate-modifying bacterial enzymes.

8) The *Campylobacter virus F358* trimeric RBP3 has a negatively charged area in its interchain groove, which might contain a catalytic site for capsular polysaccharide.

Conclusiones

1) La proteína gp20 de *Salmonella virus epsilon15* forma una espícula de la cola trimérica que está dispuesta en cuatro dominios: un dominio de unión al fago, un dominio hélice β, un dominio sándwich β y un dominio pétalo.

2) La doblez entre el dominio de unión al fago y el dominio hélice β es flexible y dificulta los ensayos de cristalización.

3) El eje formado por los dominios hélice β y el sándwich β de gp20 es estructuralmente muy homólogo a la fibra de la cola Dettilon.

4) El dominio pétalo de gp20 es un módulo α/β formado por dos subdominios: un subdominio hidrolasa α/β y otro barril β.

5) Los ácidos aspárticos en posiciones 449 y 507 de gp20 forman el sitio catalítico endoramnosidasa en un área cargada negativamente. Estos hidrolizan el antígeno O de *Salmonella enterica subsp. Enterica* serovar Anatum mediante un mecanismo de retención.

6) Según los ensayos cristalográficos, la proteína gp20 de *Salmonella virus epsilon15* tiene cuatro sitios de unión. Uno en el dominio pétalo, otro en el dominio sándwich β y dos en el dominio hélice β, antes y después del sitio endoramnosidasa.

7) La proteína RBP3 *Campylobacter virus* F358 es una hélice β corta y cristaliza tanto en estado monomérico como en estado trimérico. La estructura trimérica se asemeja a las fibras de la cola de otros bacteriófagos y la monomérica, a enzimas bacterianas que modifican carbohidratos.

8) La proteína trimérica RBP3 de *Campylobacter virus F358* tiene una zona cargada negativamente en su surco intercatenario, que puede contener un sitio catalítico para el polisacárido de la cápsula.

7. References

Abrahams, J. P., & Leslie, A. G. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallographica. Section D, Biological Crystallography*, *52*(Pt 1), 30–42. https://doi.org/10.1107/S0907444995008754

Abrishami, V., Zaldívar-Peraza, A., de la Rosa-Trevín, J. M., Vargas, J., Otón, J., Marabini, R., Shkolnisky, Y., Carazo, J. M., & Sorzano, C. O. S. (2013). A pattern matching approach to the automatic selection of particles from low-contrast electron micrographs. *Bioinformatics*, *29*(19), 2460–2468. https://doi.org/10.1093/bioinformatics/btt429

Ackermann, H.-W., & Prangishvili, D. (2012). Prokaryote viruses studied by electron microscopy. *Archives of Virology*, *157*(10), 1843–1849. https://doi.org/10.1007/s00705-012-1383-y

Ageno, M., Donelli, G., & Guglielmi, F. (1973). Structure and physico-chemical properties of bacteriophage G. II, The shape and symmetry of the capsid. *Micron (1969)*, *4*(4), 376–403. https://doi.org/10.1016/0047-7206(73)90002-2

Aidley, J., Sørensen, M. C. H., Bayliss, C. D., & Brøndsted, L. (2017). Phage exposure causes dynamic shifts in the expression states of specific phase-variable genes of Campylobacter jejuni. *Microbiology*, *163*(6), 911–919. https://doi.org/10.1099/mic.0.000470

Akoh, C. C., Lee, G.-C., Liaw, Y.-C., Huang, T.-H., & Shaw, J.-F. (2004). GDSL family of serine esterases/lipases. *Progress in Lipid Research*, *43*(6), 534–552. https://doi.org/10.1016/j.plipres.2004.09.002

Almpanis, A., Swain, M., Gatherer, D., & McEwan, N. (2018). Correlation between bacterial G+C content, genome size and the G+C content of associated plasmids and bacteriophages. *Microbial Genomics*, *4*(4). https://doi.org/10.1099/mgen.0.000168

Andres, D., Hanke, C., Baxa, U., Seul, A., Barbirz, S., & Seckler, R. (2010). Tailspike Interactions with Lipopolysaccharide Effect DNA Ejection from Phage P22 Particles *in Vitro*. *Journal of Biological Chemistry*, *285*(47), 36768–36775. https://doi.org/10.1074/jbc.M110.169003

Anthon, G. E., & Barrett, D. M. (2002). Determination of Reducing Sugars with 3-Methyl-2-benzothiazolinonehydrazone. *Analytical Biochemistry*, *305*(2), 287–289. https://doi.org/10.1006/abio.2002.5644

Arias, C. A., & Murray, B. E. (2009). Antibiotic-Resistant Bugs in the 21st Century—A Clinical Super-Challenge. *New England Journal of Medicine*, *360*(5), 439–443. https://doi.org/10.1056/NEJMp0804651

Bachtiar, B. M., Coloe, P. J., & Fry, B. N. (2007). Knockout mutagenesis of the kpsE gene of Campylobacter jejuni 81116 and its involvement in bacterium–host interactions. *FEMS Immunology & Medical Microbiology*, *49*(1), 149–154. https://doi.org/10.1111/j.1574-695X.2006.00182.x

Bacon, D. J., Szymanski, C. M., Burr, D. H., Silver, R. P., Alm, R. A., & Guerry, P. (2001). A phase-variable capsule is involved in virulence of Campylobacter jejuni 81-176. *Molecular Microbiology*, *40*(3), 769–777. https://doi.org/10.1046/j.1365-2958.2001.02431.x

Baker, M. L., Hryc, C. F., Zhang, Q., Wu, W., Jakana, J., Haase-Pettingell, C., Afonine, P. V., Adams, P. D., King, J. A., Jiang, W., & Chiu, W. (2013). Validated near-atomic resolution structure of bacteriophage epsilon15 derived from cryo-EM and modeling. *Proceedings of the National Academy of Sciences*, *110*(30), 12301–12306. https://doi.org/10.1073/pnas.1309947110

Baldvinsson, S. B., Sørensen, M. C. H., Vegge, C. S., Clokie, M. R. J., & Brøndsted, L. (2014). Campylobacter jejuni Motility Is Required for Infection of the Flagellotropic Bacteriophage F341. *Applied and Environmental Microbiology*, *80*(22), 7096–7106. https://doi.org/10.1128/AEM.02057-14

Barbirz, S., Becker, M., Freiberg, A., & Seckler, R. (2009). Phage Tailspike Proteins with β-Solenoid Fold as Thermostable Carbohydrate Binding Materials. *Macromolecular Bioscience*, *9*(2), 169–173. https://doi.org/10.1002/mabi.200800278

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R., & Leslie, A. G. W. (2011). iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta*

Crystallographica. Section D, Biological Crystallography, *67*(Pt 4), 271–281. https://doi.org/10.1107/S0907444910048675

Baxa, U., Steinbacher, S., Miller, S., Weintraub, A., Huber, R., & Seckler, R. (1996). Interactions of phage P22 tails with their cellular receptor, Salmonella O-antigen polysaccharide. *Biophysical Journal*, *71*(4), 2040–2048. https://doi.org/10.1016/S0006-3495(96)79402-X

Bertani, B., & Ruiz, N. (2018). Function and Biogenesis of Lipopolysaccharides. *EcoSal Plus*, *8*(1). https://doi.org/10.1128/ecosalplus.ESP-0001-2018

Bragg, R., van der Westhuizen, W., Lee, J.-Y., Coetsee, E., & Boucher, C. (2014). Bacteriophages as potential treatment option for antibiotic resistant bacteria. *Advances in Experimental Medicine and Biology*, *807*, 97–110. https://doi.org/10.1007/978-81-322-1777-0_7

Broeker, N. K., & Barbirz, S. (2017). Not a barrier but a key: How bacteriophages exploit host's O-antigen as an essential receptor to initiate infection: Not a barrier but a key. *Molecular Microbiology*, *105*(3), 353–357. https://doi.org/10.1111/mmi.13729

Broeker, N. K., Roske, Y., Valleriani, A., Stephan, M. S., Andres, D., Koetz, J., Heinemann, U., & Barbirz, S. (2019). Time-resolved DNA release from an O-antigen–specific *Salmonella* bacteriophage with a contractile tail. *Journal of Biological Chemistry*, *294*(31), 11751–11761. https://doi.org/10.1074/jbc.RA119.008133

Bryan, A. W., Starner-Kreinbrink, J. L., Hosur, R., Clark, P. L., & Berger, B. (2011). Structure-based prediction reveals capping motifs that inhibit β-helix aggregation. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(27), 11099–11104. https://doi.org/10.1073/pnas.1017504108

Burnham, P. M., & Hendrixson, D. R. (2018). Campylobacter jejuni: Collective components promoting a successful enteric lifestyle. *Nature Reviews Microbiology*, *16*(9), 551–565. https://doi.org/10.1038/s41579-018-0037-9

Casjens, S. R. (2005). Comparative genomics and evolution of the tailed-bacteriophages. *Current Opinion in Microbiology*, *8*(4), 451–458. https://doi.org/10.1016/j.mib.2005.06.014

Chang, J. T., Schmid, M. F., Haase-Pettingell, C., Weigele, P. R., King, J. A., & Chiu, W. (2010). Visualizing the Structural Changes of Bacteriophage Epsilon15 and Its Salmonella Host during Infection. *Journal of Molecular Biology*, *402*(4), 731–740. https://doi.org/10.1016/j.jmb.2010.07.058

Christensen, S., McMahon, R. M., Martin, J. L., & Huston, W. M. (2019). Life inside and out: Making and breaking protein disulfide bonds in *Chlamydia*. *Critical Reviews in Microbiology*, *45*(1), 33–50. https://doi.org/10.1080/1040841X.2018.1538933

Clokie, M. R. J., Millard, A. D., Letarov, A. V., & Heaphy, S. (2011). Phages in nature. *Bacteriophage*, *1*(1), 31–45. https://doi.org/10.4161/bact.1.1.14942

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D: Biological Crystallography*, *62*(9), 1002–1011. https://doi.org/10.1107/S0907444906022116

Cowtan, Kevin. (2010). Recent developments in classical density modification. *Acta Crystallographica Section D: Biological Crystallography*, *66*(Pt 4), 470–478. https://doi.org/10.1107/S090744490903947X

Daffé, M., & Marrakchi, H. (2019). Unraveling the Structure of the Mycobacterial Envelope. *Microbiology Spectrum*, *7*(4). https://doi.org/10.1128/microbiolspec.GPP3-0027-2018

Davies, G., & Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure*, *3*(9), 853–859. https://doi.org/10.1016/S0969-2126(01)00220-9

de la Rosa-Trevín, J. M., Otón, J., Marabini, R., Zaldívar, A., Vargas, J., Carazo, J. M., & Sorzano, C. O. S. (2013). Xmipp 3.0: An improved software suite for image processing in electron microscopy. *Journal of Structural Biology*, *184*(2), 321–328. https://doi.org/10.1016/j.jsb.2013.09.015

de la Rosa-Trevín, J. M., Quintana, A., del Cano, L., Zaldívar, A., Foche, I., Gutiérrez, J., Gómez-Blanco, J., Burguet-Castell, J., Cuenca-Alba, J., Abrishami, V., Vargas, J., Otón, J., Sharov, G., Vilas, J. L., Navas, J., Conesa, P., Kazemi, M., Marabini, R., Sorzano, C. O. S., & Carazo, J. M. (2016).

Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *Journal of Structural Biology*, *195*(1), 93–99. https://doi.org/10.1016/j.jsb.2016.04.010

D'Herelle, F. (2007). On an invisible microbe antagonistic toward dysenteric bacilli: Brief note by Mr. F. D'Herelle, presented by Mr. Roux. 1917. *Research in Microbiology*, *158*(7), 553–554. https://doi.org/10.1016/j.resmic.2007.07.005

Di Lorenzo, F., De Castro, C., Silipo, A., Molinaro, A. (2019) Lipopolysaccharide structures of Gram-negative populations in the gut microbiota and effects on host interactions. *FEMS Microbiology Reviews*, 43,(3) 257-272. https://doi.org/10.1093/femsre/fuz002

Durand, G. A., Raoult, D., & Dubourg, G. (2019). Antibiotic discovery: History, methods and perspectives. *International Journal of Antimicrobial Agents*, *53*(4), 371–382. https://doi.org/10.1016/j.ijantimicag.2018.11.010

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, *66*(4), 486–501. https://doi.org/10.1107/S0907444910007493

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

Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallographica. Section D, Biological Crystallography*, *69*(Pt 7), 1204–1214. https://doi.org/10.1107/s0907444913000061

Farley, M. M., Tu, J., Kearns, D. B., Molineux, I. J., & Liu, J. (2017). Ultrastructural analysis of bacteriophage Φ29 during infection of Bacillus subtilis. *Journal of Structural Biology*, *197*(2), 163–171. https://doi.org/10.1016/j.jsb.2016.07.019

Gage, M. J., & Robinson, A. S. (2003). C-terminal hydrophobic interactions play a critical role in oligomeric assembly of the P22 tailspike trimer. *Protein Science*, *12*(12), 2732–2747. https://doi.org/10.1110/ps.03150303

Gajdus, J., Kaczyński, Z., Czerwicka, M., Dziadziuszko, H., & Szafranek, J. (2008). Chemical Structure of the Salmonella Anatum (O:3,10) O-antigen. *Polish Journal of Chemistry*, *Vol. 82, nr 7*, 1393–1393.

Gandon, S. (2016). Why Be Temperate: Lessons from Bacteriophage λ. *Trends in Microbiology*, *24*(5), 356–365. https://doi.org/10.1016/j.tim.2016.02.008

Golomidova, A., Kulikov, E., Prokhorov, N., Guerrero-Ferreira, R., Knirel, Y., Kostryukova, E., Tarasyan, K., & Letarov, A. (2016). Branched Lateral Tail Fiber Organization in T5-Like Bacteriophages DT57C and DT571/2 is Revealed by Genetic and Functional Analysis. *Viruses*, *8*(1), 26. https://doi.org/10.3390/v8010026

González-García, V. A., Bocanegra, R., Pulido-Cid, M., Martín-Benito, J., Cuervo, A., & Carrascosa, J. L. (2015). Characterization of the initial steps in the T7 DNA ejection process. *Bacteriophage*, *5*(3), e1056904. https://doi.org/10.1080/21597081.2015.1056904

González-García, V. A., Pulido-Cid, M., Garcia-Doval, C., Bocanegra, R., van Raaij, M. J., Martín-Benito, J., Cuervo, A., & Carrascosa, J. L. (2015). Conformational Changes Leading to T7 DNA Delivery upon Interaction with the Bacterial Receptor. *Journal of Biological Chemistry*, *290*(16), 10038–10044. https://doi.org/10.1074/jbc.M114.614222

Grant, A. J., Coward, C., Jones, M. A., Woodall, C. A., Barrow, P. A., & Maskell, D. J. (2005). Signature-Tagged Transposon Mutagenesis Studies Demonstrate the Dynamic Nature of Cecal Colonization of 2-Week-Old Chickens by Campylobacter jejuni. *Applied and Environmental Microbiology*, *71*(12), 8031–8041. https://doi.org/10.1128/AEM.71.12.8031-8041.2005

Guichard, J. A., Middleton, P. C., & McConnell, M. R. (2013). Genetic analysis of structural proteins in the adsorption apparatus of bacteriophage epsilon 15. *World Journal of Virology*, *2*(4), 152–159. https://doi.org/10.5501/wjv.v2.i4.152

Hagiwara, S., Uetake, H., & Takeda, K. (1966). Lysogenic conversion in the structure of cellular receptor for Salmonella phage g 341. *Annual Report of the Institute for Virus Research, Kyoto University, Kyoto, Japan*, *9*, 126–129.

Holm, L. (2020). DALI and the persistence of protein shape. *Protein Science*, *29*(1), 128–140. https://doi.org/10.1002/pro.3749

Holmes, A. H., Moore, L. S. P., Sundsfjord, A., Steinbakk, M., Regmi, S., Karkey, A., Guerin, P. J., & Piddock, L. J. V. (2016). Understanding the mechanisms and drivers of antimicrobial resistance. *The Lancet*, *387*(10014), 176–187. https://doi.org/10.1016/S0140-6736(15)00473-0

Hu, B., Margolin, W., Molineux, I. J., & Liu, J. (2015). Structural remodeling of bacteriophage T4 and host membranes during infection initiation. *Proceedings of the National Academy of Sciences*, *112*(35), E4919–E4928. https://doi.org/10.1073/pnas.1501064112

Islam, M. Z., Fokine, A., Mahalingam, M., Zhang, Z., Garcia-Doval, C., van Raaij, M. J., Rossmann, M. G., & Rao, V. B. (2019). Molecular anatomy of the receptor binding module of a bacteriophage long tail fiber. *PLOS Pathogens*, *15*(12), e1008193. https://doi.org/10.1371/journal.ppat.1008193

Iwashita, S., & Kanegasaki, S. (1976). Enzymic and Molecular Properties of Base-Plate Parts of Bacteriophage P22. *European Journal of Biochemistry*, *65*(1), 87–94. https://doi.org/10.1111/j.1432-1033.1976.tb10392.x

Jault, P., Leclerc, T., Jennes, S., Pirnay, J. P., Que, Y.-A., Resch, G., Rousseau, A. F., Ravat, F., Carsin, H., Le Floch, R., Schaal, J. V., Soler, C., Fevre, C., Arnaud, I., Bretaudeau, L., & Gabard, J. (2019). Efficacy and tolerability of a cocktail of bacteriophages to treat burn wounds infected by Pseudomonas aeruginosa (PhagoBurn): A randomised, controlled, double-blind phase 1/2 trial. *The Lancet Infectious Diseases*, *19*(1), 35–45. https://doi.org/10.1016/S1473-3099(18)30482-1

Javed, M. A., Ackermann, H.-W., Azeredo, J., Carvalho, C. M., Connerton, I., Evoy, S., Hammerl, J. A., Hertwig, S., Lavigne, R., Singh, A., Szymanski, C. M., Timms, A., & Kropinski, A. M. (2014). A suggested classification for two groups of Campylobacter myoviruses. *Archives of Virology*, *159*(1), 181–190. https://doi.org/10.1007/s00705-013-1788-2

Jiang, W., Baker, M. L., Jakana, J., Weigele, P. R., King, J., & Chiu, W. (2008). Backbone structure of the infectious ε15 virus capsid revealed by electron cryomicroscopy. *Nature*, *451*(7182), 1130–1134. https://doi.org/10.1038/nature06665

Jiang, W., Chang, J., Jakana, J., Weigele, P., King, J., & Chiu, W. (2006). Structure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus. *Nature*, *439*(7076), 612–616. https://doi.org/10.1038/nature04487

Jones, M. A., Marston, K. L., Woodall, C. A., Maskell, D. J., Linton, D., Karlyshev, A. V., Dorrell, N., Wren, B. W., & Barrow, P. A. (2004). Adaptation of Campylobacter jejuni NCTC11168 to High-Level Colonization of the Avian Gastrointestinal Tract. *Infection and Immunity*, *72*(7), 3769–3776. https://doi.org/10.1128/IAI.72.7.3769-3776.2004

Juanhuix, J., Gil-Ortiz, F., Cuní, G., Colldelram, C., Nicolás, J., Lidón, J., Boter, E., Ruget, C., Ferrer, S., & Benach, J. (2014). Developments in optics and performance at BL13-XALOC, the macromolecular crystallography beamline at the Alba Synchrotron. *Journal of Synchrotron Radiation*, *21*(4), 679–689. https://doi.org/10.1107/S160057751400825X

Kabsch, W. (2010). XDS. *Acta Crystallographica Section D: Biological Crystallography*, *66*(2), 125–132. https://doi.org/10.1107/S0907444909047337

Kakasis, A., & Panitsa, G. (2019). Bacteriophage therapy as an alternative treatment for human infections. A comprehensive review. *International Journal of Antimicrobial Agents*, *53*(1), 16–21. https://doi.org/10.1016/j.ijantimicag.2018.09.004

Kanegasaki, S., & Wright, A. (1973). Studies on the mechanism of phage adsorption: Interaction between phage ϵ15 and its cellular receptor. *Virology*, *52*(1), 160–173. https://doi.org/10.1016/0042-6822(73)90406-6

Knirel, Y. A., Prokhorov, N. S., Shashkov, A. S., Ovchinnikova, O. G., Zdorovenko, E. L., Liu, B., Kostryukova, E. S., Larin, A. K., Golomidova, A. K., & Letarov, A. V. (2015). Variations in O-Antigen Biosynthesis and O-Acetylation Associated with Altered Phage Sensitivity in

Escherichia coli 4s. *Journal of Bacteriology*, *197*(5), 905–912. https://doi.org/10.1128/JB.02398-14

Kofod, L. V., Kauppinen, S., Christgau, S., Andersen, L. N., Heldt-Hansen, H. P., Dörreich, K., & Dalbøge, H. (1994). Cloning and characterization of two structurally and functionally divergent rhamnogalacturonases from Aspergillus aculeatus. *Journal of Biological Chemistry*, *269*(46), 29182–29189. https://doi.org/10.1016/S0021-9258(19)62028-4

Krissinel, E. (2015). Stock-based detection of protein oligomeric states in jsPISA. *Nucleic Acids Research*, *43*(W1), W314-319. https://doi.org/10.1093/nar/gkv314

Kropinski, A. M., Kovalyova, I. V., Billington, S. J., Patrick, A. N., Butts, B. D., Guichard, J. A., Pitcher, T. J., Guthrie, C. C., Sydlaske, A. D., Barnhill, L. M., Havens, K. A., Day, K. R., Falk, D. R., & McConnell, M. R. (2007). The genome of ε15, a serotype-converting, Group E1 Salmonella enterica-specific bacteriophage. *Virology*, *369*(2), 234–244. https://doi.org/10.1016/j.virol.2007.07.027

Kulikov, E. E., Majewska, J., Prokhorov, N. S., Golomidova, A. K., Tatarskiy, E. V., & Letarov, A. V. (2017). Effect of O-acetylation of O antigen of Escherichia coli lipopolysaccharide on the nonspecific barrier function of the outer membrane. *Microbiology*, *86*(3), 310–316. https://doi.org/10.1134/S0026261717030080

Kutter, E., De Vos, D., Gvasalia, G., Alavidze, Z., Gogokhia, L., Kuhl, S., & Abedon, S. (2009). Phage Therapy in Clinical Practice: Treatment of Human Infections. *Current Pharmaceutical Biotechnology*, *11*, 69–86. https://doi.org/10.2174/138920110790725401

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, *227*(5259), 680–685. https://doi.org/10.1038/227680a0

Lefkowitz, E. J., Dempsey, D. M., Hendrickson, R. C., Orton, R. J., Siddell, S. G., & Smith, D. B. (2018). Virus taxonomy: The database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Research*, *46*(D1), D708–D717. https://doi.org/10.1093/nar/gkx932

Limoli, D. H., Jones, C. J., & Wozniak, D. J. (2015). Bacterial Extracellular Polysaccharides in Biofilm Formation and Function. *Microbiology Spectrum*, *3*(3). https://doi.org/10.1128/microbiolspec.MB-0011-2014

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, *42*(D1), D490–D495. https://doi.org/10.1093/nar/gkt1178

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C. J., & Marchler-Bauer, A. (2020). CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Research*, *48*(D1), D265–D268. https://doi.org/10.1093/nar/gkz991

Lukianova, A. A., Shneider, M. M., Evseev, P. V., Shpirt, A. M., Bugaeva, E. N., Kabanova, A. P., Obraztsova, E. A., Miroshnikov, K. K., Senchenkova, S. N., Shashkov, A. S., Toschakov, S. V., Knirel, Y. A., Ignatov, A. N., & Miroshnikov, K. A. (2020). Morphologically Different Pectobacterium brasiliense Bacteriophages PP99 and PP101: Deacetylation of O-Polysaccharide by the Tail Spike Protein of Phage PP99 Accompanies the Infection. *Frontiers in Microbiology*, *10*, 3147. https://doi.org/10.3389/fmicb.2019.03147

Marti, R., Zurfluh, K., Hagens, S., Pianezzi, J., Klumpp, J., & Loessner, M. J. (2013). Long tail fibres of the novel broad-host-range T-even bacteriophage S16 specifically recognize *Salmonella* OmpC: T4-like *Salmonella* phage S16. *Molecular Microbiology*, *87*(4), 818–834. https://doi.org/10.1111/mmi.12134

McCallin, S., Sacher, J. C., Zheng, J., & Chan, B. K. (2019). Current State of Compassionate Phage Therapy. *Viruses*, *11*(4). https://doi.org/10.3390/v11040343

McConnell, M. R., Oakes, K. R., Patrick A. N., Mills, D. M. (2001) Two functional O-polysaccharide polymerase wzy (rfc) genes are present in the rfb gene cluster of Group E1 Salmonella enterica serovar Anatum. *FEMS Microbiology Letters*, 199, 235-240. 10.1111/j.1574-6968.2001.tb10680.x

McConnell, M. R., & Schoelz, J. E. (1983). Evidence for Shorter Average O-Polysaccharide Chainlength in the Lipopolysaccharide of a Bacteriophage Felix 01-sensitive Variant of Salmonella anatum A1. *Microbiology*, *129*(10), 3177–3184. https://doi.org/10.1099/00221287-129-10-3177

McConnell, M., Reznick, A., & Wright, A. (1979). Studies on the initial interactions of bacteriophage epsilon15 with its host cell, Salmonella anatum. *Virology*, *94*(1), 10–23. https://doi.org/10.1016/0042-6822(79)90434-3

McConnell, M., Walker, B., Middleton, P., Chase, J., Owens, J., Hyatt, D., Gutierrez, H., Williams, M., Hambright, D., Barry, M., Sage, S., Fuller, G., Birdwell, M., Rydelski, M., Risley, Sh., & Kat, B. (1992). Restriction endonuclease and genetic mapping studies indicate that the vegetative genome of the temperate, salmonella-specific bacteriophage, Epsilon 15, is circularly-permuted. *Archives of Virology*, *123*(1–2), 215–221. https://doi.org/10.1007/BF01317151

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography*, *40*(Pt 4), 658–674. https://doi.org/10.1107/S0021889807021206

McNally, D. J., Lamoureux, M. P., Karlyshev, A. V., Fiori, L. M., Li, J., Thacker, G., Coleman, R. A., Khieu, N. H., Wren, B. W., Brisson, J.-R., Jarrell, H. C., & Szymanski, C. M. (2007). Commonality and Biosynthesis of the O-Methyl Phosphoramidate Capsule Modification in Campylobacter jejuni*. *Journal of Biological Chemistry*, *282*(39), 28566–28576. https://doi.org/10.1074/jbc.M704413200

Międzybrodzki, R., Hoyle, N., Zhvaniya, F., Łusiak-Szelachowska, M., Weber-Dąbrowska, B., Łobocka, M., Borysowski, J., Alavidze, Z., Kutter, E., Górski, A., & Gogokhia, L. (2018). Current Updates from the Long-Standing Phage Research Centers in Georgia, Poland, and Russia. In D. R. Harper, S. T. Abedon, B. H. Burrowes, & M. L. McConville (Eds.), *Bacteriophages* (pp. 1–31). Springer International Publishing. https://doi.org/10.1007/978-3-319-40598-8_31-1

Mitraki, A., Barge, A., Chroboczek, J., Andrieu, J.-P., Gagnon, J., & Ruigrok, R. W. H. (1999). Unfolding studies of human adenovirus type 2 fibre trimers. *European Journal of Biochemistry*, *264*(2), 599–606. https://doi.org/10.1046/j.1432-1327.1999.00683.x

Mitraki, A., Papanikolopoulou, K., & Van Raaij, M. J. (2006). Natural Triple β-Stranded Fibrous Folds1. In *Advances in Protein Chemistry* (Vol. 73, pp. 97–124). Elsevier. https://doi.org/10.1016/S0065-3233(06)73004-2

Müller, J. J., Barbirz, S., Heinle, K., Freiberg, A., Seckler, R., & Heinemann, U. (2008). An Intersubunit Active Site between Supercoiled Parallel β Helices in the Trimeric Tailspike Endorhamnosidase of Shigella flexneri Phage Sf6. *Structure*, *16*(5), 766–775. https://doi.org/10.1016/j.str.2008.01.019

Murata, K., Liu, X., Danev, R., Jakana, J., Schmid, M. F., King, J., Nagayama, K., & Chiu, W. (2010). Zernike Phase Contrast Cryo-Electron Microscopy and Tomography for Structure Determination at Nanometer and Subnanometer Resolutions. *Structure*, *18*(8), 903–912. https://doi.org/10.1016/j.str.2010.06.006

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., & Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica. Section D, Biological Crystallography*, *67*(Pt 4), 355–367. https://doi.org/10.1107/S0907444911001314

Nakamura, Y., Gojobori, T., & Ikemura, T. (2000). Codon usage tabulated from international DNA sequence databases: Status for the year 2000. *Nucleic Acids Research*, *28*(1), 292–292. https://doi.org/10.1093/nar/28.1.292

Nobrega, F. L., Vlot, M., de Jonge, P. A., Dreesens, L. L., Beaumont, H. J. E., Lavigne, R., Dutilh, B. E., & Brouns, S. J. J. (2018). Targeting mechanisms of tailed bacteriophages. *Nature Reviews Microbiology*, *16*(12), 760–773. https://doi.org/10.1038/s41579-018-0070-8

Osawa, K., Shigemura, K., Iguchi, A., Shirai, H., Imayama, T., Seto, K., Raharjo, D., Fujisawa, M., Osawa, R., & Shirakawa, T. (2013). O-antigen chain length modulated by the *wzz* gene in

Escherichia coli O157 influenced its sensitivities to serum complement: Wzz gene in *Escherichia coli* O157. *Microbiology and Immunology*, n/a-n/a. https://doi.org/10.1111/1348-0421.12084

Pan, Y.-J., Lin, T.-L., Chen, C.-C., Tsai, Y.-T., Cheng, Y.-H., Chen, Y.-Y., Hsieh, P.-F., Lin, Y.-T., & Wang, J.-T. (2017). Klebsiella Phage ΦK64-1 Encodes Multiple Depolymerases for Multiple Host Capsular Types. *Journal of Virology*, *91*(6), e02457-16, e02457-16. https://doi.org/10.1128/JVI.02457-16

Pequegnat, B., Laird, R. M., Ewing, C. P., Hill, C. L., Omari, E., Poly, F., Monteiro, M. A., & Guerry, P. (2017). Phase-Variable Changes in the Position of O-Methyl Phosphoramidate Modifications on the Polysaccharide Capsule of Campylobacter jejuni Modulate Serum Resistance. *Journal of Bacteriology*, *199*(14), e00027-17, e00027-17. https://doi.org/10.1128/JB.00027-17

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. https://doi.org/10.1002/jcc.20084

Pickard, D., Toribio, A. L., Petty, N. K., van Tonder, A., Yu, L., Goulding, D., Barrell, B., Rance, R., Harris, D., Wetter, M., Wain, J., Choudhary, J., Thomson, N., & Dougan, G. (2010). A Conserved Acetyl Esterase Domain Targets Diverse Bacteriophages to the Vi Capsular Receptor of Salmonella enterica Serovar Typhi. *Journal of Bacteriology*, *192*(21), 5746–5754. https://doi.org/10.1128/JB.00659-10

Pires, D. P., Cleto, S., Sillankorva, S., Azeredo, J., & Lu, T. K. (2016). Genetically Engineered Phages: A Review of Advances over the Last Decade. *Microbiology and Molecular Biology Reviews*, *80*(3), 523–543. https://doi.org/10.1128/MMBR.00069-15

Pires, D. P., Oliveira, H., Melo, L. D. R., Sillankorva, S., & Azeredo, J. (2016). Bacteriophage-encoded depolymerases: Their diversity and biotechnological applications. *Applied Microbiology and Biotechnology*, *100*(5), 2141–2151. https://doi.org/10.1007/s00253-015-7247-0

Plattner, M., Shneider, M. M., Arbatsky, N. P., Shashkov, A. S., Chizhov, A. O., Nazarov, S., Prokhorov, N. S., Taylor, N. M. I., Buth, S. A., Gambino, M., Gencay, Y. E., Brøndsted, L., Kutter, E. M., Knirel, Y. A., & Leiman, P. G. (2019). Structure and Function of the Branched Receptor-Binding Complex of Bacteriophage CBA120. *Journal of Molecular Biology*, *431*(19), 3718–3739. https://doi.org/10.1016/j.jmb.2019.07.022

Prokhorov, N. S., Riccio, C., Zdorovenko, E. L., Shneider, M. M., Browning, C., Knirel, Y. A., Leiman, P. G., & Letarov, A. V. (2017). Function of bacteriophage G7C esterase tailspike in host cell adsorption: Function of bacteriophage G7C esterase tailspike. *Molecular Microbiology*, *105*(3), 385–398. https://doi.org/10.1111/mmi.13710

Puigbò, P., Guzmán, E., Romeu, A., & Garcia-Vallvé, S. (2007). OPTIMIZER: A web server for optimizing the codon usage of DNA sequences. *Nucleic Acids Research*, *35*(Web Server issue), W126-131. https://doi.org/10.1093/nar/gkm219

Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, *42*(W1), W320–W324. https://doi.org/10.1093/nar/gku316

Rohde, M. (2019). The Gram-Positive Bacterial Cell Wall. *Microbiology Spectrum*, *7*(3). https://doi.org/10.1128/microbiolspec.GPP3-0044-2018

Rohou, A., & Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*, *192*(2), 216–221. https://doi.org/10.1016/j.jsb.2015.08.008

Rohwer, F. (2003). Global Phage Diversity. *Cell*, *113*(2), 141. https://doi.org/10.1016/S0092-8674(03)00276-9

Rohwer, F., & Edwards, R. (2002). The Phage Proteomic Tree: A Genome-Based Taxonomy for Phage. *Journal of Bacteriology*, *184*(16), 4529–4535. https://doi.org/10.1128/JB.184.16.4529-4535.2002

Rohwer, F., & Segall, A. M. (2015). In retrospect: A century of phage lessons. *Nature*, *528*(7580), 46–48. https://doi.org/10.1038/528046a

Rottem, S. (2003). Interaction of Mycoplasmas With Host Cells. *Physiological Reviews*, *83*(2), 417–432. https://doi.org/10.1152/physrev.00030.2002

Sakai, K., Iwazaki, T., Yamashita, E., Nakagawa, A., Sakuraba, F., Enomoto, A., Inagaki, M., & Takeda, S. (2019). Observation of unexpected molecular binding activity for Mu phage tail fibre chaperones. *The Journal of Biochemistry*, *166*(6), 529–535. https://doi.org/10.1093/jb/mvz068

Salmond, G. P. C., & Fineran, P. C. (2015). A century of the phage: Past, present and future. *Nature Reviews Microbiology*, *13*(12), 777–786. https://doi.org/10.1038/nrmicro3564

Samson, J. E., Magadán, A. H., Sabri, M., & Moineau, S. (2013). Revenge of the phages: Defeating bacterial defences. *Nature Reviews Microbiology*, *11*(10), 675–687. https://doi.org/10.1038/nrmicro3096

Sanz-Gaitero, M., Seoane-Blanco, M., & van Raaij, M. J. (2019). Structure and Function of Bacteriophages. In D. R. Harper, S. T. Abedon, B. H. Burrowes, & M. L. McConville (Eds.), *Bacteriophages* (pp. 1–73). Springer International Publishing. https://doi.org/10.1007/978-3-319-40598-8_1-1

Scheres, S. H. W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, *180*(3), 519–530. https://doi.org/10.1016/j.jsb.2012.09.006

Scholl, D., Rogers, S., Adhya, S., & Merril, C. R. (2001). Bacteriophage K1-5 Encodes Two Different Tail Fiber Proteins, Allowing It To Infect and Replicate on both K1 and K5 Strains of Escherichia coli. *Journal of Virology*, *75*(6), 2509–2515. https://doi.org/10.1128/JVI.75.6.2509-2515.2001

Schols, H. A., Geraeds, C. C. J. M., Searle-van Leeuwen, M. F., Kormelink, F. J. M., & Voragen, A. G. J. (1990). Rhamnogalacturonase: A novel enzyme that degrades the hairy regions of pectins. *Carbohydrate Research*, *206*(1), 105–115. https://doi.org/10.1016/0008-6215(90)84010-R

Searle-van Leeuwen, M. J. F., van den Broek, L. A. M., Schols, H. A., Beldman, G., & Voragen, A. G. J. (1992). Rhamnogalacturonan acetylesterase: A novel enzyme from Aspergillus aculeatus,

specific for the deacetylation of hairy (ramified) regions of pectins. *Applied Microbiology and Biotechnology*, *38*(3), 347–349. https://doi.org/10.1007/BF00170084

Sechter, I., & Sechter-Mooreville, R. (1990). Host-range modifications of bacteriophage ε 15. *Archives of Virology*, *113*(3), 297–297. https://doi.org/10.1007/BF01316683

Seoane-Blanco, M., van Raaij, M. J., & Granell, M. (2020). Bacteriophage Tail Fibers, Tailspikes and Host Cell Receptor Interaction. In *Reference Module in Life Sciences*. Elsevier. https://doi.org/10.1016/B978-0-12-814515-9.00152-1

Sheldrick, G. M. (2010). Experimental phasing with SHELXC/D/E: Combining chain tracing with density modification. *Acta Crystallographica Section D: Biological Crystallography*, *66*(4), 479–485. https://doi.org/10.1107/S0907444909038360

Silhavy, T. J., Kahne, D., & Walker, S. (2010). The Bacterial Cell Envelope. *Cold Spring Harbor Perspectives in Biology*, *2*(5), a000414–a000414. https://doi.org/10.1101/cshperspect.a000414

Skubák, P., & Pannu, N. S. (2013). Automatic protein structure solution from weak X-ray data. *Nature Communications*, *4*(1), 2777. https://doi.org/10.1038/ncomms3777

Skubák, P., Waterreus, W. J., & Pannu, N. S. (2010). Multivariate phase combination improves automated crystallographic model building. *Acta Crystallographica. Section D, Biological Crystallography*, *66*(Pt 7), 783–788. https://doi.org/10.1107/S0907444910014642

Sorek, R., Lawrence, C. M., & Wiedenheft, B. (2013). CRISPR-Mediated Adaptive Immune Systems in Bacteria and Archaea. *Annual Review of Biochemistry*, *82*(1), 237–266. https://doi.org/10.1146/annurev-biochem-072911-172315

Sørensen, M. C. H., Gencay, Y. E., Birk, T., Baldvinsson, S. B., Jäckel, C., Hammerl, J. A., Vegge, C. S., Neve, H., & Brøndsted, L. (2015). Primary Isolation Strain Determines Both Phage Type and Receptors Recognised by Campylobacter jejuni Bacteriophages. *PLOS ONE*, *10*(1), e0116287. https://doi.org/10.1371/journal.pone.0116287

Sørensen, M. C. H., van Alphen, L. B., Fodor, C., Crowley, S. M., Christensen, B. B., Szymanski, C. M., & Brøndsted, L. (2012). Phase Variable Expression of Capsular Polysaccharide Modifications Allows Campylobacter jejuni to Avoid Bacteriophage Infection in Chickens. *Frontiers in Cellular and Infection Microbiology*, *2*. https://doi.org/10.3389/fcimb.2012.00011

Sorzano, C. O. S., Bilbao-Castro, J. R., Shkolnisky, Y., Alcorlo, M., Melero, R., Caffarena-Fernández, G., Li, M., Xu, G., Marabini, R., & Carazo, J. M. (2010). A clustering approach to multireference alignment of single-particle projections in electron microscopy. *Journal of Structural Biology*, *171*(2), 197–206. https://doi.org/10.1016/j.jsb.2010.03.011

Sorzano, C. O. S., de la Rosa Trevín, J. M., Otón, J., Vega, J. J., Cuenca, J., Zaldívar-Peraza, A., Gómez-Blanco, J., Vargas, J., Quintana, A., Marabini, R., & Carazo, J. M. (2013). Semiautomatic, High-Throughput, High-Resolution Protocol for Three-Dimensional Reconstruction of Single Particles in Electron Microscopy. In A. A. Sousa & M. J. Kruhlak (Eds.), *Nanoimaging: Methods and Protocols* (pp. 171–193). Humana Press. https://doi.org/10.1007/978-1-62703-137-0_11

Spagnolli, G., Rigoli, M., Orioli, S., Sevillano, A. M., Faccioli, P., Wille, H., Biasini, E., & Requena, J. R. (2019). Full atomistic model of prion structure and conversion. *PLoS Pathogens*, *15*(7), e1007864. https://doi.org/10.1371/journal.ppat.1007864

Steinbacher, S., Seckler, R., Miller, S., Steipe, B., Huber, R., & Reinemer, P. (1994). Crystal structure of P22 tailspike protein: Interdigitated subunits in a thermostable trimer. *Science*, *265*(5170), 383–386. https://doi.org/10.1126/science.8023158

Stirm, S., & Freund-Mölbert, E. (1971). Escherichia coli capsule bacteriophages. II. Morphology. *Journal of Virology*, *8*(3), 330–342. https://doi.org/10.1128/JVI.8.3.330-342.1971

Tao, Y., Strelkov, S. V., Mesyanzhinov, V. V., & Rossmann, M. G. (1997). Structure of bacteriophage T4 fibritin: A segmented coiled coil and the role of the C-terminal domain. *Structure*, *5*(6), 789–798. https://doi.org/10.1016/S0969-2126(97)00233-5

Taylor, N. M. I., Prokhorov, N. S., Guerrero-Ferreira, R. C., Shneider, M. M., Browning, C., Goldie, K. N., Stahlberg, H., & Leiman, P. G. (2016). Structure of the T4 baseplate and its function in triggering sheath contraction. *Nature*, *533*(7603), 346–352. https://doi.org/10.1038/nature17971

Taylor, N. M. I., van Raaij, M. J., & Leiman, P. G. (2018). Contractile injection systems of bacteriophages and related systems: Conserved features of a contractile sheath-rigid tube mechanism. *Molecular Microbiology*, *108*(1), 6–15. https://doi.org/10.1111/mmi.13921

The European Union One Health 2018 Zoonoses Report. (2019). *EFSA Journal*, *17*(12), e05926. https://doi.org/10.2903/j.efsa.2019.5926

Theuretzbacher, U., Outterson, K., Engel, A., & Karlén, A. (2020). The global preclinical antibacterial pipeline. *Nature Reviews Microbiology*, *18*(5), 275–285. https://doi.org/10.1038/s41579-019-0288-0

Trojet, S. N., Caumont-Sarcos, A., Perrody, E., Comeau, A. M., & Krisch, H. M. (2011). The gp38 Adhesins of the T4 Superfamily: A Complex Modular Determinant of the Phage's Host Specificity. *Genome Biology and Evolution*, *3*, 674–686. https://doi.org/10.1093/gbe/evr059

Twort, F. W. (1915). AN INVESTIGATION ON THE NATURE OF ULTRA-MICROSCOPIC VIRUSES. *The Lancet*, *186*(4814), 1241–1243. https://doi.org/10.1016/S0140-6736(01)20383-3

Vagin, A., & Teplyakov, A. (2010). Molecular replacement with MOLREP. *Acta Crystallographica Section D*, *66*(1), 22–25. https://doi.org/10.1107/S0907444909042589

van Alphen, L. B., Wenzel, C. Q., Richards, M. R., Fodor, C., Ashmus, R. A., Stahl, M., Karlyshev, A. V., Wren, B. W., Stintzi, A., Miller, W. G., Lowary, T. L., & Szymanski, C. M. (2014). Biological Roles of the O-Methyl Phosphoramidate Capsule Modification in Campylobacter jejuni. *PLoS ONE*, *9*(1), e87051. https://doi.org/10.1371/journal.pone.0087051

van de Putte, P., Cramer, S., & Giphart-Gassler, M. (1980). Invertible DNA determines host specificity of bacteriophage Mu. *Nature*, *286*(5770), 218–222. https://doi.org/10.1038/286218a0

Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., & Bricogne, G. (2011). Data processing and analysis with the autoPROC toolbox. *Acta Crystallographica Section D: Biological Crystallography*, *67*(4), 293–302. https://doi.org/10.1107/S0907444911007773

Walter, M., Fiedler, C., Grassl, R., Biebl, M., Rachel, R., Hermo-Parrado, X. L., Llamas-Saiz, A. L., Seckler, R., Miller, S., & van Raaij, M. J. (2008). Structure of the Receptor-Binding Protein of Bacteriophage Det7: A Podoviral Tail Spike in a Myovirus. *Journal of Virology*, *82*(5), 2265–2273. https://doi.org/10.1128/JVI.01641-07

Wang, H., La Russa, M., & Qi, L. S. (2016). CRISPR/Cas9 in Genome Editing and Beyond. *Annual Review of Biochemistry*, *85*(1), 227–264. https://doi.org/10.1146/annurev-biochem-060815-014607

Weinbauer, M. G. (2004). Ecology of prokaryotic viruses. *FEMS Microbiology Reviews*, *28*(2), 127–181. https://doi.org/10.1016/j.femsre.2003.08.001

Wen, Z., & Zhang, J.-R. (2015). Bacterial Capsules. In *Molecular Medical Microbiology* (pp. 33–53). Elsevier. https://doi.org/10.1016/B978-0-12-397169-2.00003-2

Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B., Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S., & Richardson, D. C. (2018). MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science: A Publication of the Protein Society*, *27*(1), 293–315. https://doi.org/10.1002/pro.3330

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., & Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography*, *67*(4), 235–242. https://doi.org/10.1107/S0907444910045749

Winter, G. (2010). xia2: An expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography*, *43*(1), 186–190. https://doi.org/10.1107/S0021889809045701

Winter, G., Waterman, D. G., Parkhurst, J. M., Brewster, A. S., Gildea, R. J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I. D., Sauter, N. K., & Evans, G. (2018). DIALS: Implementation and evaluation of a new integration package. *Acta Crystallographica Section D: Structural Biology*, *74*(2), 85–97. https://doi.org/10.1107/S2059798317017235

Wittebole, X., De Roock, S., & Opal, S. M. (2014). A historical overview of bacteriophage therapy as an alternative to antibiotics for the treatment of bacterial pathogens. *Virulence*, *5*(1), 226–235. https://doi.org/10.4161/viru.25991

Wright, A. (1971). Mechanism of Conversion of the Salmonella O Antigen by Bacteriophage ε34. *Journal of Bacteriology*, *105*(3), 927–936.

Xiang, Y., Morais, M. C., Battisti, A. J., Grimes, S., Jardine, P. J., Anderson, D. L., & Rossmann, M. G. (2006). Structural changes of bacteriophage φ29 upon DNA packaging and release. *The EMBO Journal*, *25*(21), 5229–5239. https://doi.org/10.1038/sj.emboj.7601386

Zechel, D. L., & Withers, S. G. (2000). Glycosidase Mechanisms: Anatomy of a Finely Tuned Catalyst. *Accounts of Chemical Research*, *33*(1), 11–18. https://doi.org/10.1021/ar970172+

Zhang, Y., Wang, Z., Zhang, J., Chen, C., Wu, Q., Zhang, L., & Zhang, X. (2011). Quantitative determination of chitinolytic activity of lysozyme using half-deacetylated chitosan as a substrate. *Carbohydrate Polymers*, *85*(3), 554–559. https://doi.org/10.1016/j.carbpol.2011.03.009

Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A. N., & Alva, V. (2018). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology*, *430*(15), 2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007

Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J., Lindahl, E., & Scheres, S. H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. *ELife*, *7*, e42166. https://doi.org/10.7554/eLife.42166