Chapter 13

Crystallographic Structure Determination of Bacteriophage Endolysins

Marta Sanz-Gaitero$^{1,2,3}$ and Mark J. van Raaij$^1$*

$^1$Centro Nacional de Biotecnologia, CSIC, Calle Darwin 3, E-28049 Madrid, Spain.
$^2$Cork Institute of Technology, Bishopstown, Cork T12 P928, Ireland.
$^3$Current address: Centre for Molecular Medicine Norway, University of Oslo, Gaustadalléen 21, 0349 Oslo, Norway.

*Corresponding author: mjvanraaij@cnb.csic.es

DOI: https://doi.org/10.21775/9781913652517.13

Abstract
Bacteriophages produce endolysins that target and cleave the hosts peptidoglycan to release their progeny at the end of the infection cycle. These proteins can be used for the eradication of pathogenic bacteria, but also for their detection. Endolysins may contain a single catalytic domain or several domains, including a cell wall binding domain. To understand their function in detail and design mutated or chimeric molecules with novel properties, knowledge of their structures and detailed mechanisms is necessary. X-ray protein crystallography is an excellent method to obtain high-resolution structures of biological macromolecules, and here we describe the method and the folds of known endolysin domains.

Introduction
Bacteriophages (phages) are a promising treatment for antibiotic-resistant pathogens (see chapter by Hoyle et al., this book). However, they also present
products, such as endotoxins, due to incomplete purification. Specific self-replication of phages at the infection site is a major advantage. However, a self-

may be better to use isolated phage proteins.

glycosidases, peptidases or amidases (Loessner, 2005).

(O’Sullivan et al., 2016; see chapter by Gutiérrez and Briers, this book).
Endolysin applications are varied. They can be used as antimicrobial agents, for bacterial identification or for surface sterilization. Each of these applications requires the use of endolysins with different characteristics. For some applications, a wide lytic range may be necessary, for others a high specificity may be desired. Resistance to different temperature, pH, salinity or humidity ranges may also be required.

To obtain custom-designed endolysins, several techniques can be applied. Mutagenesis and protein truncation can be used to modify the lytic and antibacterial activity, cell-wall binding domain dependence, specificity, half-life and thermostability of endolysins. Recombining domains or domain swapping has been used successfully to create endolysins with altered properties (chimeolysins), such as activity against a certain type of bacteria or increased solubility (Yang et al., 2014; Gerstmans et al., 2018). Sequences can be added to these domains to increase their solubility or hydrophobicity, as in the case of artilysins, where an outer membrane permeabilizing region is added to allow them to pass through the bacterial outer membrane, allowing them to be effective even when applied exogenously. Cell wall binding domains of endolysins have been fused to fluorescent proteins or antibodies for bacterial detection purposes.

To reliably design tailor-made endolysins, knowledge about their characteristics, catalytic mechanisms, regulation is very advantageous. High-resolution structures of endolysins help to determine these properties. Determining the structure not only helps to elucidate the mechanism of action of that specific endolysin, but also allows to predict the structure of closely related enzymes.

**Structure determination methods**

Gauto et al., 2019). (Everett et al., 2015; Wang et al., 2017).
Nuclear Magnetic Resonance (NMR) spectroscopy
Atomic nuclei are charged spinning particles. When exposed to a strong magnetic field and radiofrequency radiation, they will produce an electromagnetic signal with a frequency characteristic of the magnetic field at the nucleus. This signal will be different depending on the chemical environment, and the magnetic properties of every atom, making it useful for structure determination (Krishnan, 2019). The main advantages of NMR spectroscopy are that structures are determined in solution and the possibility of investigating protein dynamics and ligand binding properties. The main disadvantages are the requirement of high amounts of isotopically labelled sample and the fact that for large proteins and protein complexes, signals start to overlap, limiting the technique to smaller proteins (generally up to 30-50 kDa).

Cryo-Electron Microscopy (Cryo-EM)
In cryo-electron microscopy, small samples, in which the macromolecular particles are randomly oriented, are vitrified and kept at very low temperature. In the microscope, they are exposed to electrons to produce images of individual particles. These images are averaged to produce several two-dimensional classes that will be used to reconstruct the three-dimensional structure of the macromolecule. The main advantages of cryo-EM are the small amount of sample required and the fact that it is not necessary to obtain protein crystals. It is also suitable for large complexes, such as phage particles. On the other hand, cryo-EM does not work so well for smaller proteins (less than around 150 kDa). It is also not routinely possible to achieve atomic resolution structures.

X-ray crystallography
Crystallography uses diffraction to determine the position and arrangement of the atoms in a crystal. Using X-rays, atomic or near-atomic resolution can be achieved for small molecules, proteins and, in favourable cases, large complexes such as whole viruses. The main drawback of this technique is that crystals are necessary, and crystallisation of proteins requires high amounts of very pure and homogeneous sample and time. Often, a well-diffracting crystal can simply not be obtained. In addition to that, a static crystalline structure is obtained, although molecular dynamic techniques may provide clues about the
X-ray crystallography is often the technique of choice.

Main steps in X-ray crystallography

the flexible loops, are often useful.

concentration decreases and gets below the supersaturation level.
aggregation occur, leading to nucleation and crystal growth (Rupp, 2009).

Figure 1: Main steps for protein structure determination using X-ray crystallography. See text for more details.
molecules (Als-Nielsen et al.,)

collected at different angles, allowing for a complete diffraction data set.

unit, which may contain one or more protein molecules.
Unfortunately, only the intensity of the diffraction spots can be measured experimentally, as the phase is lost. This is known as the phase problem (Cowtan, 2003).

To solve the phase problem, several approaches can be used (Taylor, 2010). In molecular replacement, we use the phases calculated from a protein model derived from a structure with high sequence similarity to ours (greater than 30% sequence identity as a rule of thumb), assuming their structures will be similar.

For molecular replacement, first the model has to be placed in the correct location and orientation in the unit cell, using sophisticated software. In experimental phasing, a limited number of heavy atoms incorporated at specific places into the protein structure (for example mercury ions bound to accessible cysteine residues). This is usually done by trial and error soaking of heavy atom compounds into pre-formed crystals. A commonly used alternative method is to incorporate seleno-methionine residues into the protein by expressing the protein in special growth media and crystallising this seleno-methionine derivative version of the protein. The substructure of the heavy atoms is simpler and easier to calculate than the complete protein structure and can be used for phase determination.

Once the phases have been estimated, an electron density map of the protein can be obtained. A model that fits this data is then built, and if built correctly, this model then provides improved phases for calculation of an improved electron density map. The model is refined against the data, incorporating restraints limiting excessive variation of bond lengths, bond angles and other known protein structural parameters. As a final step, the structural model of the protein will be validated. The geometry of the final model will be analysed to detect potential outliers in rotamer prevalence, Ramachandran configurations, close contacts and other parameters that might have been introduced by the building or refinement process (Kleywegt, 2000). If the geometry of the model is correct, and the model to data statistics are reasonable, the model is deposited in the PDB, the Protein Data Bank (www.pdbe.org).
**Endolysin structure**

Catalytic domains of endolysins belong to one of the following enzyme families (Loessner, 2005; Fenton et al., 2010): N-acetylμuramidases (lysozymes), N-acetyl-β-D-glucosaminidases (glycosidases), N-acetylμuramoyl-L-alanine amidases, L-alanoyl-D-glutamate endopeptidases and interpeptide bridge-specific endopeptidases (Figure 2). Single-domain endolysins are exclusive to phages infecting Gram-negative bacteria and consist of just one catalytic domain. However, most endolysins exhibit a modular configuration with one or several catalytic domains and a cell wall binding domain. These domains are connected by flexible linkers. The lytic spectrum of an endolysin depends on its cell wall binding domain and on the conservation among bacteria of the cell wall motif that the cell wall binding domain recognises (Nelson et al., 2012, Navarre et al., 1999; O’Flaherty et al., 2005; Fernandes et al., 2018).

![Figure 2](image)

**Figure 2**: Enzymatic activities of peptidoglycan degrading enzymes. The basic structure of Gram-negative cell wall peptidoglycan, indicating the possible enzymatic activities of peptidoglycan-degrading enzymes and the bonds they cleave.
**N-acetylglucosaminidases or glycosidases**

N-acetylglucosaminidases hydrolyse the β-(1→4) glycosidic bond between N-acetylmuramic acid (MurNAc) and N-acetylglucosamine (GlcNAc). The only structure of a phage endolysin showing a N-acetylglucosaminidase domain is PlyC, the lysin from *Streptococcus* phage C1 (McGowan et al., 2012; Broendum et al., 2018). PlyC is a powerful multimeric endolysin that consists of two separate proteins: a single copy of the 50 kDa PlyCA subunit and at least eight copies of the 8 kDa PlyCB subunit. The eight PlyCB subunits form the cell wall binding domain of the protein and are arranged in a planar octameric ring 80 Å in diameter and 20 Å high. The PlyCA subunit contains two different domains, a carboxy-terminal cysteine, histidine-dependent amidohydrolase/peptidase (CHAP) domain and an amino-terminal N-acetylglucosaminidase domain. The N-acetylglucosaminidase domain is structurally related to N-acetylglucosaminidase domains from the glycosyl hydrolase 73 (GH73) family, with a highly conserved core formed by six α-helices. This core contains a conserved glutamic acid residue (glutamate 78) essential for catalytic activity and a substrate-binding groove (Figure 3A).

**N-acetylmuramidases (lysozymes or muramidases)**

N-acetylmuramidase domains hydrolyse the β-(1→4) glycosidic bond between N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) in the sugar moiety of the cell wall. Phage muramidase domains belong to at least three different families of glycosyl hydrolases called GH19, GH24 and GH25. Most belong to the GH25 family, including Cpl-1, the endolysin from the *Streptococcus pneumoniae* phage Cp-1 (Perez-Dorado et al., 2007). The Cpl-1 structure is formed by two domains connected to each other with a linker. The cell wall binding domain presents six choline-binding repeats. The muramidase domain forms an irregular (β/α)_5β_3 barrel, in which the catalytic residues are located. All members of the GH25 family share this barrel fold. Hydrolysis in this group of muramidases takes place via a general acid/base mechanism that requires two acidic amino acids. One acidic residue behaves as the proton donor and the other acts as the nucleophile, promoting hydrolytic attack by a water molecule. These residues are located in a catalytic groove and are
aspartate 10 (Figure 3B, left).

**Figure 3.** Representation of the variety of enzymatically active domains in phage endolysins. The proteins are shown in cartoon representation, coloured according to their secondary structure. α-helices are represented in red, β-sheets in yellow and loops in green. Carbon atoms of catalytic residues are depicted in dark blue, carbon atoms of ligands in cyan, zinc ions as dark grey or black spheres and calcium ions as orange spheres. A. N-acetyglucosaminidase domain from the streptococcal phage C1 lysin PlyC (PDB entry 4F88). B. Different families of N-acetylmuramidases. Left: Cpl-1, endolysin from the *Streptococcus pneumoniae* phage Cp-1, representative of the GH25 family, (PDB entry 2IXU). Centre: T4 lysozyme, representative of the GH24 family (PDB entry 4LZM). Right: Endolysin from *Salmonella* phage SPN1S, representative of the GH19 family (PDB entry 4OK7). C. Different families of N-acetylmuramoyl-L-alanine amidases. Left: T7 lysozyme, representative of the Pfam amidase_2 family (PDB entry 1LBA). Right: CD27L, *Clostridium difficile* φCD27 phage endolysin, representative of the VanY d-Ala-d-Ala carboxypeptidase subfamily (PDB entry 3QAY). D. Different families of peptidases. Left: CHAP\_K, CHAP domain of the staphylococcal phage K endolysin LysK (PDB entry 4CSH). Right: pLy500, L-alanoyl-D-glutamate peptidase from *Listeria* phage A500 (PDB entry 2VO9), representative of the M15C family of endopeptidases.

strongly conserved. In the case of Cpl-1, the residues are glutamate 94 and
One of the most studied phage muramidases is T4 lysozyme (Bell et al., 1991). T4 lysozyme has been studied in depth and exhaustive mutational analyses have been performed to infer the function of every residue (Rennell et al., 1991). This enzyme is an inverting glycoside hydrolase that belongs to the GH24 family. T4 lysozyme is formed by two domains, that can move with respect to each other. The active-site cleft is located in the interface between the domains. Without the peptidoglycan substrate, T4 lysozyme remains in its open conformation. After peptidoglycan is added, the lysozyme transitions to a closed processive conformation. The change of conformation is caused by an 8 Å hinge motion that brings the two lobes of the enzyme closer to each other (Akhterov et al., 2015). Once the enzyme is in a closed conformation, a nucleophilic water molecule, activated by aspartate 20, attacks the \( \beta-(1\rightarrow4) \) linked anomeric C1 carbon of the peptidic N-acetyl muramic acid, while the general acid glutamate 11 donates a proton to the leaving aglycone, N-acetyl glucosamine (Kuroki et al., 1995). Mediated by hydrogen bonds, threonine 26 helps to position the water molecule that performs the nucleophilic attack (Figure 3B, centre).

The endolysin from *Salmonella* phage SPN1S (Park et al., 2014) belongs to the GH19 chitinase family, although it has muramidase activity. It is a helical monomer with a Pacman-like shape, formed by two domains: a large domain made up of the amino-terminal and carboxy-terminal regions of the protein and a smaller domain made up of the central part of the protein. The large domain is formed by a total of eight \( \alpha \)-helices. Helix \( \alpha 3 \) is located in the central part of the domain, surrounded by helices \( \alpha 1, \alpha 2 \) and \( \alpha 10-\alpha 13 \). The small domain is formed by the five central \( \alpha \)-helices (\( \alpha 4-\alpha 8 \)) and a \( 3_{10} \)-helix. Helices \( \alpha 4, \alpha 6 \) and \( \alpha 7 \) form an antiparallel three-helix bundle whose axis is perpendicular to one of the bundles formed by the helices in the large domain. These two regions are connected by two long loops. As a result, a deep wide groove is formed between both domains, where the catalytic cleft is located. The catalytic residues from family 19 chitinases are two glutamates. A glutamate located at the end of the third \( \alpha \)-helix (glutamate 49 in SPN1) acts as the catalytic general acid, whereas another glutamate located in a more variable loop-like structure (glutamate 58 in SPN1S) acts as the catalytic general base (Figure 3B, right).
Acetylmuramoyl-L-alanine amidases

Amidases cleave the amide bond connecting the sugar and peptide moieties of the bacterial cell wall. According to the classification performed by Broendum et al. (2018), two different zinc-dependent N-acetylmuramoyl-L-alanine amidase folds exist. The most common amidase domain belongs to the Pfam amidase_2 family (Firczuk et al., 2007). Although several phage endolysins contain this kind of fold, the more studied one has been T7 lysozyme (Cheng et al., 1994). These enzymes contain a five-stranded, mostly parallel β-sheet with conserved topology, flanked by a varying number of α-helices. The active site zinc ion is located on one of the sides of the β-sheet, at the bottom of the peptidoglycan binding groove. The zinc ion is tetrahedrally coordinated to a water molecule, two histidine residues and a cysteine residue. These residues are highly conserved (Figure 3C, left).

Other amidases, like the Clostridium difficile φCD27 phage endolysin CD27L (Mayer et al., 2011), belong to the VanY d-Ala-d-Ala carboxypeptidase subfamily. Despite belonging to this subfamily, they have N-acetylmuramoyl-L-alanine amidase activity. They exhibit an α/β fold, with a core of six beta strands surrounded by five alpha helices. A zinc ion is present in the catalytic groove, coordinated by two histidine residues and a glutamic acid. In the case of CD27L, these residues are histidine 9, histidine 84 and glutamate 26 (Figure 3C, right).

Peptidases

Peptidases cleave the peptide moiety of the peptidoglycan. Depending on the links in the peptidoglycan they target, there are two types, L-alanoyl-D-glutamate endopeptidases and interpeptide bridge-specific peptidases. Most of the domains with peptidase activity found in phage endolysins are histidine-dependent amidohydrolase/peptidase (CHAP) domains. CHAP domains show a great versatility and can have N-acetylmuramoyl-L-alanine amidase or D-alanyl-glycyl endopeptidase activities (Bateman et al., 2003). CHAP domains are formed by two α-helices clustered against a six-stranded antiparallel β-sheet, accommodating a long and deep hydrophobic catalytic groove between them. The catalytic triad is formed by a cysteine, a histidine and
such as CHAP_{K} LysK (Sanz-Gaitero et al., 2014), contain a calcium ion bound to an EF-hand from the long loop connecting the first and second \alpha

phage C1, do not contain a calcium binding motif (McGowan et al., 2012).

CHAP domain is pLy500, the L-alanoyl-D-glutamate peptidase from resembles a sofa, with an \alpha-helix and the three antiparallel \beta the ‘seat’. Another \alpha additional loop conforms to the 'left armrest' of the sofa. A catalytically zinc

a groove formed between the ‘seat’ and the ‘backrest’ (Figure 3D, right).

Cell wall binding domains

lower (Fischetti, 2008).

neighbouring cells that have not yet produced phage particles (Loessner et al. 2002).

(Broendum et al., 2018):
1. SH3b domains, with a β-barrel fold, which can contain between five and seven β-strands connected by linkers. PlyPSA, the endolysin from the Listeria phage PSA (Korndoerfer et al., 2006) contains two SH3b repeats in its cell wall binding domain. These copies are held together by means of swapped β-strands (Figure 4A).

2. Choline binding modules (ChBr), often found in pneumococcal endolysins such as Cpl-1, which recognise the teichoic acids of the pneumococcal cell wall (Perez-Dorado et al., 2007). They are formed by multiple tandem copies of a short amino acid sequence. The Cpl-1 cell wall binding domain contains an amino-terminal super-helical subdomain formed by the first four ChBr repeats, where the choline-binding pockets are located, and a carboxy-terminal β-sheet that interacts with the catalytic domain (Figure 4B).

3. Three-helix bundles. In endolysins targeting Gram-positive bacteria, such as the Streptococcus phage CP-7 endolysin Cpl-7 (Bustamante et al., 2017), several copies of this fold compose the cell wall binding domain. Endolysins targeting Gram-negative bacteria only contain one copy of the three-helix bundle (Figure 4C).

4. α/β multimers. Several endolysin cell wall binding domains show a similar fold consisting of a central parallel or antiparallel four-stranded β-sheet flanked by two helices. One example of this kind of fold is PlyC, the lysin from the streptococcal phage C1 (McGowan et al., 2012). These cell wall binding domains oligomerize to form a functional assembly (Figure 4D).

Endolysin variability
Phage-encoded endolysins show a lot of diversity. They can contain a single globular catalytically active domain or combine different cell wall binding domains and enzymatically active domains to form a multi-modular endolysin. This flexibility in their design makes phage endolysins highly versatile proteins, with different catalytic activities, host ranges and stabilities. A comparative genomic analysis of 723 endolysins identified 24 different enzymatically active domains and 13 cell wall binding domain types arranged in 89 different architectures, highlighting the high variety among phage endolysins (Oliveira et al., 2013).
**Figure 4.** Representation of the variability of cell wall binding domains in phage endolysins. The proteins are shown in cartoon representation, coloured according to their secondary structure. α-helices are represented in red, β-sheets in yellow and loops in green. **A.** PlyPSA, endolysin from the *Listeria* phage PSA (PDB entry 1XOV) contains two SH3b copies in its cell wall binding domain. **B.** Choline binding module (ChBr) in the cell wall binding domain of Cpl-1, the endolysin from the *Streptococcus pneumoniae* phage Cp-1 (PDB entry 2IXU). **C.** *Streptococcus* phage CP-7 endolysin Cpl-7 cell wall binding domain showing three copies of a three-helix bundle (PDB entry 5I8L). **D.** Dimer of PlyCB, the cell wall binding domain of the endolysin from the *Streptococcus* phage C1 (PDB entry 4F87), forming α/β multimers.

The multi-modular nature of many endolysins offers the opportunity to design tailor-made proteins with properties adapted for different applications, by combining, truncating or mutating these domains. Due to the flexibility between modules, not many full-length multi-modular endolysins structures have been published. A common approach in structure determination of multi-modular
and at least eight copies of the 8 kDa PlyCB subunit (see the section N-acetylglucosaminidases). Enterococcus

ribosomal binding site within the gene.

Clostridium difficile phage φCD27 endolysin CD27L (Dunne et al.

independently of the cell wall binding domain.

The structures from two different ‘SAR’ endolysins have been solved using X-ray crystallography, Lyz of coliphage P1 (Xu et al., lambdoid phage 21 (Sun et al., 2009). These enzymes contain an amino-
form. Dissipation of the proton motive force causes the transmembrane domain to be released from the cell membrane. After leaving the bilayer, the protein refolds into a periplasmic active form (Fernandes et al., 2018). The holins facilitate this process by mediating the collapse of the proton motive force (Xu et al., 2004; Xu et al., 2005; Sun et al., 2009).

**Conclusion**

X-ray crystallography has contributed significantly towards the knowledge of endolysin structure and mechanism. Structures of catalytic domains have provided clues as to their enzymatic mechanisms, while structures of cell wall binding domains have informed us as to how they specifically recognise the peptidoglycan of their bacterial hosts. However, structures of endolysin domains complexed with peptidoglycan substrates are still scarce. Future improvements in the organic synthesis of peptidoglycan analogues will hopefully fill this void. In any case, structural biology in general, and X-ray crystallography in particular, will continue to play a role in determining structures of endolysins, endolysin domains and their complexes with substrate analogues. In turn, this structural information will guide the design of more stable and more efficient enzybiotics to eradicate pathogenic bacteria and also positively influence the use of endolysin-based bacterial detection methods.

**Acknowledgements**

This work was supported by grant BFU2017-82207-P from the Spanish Ministry of Science and Innovation to MJvR and by a RISAM PhD fellowship from the Cork Institute of Technology to MS-G.

**References**


Everett, J.K., Tejero, R., Murthy, S.B., Acton, T.B., Aramini, J.M., Baran, M.C., Benach, J., Cort, J.R., Eletsky, A., Forouhar, F., Guan, R., Kuzin, A.P., Lee,


application to structural biology (New York, USA: Garland Science).


