| Title | A novel activation mechanism for Clostridial bacteriophage endolysins |
|---|---|
| Author(s) | Dunne, Matthew Stephen |
| Publication date | 2014 |
| Original citation | Dunne, M. S. 2014. A novel activation mechanism for Clostridial bacteriophage endolysins. PhD Thesis, University College Cork. |
| Type of publication | Doctoral thesis |
| Rights | © 2014, Matthew S. Dunne.<br>http://creativecommons.org/licenses/by-nc-nd/3.0/<br><br>![CC BY NC ND] |
| Item downloaded from | http://hdl.handle.net/10468/1925 |

University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

# A novel activation mechanism for *Clostridial* bacteriophage endolysins

A thesis submitted to the National University of Ireland, Cork, in fulfilment of the requirements for the degree of

**Doctor of Philosophy**
by

**Matthew Stephen Dunne**

The European Molecular Biology Laboratory (EMBL), Hamburg Outstation and the Department of Biochemistry, University College Cork

**April 2014**

Supervisors:
Dr. Rob Meijers, EMBL Hamburg
Dr. Thomas Moore, University College Cork

Head of Department: Professor Thomas Cotter

Coláiste na hOllscoile Corcaigh, Éire
University College Cork, Ireland

EMBL
4O YEARS I 1974–2014

# Table of Contents

# Declaration

I hereby declare that the content of this thesis is the result of my own work and has not been submitted for another degree, either at University College Cork or elsewhere.

Signed: _____

       Matthew Stephen Dunne

*The work in this thesis is dedicated to Mum, Dad and Greg   x*

# Thank you

Firstly I want to thank the boss, Dr. Rob Meijers, for his continued guidance during my studies and for providing me with what has been a very interesting project. Thanks to Tom Moore, my supervisor from University College Cork, for providing his knowledge and ideas to help guide the research. Thanks go to my PhD examiners, Prof. David Sheehan and Prof. Martin Loessner, for the generous 3 hour 'grilling' during the defence. Thanks to Dr. Melinda Mayer for the great collaboration and all the lysis assays during the PhD. Thanks to UCC and the EMBL PhD programme for allow me to pursue a PhD. Although challenging at times, the freedom to develop as a scientist within EMBL has been fun and rewarding and the inventive science being performed by my peers at EMBL has always acted as a motivating factor towards my own research. Thanks to The University of Birmingham for a fantastic introduction into the world of biochemistry and a particular thank you to my Bachelor's dissertation supervisor Dr. Scott White for introducing me to structural biology and recommending I pursue a PhD.

A big thank you goes to Dr. Nina Krueger for all her help, advice and laughter throughout my PhD and for teaching me how to make the 'perfect' SDS-PAGE gels. Thanks also to the rest of the group: Rebecca, Heidi, Anna, Ben, Magda, Alex and Maria for their support, entertainment and boozy nights out. Thanks to Stefan, Cy and Haydyn for helping contribute to the first (of hopefully many) publications. Thanks to the SPC team: Sandra, Stephane, Vanessa, Gry, Eleonora and Jessi for always finding time for that 'one last' MS sample, even on a Friday afternoon just before the beer session starts! Thanks as well to everyone else from the Hamburg outstation for providing a fun and energetic place to work in.

Thanks to everyone from the VAF Triathlon group for the entertaining training sessions and two great summers competing in the Landesliga, whether it was raining, storming, snowing or occasionally a spot of sunshine we were out on the bikes, running around Volkspark or diving into the AppenSee. *Wunderbar*! A huge Hurricanes's *'Prost'* to all the guys from Hamburg Hurricanes S.V. for the great nights out and a well fought couple of seasons in the Freizeitliga.

Last, but definitely not least, a big thanks to the support from friends and family back home: Mum, Dad, Greg, Grandma, Nana, Granddad, Hannah for always being there; Angela, thanks for the proofread; Luke, Dean & Mark for the annual S.Head trips; Ed, Stu & Al for being great mates and of course all the guys and girls from the annual King's 12 Pubs of Xmas Crawl ;). Spot on.

# Abstract

Bacteriophage-encoded endolysins are produced at the end of the phage lytic cycle for the degradation of the host bacterial cell. Endolysins offer the potential as alternatives to antibiotics as biocontrol agents or therapeutics. This thesis investigates the lytic mechanism for three bacteriophage endolysins that target *Clostridium* species living under different conditions. For these three *Clostridial* endolysins a trigger and release mechanism is proposed for their activation. During host lysis, holin lesion formation suddenly permeabilises the membrane which exposes the cytosol-sequestered endolysins to a sudden environmental shock. This shock is suggested to trigger a conformation switch of the endolysins between two distinct dimer states. The switch between dimer states is proposed to activate a novel autocleavage mechanism that cleaves the linker connecting the N-terminal catalytic domain and the C-terminal domain. The release of the catalytic domain is proposed to enable more efficient digestion of the bacterial cell wall for more effective release of viral progeny. Crystal structures of autocleaved fragments of CD27L and CTP1L were previously obtained. In these structures, cleavage occurs at the stem of the linker connected to the C-terminal domain. Despite a sequence identity of only 22% between 81 residues of the C-terminal domains of CD27L and CTP1L, they represent a novel fold that is identified in a number of different lysins that target *Clostridia*. Within the crystal structures, the two distinct dimerisation modes are represented: the elongated head-on dimer and the side-by-side dimer. Introducing mutants that inhibit either of the dimer states causes a decrease in efficiency of both the autocleavage mechanism and the lytic activity of the endolysins. The two dimer states were validated for the full-length endolysins in solution by using right angle light scattering, small angle X-ray scattering and cross-linking experiments with the photo-activated cross-linker p-benzoyl-phenylalanine. Turbidity assays with CTP1L and the dimer mutants indicated that a decrease in lytic activity was proportional to a decrease in cleavage efficiency. The autocleavage mechanism of CD27L involves a peptide hydrolysis reaction of the peptide bond between methionine and glutamate residues. Interestingly, mass spectrometry revealed an additional mass of 32 Da on the cleaved C-terminal domains of CTP1L and CS74L. Using anomalous sulphur scattering of the CTP1L crystal, the 32 Da mass was identified to be due to incorporation of a methionine residue, which is proposed to be a post-proteolytic modification of the cleaved C-terminal domains. Altogether the data represents a new type of regulation for these endolysins that is governed by the C-terminal domains.

# Abbreviations

| | |
|---|---|
| BLAST | Basic Alignment Search Tool |
| CBD | Cell wall binding domain |
| CDI | *C. difficile* infections |
| CWM | Cell wall material |
| DTNB (Ellman's reagent) | 5,5'-dithiobis-(2-nitrobenzoic acid) |
| E value | Expected value |
| EAD | Enzymatically active domain |
| ESI-LC-MS/MS | Liquid chromatography coupled to electrospray ionisation tandem mass spectrometry |
| GlcNAc | N-acetylglucosamine |
| HEK cells | Human embryonic kidney cells |
| LC-MS/MS | Liquid chromatography tandem mass spectrometry |
| MALDI-TOF | Matrix-assisted laser desorption/ionisation time-of-flight |
| MRSA | Methicillin-resistant *Staphylococcus aureus* |
| MS | Mass spectrometry |
| MurNAc | N-acetylmuramic acid |
| MW | Molecular weight |
| MWC mechanism | Monod-Wyman-Changeux mechanism |
| Ni-NTA | Nitrilotriacetic acid (affinity purification) |
| pBpa | *p*-benzoyl-l-phenylalanine |
| pmf | proton motive force |
| RALS/RI/UV measurements | Right-angle light scattering/refractive index/ultraviolet measurements |
| SAR domain | Signal arrest release domain |
| SAXS | Small-angle X-ray scattering |
| SDS-PAGE | Sodium dodecyl sulfate polyacrylamide gel electrophoresis |
| SEC | Size exclusion chromatography |
| SUMO3 | Small ubiquitin-related modifier 3 |

# Chapter 1: General Introduction

*A brief history of bacteriophage therapy*

Nearly a century has passed since the first published accounts of bacteriophage as agents of bacterial pathogenicity. In 1915 Fredrick Twort, working at the Brown Institution of London, published the first result of bacteriophages (also known as phages). Twort noticed that colonies of *Micrococcus* infected with filter passed viruses became transparent, an indicator of bacterial cell death. Twort also reported that this phenomenon was transferable between bacterial colonies (Twort 1915). In 1917, two years after Twort's observations, Felix d'Herelle independently reported on bacteriophage as exogenous agents of immunity while at the Pasteur Institute in Paris (d'Herelle 1917). D'Herelle proposed the term bacteriophage which means "bacteria eater," to describe their bactericidal action. In 1923 d'Herelle helped George Eliava create the Bacteriophage Institute in Georgia (now the Eliava Institute), which continues to research phage therapy and its applications, as well as producing phage preparations for treatment against a variety of bacterial infections (O'Flaherty *et al.* 2009). Between 1920 and 1940 antibiotics were still in their developmental stages allowing phage therapy to take centre stage for the treatment of bacterial infections. The Bacteriophage Inquiry, between 1927 and 1936, was one of the largest investigations engaged between numerous research groups and hospitals into the prophylactic use of phage to control cholera outbreaks in India (Summers 1993; Kutter & Sulakvelidze 2004; Nelson *et al.* 2009). Even with numerous positive results the enthusiasm for phage therapy failed to progress as the treatments did not live up to the earlier hype surrounding phage therapy.

A major reason for the decline in phage therapy research was the advent of sulfonamides in the 1930s and the arrival of penicillin and other antibiotic treatments in the 1940s (Deresinski 2009). Antibiotics became the routine treatment for bacterial infections and interest in phage therapy research significantly declined. Western medicine still continued to moderately research phage therapy. In 1982 Smith & Huggins showed intramuscular doses of a bacteriophage (anti-K1 phage) had improved *in vitro* and *in vivo* activity against *Escherichia coli* when compared to doses of antibiotics (ampicillin, tetracycline, streptomycin and trimethoprim/sulfamethoxazole) (Smith & Huggins 1982). Phage therapy has continued to be employed in Eastern Europe, particularly at the Eliava institute, which has reported success with phage therapy, treating a range of infections such as methicillin-resistant *Staphylococcus aureus* (MRSA) (Parfitt 2005).

### *The reemergence of bacteriophage therapy*

Due to the heavy use of broad-spectrum antibiotics, reports of antibiotic resistance in bacteria, particularly in hospital environments, are widespread and increasing every year (Davies 2006). Antibiotic resistance has been reported in a number of pathogenic bacteria including MRSA, *Salmonella*, *Mycobacterium tuberculosis*, *E. coli* and *Clostridium difficile* (O'Flaherty *et al.* 2009). Accelerated distribution of resistance genes within pathogenic and commensal bacterial communities is associated with the lack of antibiotic moderation for minor infections, antibiotic exploitation for non-therapeutic applications such as agriculture and the fact pharmaceutical companies are investing less in discovering novel antibiotics due to short returns on investments (Davies 2006).

Over the past 15 years there has been a resurgence of interest towards the use of phage not just as therapeutics but also integrated into modern biotechnology, food science and agriculture. Combined with the high throughput sequencing of phage genomes, our knowledge of phage biology, particularly concerning phage-bacteria interactions and their ability to target bacterial infections, has been used to inspired the reemergence of phage as potential therapeutics and biocontrol agents (O'Flaherty *et al.* 2009). PhagoBioDerm® (PolymerPharm) is a biodegradable dressing used as a topical treatment which consists of a mix of antibiotics and virulent phages that target *E. coli, P. aeruginosa, S. aureus* and Streptococcus infections (Kutter & Sulakvelidze 2004). Phages are also being used for agricultural applications, where phage use is not as regulated as medical applications (O'Flaherty *et al.* 2009). The contamination of food by bacteria such as *Salmonella*, *Campylobacter jejuni*, *E. coli* and *Listeria* remains a persistent health problem worldwide where phage therapy has the potential to control pathogenic bacteria, improving food safety and reducing food waste. Listeriosis, primarily an infection of the central nervous system caused by *Listeria monocytogenes* can be fatal. In 2006 ListShield™ (Intalytix, Inc.) was approved by the food and drug administration as a phage cocktail that targets *L. monocytogenes* contaminants in meat and poultry products (Ledford 2009).

As bacteriophages have co-evolved with their bacterial hosts for over 1 billion years, the continuous adaptations made by phage and host bacteria can limit phage lethality. Bacteria can modify their cell surface receptors, produce exopolysaccharides that coat the bacteria to block adsorption, inhibit DNA injection via the superinfection exclusion (Sie) system or degrade viral DNA by employing restriction enzymes (Labrie *et al.* 2010). Therefore, many bacteriophages isolated from the host environment are not efficient in the rapid

eradication of pathogenic bacteria. The major drawback of whole phage therapy occurs if bacteria develop resistance against phage attachment and phage production (Fischer *et al.* 2013).

The possibility remains either to genetically engineer more potent bacteriophages to tackle viral-resistance of bacteria or to use the individual phage components that contribute to bacterial lysis. Phages can influence critical metabolic processes within bacteria to direct the production of their own viral progeny. These individual systems of viral manipulation offer the potential to be developed into antimicrobial agents (Fischetti 2006). As more viral genomes are characterised so are their respective lytic mechanisms. The characterisation and isolation of proteins involved in bacterial lysis, particularly the holin-endolysin systems, are the most promising viral components that could be developed into antimicrobials. Normally endolysins are release from within infected cells during the lytic cycle, however, they also demonstrate a high lytic activity even when small amounts are applied to dense suspensions of Gram-positive cells, rapidly lysing cells within seconds (Loessner 2005). The development of endolysins as antimicrobials is accelerating, especially with the various lytic systems being investigated. Furthermore, the effectiveness of recombinant endolysins as antimicrobials has been demonstrated with *in vivo* murine studies against model infections of *Streptococcus pneumoniae* (Loeffler *et al.* 2001; Witzenrath *et al.* 2009).

***Composition and Morphology of Bacteriophage***

Phages are abundant and ubiquitous self-replicating units, with estimations of $10^{31}$ phage particles in the biosphere, ten-fold higher than the estimated number of prokaryotes (Whitman *et al.* 1998). Phages have a large taxonomic diversity, there are at least 17 known families of phages that infect 154 genera of host bacteria, with each phage species specific to certain bacterial hosts (Ackermann 2007). Over 5560 phages have been examined by negative staining electron microscopy to reveal their morphological diversity (Ackermann 2007). The type of genetic material and morphology varies depending on the phage species. The majority of phage have genomes consisting of double stranded (ds) DNA, although there are groups of phage with single stranded (ss) DNA, ssRNA or dsRNA (Kutter & Sulakvelidze 2004). Phage genomes are enclosed within a protein capsid to help protect the genetic material before insertion into the host cell. It is estimated that 96% of these phages have tails protruding from the capsid and belong to the order of

*Caudovirales* and the remaining 4% are either filamentous, polyhedral or pleomorphic (Ackermann 2007). *Caudovirales* have icosahedral heads (capsids) and a protruding tail, which acts as a tube through which the viral genome passes during infection of the host cell. *Caudovirales* can be sub-divided into three families depending on the composition of the tail. *Myoviridae* have contractile tails, *Siphoviridae* have long non-contractile tails and *Podoviridae* have short tail structures. The *Clostridial* phage examined in this thesis belong to the order of *Caudovirales,* described in more detail in Figure 1.
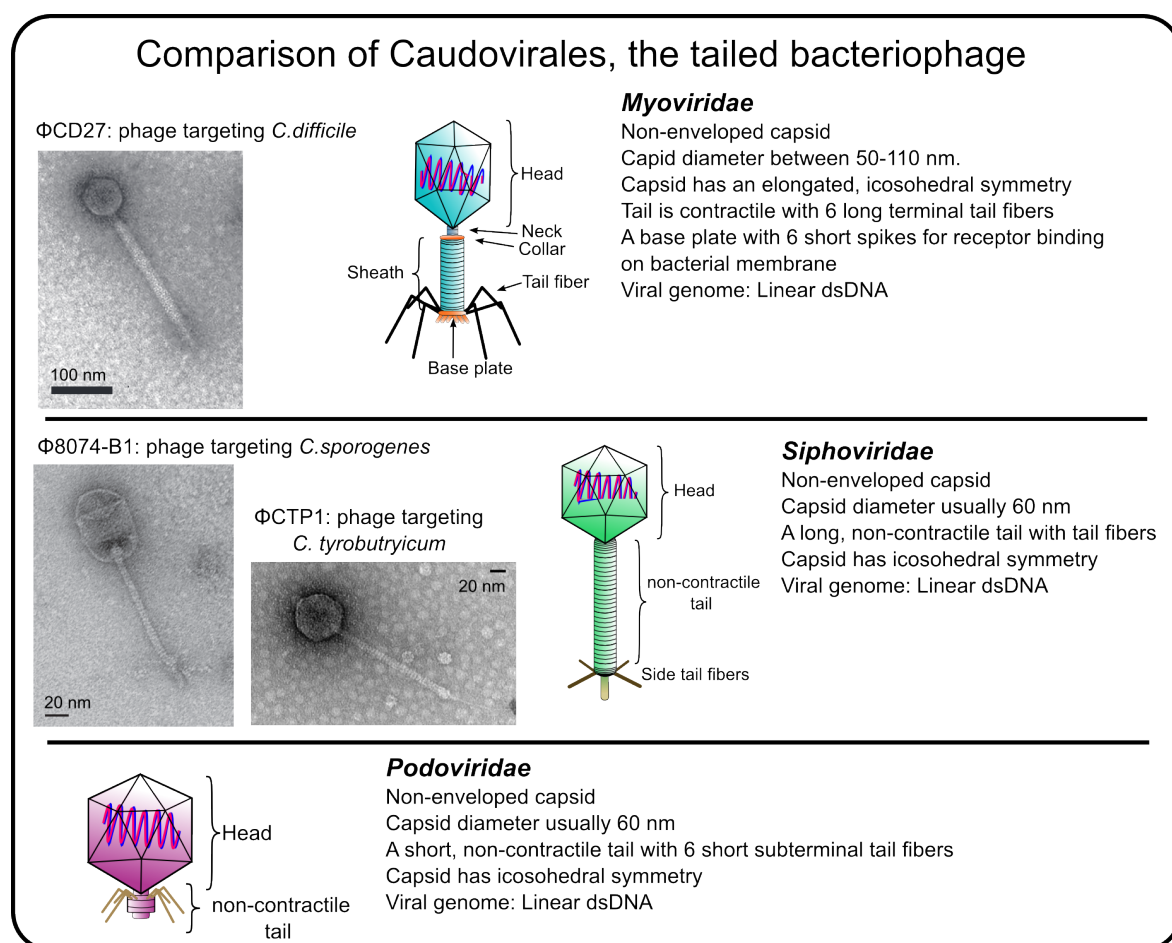


**Figure 1 The morphology of the three subfamilies of *Caudovirales*: *Myoviridaes*, *Siphoviridaes* and *Podoviridaes*.** *Clostridial* phages have been found belonging to all three subfamilies. For instance, ΦCD27 is a *Myoviridae* that infects *C. difficile* (Mayer *et al.* 2008) while ΦCTP1 and Φ8074-B1 are *Siphoviridaes* that infect *C. tyrobutyricum* and *C. sporogenes* respectively (Mayer *et al.* 2010; 2012). The morphology of ΦCD27 is distinct from ΦCTP1 and ΦCS74 due to a larger tail diameter (17 nm compared to ~ 7 nm) as a result of a contractible sheath on the tail, which allows the tail to act like a syringe when piercing the cell wall of the host. All three genomes are double-stranded DNA, similar to other *Clostridial* bacteriophage, and consist of approximately 50, 000 base pairs each. Electron micrograph images were reproduced with consent from Journal of Bacteriology (American Society for Microbiology) from Mayer *et al.* (2008; 2010; 2012).

***Clostridium: notorious toxin producers and clinical targets***

The *Clostridia* are anaerobic, Gram-positive and environmental bacilli commonly found in sewage, soil and the intestines of humans and animals (Hatheway 1990). *Clostridial* infections are notorious due to the range of diseases and toxins produced by certain members: *Clostridium perfrigens*, *Clostridium botulinum, Clostridium tetani and C. difficile* are the respective causative agents of gas gangrene, botulism, tetanus and colitis (Hatheway 1990).

*Clostridium tyrobutyricum* has been identified as the main species of bacterium associated with the late blowing defect of cheese, a major cause of food wastage (Klijn *et al.* 1995). *C. tyrobutyricum* spores present in raw milk can proliferate in the anaerobic conditions present during cheese production where the *C. tyrobutyricum* bacteria produce butyric acid and hydrogen gas, causing expansions within the cheese and defects in texture and taste (Mayer *et al.* 2010).

*Clostridium. sporogenes* is the name given to isolates of *C. botulinum* that lack the gene for the production of the botulinum neurotoxin, the cause of the potentially fatal botulinum. *C. sporogenes* is commonly used as a nontoxigenic surrogate for studying *C. botulinum* contaminations in food and for validating sterilisation procedures (Bradbury *et al.* 2012). *C. sporogenes* behaves in a similar manner to *C. tyrobutyricum* and is also associated with a range of food spoilage problems (Mayer *et al.* 2012). For instance, *C. sporogenes* was shown to enhance the butyric fermentation and cheese blowing associated with *C. tyrobutyricum* (Le Bourhis *et al.* 2007).

*C. difficile* infections (CDIs) are the cause of 20-30% of antibiotic associated diarrhea cases and nearly all incidents of pseudomembranous colitis, a potential life-threatening inflammation of the colon (Rea *et al.* 2013). The use of broad-spectrum antibiotics can result in an imbalance in the gastrointestinal tract microbiota, predisposing already susceptible patients to the risk of CDIs (Rea *et al.* 2013). CDI treatment typically involves using different combinations of antibiotics which can further reduce the intestinal defenses of patients, for instance the antibiotic metronidazole is commonly used to combat anaerobic bacteria such as *C. difficile*, however, in murine studies it actually predisposed the gastrointestinal tract to other colitis inducing bacteria such as *Citrobacter rodentium* (Wlodarska *et al.* 2011). Endospores of *C. difficile* can be released in faecal matter from infected patients and spread between susceptible patients to become an exogenous source

of CDIs (Rea *et al.* 2013). *C. difficile* endospores are of high-risk in a hospital environment, especially as endospores are resistant to fluctuations in temperature, oxygen acidity, deprivation of nutrients and chemical treatment, making them difficult to eradicate (Gerding *et al.* 2008).

### *Use of phage and endolysins against Clostridial colonisations and infections*

ΦCD27 and other *C. difficile* phage are unsuitable for phage therapy, as they are generally temperate (Rea *et al.* 2013), resulting in low phage production and prone to horizontal genetic transfer of phage DNA due to prophage incorporation into the host genome (Canchaya *et al.* 2003). As endolysins are protein therapeutics that do not rely on self-replication (O'Flaherty *et al.* 2009), an alternative approach is to use the phage endolysins, such as CD27L, to treat *C. difficile* infections. CD27L was successfully expressed in *Lactococcus lactis* cells (Mayer *et al.* 2008), however, for the successful application of endolysins they required secretion to target their specific hosts. Expression of *L. monocytogenes* endolysins containing a signal peptide were shown to be secreted from *L. Lactis* cells (Gaeng *et al.* 2000). Similarly, *L. lactis* cells have been engineered to express and secrete the LysH5 endolysin to kill *S. aureus* cells (Rodríguez-Rubio *et al.* 2012), although the technique required improved secretion, the same system could be used to eliminate *C. tyrobutyricum* cells from raw milk or added during cheese manufacturing to inhibit the colonisation of *C. tyrobutyricum*. Introducing the genes for endolysin expression into commensal or non-pathogenic bacteria for secretion in the GIT or in food preparation can be used to fight specific pathogenic bacteria.

*Bacteriophage life cycle*
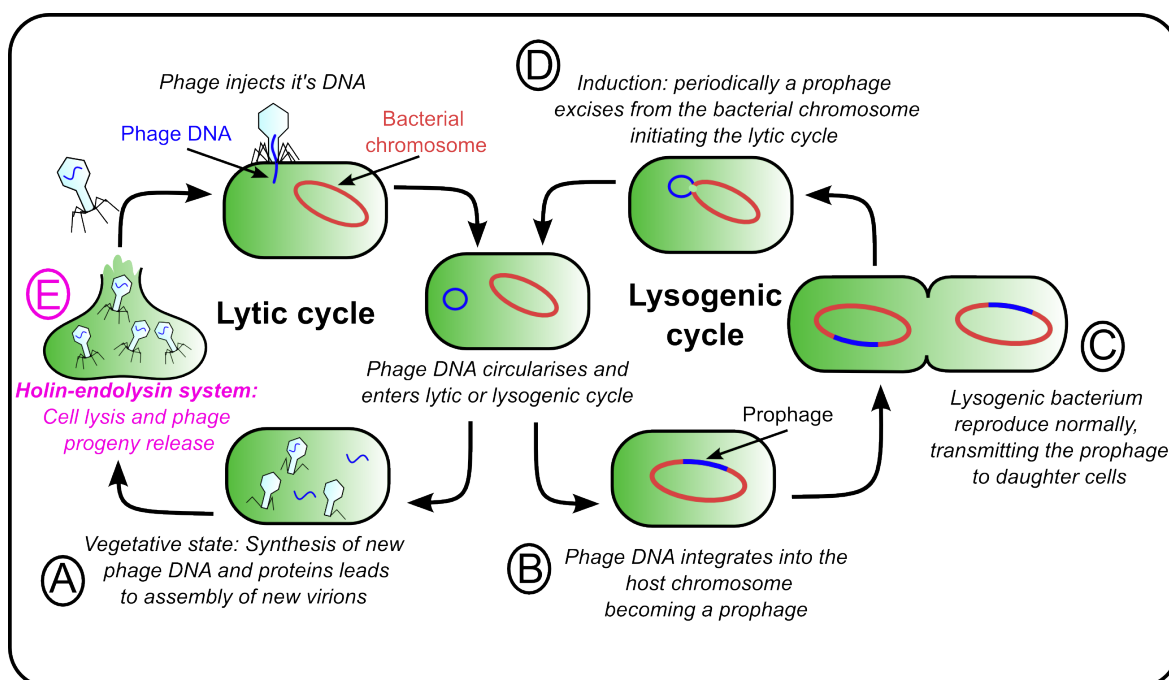


**Figure 2 The lytic and lysogenic life cycles of bacteriophages.** The different steps are described in the text below.

The first step in any phage life cycle is adsorption of the phage to the bacterial cell by binding of the tail fibres to exposed receptors, such as teichoic acid polysaccharides, found on the surface of bacterial cell walls (Rakhuba *et al.* 2010). Phage adsorption is species-specific, as phages recognise differences between receptor morphology, localisation and density of receptors on the bacterial cell wall (Braun & Hantke 1977). After binding to the bacteria, the hollow tail fibre penetrates the bacterial envelope and the viral DNA is injected into the bacterial cell.

Phage can be split into two distinct groups depending on their life cycle. Phage can have either a lytic cycle, used by virulent and temperate phage, which leads directly to viral replication and cell death, or a lysogenic cycle used only by temperate phages (Figure 2). The lysogenic cycle involves integration of the phage genome (prophage) into the bacterial host's genome (Figure 2, A). The bacterial host continues normal growth and during replication the prophage gets copied with the host genome, transferring the prophage to daughter cells, which also transfer the prophage when they replicate (Figure 2, B). The prophage can excise itself from the host genome in a process called Induction and the prophage enters the lytic cycle (Figure 2, C). Mitomycin C can be used experimentally to

initiate prophage induction, the phage genome can then be isolated from host bacteria for sequencing (Otsuji *et al.* 1959).

During the lytic cycle the phage remains in a vegetative state. Phage DNA takes over the host's cellular machinery, making copies of the phage DNA and directing ribosomes to produce phage proteins for the self-assembly of phage progeny (Figure 2, D) (Lee 1996). The phage progeny remain sequestered within the cytoplasm until cell lysis by the dissolution of the bacterial cell wall and release of the phage progeny from the cell. Cellular lysis is an elegant and precisely timed event governed by the holin-endolysin system (Young 2002) (Figure 2, E).

### *The holin-endolysin system is localised on the lytic cassette*

Holin and endolysin genes are adjoined in the majority of dsDNA phage genomes, their location in the genome is referred to as the lytic cassette (Young *et al.* 2000). Holin is normally located upstream and in the same orientation as endolysin. Typically clustered within the lytic cassette of phage that infect Gram-negative bacteria are other lytic genes, including antiholin, that help regulate lytic function (Young 2002). The *Clostridial* phage genomes of ΦCD27 and ΦCTP1 have two component cell lysis cassettes comprising of genes for holin and endolysins, with the holin gene immediately upstream of the endolysin (Mayer *et al.* 2008; 2010). The genome of ΦCS74 also contains a gene for an endolysin (CS74L), however, no putative gene has yet been recognised for a holin protein (Mayer *et al.* 2012). Holin and endolysin are both under late gene transcriptional regulation and are produced near the end of the vegetative state. This was shown for *Listeria* endolysins by measuring relative mRNA levels for endolysin across the vegetative state of infected cells, lysis occurred after 60 to 70 minutes and endolysin mRNA was detected 15 to 20 minutes after phage infection (Loessner *et al.* 1995).
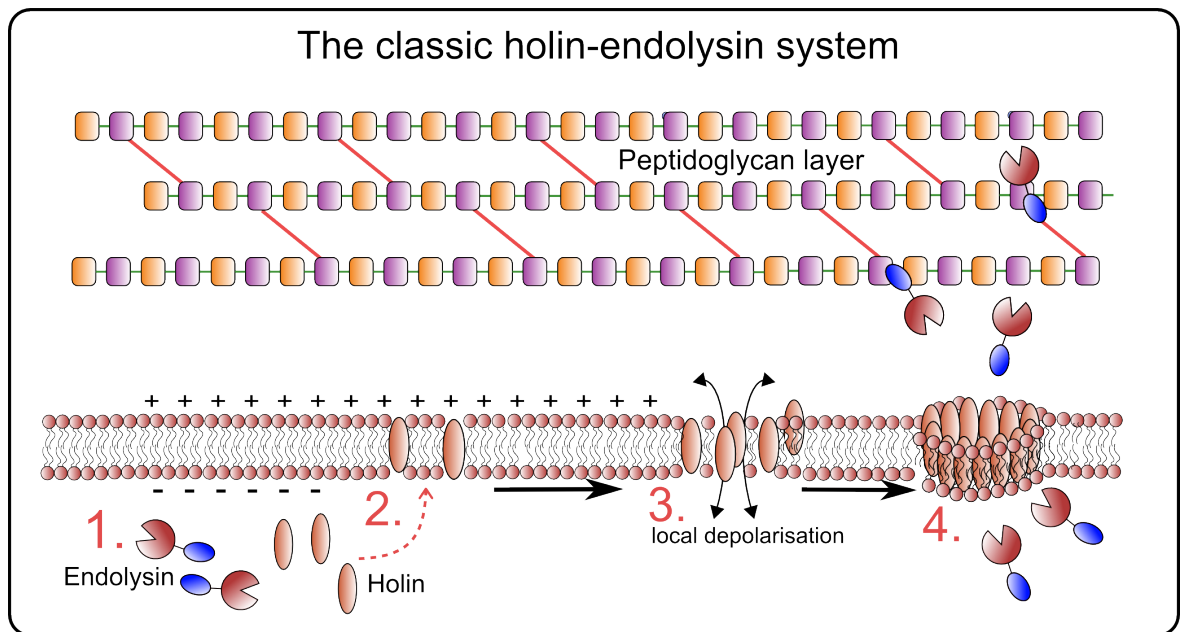
**Figure 3 The classic holin-endolysin activation mechanism based on the λ and T4 lysis paradigm**. The lytic process is described below.

### *The classic holin-endolysin system*

Holin and endolysin are expressed in parallel with the final viral progeny assembly at the end of the vegetative state (Figure 3, step 1). Endolysins remain folded but sequestered within the cytosol, while the holin proteins begin to accumulate throughout the cytoplasmic membrane (Figure 3, step 2). Within the membrane, holin proteins exist in a dilute phase as isolated proteins, free to move in the membrane with no effect on the membrane or its morphology. Holin proteins keep harmlessly entering the membrane until a critical concentration of holin is reached, represented as the time of lethal triggering (White *et al.* 2010) (Figure 3, step 3) At this point the holin proteins suddenly and spontaneously accumulate into large oligomeric aggregates, also described as rafts (Gründling *et al.* 2001). These large rafts cause depolarisation of the membrane and the collapse of the proton motive force (pmf) between the cytosol and the periplasm. The membrane stops supporting respiration due to the lack of a pmf and becomes permeable to ATP, ions and small molecules (Ziedaite *et al.* 2005). The collapse of the pmf is the trigger to initiate a conformational change within the holin rafts and the holin proteins rearrange to form lesions of varying size up to 300 nm in diameter (Figure 3, step 4) (White *et al.* 2010). These lesions are permeable to endolysins and other large molecules (Wang *et al.* 2003), which pass from the cytosol to the cell wall and begin to degrade peptidoglycan (Figure 3, step 4). Peptidoglycan degradation causes hypotonic lysis of the bacteria, as the cell wall can no longer withstand the osmotic internal pressure of the cell. Recently Dewey *et al.* (2010) used cryo-electron tomography to visualise holin lesions, which were highly

variable in size with an average diameter of 340 nm (± 35 nm), but were uniformly distributed and remained stable within the membrane. Essentially, holin lesion formation dictates the length of the vegetative state by controlling the release of the cytoplasmically sequestered endolysins into the cell wall (Wang *et al.* 2000).

### *Holin-independent export mechanisms*

The classic holin-endolysin system described for the paradigm systems of the Enterobacterial phage λ and T4 was thought to be the universal mechanism of cell lysis (Ramanculov & Young 2001). However, recently a number of novel lysis systems have been reported for Gram-positive, Gram-negative and mycobacterial targeting phages, whereby endolysins are transported across the membrane without passing through holin lesions.

### *Phages that target Gram-positive bacteria can secrete their endolysin using the Sec transportation system*

Endolysin Lys44, from phage fOg44 that targets the Gram-positive *Oenococcus oeni* encodes an endolysin (Lys44) that contains a hydrophobic N-terminal region similar to a Sec-type secretion peptide. Lys44 gets secreted and the secretion signal processed by the leader peptidase after translocation to the cell wall (Sao-Jose *et al.* 2000) (Figure 4). The 46 kDa pre-protein of Lys44 is only processed to the mature 43 kDa endolysin in the presence of the signal peptidase (LepB) and the translocase-associated ATPase (SecA), both essential components of the general secretion pathway (Sao-Jose *et al.* 2000). Processed Lys44 could be detected during the vegetative state of phage infection while the viral progeny are still self-assembling. Therefore, the secreted Lys44 has a regulatory system to keep it inactive once it passes across the membrane. A model was suggested whereby Lys44 is regulated in a similar manner to autolysins, whereby the ionised state of the membrane regulates autolysin activity (Calamita & Doyle 2002). The C-terminal domain of Lys44 is comprised of two LysM modules that mediate peptidoglycan binding and are suggested to keep the endolysin inactive by electrochemical gradients within the cell wall until holin-induced depolarisation (Parreira *et al.* 1999; Nascimento *et al.* 2008).
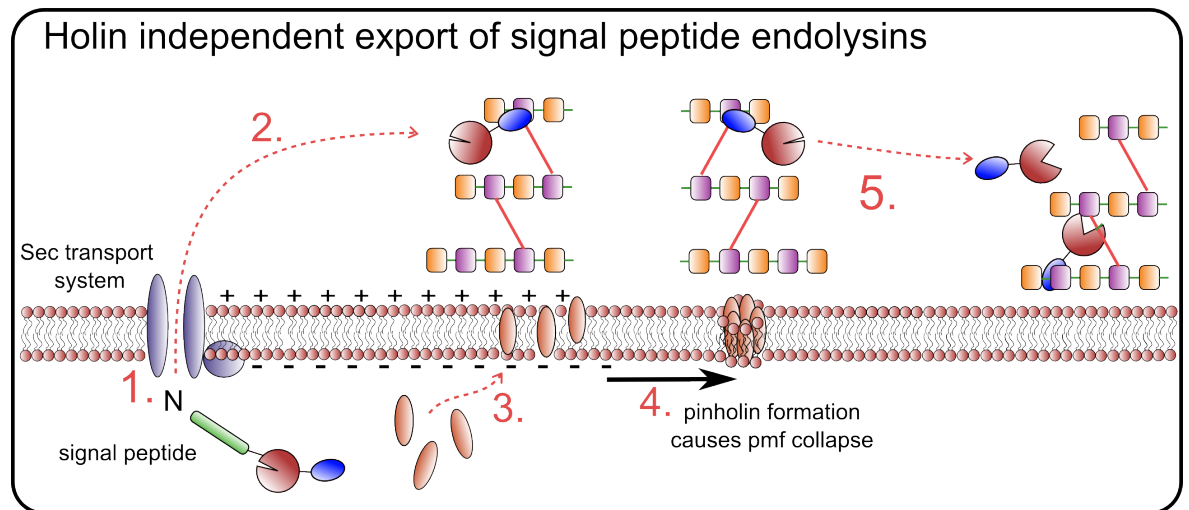
**Figure 4 Export of endolysins using signal peptides**. After Sec transportation (1) the endolysins bind to their respective cell wall ligands in an inactive state (2) such as Lys44, which has C-terminal LysM domains that bind peptidoglycan. (3) Holins or pinholins accumulate within the membrane until they reach the critical concentration leading to raft formation and progressing to lesion formation (4). Holin lesions permeabilise the membrane, disrupting electrochemical gradients across the cell wall and activating endolysin release and activation.

Activation of Lys44 was shown to require depolarisation of the pmf across the membrane (Nascimento *et al.* 2008). Nisin is a polypeptide antibiotic that forms 2 nm holes in cell membranes, allowing the efflux of ions across the bacterial membrane to mimic the effect of holin formation (Ruhr & Sahl 1985). *O. oeni* cells incubated with externally applied Lys44 could only be lysed upon addition of nisin, similarly in an *L. lactis* model, where Lys44 was secreted by the cells to the cell wall, lysis was only achieved after addition of nisin, the electrical gradient of the cellular membrane was therefore postulated to regulate activation of endolysin Lys44 (Nascimento *et al.* 2008) (Figure 4).

### *The Signal-Anchor-Release endolysins*

Similar classes of endolysins to the signal peptide containing Lys44 have been discovered in phages that target Gram-negative bacteria, that also use the Sec transportation system. Lyz[P1] and R[21], the respective lysozymes from coliphage P1 and 21, use an atypical N-terminal transmembrane domain that behaves as a <u>s</u>ignal <u>a</u>rrest <u>r</u>elease (SAR) signal to export the endolysins via the host Sec-dependent translation pathway (Xu *et al.* 2004; 2005). However, after translocation the signal peptide is not removed by signal peptidases, instead it behaves as a transmembrane domain and anchors the endolysin to the membrane in an inactive state, with the catalytic residues protruding into the periplasm (Kuty *et al.* 2010).

This unusual N-terminal SAR domain has an elevated amount of weakly hydrophobic residues (glycine, alanine, serine, and threonine) compared to normal transmembrane domains that allows a low spontaneous release of the endolysin from the lipid bilayer, that is proposed to gradually accomplish cell lysis (Park *et al.* 2006). Interestingly, the Lyz gene is expressed independently of the holin gene and is transported and membrane-tethered outside of the cell ahead of complete progeny phage assembly. Therefore, similar to Lys44, it has a regulatory mechanism to maintain Lyz$^{P1}$ in an inactive state until the trigger for lysis (Kuty *et al.* 2010) (Figure 5, A).

Sequence analysis showed Lyz$^{P1}$ was homologous to the T4 canonical lysozyme E, apart from a single modification within the catalytic triad where an aspartate residue was modified to a cysteine (Cys51), reverse mutation of this cysteine back to aspartate still demonstrated lytic activity (Kuty *et al.* 2010). Xu *et al.* showed that when the SAR domain is tethered within the membrane a disulphide bond forms between the catalytic residue Cys51 and another cysteine residue (Cys44), maintaining the enzyme in an inactive state. Contained in the SAR domain is another cysteine residue (Cys13) in its sulfhydryl form, upon membrane depolarisation the SAR domain gets released from the membrane and Cys13 reduces the disulphide bond between Cys51-Cys44. Disulphide isomerisation forms a new bond between Cys44-Cys13, releasing Cys51 in the sulfhydryl form, which activates the Lyz$^{P1}$ endolysin (Xu *et al.* 2005) (Figure 5, B).

The structural rearrangement required to activate Lyz$^{P1}$ was also observed. Crystallisation of full-length Lyz$^{P1}$ represented the active state, whereas crystallisation of the periplasmic domain (without the SAR domain) of Lyz$^{P1}$ represented the inactive state when the SAR domain would be tethered within the membrane (Xu *et al.* 2005). In the inactive state the Cys44-Cys51 bond could be observed, whilst the full-length active Lyz$^{P1}$ showed disulphide isomerisation, which involved a large structural rearrangement of three helices into a three beta strand sheet that repositioned residues within the catalytic cleft, activating the endolysin. Lyz$^{P1}$ could still be translocated across the membrane when its SAR domain was replaced with a signal sequence, however, it only became active upon the addition of reducing agents (Xu *et al.* 2004). SAR domains are an interesting example of transmembrane domains as they maintain a hydrophobic nature when tethered to the membrane but remain soluble once released to refold and activate the endolysin (Kuty *et al.* 2010).
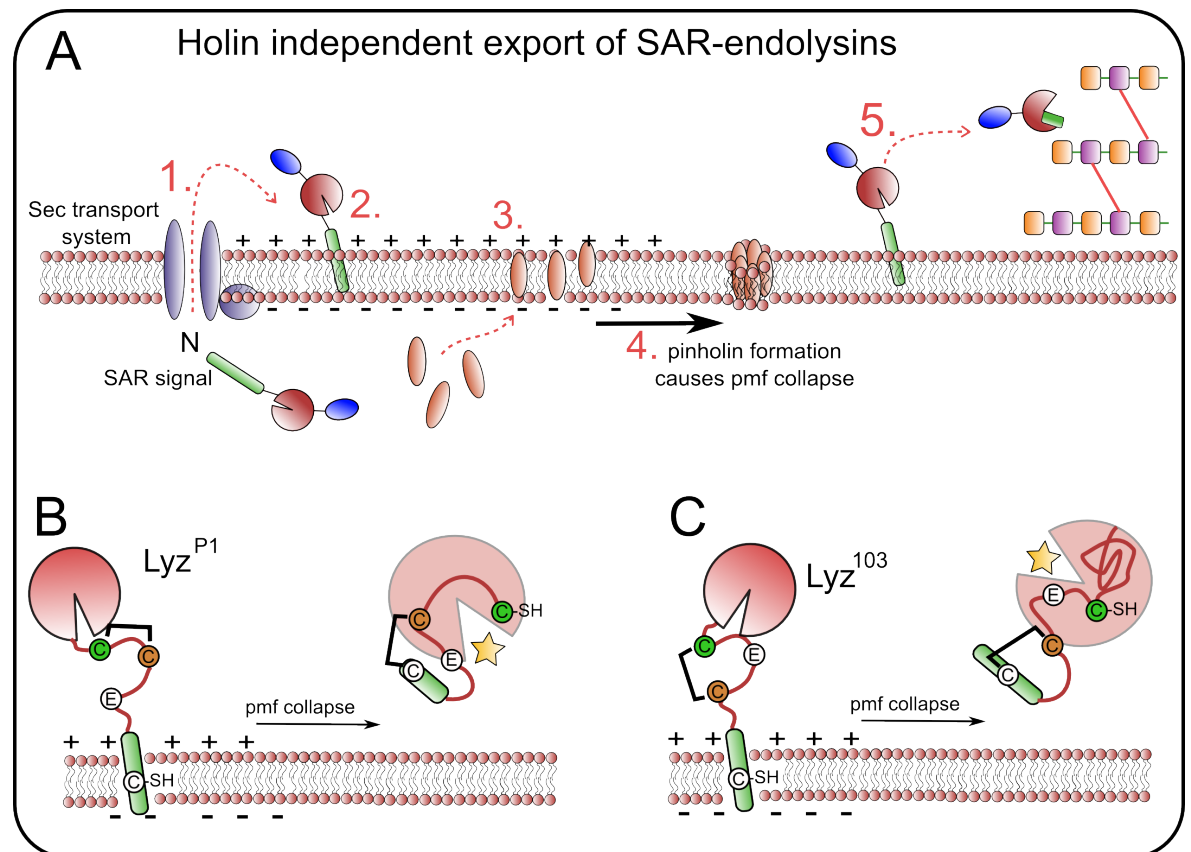
**Figure 5 Holin independent export of SAR-endolysins**. **A:** (1) Sec-transported endolysins are secreted with the SAR domain still attached. (2) Endolysin remains tethered to the membrane by the SAR domain. (3) Holin is expressed and accumulates in the membrane. (4) At a genetically pre-determined time holin accumulated in the membrane forms rafts leading to depolarisation and pinholin formation, collapsing the membrane potential. (5) Electrochemical restraints on the SAR domain are removed, the catalytic domain refolds and incorporates the SAR domain to become active and begins peptidoglycan degradation. **B: The release mechanism of Lyz[P1].** After the SAR domain is released disulphide isomerisation releases the constraint on an active site cysteine (green), allowing the endolysin to become active. **C: The release mechanism of Lyz[103].** An active site glutamate (green) is blocked by disulphide caging of neighbouring cysteines. Disulphide switching after SAR domain cause conformational changes and unblocks the glutamate and Lyz[103] becomes active. Figures B and C were adapted from Xu *et al.* (2005) and Kuty *et al.* (2010).

A similar mechanism to Lyz[P1] has been reported for the P21 phage endolysin R[21], which represents a second class of SAR-endolysins. R[21] is also regulated by conformational rearrangement of the endolysin after release of the SAR domain, except it does not involve thiol rearrangement. Instead, when R[21] is tethered to the membrane an active site glutamate residue is subject to steric hindrance by the close proximity to the membrane bilayer. Upon depolarisation and release, the SAR domain refolds into the protein domain and the glutamate residue repositions to complete a Glu/Asp/Thr catalytic triad within the active site, activating the endolysin (Sun *et al.* 2009). Replacement of the SAR domain of Lyz[P1]

with the cysteine-less SAR domain of R[21] rendered the endolysin inactive, due to the lack of disulphide isomerisation (Xu *et al.* 2004).

Recently, a third mode of SAR endolysin regulation has been described for lysozyme Lyz[103] from the *Erwinia amylovora* phage ERA103 (Kuty *et al.* 2010). The catalytic glutamate residue (Glu43) of Lyz[103] is kept in an inactive state by a disulphide bond formed between flanking cysteine residues (Cys42-Cys45) which cages the enzyme's active site. Upon membrane depolarisation and SAR domain release, a free thiol group of a cysteine (Cys12) within the SAR domain disrupts the disulphide caging by switching the disulphide bonding residues (Cys12-Cys42) (Figure 5, C). This releases constraints on the glutamate residue that forms the active site cleft, activating the endolysin (Kuty *et al.* 2010). Examination of the protein sequences of related bacteriophage endolysins suggests that the presence of an N-terminal SAR sequence is not uncommon and it is estimated from sequence comparisons that 25% of phages possess SAR containing endolysins (Park *et al.* 2007).

The SAR containing endolysin KMV45, from the Gram-negative *P. aeruginosa* targeting phage ϕKMV, exhibits similar characteristics to the P1 phage endolysins. KMV45 also gets transported and membrane tethered due to an N-terminal SAR signal, until membrane depolarisation driven by a pinholin (Briers *et al.* 2011). However, KMV45 does not demonstrate disulphide caging (Lyz[103]) or disulphide isomerisation (Lyz[P1]) to keep it inactive while membrane bound. Instead it is suggested that the positively charged N-terminal SAR domain and the negatively charged inner side of the cytoplasmic membrane is enough to keep KMV45 inactive when inserted into the membrane. The blocking of enzymatic activity is due to steric hindrance on a catalytic glutamate that would be positioned in close proximity to the membrane when tethered (Briers *et al.* 2011).

Ultimately, the trigger that regulates the efficient release of these SAR containing endolysins from the membrane is depolarisation by pinholin lesion formation and the consequent relief of ionic interactions. First described for the Lamboid phage 21, the pinholins function by disrupting the pmf without the formation of large, macromolecular lesions (Park *et al.* 2007).

*Pinholins*

Pinholins can mediate host cell lysis only when co-expressed with the SAR-endolysins, but not with canonical cytoplasmic endolysins lacking the SAR signal. Pinholins behave like canonical holins, accumulating harmlessly within the bilayer until at the point of lysis they rapidly form lethal membrane lesions with central pores < 2 nm in diameter (Pang *et al.* 2013). The pinholin pore size is too small to deliver cytoplasmic endolysins to the periplasm but they do cause membrane depolarisation which activates the release of the membrane attached SAR-endolysins. Park *et al.* proposed that the pinholin & SAR-endolysin could represent an intermediate stage in the evolution of holin-endolysin systems towards the better-known classic system (Park *et al.* 2007).

*Mycobacteriophage endolysins can be transported using molecular chaperones*

Mycobacteriophage are phage that infect mycobacterial cells, bacteria with a complex cell wall involving a lipid shell around the bacteria containing mycolic acids and lipids (Brennan 2003). More than 2500 mycobacteriophage have been isolated, mainly targeting *Mycobacterium smegmatis*, but until recently little was known about the mechanisms of lysis for these phages (Catalão *et al.* 2013). The lysis cassette of mycobacteriophage Ms6 encodes 5 genes with features not seen in other common phage, including a novel endolysin (LysB) with lipolytic activity that specifically hydrolyses the mycolic acids of the mycobacterial outer membranes (Gil *et al.* 2008; 2010).

Interestingly, phage Ms6 also encodes a peptidoglycan amidase endolysin (LysA) that accesses the cell wall independently of holin formation (Catalao *et al.* 2010). LysA differs from the above examples of holin independent secretions, as it does not have an N-terminal SAR or signal peptide domain. In the Ms6 lytic cassette, immediately upstream of the gene for LysA (*lysA*) is a gene (*gp1*) encoding a chaperone-like protein (Catalao *et al.* 2010). The export of LysA was shown to be dependent on the presence of Gp1, which together get secreted into the cell wall. Deletion of *gp1* resulted in a 70% decrease of the cell burst size (the average number of phage released by a lysed cell) (Catalao *et al.* 2010). The chaperone Gp1 was shown to interact with the first 60 amino acids of LysA, which then is proposed to use the host SecA secretion system to export LysA across the membrane. In an *E. coli* model expressing Gp1 and LysA, a defective Sec secretion system or addition of sodium azide an inhibitor of Sec activity, inhibited lysis. Furthermore, deletion of *gp1*

resulted in no detectable lysA within the cell wall and resulted in a loss of lytic activity (Catalao *et al.* 2010).

In Gp1-LysA expressing *M. smegmatis* cells, adding nisin could trigger lysis. Nisin induced lesions are too small for LysA secretion and Catalao *et al.* suggested that after secretion LysA binds to cell wall epitopes in an inactive conformation until holin-induced depolarisation activates the endolysin. The mechanism of cell wall binding and regulation still remains to be elucidated for this class of endolysins, as well as the role of Gp1 within the cell wall. An analysis on mycobacteriophage genomes by Hatfull *et al.* (2010) revealed that just 18 out of 60 known mycobacterophage genomes have Gp1 homologues suggesting that, at least for mycobacteriophage, different mechanisms of lysin delivery and endolysin activation could exist.

*Endolysins can hijack cell wall synthesis pathways to enter the cell wall*

The pneumococcal phage SV1 encodes an endolysin (Svl) lacking any known signal peptide or SAR domain and demonstrates another novel secretion mechanism independent of holin or the SecA system. Svl export is dependent on choline containing teichoic acids found within the cell walls of pneumococci. Endolysin Svl is structurally similar to the pneumococcal autolysin LytA as both have C-terminal choline binding domains that recognise teichoic acids within pneumococcal cell walls that are decorated with choline residues (López & García 2004).

The lytic activity of autolysin LytA is inhibited when choline-bound and the endolysin Svl is believed to behave in a same manner, remaining inactive in the choline bound state until holin-induced lysis and depolarisation releases the endolysin to begin cell wall degradation (Frias *et al.* 2013). Interestingly, the lytic cassettes of SV1 and other pneumococcal phage contain no other functional proteins apart from endolysin and holin, and no Gp1 related chaperone proteins exist as seen for Ms6 mycobacteriophage. Therefore SV1 is exported without the use of SAR or signal peptides or chaperone proteins for Sec secretion (Frias *et al.* 2013). Instead Svl is proposed to bind choline molecules during their processing into teichoic acid precursors within the cytoplasm. Precursors of teichoic acids are transported across the membrane by TacF, a member of the polysaccharide transmembrane transporters, otherwise known as flippases (Damjanovic *et al.* 2007). Frias *et al.* hypothesised that Svl can bind choline residues during teichoic acids translocation to gain access to the cell wall, where they remain inactively bound to the teichoic acid waiting for holin-induced depolarisation for activation. Therefore Svl behaves in a similar manner to

the other holin-independent systems described above (Xu *et al.* 2004; Nascimento *et al.* 2008; Catalao *et al.* 2010), gaining access to the cell wall without passing through holin lesions.

Svl was continuously detected in the cell walls of pneumococcal cells with a holin gene deletion (Δ*hol*) indicating endolysin translocation into the cell wall was holin independent. Pneumococcal cells grown with choline substituted for ethanolamine decorate their cell wall teichoic acids with ethanolamine, which inhibits choline-binding proteins from attaching to the cell wall. Endolysin Svl was undetectable in both cell wall fractions and cell media. Instead Svl was only detectable in cell fractions indicating that it was no longer translocated out of the cells without choline present (Frias *et al.* 2013). Choline is therefore essential for Svl endolysin translocation.

LytA, the autolysin of *S. pneumococciae*, also has no secretion signal and transport of this host protein could also occur in a similar manner to the phage endolysin. Mellroth *et al.* previously reported that LytA mainly resided within the cytoplasm of pneumococcal cells with only 5% of expressed LytA entering the cell wall during growth. It was suggested that translocation of LytA was achieved by cellular lysis and not active secretion, where LytA released from lysed cells binds neighbouring cells (Mellroth *et al.* 2012). The mechanism of LytA and Svl secretion needs more comprehensive research, particularly to see if they can interact with the teichoic acid translocation system within the cytoplasm or if they can be detected in a bound state to precursor teichoic acid within growing cells. The lack of signal peptides and the existence of analogous choline-binding domains is prevalent in other pneumococcal phage endolysins (López & García 2004), indicating that the above lytic mechanism could be another demonstration of the diversity of phages exploiting the host cellular machinery for endolysin targeting.

### *Depolarisation of the membrane and loss of proton motive force activates endolysins.*

Both SAR-endolysins and the signal peptide secreted endolysins are kept inactive before holin lesion formation by tethering of the SAR domain into the membrane or kept in an inactive state due to electrochemical gradients within the cell wall. The application of exogenous nisin (as described for Lys44) can initiate lytic activity due to the formation of 2 nm holes that are nonselective for the flow of ions and small metabolites, imitating the role of the holin or pinholin lesions and depolarising the membrane (Nascimento *et al.* 2008). Before depolarisation the negative regulation of endolysins, and their related autolysins, by the cell wall is suggested to be due to binding to teichoic acids due to their

polyanionic composition and their ability to change conformation depending on their ionisation state (Neuhaus & Baddiley 2003) (Figure 6, A). Teichoic acids are commonly decorated with cationic glycosyl or d-alanyl ester attachments that help regulate the overall net anionic charge of teichoic acid and the cell wall during normal cell metabolism and have been shown to regulate autolytic activity (Neuhaus & Baddiley 2003). Inhibition of normal teichoic acid metabolism can infer an increase in lytic activity for autolysins and endolysins. For instance, wall teichoic acids were shown to be non-essential for *B. subtilis* viability, however, removing the ability of *B. subtilis* to attach d-alanyl esters to teichoic acid increased autolytic activity endogenously or after chemical induction (Wecke *et al.* 1996). Activation of the major autolysin AtlA of *S. aureus* is dependent on the pH of the cell wall, whereby an acidic environment generated by proton binding of teichoic acid greatly reduced autolytic activity (Biswas *et al.* 2012). The ionic capacity of the cell wall could also play a role in regulation. The binding of the *Listeria* phage endolysin Ply500 has been shown to depend on ionic interactions for cell wall binding, whereby Ply500 binds the cell wall with optimal affinity at 100-150 mM NaCl and shows decreasing affinity as the ionic strength increases (Loessner *et al.* 2002).

Frias *et al.* demonstrated that in the absence of phage endolysins the activation of the host autolysin, LytA, in *S. pneumoniae* at the end of the phage lytic cycle was sufficient to allow viral progeny release, although at a lower amount than when endolysin was also present. The utilisation of endolysins and autolysins would boost lytic activity and increase overall phage fitness, providing another example of how phage can exploit the host enzymes for their own benefit (Frias *et al.* 2009).

*The enzymatic activity of the endolysins*

There are several mechanisms used by phage to control the release and activation of endolysins. Nevertheless, once endolysins are activated within the cell wall they lyse cells by sabotaging the structural integrity of the cell wall by degrading peptidoglycan. Endolysins are peptidoglycan hydrolases that have two tasks during peptidoglycan digestion: substrate recognition and peptidoglycan hydrolysis (Loessner 2005).
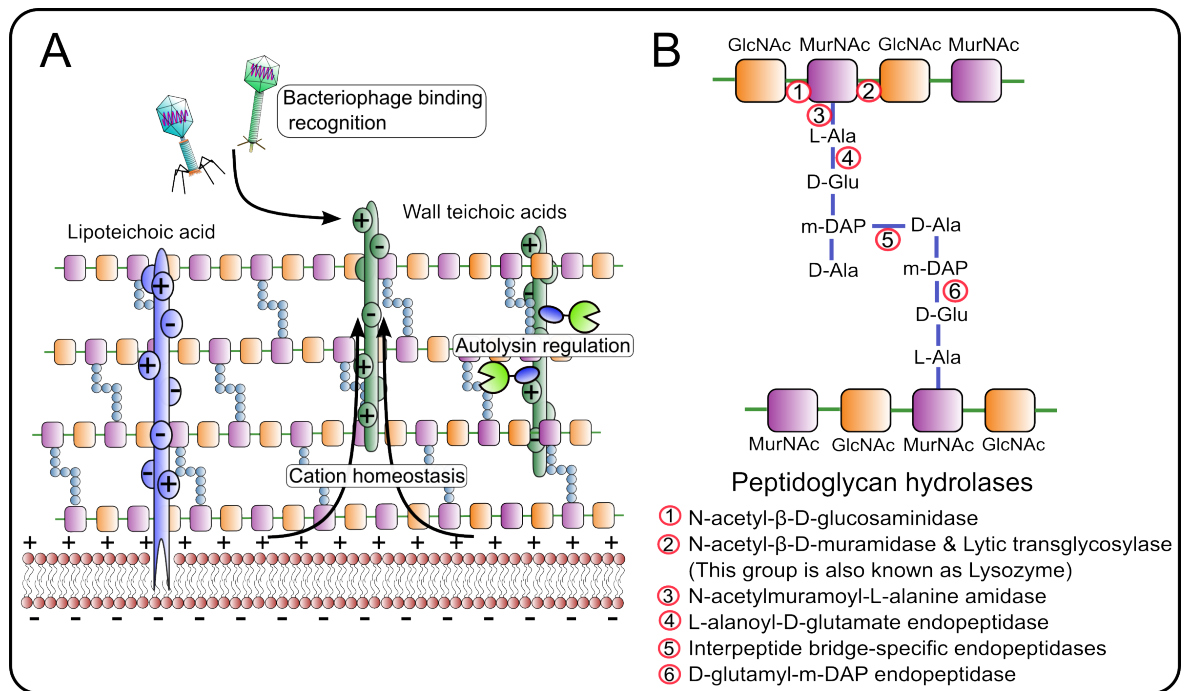


**Figure 6 A: The various roles of teichoic acid within the Gram-positive cell wall**. Gram-positive bacteria cell walls are between 15-80 nm in width, consisting of a repeating polymer of the amino sugars N-acetylmuramic acid (MurNAc) and N-acetylglucosamine (GlcNAc) linked together by a peptide chain attached to a hydroxyl group from the MurNAc subunit. The length of the glycan strands, amount of cross-linking and the composition of the peptide chain varies between bacterial species and growth conditions (Vollmer & Seligman 2010). Within the cell wall are lipoteichoic acids bound to the membrane by a terminal diacylglycerol and wall teichoic acids that are covalently linked to the peptidoglycan layers. Teichoic acids are responsible for phage binding and adsorption, autolytic regulation, cation homeostasis, antimicrobial peptide binding and recognition by other bacteria (Weidenmaier & Peschel 2008).

**Figure 6 B: Peptidoglycan hydrolases cleave different bonds.** (1) N-acetyl-β-D-glucosaminidases cleave the β 1-4 glycosidic bond linking GlcNAc to MurNAc subunits of the peptidoglycan polymer. (2) N-acetyl-β-D-muramidases and Lytic transglycosylases are collectively termed lysozymes, these cleave the β 1-4 glycosidic bond linking MurNAc to GlcNAc subunits. (3) N-acetylmuramoyl-L-alanine amidases hydrolyse the amide bond between MurNAc and the a-position L-alanine of the peptide linker. (4) L-alanoyl-D-glutamate endopeptidases. The peptidoglycan bonds sensitive to 1,2 & 3 are conserved in nearly all bacteria whereas the amino acid sequence of the peptide linker between MurNAc subunits, targeted by endopeptidases (5), can vary depending on species or genera (Schleifer & Kandler 1972). (6) D-glutamyl-m-DAP endopeptidase. *Figure 6 A adapted from Weidenmaier & Peschel (2008). Figure 6 B adapted from Loessner (2005).*

***The modular organization of endolysins from Gram-positive bacteria infecting phages.***
The endolysins of Gram-positive bacteria targeting phage have evolved to use a modular design whereby the two tasks of substrate recognition and peptidoglycan hydrolysis are divided into two domains: an enzymatically active domain (EAD) and a cell wall binding domain (CBD), separated by a short linker (Loessner 2005). The EAD is normally located at the N-terminus and the CBD at the C-terminus, albeit there are endolysins with different modular organisations (Figure 7). While the EAD catalyses cell wall degradation, the CBD targets endolysins to cell wall epitopes characteristic to certain bacterial strains, which imposes a further level of specificity to endolysins. The high affinity binding of the CBD to the cell wall is proposed to trap endolysins within the cell wall and inhibit their diffusion after lysis. Without cell wall binding the endolysins could enter the surrounding media and compromise phage survival by degrading possible future bacterial hosts (Schmelcher *et al.* 2012).

It is apparent that host autolysins and phage endolysins share common ancestry, phage genomes are modular in organisation and display a highly mosaic nature, therefore it is not unusual that phage endolysins have evolved to combine various enzymatic and binding domains into their configuration (Casjens 2005).
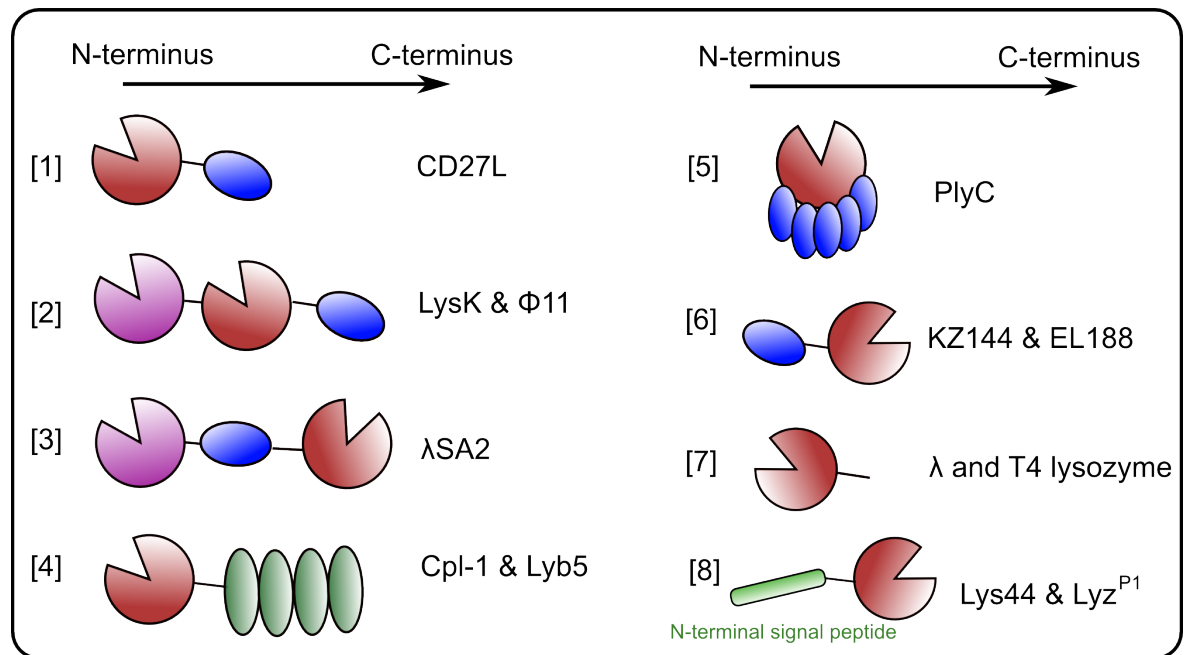
**Figure 7 Currently known modular organisations of endolysins as described in the literature**. [1] CD27L represents a typical Gram-positive infecting phage endolysin with a single N-terminal EAD and C-terminal CBD (Mayer *et al.* 2011). [2] Double N-terminal EADs are common among staphylococcal phage endolysins such as LysK and Φ11 (Nelson *et al.* 2012). [3] Another streptococcal phage endolysin λSA2 has two central CBDs flanked by two EADs (Pritchard *et al.* 2007). [4] Endolysins can have tandem repeats of CBDs such as choline-binding motifs for Cpl-1 (Hermoso *et al.* 2003) or multiple lysin motif (LysM) repeats such as Lyb5 (Hu *et al.* 2010). [5] PlyC is a multimeric enzyme and the largest endolysin found to date (Nelson *et al.* 2006). [6] CBDs are not common among Gram-negative infecting phage endolysins, however, Gram-negative *P. aeruginosa* targeting phage endolysins KZ144 and EL188 contain N-terminal CBDs (Briers *et al.* 2007). [7] A typical Gram-negative bacteria targeting phage endolysin consists of a single EAD with no CBD such as λ and T4 endolysins; [8] Endolysins secreted independently of holin either using a signal peptide, such as Lys44 (Sao-Jose *et al.* 2000) or a signal anchor release domain such as Lys[P1] (Xu *et al.* 2004). *Figure 7 adapted from Schmelcher et al. (2012).*

For most endolysins with a modular organisation, an interdomain linker typically seperates the EAD and the CBD. This linker imparts flexibility to these enzymes and is suggested to be one of the reasons why full-length crystallisation of these proteins is challenging (Nelson *et al.* 2012). Only a few complete endolysin structures have been resolved to date including Cpl-1 (Hermoso *et al.* 2003), PlyPSA (Korndörfer *et al.* 2006) and recently the C. *perfringens* phage endolysin Plm (Tamai *et al.* 2014). PlyC (Figure 7, [5]) is the only multimeric endolysin to be identified that is the product of two separate genes (Nelson *et al.* 2006). PlyC was recently shown to rapidly degrade biofilms of streptococcal cells with a higher activity than conventional antibiotics and with no signs of developing bacterial resistance (Shen *et al.* 2013). The potent lytic activity of PlyC is likely due to its unique arrangement, the structure of which was also recently revealed (McGowan *et al.* 2012). PlyC consists of a single enzymatic PlyCA domain bound to a ring of eight PlyCB

domains. PlyCB contains a binding cleft which guides the attachment of the complex to the cell wall, while PlyCA has two peptidoglycan degrading domains: a cysteine, histidine-dependent amidohydrolases/peptidase domain and a glycoside hydrolase domain (McGowan *et al.* 2012).

### *The majority of endolysins from phages that infect Gram-negative bacteria have single enzymatic domains*

As a comparison, Gram-negative bacteria infecting phage mostly consist of a single EAD without any particular CBD module, the traditional example being the λ phage lysozyme (Figure 6, [7]). Substrate recognition and specific binding of the endolysin to the cell wall may not be necessary for these endolysins due to a relatively thin peptidoglycan layer and the presence of the outer membrane in Gram-negative bacteria, which limits access of the endolysin to the outside even after lysis (Schmelcher *et al.* 2012). However, there are exceptions. The Gram-negative *P. aeruginosa* infecting phage endolysins KZ144 and EL188 use a modular structure with N-terminal CBDs that were shown to bind with high affinity to the *P. aeruginosa* peptidoglycan (Briers *et al.* 2007; 2009).

### *The Enzymatically Active Domain*

Typically the EADs found in endolysins belong to the same families as the host autolysins that break down and remodel the peptidoglycan during replication, growth and sporulation (Regulski *et al.* 2013). Peptidoglycan and the peptide chain arrangement between N-acetylmuramic acid (MurNAc) subunits and secondary cell wall polymers will vary between bacterial species, but they still maintain the centrally conserved cross-linked peptidoglycan structure and therefore provide a limited amount of bonds available for hydrolysis. Depending on the type of peptidoglycan bond attacked there are six classes of peptidoglycan hydrolases that EADs employ (Nelson *et al.* 2012), these are reviewed in Figure 6, B.

Endolysins that recognise peptidoglycan bonds conserved within a genus are prone to cleave the same bond present in a broader range of bacterial species than just the original host (Donovan & Foster-Frey 2008). For instance, broad activity has been reported for the phage B30 endolysin hosted by a small subgroup of streptococcus (type III) that demonstrated an extensive and improved rate of lysis against whole groups of streptococci,

even though these strains were not susceptible to phage B30 and production of this particular endolysin (Pritchard *et al.* 2004).

Endolysin EADs can be engineered to increase their host range of activity. The EAD of endolysin CD27L, the activation mechanism of which forms a focus of this study, consists of a zinc-dependent *N*-acetylmuramoyl-l-alanine amidase domain (Mayer *et al.* 2011). Alignment of CD27L with other amidase endolysins revealed conserved residues within the EAD, suggesting that the EAD contained key features for specific cell wall ligands (Mayer *et al.* 2011). The backbone structure of CD27L resembled the *L. monocytogenes* endolysin PlyPSA, and certain residues were conserved within their active sites. Between these conserved residues a single residue was observed that appeared highly variable between similar sequences. Mutation of this residue, leucine in the CD27L EAD active site to tryptophan, as found in PlyPSA, increased the host range of the truncated CD27L EAD to develop lytic activity against *L. monocytogenes* cells (Mayer *et al.* 2011), demonstrating that the endolysin EADs are also responsible for substrate specificity.

### *Specificity is linked to the selective targeting of cell wall substrates by the cell wall binding domain*

Whilst the EAD alone cleaves specific cell wall ligands, it is the CBD that is believed to play the essential role in endolysin specificity, by directing the binding of endolysins to the cell wall. The CBDs of *L. monocytogenes* phage endolysins have been reported to associate with their cell wall ligands with nanomolar affinities as determined by surface plasmon resonance, with values comparable to high affinity antibodies (Loessner *et al.* 2002; Schmelcher *et al.* 2010). The CBD targets the endolysin to the cell wall in order to bring the EAD within close proximity of its substrate for efficient cell wall lysis.

### *Diversity of Gram-positive cell wall binding domain epitopes*

CBDs can associate with a broad range of species-specific motifs present in the peptidoglycan or cell wall polymers of the cell wall. Lyb5 lysin from phage ΦPYB5 targets it's host *Lactobacillus fermentum* using three C-terminal lysin motif domains (LysM) (Hu *et al.* 2010). LysM is a motif between 44-65 amino acids, commonly linked into multiple domains, LysM is suggested to bind the GlcNAc subunits in the backbone of peptidoglycan (Buist *et al.* 2008). The CBDs of the Lc-Lys and Lc-Lys2 endolysins from the *L. casei* phage BL23 specifically target the interpeptide bridge of peptidoglycan, only when it contains D-Asn residues (Regulski *et al.* 2013). The C-terminal repeats of SH3 domains for

the *C. perfringens* phage endolysin Psm are proposed to recognise the peptide side chain of peptidoglycan (Tamai *et al.* 2014). CBDs of *B. anthracis* phage endolysins PlyL and PlyG bind specifically to cell wall polymers depending on their pattern of galactosylation with µM affinity, similar to other protein-carbohydrate interactions (Mo *et al.* 2012; Ganguly *et al.* 2013). Pneumococcal phage lysins have modular C-terminal choline binding domains. Cpl-1 was crystallised with choline bound within it's C-terminal repeats of a choline binding motif (Hermoso *et al.* 2003), Cpl-1 was also separately crystallised with the EAD bound to the substrate peptidoglycan (Pérez-Dorado *et al.* 2007), which provided the first structural evidence of peptidoglycan recognition by the EAD. Choline binding motifs are common amongst pneumococcal phage endolysins, possibly as choline is a common and necessary decoration of teichoic acid for their pneumococcal hosts. *Listeria* phage endolysin PlyP35 only binds to teichoic acids that contain terminal GlcNAc residues. *Listeria* cells that had been mutated to inhibit GlcNAc decoration of teichoic acids were resistant to PlyPSA binding (Eugster *et al.* 2011), however, three other *Listeria* phage endolysins Ply118, Ply511, and PlyP40 were shown to be restricted from binding areas of the cell wall by the presence of teichoic acids and instead were proposed to bind the peptidoglycan backbone directly (Eugster & Loessner 2012).

### *Endolysin domain shuffling can improve lytic activity when applied externally to cells*

In a number of studies the exogenous application of the recombinant endolysin truncated after the EAD gave a higher rate of lysis than the full-length endolysin, questioning the rationale for CBD incorporation. For instance, the phage lysin LysK consists of two EADs and a CBD, truncation of the endolysin to just the N-terminal EAD still showed high lytic activity against staphylococcal cells, which also demonstrates the modular nature of endolysins as the N-terminal EAD was correctly folded and stable when separated from the other two domains (Horgan *et al.* 2009).

When applied externally, the *Clostridial* endolysins CS74L and CD27L are equally active as full-length endolysins or separate EADs, however, the related CTP1L is only active as the full-length protein (Mayer *et al.* 2010; 2011; 2012). The endolysin PlyB behaves like CTP1L and is only active against *B. anthracis* as the full-length endolysin, removal of the CBD inhibited lytic activity (Porter *et al.* 2007). An EAD truncated construct of the phage endolysin PlyL showed higher lytic activity than full-length endolysin against *B. anthracis* and two other Bacillus strains (Low *et al.* 2005). From this observation Low *et al.* proposed that the C-terminal domain could regulate the catalytic domain allosterically via

intermolecular interactions, until the CBD binds the cell wall, which releases constraints on the EAD to become active. However, it has now been shown that the two domains of PlyL have no significant interactions (Low *et al.* 2011).

Low *et al.* therefore proposed that the surface charge of the domains could have a role in endolysin activation. A positive surface charge on the EAD of an endolysin, such as the *B. anthracis* phage endolysin PlyL, correlated with a high lytic activity when the CBD was removed. Compared to the EAD of XlyA from a *B. subtilis* prophage, the EAD of which has a net negative charge and is inactive when expressed as a separate domain (Low *et al.* 2011). Surface mutations were introduced to modify the charge of XlyA from a net charge $Z = -3$ to $+3$, the EAD was stable and folded with the change in surface charge and was transform into an active domain with the same activity as the native full-length XlyA (Low *et al.* 2011). Considering the previous evidence that pmf dissociation after holin formation is related to endolysin activation, surface charge most likely plays a part in activation and could provide another factor to consider for enhancement of the lytic activity of engineered endolysins.

Given the diversity of endolysin modulation, the mechanics of CBD cell wall recognition and endolysin regulation must vary across all families of phages. As more structures of endolysins are elucidated the more we can probe into their function, and begin to engineer endolysins for biotechnological use. Regarding the classic holin-endolysin system it is still not clear how the endolysins become active, in fact due to the diversity in modular organisation of endolysins there are probably multiple ways to trigger endolysin activity. Currently, the *in vivo* function of the CBD remains obscured by the fact these endolysins from Gram-positive bacteria targeting phage have mainly been studied as recombinant proteins applied exogenously to bacterial cells. To fully understand the native requirements of the CBD they need to be expressed and monitored in actively propagating phage systems.

The CBD controls specificity but this can be altered to target EADs to different strains or species by using the diverse enzymatic activities and modular organisation of endolysins to generate endolysin chimeras with improved activity against different bacterial types. Molecular engineering by removal, addition or repetition of domains has been used to generate and investigate different modular organisations of endolysins. For instance, duplication of the CBD of Ply500 from *Listeria* phage increased the binding affinity of the

domain by 50 times, however, CBD duplication also reduced lytic activity against cells in suspension, indicating that stronger binding of endolysins did not increase activity in this case (Schmelcher *et al.* 2011). Recently, a chimeric lysin, ClyH, was generated by combining the EAD and the CBD from two different *S. aureus* phage endolysins. ClyH was designed to target MRSA infections and was demonstrated to be more effective than its parental form against multiple MRSA strains *in vitro* and with an *in vivo* murine model. The engineered ClyH has potential as a future antimicrobial to treat MRSA infections (Yang *et al.* 2014).

### *Use of high-affinity cell wall binding domains as biosensors*

Owing to their substrate specificity and observed high rates of activity, endolysins have a potential biotechnological use, such as for food preservation or diagnostic markers for pathogens. The high affinity binding observed with CBDs has the potential to replace antibodies for the detection of specific bacterial strains, for instance in pathogen contaminated food (Kretzer *et al.* 2007) or for patient samples to detect sources of infection. Recombinant CBDs of phages that target *Listeria* were coupled to magnetic beads and mixed with contaminated food to selectively bind contaminating *Listeria* cells by recognising the specific epitopes present on *Listeria* cell walls. The separation and concentration of target cells could then be used for diagnostic purposes (Kretzer *et al.* 2007). As binding of CBDs can be specific for certain strains or much broader and cover entire bacterial genera there is much scope for CBDs as detectors. Attachment of fluorophores to CBDs was demonstrated as a future application as biomarkers for the quantitative detection of bacterial strains, even in mixed cultures it would be possible to simultaneously detect different strains with different fluorophores (Schmelcher *et al.* 2010; Schmelcher & Loessner 2014).

### *Uses of endolysins as antimicrobials*

Antibiotics have a broad range activity against pathogenic and commensal bacteria, which due to selective pressure can accelerate the spread of resistance genes in bacterial communities (Fenton *et al.* 2010). Therefore the unique strain and species specificity observed for endolysins provides an advantage over broad range antibiotics, as it should be possible to target specific pathogenic bacteria with a high potency. The use of endolysin antimicrobials could reduce antibiotic associated diarrhoea infections, as specific endolysins produced to inhibit *C. difficile* colonisation would not affect the GIT

microbiota. Endolysins can be used against antibiotic resistant bacteria as they target different bacterial components with different mechanisms of action to antibiotics. Bacteria are also less likely to gain resistance against endolysins, as the basic components that endolysins target, such as the MurNAc-GlcNAc peptidoglycan bond, cannot be easily modified by bacteria (Schuch *et al.* 2002).

Due to the lack of an outer membrane in Gram-positive bacteria, externally applied endolysins are able to gain access to the peptidoglycan layer and due to the peptidoglycan network acting as an open mesh, externally applied endolysins would be free to diffuse and exert their effect throughout the cell wall (Low *et al.* 2011). One of the first *in vivo* demonstrations of the possibility of endolysins as antimicrobials was by Loeffler *et al.* who showed the successful treatment of mice colonised with S. *pneumoniae* by administering the endolysin Pal. 5 hours after a single dose of the Pal endolysin the colonised pneumococcal cells were no longer detectable in the mice. Furthermore, the Pal treatment was shown to have no effect on commensal bacterial strains (Loeffler *et al.* 2001).

As endolysins target structures conserved only in bacterial cells they should have no effect on mammalian cells. However, as both phage and their respective endolysins are proteinacious molecules, a likely pharmacological barrier against their use as systemic antimicrobials is the potential immunogenic response. Loeffler *et al.* reported the lytic activity of Cpl-1 lysin raised against rabbit hyperimmune serum was slowed but not blocked by the immunogenic response and they detected Immunoglobulin G had been raised against Cpl-1, but this only had a modest inhibitory effect on lytic activity (Fenton *et al.* 2010; Loeffler *et al.* 2003). Endolysins could be used as supplementary treatment with antibiotics, the synergistic use of antimicrobials involves using two or more different antimicrobials to increase treatment efficacy by (1) reduce the time of treatment, (2) reduce the antimicrobial dose required and (3) decrease the risk of the pathogen from acquiring resistance (Schmelcher *et al.* 2012). Successful applications of synergy between endolysins includes the combination of the *S. pneumoniae* endolysins Cpl-1 and Pal-amidase, which together enhanced bacteriolytic activity compared to the endolysins alone, possibly as the lysins have different catalytic EADs and therefore target different cell wall bonds providing a two-pronged attack (Loeffler & Fischetti 2003).

Paradoxically, the advantage of species specificity for both phage therapy and endolysins could be a disadvantage in a clinical setting, as identification of the etiological pathogen is

required to decide the correct type of endolysin or phage therapy to use. Diagnostics is moving away from classical diagnostics such as microscopy, antigen detection and immunoserology, to the use of real-time PCR-based detection methods (Ince & McNally 2009). Detection of etiological agents of infection by DNA-based tests provides more than just an instant diagnosis; they also make differentiating between bacterial species more specific and sensitive, therefore endolysins could be selected depending on the type of infection. The combination of rapid diagnostics and the growing repertoire of exploitable antimicrobial endolysins combined with the synergistic use of antibiotics, could pioneer the future of infectious treatment.


### *In conclusion*

The engineering of recombinant endolysins has opened up a range of new applications for these proteins, from biotechnological applications such as biomarkers to antimicrobials for food safety, decontamination and most importantly as potential therapeutics alongside antibiotics. As more bacteriophage are being characterised, so are the various novel lytic systems that they employ. However, the future use of endolysins as antimicrobials could be hindered by a limited understanding, for most lytic systems, of the mechanics that relate the catalytic activity of the N-terminal EAD to the role of the C-terminal CBD, particularly for the classic holin/endolysin system where the mechanisms of endolysin activation are still unknown.

# Project aims

The holin/endolysin system is essential for host cell lysis by bacteriophage. Several lytic systems have been identified and their activation mechanisms elucidated, such as the SAR endolysins (Xu *et al.* 2004; Sun *et al.* 2009; Kuty *et al.* 2010) and the secreted endolysins (Sao-Jose *et al.* 2000). However, it is still not clear if an activation mechanism exists for the classic holin/endolysin system. At the end of the lytic cycle, the canonical endolysins are sequestered within the cytosol and require holin-mediated disruption of the membrane to gain access to the cell wall. The entry of these endolysins follows a well-timed trigger mechanism, directed by the holin protein (White *et al.* 2010). When the endolysins pass into the cell wall, they do so in an activated state to begin cell wall degradation. The molecular mechanisms underlying this activation are unknown. The goal of this research is to investigate the activation mechanism for endolysins that target three different species of *Clostridial* bacteria: *C. difficile* (CD27L endolysin), *C. tyrobutyricum* (CTP1L endolysin) and *C. sporogenes* (CS74L endolysin).

As more cell wall binding domain structures are interpreted, more needs to be known regarding their interactions with cell wall ligands and the specific cell wall decorations endolysins recognise. For instance, different peptidoglycan chemotypes are prone to species-specific modifications that could play a crucial role in endolysin recognition. As a side-project, I aim to develop a methodological approach to gain insight into the ligand-endolysin interaction during cell wall recognition. Using unnatural amino acid incorporation (Chin *et al.* 2002), a photo-cross-linking amino acid could be inserted into the cell wall binding domain of an endolysin for the specific covalent attachment to its respective ligand. Mass spectometry could be used to analyse the endolysin-ligand complex to provide a detailed view of the ligand, complete with specific modifications.

# Chapter 2: Results. The structural basis for the activation mechanism of the *Clostridial* endolysins

*Contributions from other authors to the following work.*

The crystal structures of the CD27L and CTP1L endolysins were solved prior to my involvement with the project. Whilst at the Synchrotron Soleil, Saint Aubin, Dr. R Meijers performed the crystallisation trials and structure determination of CTP1L. Vasiliki Garefalaki and Dr. R Meijers performed the crystallisation and structure determination of CD27L together at EMBL, Hamburg. Full details of expression, purification and crystallisation of CD27L and CTP1L are described in materials and methods. The following chapter describes the key features regarding the crystallisation, structure determination and structural analysis of CD27L and CTP1L, which form the basis for the investigation into the regulation and activation of the three *Clostridial* endolysins. Throughout the text it has been specified where other authors contributed to the work. Results from this chapter have in parts been modified from the submitted article "The CD27L and CTP1L endolysins targeting Clostridia contain a built-in trigger and release factor" (submitted to PLOS Pathogens, 2014). The full manuscript can be found in the Appendix, page 170.

*Full-length CD27L crystallisation revealed just the C-terminal domain*

Freshly purified full-length CD27L endolysin crystallised by hanging drop within 24 hours. Any delays in purification or crystallisation setup prevented crystal formation. An X-ray data set was collected to 2.3 Å from a fresh crystal and molecular replacement with the previously determined CD27L EAD (Mayer *et al.* 2011) was unsuccessful. It was suspected that the crystal possibly contained just the C-terminal domain of the endolysin. To determine the structure, the C-terminal domain of CD27L (ΔN-CD27L) was independently cloned, expressed, purified and crystallised. The presence of cysteine residues within the C-terminal domain was exploited and the structure was determined by singular anomalous diffraction using a mercury derivative. Once the structure of the CD27L C-terminal was elucidated it was discovered that the full-length protein had degraded and the crystal contained only six copies of the C-terminal domain. Data collection and refinement statistics can be found in Appendix Table 4.

*Structure determination of CTP1L showed partial proteolysis*

Immediate purification and crystallisation of full-length CTP1L also resulted in crystals following 24 hours incubation. Akin to crystallisation of CD27L, any delay in purification or crystallisation prevented crystal formation. A fresh CTP1L crystal diffracted to a resolution of 2.1 Å, and was solved by molecular replacement for the catalytic domain,

which consisted of an a/b fold, which was similar to the peptidoglycan hydrolase (endo-N-acetylmuramidase cellosyl) from *Streptomyces coelicolor* (Rau *et al.* 2001) (PDB code 1JFX). The structure of the full length CTP1L endolysin is comprised of the catalytic domain (residues 1 to 190) connected by a linker of 6 residues to a C-terminal domain (residues 195 to 273). The electron density for the residues within the linker was weaker than the rest of the molecule and these residues had to be refined with partial occupancy. Close to the linker, an additional C-terminal domain was found in the asymmetric unit that lacked the N-terminal catalytic domain. This second C-terminal domain was truncated at a valine residue (V195) and remained closely associated with the C-terminal domain of the full-length protein. Interestingly, a few crystals appeared after four days crystallisation that turned out to be of the C-terminal domain alone, a similar result to the crystallisation of full-length CD27L. A single crystal diffracted to 1.2 Å resolution and the structure was resolved. The orientation of the C-terminal domains was identical to the two C-terminal domains found in the mixed crystal form. Data collection and refinement statistics can be found in Appendix Table 5.

### *The C-terminal domains of CD27L and CTP1L adopt a novel protein fold*

Despite a sequence alignment of the cell wall binding domains of CTP1L and CD27L indicating a sequence identity of only 22 % for 81 residues, the folds of the C-terminal domains of CD27L and CTP1L are very similar. The fold consists of a platform of four parallel beta strands with two alpha helices ($\alpha$1 and $\alpha$3) mounted on top (Figure 8, C). The DALI server was used to compare the C-terminal domains of CTP1L and CD27L with structures in the Protein Data Bank (PDB) to search for similar folds, but this did not result in any significant similarities (Holm & Rosenström 2010).
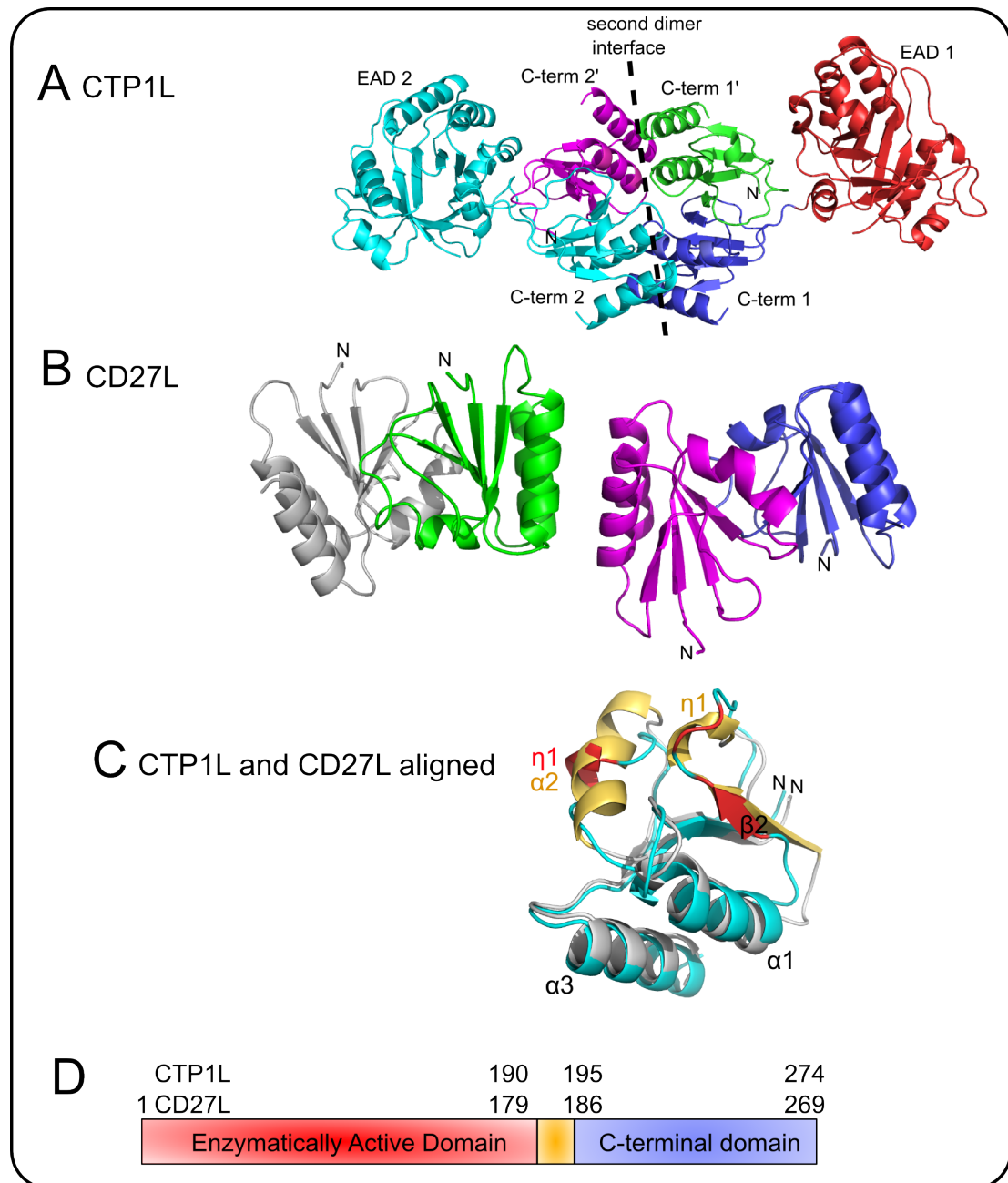
**Figure 8 Structural overview of the partially cleaved CTP1L and the fully cleaved CD27L A:** Cartoon representation of the CTP1L crystal structure. The asymmetric unit consists of a full-length monomer and a cleaved C-terminal domain, with a stable interface between the two C-terminal domains. Between the crystal interfaces both C-terminal domains form head-on contacts with the opposing C-terminal domains. **B:** The tetrameric assembly observed in the CD27L crystal structure. Two dimerisation states are visible, a side-by-side dimer (grey and green or magenta and blue) where the N-termini point in the same direction or a head-on dimer (green and magenta) where the N-termini point in opposite directions. **C:** In CTP1L the C-terminal domain consists of a platform of four parallel beta strands with two alpha helices (α1 and α3) on top. The N-terminus consists of beta strand ß1 at the center of the beta sheet is connected to helix α1, which is followed by beta strand ß2 (red) on the outside of the sheet that connects through an extended loop to beta strand ß3 at the centre of the beta sheet. A single 310 helical turn (η1)(red) connects beta strand ß3 and ß4 and the fold ends with an alpha helix α3 at the C-terminus of the protein sitting parallel to the α1 helix. CD27L contains small changes when compared to CTP1L, including an elongated beta strand ß2 (gold), the insertion of an extra single 310 helical turn (η1) (gold) and the 310 helical turn in CTP1L is modified to a 3 turn helix α2 (gold). **D:** Linear map of the domain organisation of CTP1L and CD27L with residue numbers.

*Sequence analysis of related phage endolysins with similar C-terminal domains*

Sequence similarity searches were carried out using the NCBI Basic Alignment Search Tool (BLAST) (Johnson *et al.* 2008) against the non-redundant protein sequence databases. Searches using the full-length sequence of CD27L, CTP1L or CS74L recognised related sequences based only on their similarity to the N-terminal catalytic domains. Searches were consequently restricted to the proteolytic C-terminal fragments of CTP1L (residues V195 to K274) or CD27L (residues M186 to R270). A sequence similarity search using the CD27L C-terminal domain found 14 unique sequences from different *C. difficile* autolysins or putative endolysins from *C. difficile* targeting phage with Expected (E) values <0.01. The E-value parameter indicates the number of hits with a given score that would be expected to occur at random, within a database of a particular size. As a result, the lower the E-value the more significant the identified sequence (Pevsner 2013).

A sequence similarity search with the C-terminal domain of CTP1L identified the C-terminal domain of CS74L and 19 other proteins with E values < 0.01. The 19 sequences were from other putative *Clostridial* phage endolysins or *Clostridial* autolysins, as well as four peptidoglycan hydrolases identified from the desulfitobacterium family (Figure 9, B). Weblogo (Crooks *et al.* 2004) was used to produce a graphical representation of the multiple sequence alignment between the C-terminal domains of CTP1L and CS74L with the 19 other identified sequences to show more specifically the sequence conservation. The main conservation appeared to reside on the α1 and α3 helices (Figure 9, C).

Even with a low overall sequence similarity, the separate sequence alignments of CD27L and CTP1L reveal three totally conserved residues. D198, T260 and G235 in CD27L and the corresponding residues D206, T262 and G240 in CTP1L, which form a hydrogen-bonding network together, involving a single water molecule in both crystal structures. Due to their central location and conservation within the core of the domain, these three residues could be crucial for maintaining the tertiary structure of the domain.
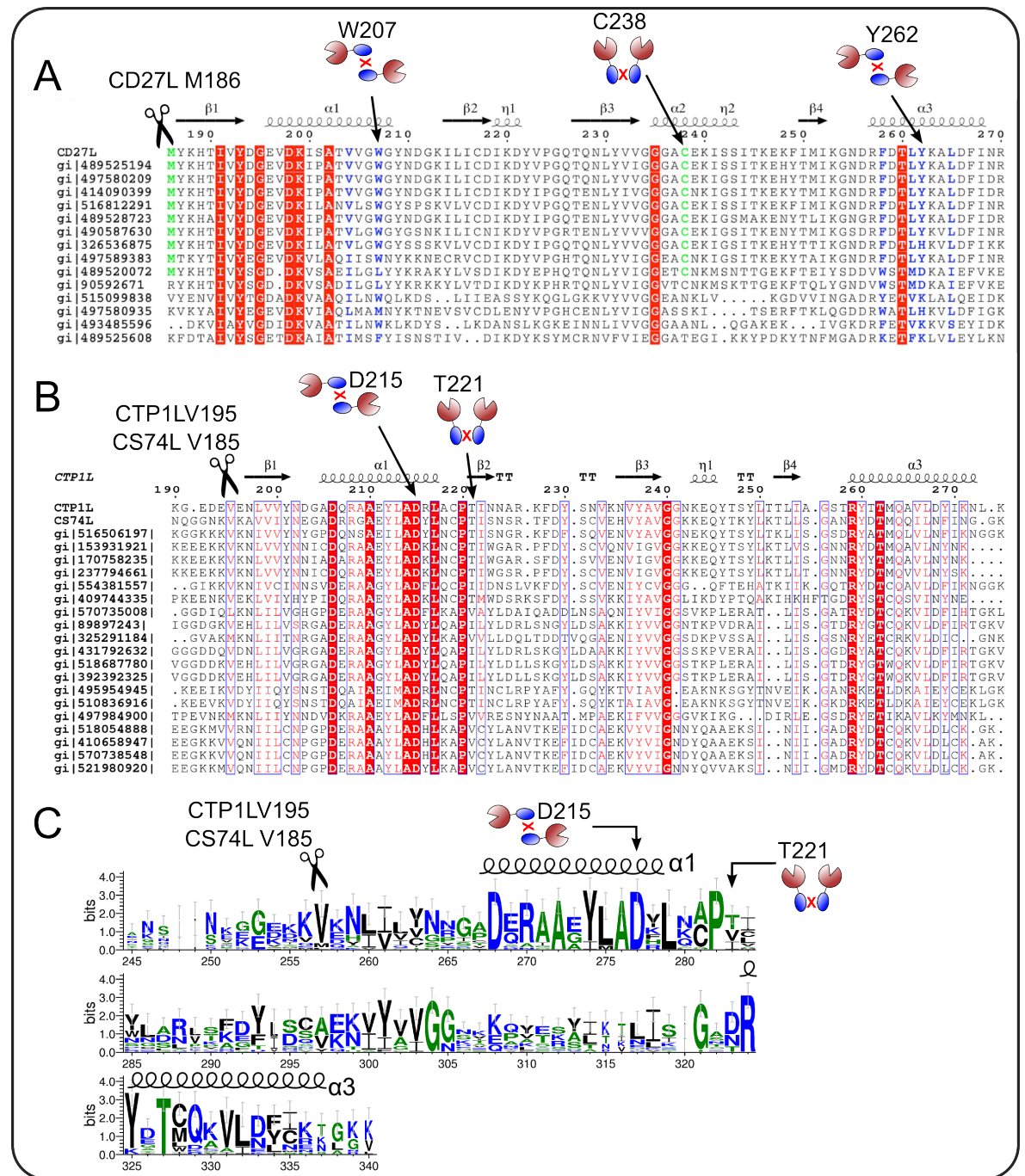
**Figure 9 Sequence alignments of CD27L and CTP1L indicating that the fold is prevalent for other lysins that target Clostridia A:** Sequence alignment of the C-terminal domain of CD27L and sequences with a significant BLAST score (E < 0.01) was generated using ESPRIPT (Gouet *et al.* 2003). The secondary structure of the C-terminal domain of CD27L is depicted with arrows for beta strands and curls for alpha helices. Conserved residues are red, hydrophobic residues involved in the head-on dimer interfaces are blue and the cleavage site M186 and the C238 involved in the side-by-side dimer interface are green. ✂ indicates the cleavage positions at the N-terminus of each C-terminal domain **B:** Sequence alignment of the C-terminal domain of CTP1L and CS74L with the 19 top sequences with BLAST scores E <0.01 generated using ESPRIPT with secondary structure elements of CTP1L. **C:** Weblogo representation of amino acid conservation for the sequence alignment for CTP1L, CS74L and the 19 top sequences from Figure B. At each position, the height of the stack is a representation of sequence conservation at that position. For all three endolysins the most conserved residue are positioned on alpha helices α1 and α3. In **A, B & C** the key residues that have been mutated during the project and demonstrated in the thesis to be required for side-side or head-on dimerisation have been indicated. *Image A is modified from the submitted article, Appendix page 170.*

*SDS-PAGE analysis identified a continuous proteolytic mechanism*

Expression of all three endolysins, CD27L, CTP1L and CS74L, was susceptible to continuous proteolysis. SDS-PAGE analysis of freshly nitrilotriacetic acid (Ni-NTA) affinity purified CTP1L, CS74L and CD27L always produced a protein band with a molecular weight corresponding to the C-terminal domain (Figure 10, A). This proteolysis could not be inhibited by the addition of protease inhibitors. The type of lysis protocol applied also had no effect, mechanical lysis by sonication, repeating freeze-thaw cycles or chemical digestion involving lysozyme all presented the same cleaved product by SDS-PAGE analysis. These observations are not extraordinary and similar proteolytic processes can be explicated from studies on other unrelated phage endolysins. For instance, the catalytic domains of PlyL (Low *et al.* 2005) and PlyB (Porter *et al.* 2007) had to be cloned and crystallised separately due to degradation of the full-length proteins. The linkers between the domains of these structures are always extended and solvent exposed, comparable to the linker in the crystal structure of the full-length CTP1L monomer.

Matrix-assisted laser desorption/ionisation (MALDI-TOF) mass spectrometry (MS) revealed a protein product that corresponded to the theoretical masses of the C-terminal domains for all three endolysins (Figure 10, B). Combined with the C-terminal domain structures of the degraded full-length CTP1L and CD27L a proteolytic mechanism was evidently occurring at the beginning of the β1- beta strand that joins the C-terminal domain to the linker. For CD27L the calculated MW for the cleaved product corresponded closely to the theoretical mass of the C-terminal domain after peptide hydrolysis of the backbone between a glutamine (Q185) and methionine (M186). Interestingly, for CTP1L and CS74L the experimental masses could not be fitted for conventional peptide hydrolysis. The experimental value of 9032 Da for CTP1L was 40 Da higher than the theoretical mass of 8990 Da for hydrolysis between the glutamate (E194) and valine (V195) cleavage site observe in the crystal structure. Similarly for CS74L, MS analysis revealed a MW of 9075 Da, also 42 Da greater than the calculated mass after peptide hydrolysis at the equivalent cleavage site between lysine 184 and valine 185 in CS74L (9035 Da). The proteolytic products that corresponded to the C-terminal domains were isolated after SDS-PAGE and analysed by tryptic digestion followed by mass spectrometry. It was confirmed that these fragments were the intact C-terminal domains for all three endolysins. As there are no known proteases from *E. coli* that could cleave these scissile bonds (M-Q, E-V or K-V), this indicated that the cleavage could be from a yet unknown group of proteases or an autocatalytic mechanism.

The additional mass observed on the C-terminal domains of CTP1L and CS74L could resemble a post-proteolytic modification for these endolysins. The identification of this mass change is addressed in detail later in Chapter 3: "Probing the autocleavage mechanism of endolysin activation".
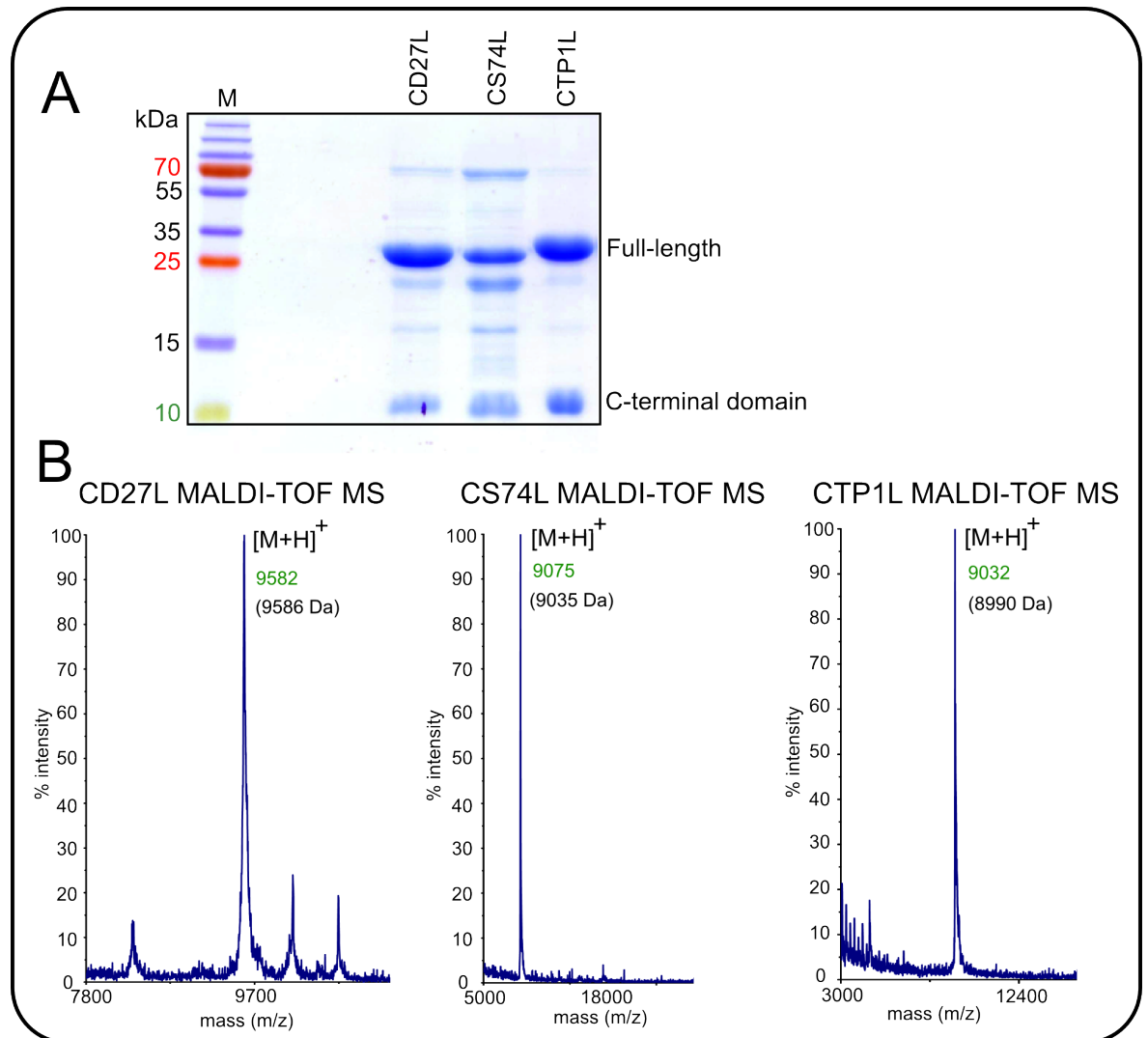


**Figure 10 Ni-NTA purification and MALDI-TOF analysis of CD27L, CTP1L and CS74L A:** SDS-PAGE analysis established cleavage of the C-terminal domain from the full-length proteins for the three endolysins. M: marker. **B:** MALDI-TOF spectra for the cleaved C-terminal domains of the three endolysins. In green are the MS detected masses and in parentheses are the expected masses if peptide hydrolysis was to occur at the cleavage sites. For CD27L the mass over charge ratio (m/z) value of the cleaved C-terminal domain agreed with the expected mass for peptide hydrolysis between Q185-M186. For CS74L and CTP1L an extra mass of 42 and 40 Da was respectively revealed for the C-terminal domains, suggesting a post-proteolytic modification. *The identification of this additional mass is investigated in chapter three.*

*The effect of ionic strength during size exclusion chromatography of the endolysins*

Size exclusion chromatography (SEC) was couple with right-angle light scattering (RALS) / refractive index (RI) / UV measurements (denoted here as SEC-RALS) to determine the absolute molecular mass of the three endolysins in solution and establish the oligomeric states of the endolysins. SEC was performed using the same analytical S75 (10/300) (Tricorn) size exclusion column with the same buffer (20 mM HEPES 500mM NaCl, pH7.4) to allow comparisons between oligomeric states of the different endolysins. CD27L was sensitive to the ionic composition of the buffer used during SEC and could only be easily size excluded in buffers with high ionic strength, (buffers containing ≥ 500 mM NaCl). When lower salt concentrations were used during SEC the recovered amounts of protein were always considerably lower than the amount loaded, suggesting the protein was aggregating in low salt solutions. This is not a unique observation and numerous proteins are sensitive to buffer compositions and ionic strength in particular. However, CD27L was always stable and did not aggregate when dialysed after Ni-NTA purification into low ionic strength solutions (for instance 10 mM HEPES, pH 7.4, the buffer used for Circular Dichroism), suggesting the effect of ionic strength was restricted to chromatography.

The elution profile of wildtype CD27L was complex and suggested that the protein existed as a number of different conformations in solution (Figure 11, A & B). Considering the monomer mass of CD27L is 32 kDa, the calculated molecular weight (MW) of Peak 1 (68 ± 4 kDa) is interpreted as the CD27L dimer (theoretical MW of 64 kDa). For peak 2 the calculated MW of 43 ± 2 kDa was interpreted as a mixture of monomer in complex with the cleaved C-terminal domain (theoretical MW of 42 kDa) and peak 3 gave a calculated MW of 33 ± 7 kDa, which was interpreted to consist of the monomer. The reason for the larger error for the monomer peak MW estimation and a larger range of MW values: minimum MW: 26 kDa, maximum MW: 53 kDa) was due to a lack of resolution on the column when the MW was estimated. Aggregation or the species of a preceding peak can bleed into the peak of interest and influence the right angle light scattering, which produces a broader MW distribution. RALS analysis indicated that CD27L existed in solution with different oligomeric states as a mix between monomer and dimer with remnants of a proteolytic cleavage mechanism in the form of the monomer + cleaved C-terminal domain also being present. To confirm the presence of the monomer + C-terminal domain peak (peak 3) the elutions from SEC were analysed by SDS-PAGE (Figure 11, C). Samples taken from the elutions of peak 3 contained the C-terminal domain and full-length CD27L,

consistent with the MW calculated for this peak. Unfortunately, the peaks are not totally resolved and the C-terminal domain is faintly visible throughout the SEC profile, an unavoidable circumstance due to the resolution limits of SEC. However, peak 3 contained the highest concentration of cleaved C-terminal domain.
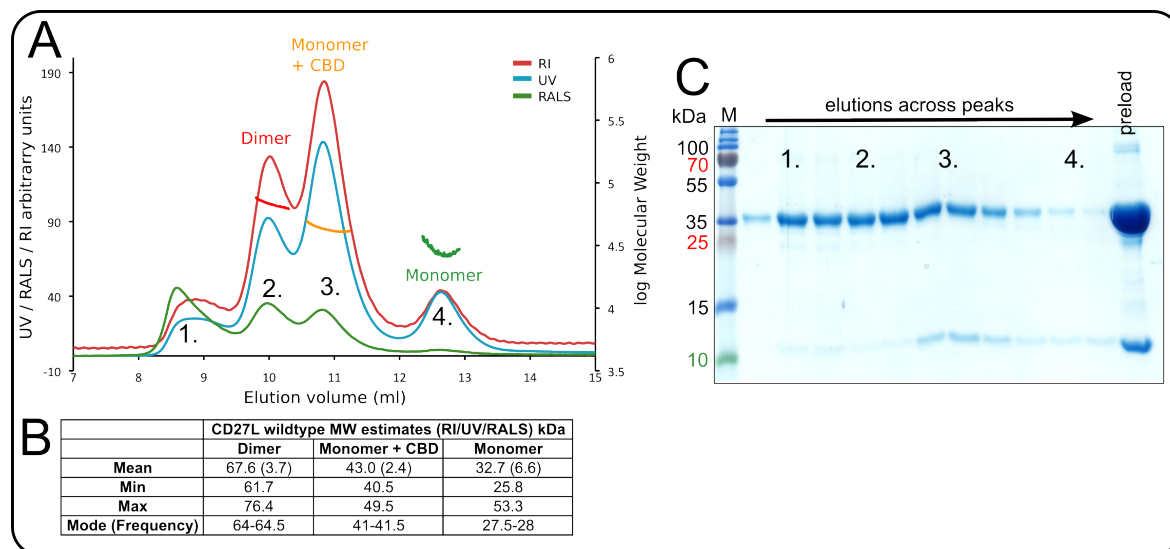


**Figure 11 Oligomerisation of wildtype CD27L in solution A:** UV/RI/RALS size exclusion traces and MW correlation estimates (coloured lines through the peaks) for wild-type CD27L. The dimer (peak 2) and monomer + CBD (peak 3) states of CD27L have narrow MW distributions to give MW calculations close to their theoretical values. Peak 4 is affected by bleeding from previous peaks influencing the RALS intensities, but MW calculations suggested a monomer species. **B:** Summary table showing the MW correlation estimates for wildtype CD27L. Both dimeric (peak 2) and monomer + C-terminal domain (peak 3) states of the protein have narrow MW distributions through their respective peaks. MW distribution of the monomer (peak 4) is broader, caused by resolution limits of the SEC column. This peak may also contain the C-terminal domain and other contaminations. Numbers in brackets refer to ± standard deviation from the mean molecular weight. The mode refers to the most common MW range encountered through the MW correlations **C:** SDS-PAGE analysis shows the preloaded sample before SEC-RALS and the elutions going across the elution profile as collected from a separate SEC run. The lower MW C-terminal domain is present throughout the spectra but most concentrated within peak 3, correlating with the MW of monomer + CBD calculated by RALS. *Image A and table B are reproduced from the submitted article, Appendix page 170.*

The SEC profiles of CTP1L and CS74L also indicated a relationship between the ionic strength of the buffer and the corresponding retention volumes of the endolysins. CTP1L and CS74L both eluted as sharp single peaks at 17.95 ml and 18.4 ml respectively in low ionic strength buffers (e.g. HEPES 20 mM, pH7.4) (Figure 12 A & C). However, upon the addition of 500 mM NaCl to increase ionic strength (e.g. HEPES 20 mM, NaCl 500 mM, pH7.4, as used for SEC-RALS measurements) the endolysins eluted as broader peaks and at a much earlier volume: 10.16 ml for CTP1L and 11.02 ml for CS74L. For both endolysins the elution profiles shifted by 7 ml. This could be due to the change in ionic

strength affecting the configuration of the endolysins, as an increase in NaCl concentration decreases the propensity of the endolysins to interact with the column matrix and therefore producing a more immediate elution. Molecular weight estimations were attempted for CTP1L and CS74L in the low ionic strength buffers, however, the proteins eluted at the end of the column volume where the RI and RALS suffered inconsistent scattering due to the exchange of contaminants initially present in the injected sample.

RALS was used to estimate the MW of CTP1L in the same high ionic strength buffer as CD27L (Figure 12 E). CTP1L eluted at the same volume as the CD27L dimer peak (~10 ml). The calculated MW of 66 ± 5 kDa also corresponded to a MW for the CTP1L dimer (theoretical MW of 65.6). RALS was attempted with the third endolysin CS74L, however, the propensity of CS74L to aggregate (a common occurrence when purifying CS74L) caused erratic, large scattering during RALS measurements and the MW could not be estimated (Figure 12, C). Visual inspection of the SEC profile for CS74L indicated a broad elution with a main peak at 11 ml and a preceding shoulder beginning at 10 ml, similar to the CTP1L profile, except for the large aggregation.

SDS-PAGE analysis on the elutions for both CTP1L and CS74L indicated the presence of the full-length protein and the cleaved C-terminal domain throughout all the elution peaks and was true with or without NaCl in the SEC buffer (Figure 12, B & D). As the full-length protein and the cleaved C-terminal domain eluted as a single peak, the C-terminal domain may still be interacting with the full-length protein after cleavage. The influence of ionic strength on endolysin activity within the cell wall has previously been reported, for instance the activity of *Listeria* phage endolysin Ply500 was shown to decrease as ionic strength increased (Loessner *et al.* 2002). The sudden assembly of holin lesions during host lysis would provoke a large environmental disturbance in ionic composition, which could influence endolysin activation or induce a conformational shift. Nevertheless, in solution it is apparent that CD27L exists as different oligomerisation states including the dimer conformation, whereas CTP1L is dominated by the dimer conformation.
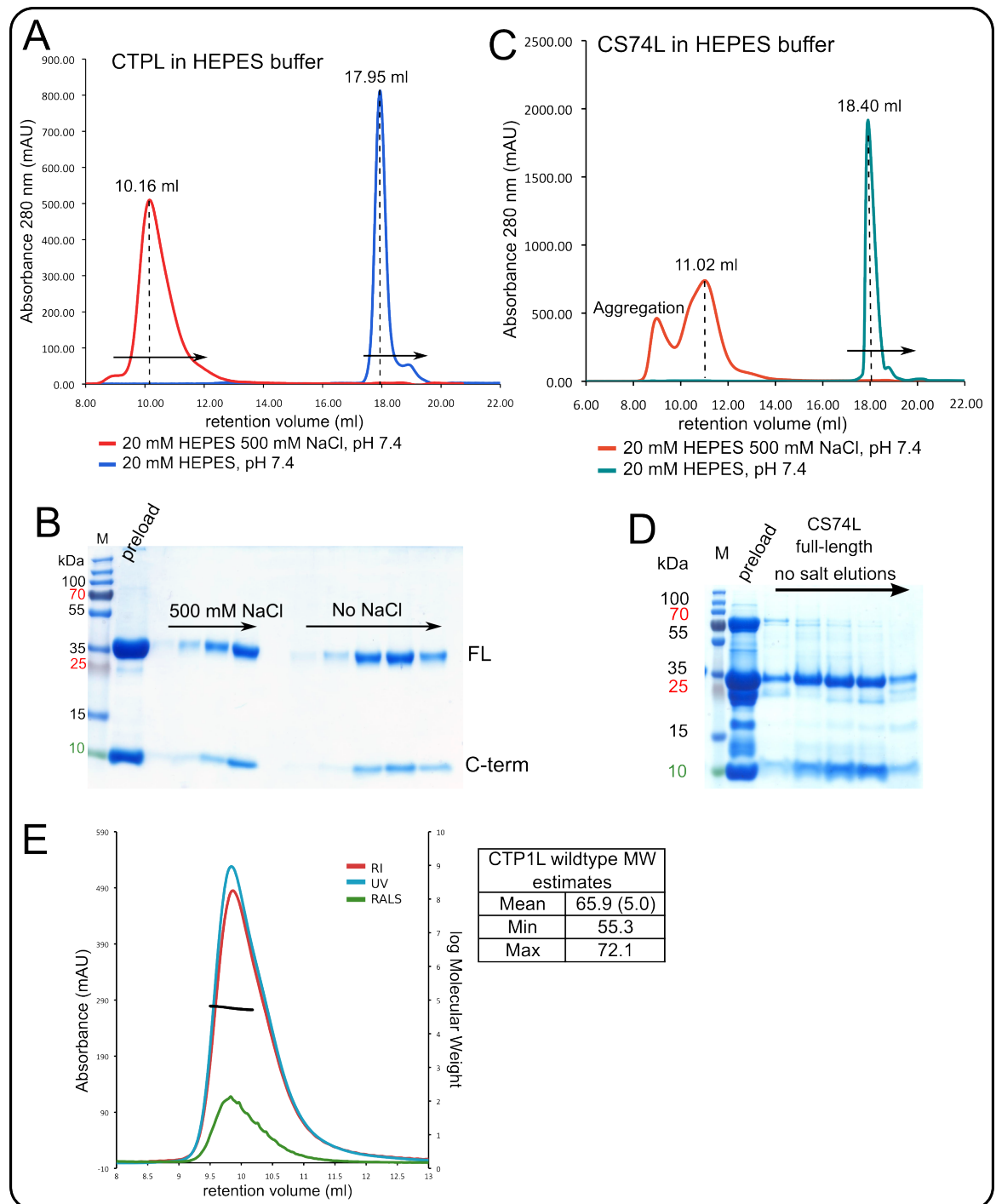
**Figure 12 Ionic strength affected the elution profiles of CTP1L and CS74L A:** SEC profiles of CTP1L in high (red) or zero (blue) salt concentrations (20 mM HEPES ± 500 mM NaCl, pH 7.4). **B:** SDS-PAGE analysis of the two peaks from the SEC profiles of CTP1L in panel A. Both peaks contain the full-length and C-terminal domain. **C:** SEC profiles of CS74L in high (orange) or zero (green) salt buffer as used for CTP1L. In high salt conditions CS74L had a higher tendency to aggregate. **D:** SDS-PAGE analysis on the CS74L peak eluting in the low salt buffer (green line) indicated the presence of the full-length and the C-terminal domain of CS74L. **E:** UV/RI/RALS size exclusion traces and MW correlation estimates for wild-type CTP1L and the summary table of the MW correlation estimated for the single peak, mean MW was 65.9 kDa indicating a dimer in solution. Also shown are minimum and maximum values calculated from the peak. Numbers in brackets refer to ± standard deviation from the mean molecular weight.

### *The head-on dimer state of CTP1L and CD27L*

The crystal structure of the degraded product of the full-length CD27L protein consisted of six C-terminal domain molecules all involved in head-on dimers (Figure 8, B). The head-on dimer is also present between the unit interfaces present in the crystal lattice of the CTP1L crystal (Figure 8, A). For both CTP1L and CD27L the head-on dimer is remarkably similar, the dimer interface is formed between opposing alpha helices α1 and α3 which align parallel to each other and engage in hydrophobic stacking with their symmetry mates from the second C-terminal domain. Using the PyMOL alignment tool the head-on dimers of CD27L and CTP1L can be superimposed to give a root-mean-square deviation (RMSD) of 2.273 Å for 160 residues (Figure 13, A). Interestingly, when engaged in this head-on dimerisation mode the N terminal EADs for both endolysins would point in opposite directions. The buried surface area of the hydrophobic core for the CD27L head-on dimer was calculated between 1200 and 1300 $Å^2$ by the PISA server (Krissinel & Henrick 2007). The core consists of hydrophobic residues valine 204, leucine 261 and leucine 265 and three aromatic residues tryptophan 207, phenylalanine 258 and tyrosine 262 (Figure 13, B). The strong hydrophobic core and stacking of aromatic rings indicate the head-on dimer would be stable in solution.

The CTP1L head-on dimer interface is also dominated by hydrophobic interactions. Along the symmetry axis between helices α1 and α3 of CTP1L, two tyrosines (Y212 and Y260) stack close together with tyrosine 212 forming a hydrogen bond with the carbonyl group of alanine 205 on the opposite domain (Figure 14, A). The side chain of Y260 points away from the dimer interface but is kept in position by interacting with a water molecule. Methionine 263 sits in the centre of the α1 and α3 helices and stacks against its symmetry mate. Methionine at this position is semi-conserved but the sequence alignment indicated that the other hydrophobic amino acids cysteine or leucine can also be found at this position (Figure 9, B & C). Hydrogen bonding between symmetry mates of Glutamine 264 and Leucine 267 further stabilise the dimer interface. Although hydrophobic residues are favoured at these positions for CD27l and CTP1L, sequence alignment indicates that the type of hydrophobic or aromatic residues at these positions is not conserved.

Whilst hydrophobic interactions dominate the centre of the CTP1L head-on dimer, the edges of the dimer interface are involved in salt bridging between the two monomers (Figure 14, A). Aspartate 215 (D215) forms salt bridges with the terminal aliphatic side chains of two arginine residues (R208 and R259) on the opposite domain. The same salt

bridge network exists for the symmetry mates of these residues at the other end of the dimer interface. The presence of salt bridges between the two domains further establishes that the head-on dimer would be stable in solution.


### *Inhibition of the head-on dimer reduces proteolysis*

To test the significance of the CTP1L head-on dimer, and the salt-bridge network in particular, D215 was mutated to alanine (D215A). D215A had a surprising effect as it totally inhibited the proteolytic mechanism that we observed for wildtype CTP1L, as no C-terminal domain could be detected by SDS-PAGE (Figure 14, C). The MW of CTP1L D215A was measured using SEC-RALS (Figure 14, B). D215A eluted as a single peak and gave a calculated MW of 33 ± 1.2 kDa, corresponding to the MW of a CTP1L monomer (theoretical MW: 32.8 kDa). Circular dichroism was used to ensure the inhibition of cleavage was not due to incorrect folding and the spectra closely resembled that of wildtype CTP1L (Appendix Figure 7). As D215A remained as a monomer in solution the salt bridging observed in CTP1L between D215 and R208 and R259 must be critical for the formation of the head-on dimer. This also indirectly implied that without formation of the head-on dimer the proteolytic mechanism that cleaves the C-terminal domain was also inhibited.

R208, D215 and R259 are highly conserved among CTP1L, CS74L and the 19 other aligned sequences, but compared to CD27L only R259 is conserved between the two sets of sequence alignments. In CD27L this arginine is residue R257 (Figure 13, A) and its side chain points away from the centre of the stacked helices of the head-on dimer and is not involved in bonding between the two monomers of the head-on dimer. Without a salt bridging network as seen in CTP1L, the head-on dimer of CD27L is maintained solely by hydrophobic interactions between the two domain and thus would not be as stable as the head-on dimer in CTP1L. For CD27L mutants of the aromatic residues tryptophan 207 (W207A and W207R) and tyrosine 262 (Y262A) that form part of the head-on hydrophobic interface were explored to see if interference of the head-on dimer also affected CD27L (Figure 13, B & D).

These mutants did not totally inhibit cleavage of the C-terminal domain as seen for CTP1L D215A, however, they significantly inhibited proteolysis and all showed a large decrease in the amount of C-terminal domain detected by SDS-PAGE (Figure 13, D). Mutants Y262A and W207A showed a large inhibition in cleavage whereas mutation W207R

showed a moderate decrease in cleavage. For W207R this was surprising, as the additional charge repulsion when arginine is incorporated was predicted to hinder head-on dimerisation to a greater degree than alanine incorporation.
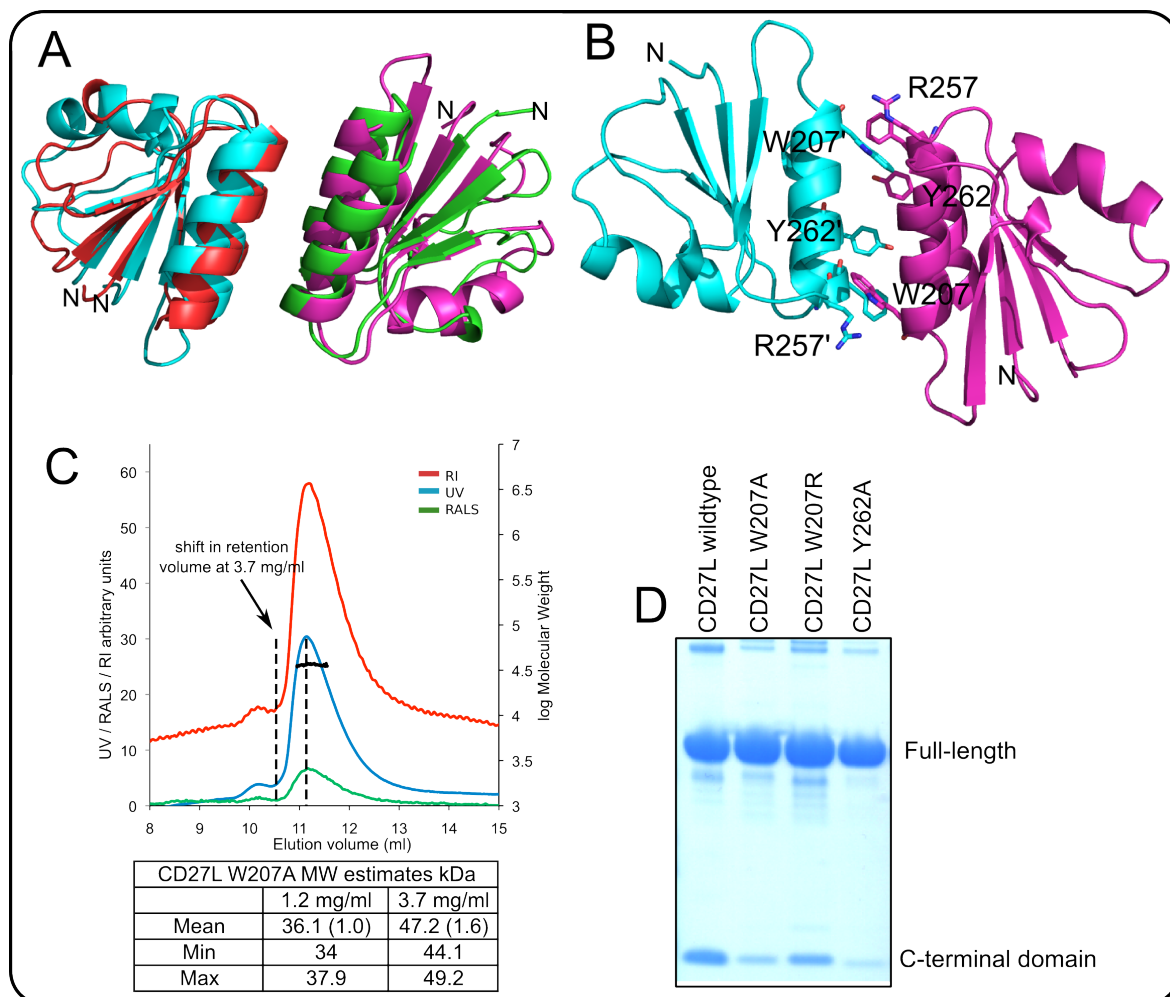


**Figure 13 Mutations within the head-on dimer interface of CD27L A:** Superposition of the C-terminal domains of CTP1L (red and green) and CD27L (cyan and magenta) in the head-on dimer conformation. **B:** Cartoon representation of the C-terminal domain head-on interface of CD27L. Aromatic residues W207 and Y262 are part of the hydrophobic interface that maintains the dimer interface in solution. Also highlighted is R257, which is the only conserved residue in the head-on dimer that shows conservation between CD27L, CS74L and CTP1L. In CTP1L the equivalent R259 is involved in a salt bridging network at the dimer interface, however, R257 forms no polar contacts within the crystal structure of CD27L. In the head-on dimer state the N-terminal linkers are pointing in opposite directions. **C:** UV/RI/RALS SEC trace for the low concentration (1.2 mg/ml[-1]) of CD27L W207A. The 3.7 mg/ml[-1] sample had the same profile, except the retention volume of the peak shifted by 0.5 ml from 11.3 ml to 10.8 ml (indicated by dotted line). MW correlation estimates for the mutant W207A at the different concentrations were calculated. The estimated MW of 36 kDa at low concentration increases to 47 kDa with the higher concentration. **D:** SDS-PAGE analysis of CD27L wildtype and the three head-on mutants, W207A, W207R and Y262A. Head-on mutants showed a decrease in the amount of cleaved C-terminal domain. All samples were loaded fresh after Ni-NTA purification at a concentration of 2 mg.ml[-1].
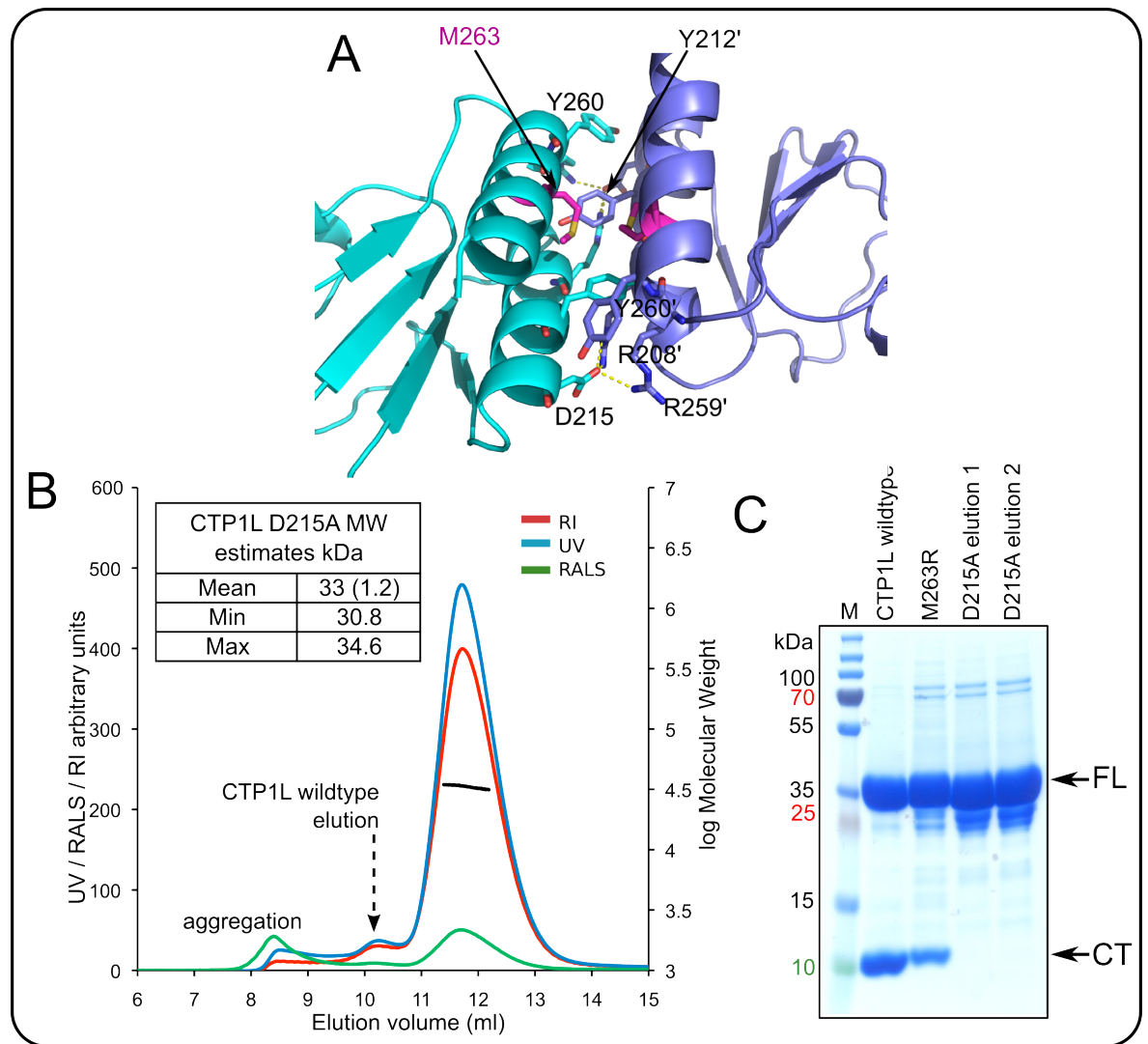
**Figure 14 Mutation within the head-on dimer interface of CTP1L A:** Cartoon representation of the C-terminal domain head-on interface of CTP1L. Highlighted residues D215, R208 and R259 form a salt bridging network at either end of the head-on interface and the N-terminal linkers are pointing in opposite directions. M263 residues are also indicated (magenta) and proposed to be part of the cell wall binding site of the C-terminal domain. Also indicated are Y212 and Y260, aromatic residues involved in the head-on interface **B:** UV/RI/RALS SEC traces and MW correlation estimates for the head-on interface mutant D215A, the peak eluted at a higher retention volume than the wildtype CTP1L (indicated by dotted line) and the MW estimation of 33 kDa corresponded to a monomer in solution. **C:** SDS-PAGE analysis of wildtype CTP1L, M263R and low and high concentrations of D215A. Mutation M263R is detailed on page 47. Mutant D215A totally inhibited cleavage of the C-terminal domain. All samples were loaded fresh after Ni-NTA purification at a concentration of 2 mg.ml$^{-1}$.

W207A was chosen for SEC-RALS measurements to investigate the effect this mutation had on the dimer state of the endolysin (Figure 13, C). Compared to the complicated elution profile of wildtype CD27L, W207A eluted as a single peak and gave a calculated MW of 36 ± 1.0 kDa, close to the theoretical mass of the CD27L monomer (32 kDa). To improve the quality of the RI/RALS traces the concentration of the injected sample of W207A was increased from 1.2 mg.ml$^{-1}$ to 3.7 mg.ml$^{-1}$. Interestingly, when a higher

concentration of W207A was used the calculated MW increased to $47 \pm 2$ kDa and the retention volume of the peak shifted from 11.3 ml (for 1.2 mg.ml$^{-1}$) to 10.8 ml (for 3.7 mg.ml$^{-1}$), which was closer to the retention volume of the dimer peak observed for wildtype CD27L (10.6 ml). This change in mass was not due to a dilution effect, calculation error or bleeding of peaks skewing the calculations. Considering RALS calculates the average MW across a peak, W207A could exist as a dynamic mixture of monomer and dimer in solution, whereby the increase in concentration shifted the equilibrium towards the dimer state, observed by the increase in the calculated MW.

### *A single mutation within the proposed cell wall binding site inhibits lytic activity*

Methionine 263 resides on helix α3 of the C-terminal domain of CTP1L and points towards the interface of the proposed head-on dimer of the C-terminal domain (Figure 14, A). Mutation to arginine (M263R) replaced the hydrophobic side chain with a positive charge, which was expected to change the dynamics in the dimerisation mode and affect the ability of CTP1L to form the head-on dimer. However, the C-terminal domain was still cleaved to the same degree as the wildtype CTP1L and indicated no modification of the autocleavage mechanics (Figure 14, C). Interestingly, this mutant totally abolished CTP1L lytic activity against *C. tyrobutyricum* cells (Figure 22, A). Applying an excessive amount of M263R (30 μg compared to 1 μg and 10 μg normally used) also could not establish a lytic response.

BLAST sequence analysis indicated methionine at this position is not conserved but is present in 6 of the 21 aligned sequences, with cysteine as the most common residue found at this position. Mutation of M263R provided indirect evidence that the C-terminal domain of CTP1L could be involved in cell wall binding. As M263R abolished cell lysis but did not affect the proteolytic cleavage mechanism, the introduction of a positive charge was suggested to modify the normal behaviour of the binding site and prevented attachment of the endolysin to the cell wall. Without cell wall attachment the lytic activity was inhibited.

### *The side-by-side dimerisation mode*

A second dimerisation mode exists in the crystal structures of CD27L and CTP1L. In the asymmetric unit of CTP1L the C-terminal domain of the full-length monomer forms a close interaction with the cleaved C-terminal domain, denoted as a side-by-side dimer (Figure 15, A). A similar side-by-side dimer exists for two of the C-terminal domains in

the crystal structure of the degraded full-length CD27L (Figure 15, B). Attempts to superimpose the two side-by-side dimers of CTP1L and CD27L failed and visual inspection of the two endolysins showed that are comprised of different side-by-side conformations. Even though the side-by-side dimers are comprised of different arrangements of the domains, the positions of the linkers that extend from the C-terminal domains to the N-terminal EADs are parallel for both CD27L and CTP1L. Additionally, in both side-by-side dimers the alpha helices α1 and α3, which form the head-on dimer interface, are exposed and point away from the centre of the dimer interfaces. For CD27L the pairs of helices point 180° away from each other, whereas in CTP1L the pairs of helices are facing 90° away from each other (Figure 15, A). If the C-terminal domains do interact as cell wall binding domains it would be in the side-by-side formation where they would most likely be able to bind to substrates when the proposed binding sites, including M263, are facing outwards.

### *Mutagenesis to prevent the side-by-side dimer formation*

In the side-by-side dimer of CD27L the extended α2 helices of opposing monomers form a buried surface area of 1216 $\text{Å}^2$ together, a comparable value to the head-on dimer interface (1200-1300 $\text{A}^2$). Within the interface the side chain sulfhydryl of a semi-conserved cysteine residue (C238) protrudes into the centre of the dimer interface, coming within 3.4 Å of its symmetry mate. This distance is too large to qualify for a disulphide bond. Around C238 are residues methionine 251, lysine 253 and their respective symmetry mates. All their side chains point in the direction of the C238 sulfhydryl groups and are proposed to destabilise the formation of a possible disulphide bond between the C238 residues. The exterior of the bacterial cell would favour the formation of disulphide bonds, which would strengthen the side-by-side dimer so that the endolysin becomes accessible to bind to the cell wall as the C-terminal domains are facing outwards and not constrained as seen in the head-on dimer.

To test if disulphide formation contributed to the side-by-side dimerisation, cysteine 238 was mutated to serine (CD27L C238S) replacing the terminal sulfhydryl group with a hydroxyl. SDS-PAGE analysis showed that C238S was still cleaved to the same extent as wildtype CD27L (Figure 15 E), indicating that side-by-side formation is not governed by disulphide bond formation, instead the side-by-side formation is most likely maintained by electrostatic interactions between the two monomers. Therefore, to test if the repulsion of opposing charges could inhibit side-by-side dimerisation, C238 was mutated to arginine

(C238R). C238R had the same effect on proteolytic cleavage as the head-on mutants (W207A, W207R and Y292A) and significantly reduced the production of the cleaved C-terminal domain as observed by SDS-PAGE (Figure 15 E).

The side-by-side dimer of the CTP1L C-terminal domains has a buried surface area of 1120 $Å^2$, a similar value to the side-by-side dimer of CD27L. The interface consists of main chain hydrogen bonding between cysteine 219, proline 220, threonine 221 and isoleucine 222. There is also salt bridging between symmetry mates of arginine 226 and glutamate 211 at both ends of the side-by-side dimer interface. R226 also interacts by salt bridging to D215, which is also involved in salt bridging in the head-on dimer interface. This suggests that D215 could have a dual role connecting these two dimer states. The side chain of the cleavage site V195 for the full-length monomer also closely interacts with the edge of the side-by-side dimer interface. Within the centre of the side-by-side interface, the closest interacting residues consist of hydrogen bonding between the threonine 221 symmetry mates. Even though the side-by-side interfaces differ between CD27L and CTP1L, in an early attempt to mimic the latent disulphide bond theorised to exist between the C238 residues in CD27L and promote side-by-side dimerisation, threonine 221 was mutated to cysteine (T221C). Interestingly, the mutation had the opposite effect and instead of stabilising the side-by-side dimer it totally inhibited proteolysis and no cleaved C-terminal domain could be detected by SDS-PAGE (Figure 15, E). The introduction of opposite charges, as performed for the C238R mutant of CD27L, was replicated into CTP1L (T221R), which also totally inhibited proteolysis (Figure 15, E).

Sequence alignment between CTP1L and CS74L indicated that T221 was also positioned at the same location in the C-terminal domain of CS74L. To test the conservation of this side-by-side interface, I introduced the same arginine mutation into CS74L (T211R). T211R demonstrated the same effect as CTP1L T221R and T221C and totally inhibited proteolysis (Figure 15 E). Interestingly, both CTP1L T221R and CS74L T211R showed signs of degradation to a fragment with a molecular mass just below the mass of the full-length proteins. The degradation of CS74L T211R was more pronounced and the entire full-length protein eventually formed the degradation product (Figure 15, E). MALDI-TOF MS analysis revealed the mass of these fragments to be 21.8 kDa for CTP1L T221R and 24.2 kDa for CS74L T211R (Appendix Figure 4). Tryptic digestion linked with liquid chromatography tandem mass spectrometry (LC-MS/MS) identified peptides within the C-terminal domain, beyond the proposed site of autocleavage at position 185, indicating the

C-terminal domain was internally degrading, which was not due to autocleavage. Furthermore, the masses of these degradation products do not correspond with the MW of the EAD (MW of CTP1L EAD 23.8 kDa, EAD of CS74L 22.1 kDa) or from degradation within the 6 residue linker regions for the two endolysins. Therefore, without the formation of a stable side-by-side dimer these endolysins are susceptible to degradation. Degradation products could not be detected for CTP1L T221C and it was subsequently submitted for crystallisation trials, although unfortunately no positive crystal hits could be detected.

### *Right angle light scattering on the side-by-side mutants*

The oligomeric states of the side-by-side mutants were also assessed by SEC-RALS, using the same measurement conditions as the wildtype endolysins. CD27L C238R eluted as a single peak at the same retention volume as the dimer peak observed for wildtype CD27L (Figure 15 D). MW estimations indicated a mass of $60 \pm 4$ kDa, which closely matched the theoretical mass of the CD27L dimer (64 kDa). CTP1L T221C eluted as two peaks during SEC-RALS measurements. The first peak with a retention volume of 10.4 ml aligned with the same retention volume of the dimer peak for wildtype CTP1L. The second peak eluted at 11.7 ml with the same retention volume as the monomer peak observed for the head-on mutant D215A. Unfortunately, due to the lack of resolution between the two peaks and the aggregate peak bleeding into both peaks, MW calculations using RALS were not possible. The presence of monomer and dimer species in solution for CTP1L T221C suggested that inhibition of the side-by-side dimer increases the propensity for the endolysin to form an equilibrium state in solution.
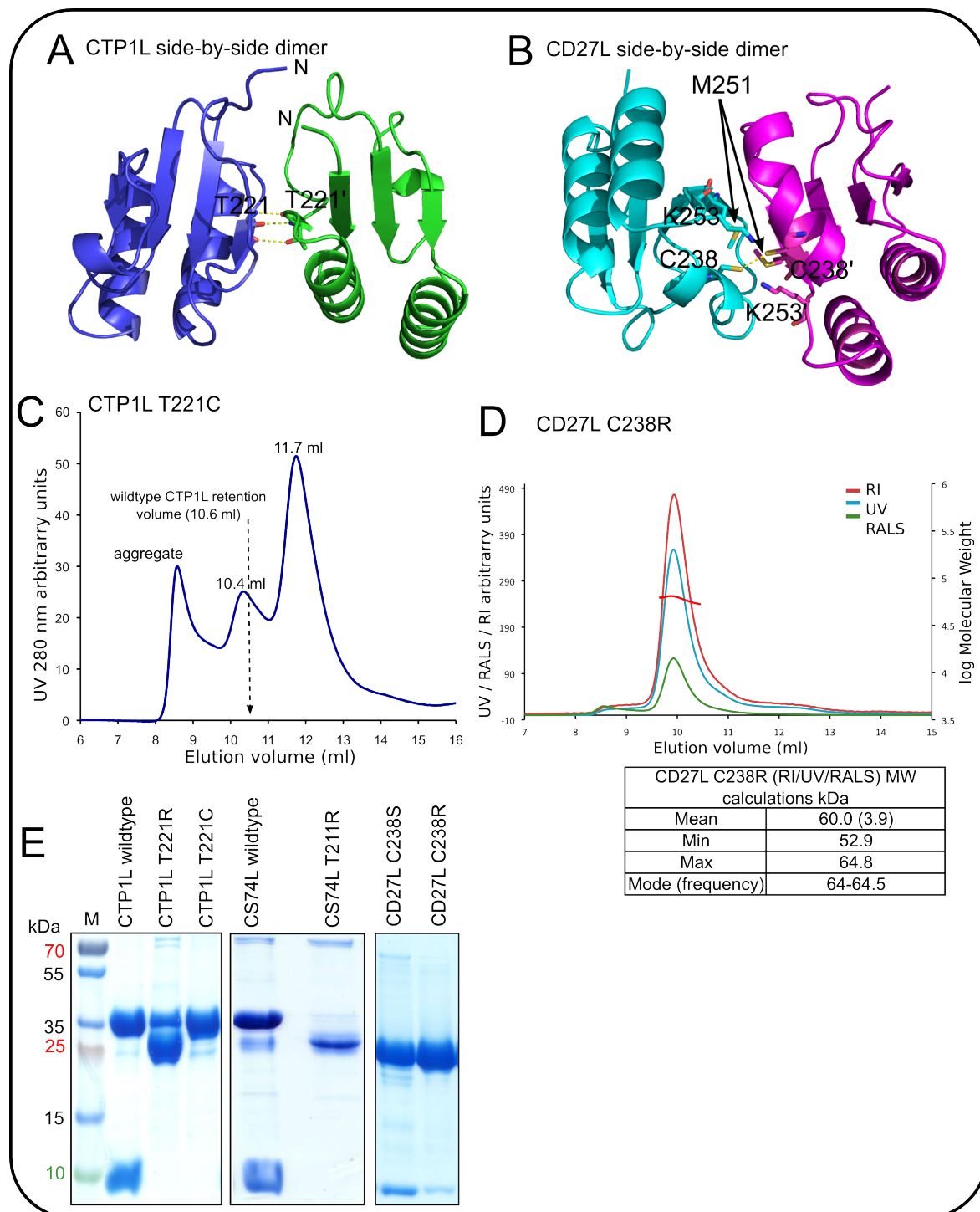
**Figure 15 Mutations of the side-by-side dimer of CD27L and CTP1L. A:** Cartoon representation of the side-by-side dimer of CTP1L with residue T221 highlighted which was modified to arginine (T221R) or cysteine (T221C). **B:** Cartoon representation of the side-by-side dimer interface of CD27L indicating the close arrangement of residues around cysteine 238, modified to arginine (C238R) in Figure D. **C:** SEC UV 280 nm profile of CTP1L T221C (Buffer: 20 mM HEPES, 500 mM NaCl, pH 7.4) eluting in two peaks, the 10.4 ml elution corresponds to the dimer peak observed for wildtype CTP1L and the elution at 11.7 ml corresponds to the monomer peak of CTP1L D215A. The aggregation peak affected MW calculations using RI/RALS and are not shown. **D:** UV/RI/RALS SEC traces and MW correlation estimates for CD27L C238R. The protein eluted as a single main peak corresponding to a MW of 60 kDa, close to the theoretical dimer MW. The elution profile shows a long tail spread between 10.5 ml to 13 ml suggesting other species such as the monomer are also present during elution. *Image D is reproduced from the submitted article Appendix page 170.*

*Small-angle X-ray scattering to determine low resolution shape in solution*

RALS measurements indicated the presence of an equilibrium for the endolysins between a monomer and the head-on and side-by-side dimer states. Small-angle X-ray scattering (SAXS) was used to determine the low-resolution global shape and determine the oligomeric state of the endolysins in solution. The SAXS analysis was performed with the assistance of Haydyn Mertens (EMBL, Hamburg), who analysed the data and produced the scattering curves and DAMMIF models shown in Figures 16 and 17. Full details of experimental conditions and analytical processes can be found in materials and methods and Appendix table 5. SAXS data was collected for all three wildtype endolysins CTP1L, CS74L, CD27L and also the C238R mutant of CD27L in concentration series ranging from 0.5 to 4 mg/ml. Data were merged and extrapolated to infinite dilution to obtain SAXS profiles with minimal inter-particle interference.

*The shape of the dimers for the oligomer state of CD27L and CTP1L*

A full-length model of CD27L was generated by combining the previously solved catalytic domain of CD27L (Mayer *et al.* 2011) (PDB: 3QAY) combined with the CD27L C-terminal domain and using the structure of the intact PlyPSA endolysin (PDB: 1XOV) for domain placement. From the forward scattering intensity the molecular mass of wildtype CD27L was estimated to be $42 \pm 4$ kDa, considerably lower than the theoretical mass of 64 kDa for the CD27L dimer. This suggested the existence of an equilibrium between the head-on and side-by-side dimerisation states combined with dissociation products. SAXS curves were generated for the head-on and side-by-side dimers using CRYSOL, however, they produced poor fits with respective discrepancies of $\chi = 1.8$ and $4.0$ when compared to the experimental data from wildtype CD27L (Figure 16, A). Interestingly, the experimental data of CD27L C238R fitted a scattering curve calculated for the head-on dimer configuration with a discrepancy approaching $\chi = 1.0$. A model of the side-by-side dimer could not be fitted to the scattering curve of C238R and gave a discrepancy of $\chi = 3.3$. This indicated that the solution state of the C238R mutant exists as the head-on dimer and correlated well with the dimer species calculated from the RALS measurements. The real space probability distribution *p(r)* function for atomic pair distances of the C238R mutant (Figure 16, B, green line) displayed two distinct peaks. The distribution was characteristic of an elongated protein with a large peak at 70 Å that matched the distance between the two centres of the catalytic domains when the endolysin is engaged in the head-on dimer. The *p(r)* function of the wildtype protein (Figure 16, B, blue line) lacks this peak at 70 Å and instead displays a smaller maximum size, again suggesting an equilibrium between

dimers and dissociated products. The model volume constructed for the wild-type data ($V_{ex}$=94000 ± 10000 Å$^3$) is consistent with a mixture and for the C238R data ($V_{ex}$ = 123000 ± 10000 Å$^3$) a dimeric conformation (Appendix Table 5). Additionally, the extended envelope reconstructed from the experimental SAXS *ab initio* for the C238R mutant overlays well with the head-on dimer model (Figure 16, C & D).
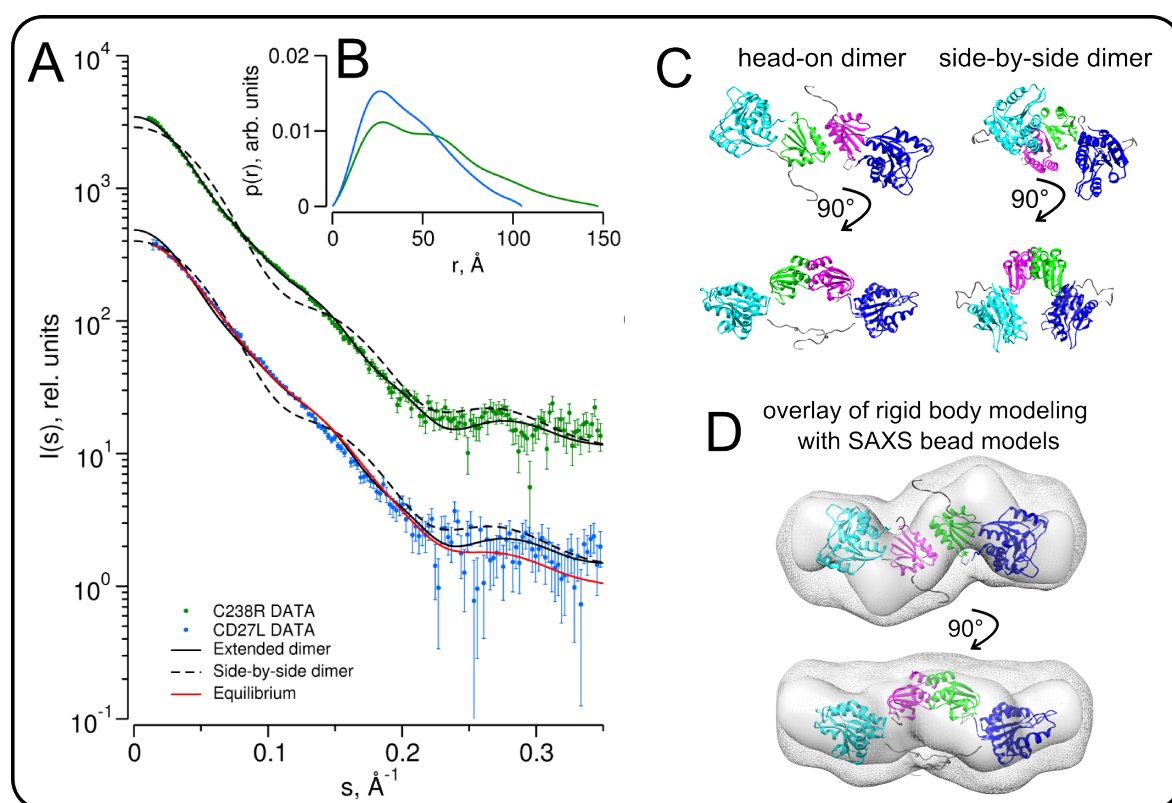


**Figure 16 SAXS analysis on CD27L and CD27L C238R. A:** Experimental scattering curves for CD27L (blue) and C238R (green) with calculated scattering curves of different dimer configurations as indicated in the legend. Missing regions of the structure (n-terminal polyhistidine tags and the interdomain linker) were refined against the SAXS data, using the program CORAL to keep the domains fixed. The fit to the curve using the equilibrium model generated by OLIGOMER is shown as the red line. **B:** Real-space distance-distribution functions *p(r)* determined by indirect Fourier transformation (CD27L: blue, C238R: green) **C:** Reconstructed domains of CD27L C238R refined by rigid body modeling (cartoon) are overlaid with the best (surface) and average (mesh) DAMMIF bead models. In both reconstructions C238R exists as the elongated head-on dimer with the C-terminal domains in the centre and the EADs facing in opposite directions on the exterior. *Figure 16 is reproduced from the submitted article Appendix page 170.*

To further investigate oligomerisation states and clarify the appearance of degradation products for the CD27L samples, the program OLIGOMER (Petoukhov *et al.* 2012) was used to fit the experimental scattering curves of wildtype CD27L and the side-by-side mutant C238R to model potential multi-component mixtures of proteins and find the fractions of each component in solution. Form factor files were generated for the head-on

and side-by-side dimer modes and individual domains. These served as input for a fitting procedure where the discrepancy between the theoretical scattering of the individual components and the experimental data is minimised through adjustment of component volume fractions (Table 1). The contribution from potential degradation products (for instance, the side-by-side dimer without the EADS, dimers of C-terminal domains alone and individual domains) were pooled together as a single component.

| | Volume fractions | | | |
| Sample | Head-on dimer | Side-by-side dimer | Degradation products | Fit, $\chi$ |
|---|---|---|---|---|
| CD27L | $0.41 \pm 0.02$ | $0.24 \pm 0.02$ | $0.35 \pm 0.02$ | 1.5 |
| C238R | $0.98 \pm 0.01$ | $0.02 \pm 0.01$ | 0.0 | 1.1 |

**Table 1 Equilibrium analysis of the SAXS data for CD27L and CD27L C238R using OLIGOMER.** Values reported for merged data sets (CD27L: 0.9 & 4.0 mg.ml$^{-1}$, CD27L C238R: 1 & 8.4 mg.ml$^{-1}$)

For wildtype CD27L the head-on dimer is the dominant species in solution, but other components (side-by-side dimer and degradation products) are also present, explaining the low apparent molecular mass determined from the wildtype SAXS data and why individual components could not be fitted. For the side-by-side mutant C238R, scattering is described exclusively by the head-on dimer state scattering, indicating that inhibition of the side-by-side dimer contributes to maintaining CD27L in the head-on dimer state. The fit of the equilibrium state of CD27L wildtype is depicted as a red line in Figure 16 A.

*Oligomeric states of CTP1L and CS74L*

SAXS was also used to detect the oligomeric states of CTP1L and CS74L in solution. During SEC the salt content of the buffer affected the retention volume of CTP1L and CS74L, suggesting ionic strength influences oligomeric state. Both proteins were purified using a buffer with no salt (20 mM HEPES, pH 7.4) and just before SAXS measurements NaCl was added to a final concentration of 500 mM. However, the only noticeable difference when salt was added during SAXS experimentation was a reduction in scattering contrast between sample and solvent and no effect on the oligomeric state was observed. Unfortunately, due to experimental error with concentration measurements, the molecular masses calculated from the forward scattering intensities for CTP1L and CS74L were affected and underestimated. Nevertheless, the molecular mass calculated for CTP1L (62 ± 5 kDa) suggested a dimer and the molecular mass calculated for CS74L of 46.5 ± 4

kDa, lower than the expected dimer MW (62 kDa), suggested the existence of an equilibrium in solution as previously observed for wildtype CD27L SAXS measurements.

When dimeric models of the head-on and side-by-side dimer were generated based on the CTP1L crystal structure they provided poor fits to the data (3.1 < χ < 5.1) (Figure 17, A). For CTP1L and CS74L it was subsequently observed that the *ab initio* envelope reconstructed from the SAXS scattering by DAMMIF displayed elongated volumes with additional density in the centre of the envelopes. This extra density could be occupied by a head-on dimer with two C-terminal domains associated in the centre (Figure 17 B). Indeed, such a structure provided a very good fit to both scattering curves (χ = 1.4 for CTP1L and χ = 1.5 for CS74L) (Figure 17, A). Thus both endolysins are best described in solution by a model of the head-on dimer with the cleaved C-terminal domains attached in a side-by-side orientation, as observed in the crystal lattice of CTP1L (Figure 8, A). OLIGOMER was used to investigate the potential multi-component mixtures of CS74L and CTP1L and find the fractions of each component in solution. This was performed in a similar approach to the analysis on CD27L and CD27L C238R, except the generated form factor files used the crystal structure of CTP1L as a model instead of CD27L. Generated models included the head-on dimer and the head-on dimer with the two cleaved C-terminal domains still attached that gave the best fit to the CTP1L and CS74L scattering curves. The contributions from potential degradation products were pooled together as a single component and included the CTP1L monomer, dimers of C-terminal domains alone and the CTP1L EAD alone). The volume fractions of all individual components that minimised the discrepancy between theoretical scattering and experimental data are described in Table 2.

| | Volume fractions | | | |
|---|---|---|---|---|
| Sample | Dimer + 2x C-terminal domains | Head-on dimer | Degradation products | Fit, χ |
| CTP1L | 0.85 ± 0.01 | 0.14 ± 0.01 | 0.02 ± 0.01 | 1.36 |
| CS74L | 0.70 ± 0.02 | 0.02 ± 0.02 | 0.29 ± 0.02 | 1.1 |

**Table 2 Equilibrium analysis of the SAXS data for CD27L and CD27L C238R using OLIGOMER.** Values reported for merged data sets (CTP1L 0.24 to 4.0 mg.ml$^{-1}$, CS74L C238R: 1.3 to 6.8 mg.ml$^{-1}$)

For CTP1L the fit did not change when a multicomponent state was fit to the data ($\chi$ = 1.36, compared to 1.40) and suggested that the head-on dimer with the two retained cleaved C-terminal domains was the dominant species in solution (85%) with the head-on dimer alone as the other component (14%). The equilibrium mixture calculated by OLIGOMER for CS74L was checked against the original scattering curve and indicated an even better fit to the data ($\chi$ = 1.1) than the head-on dimer plus the two C-terminal domains alone (Figure 17, A, blue line). For CS74L, the same head-on dimer with the two retained C-terminal domains was also the dominant species in solution (70%) and the rest of the equilibrium mixture (29%) consisted of the accumulation of degradation products. For CS74L this was a similar result to the OLIGOMER analysis of wildtype CD27L and explained why the molecular mass calculated from the forward scattering of CS74L was lower than expected, as lower MW components were also present in solution.
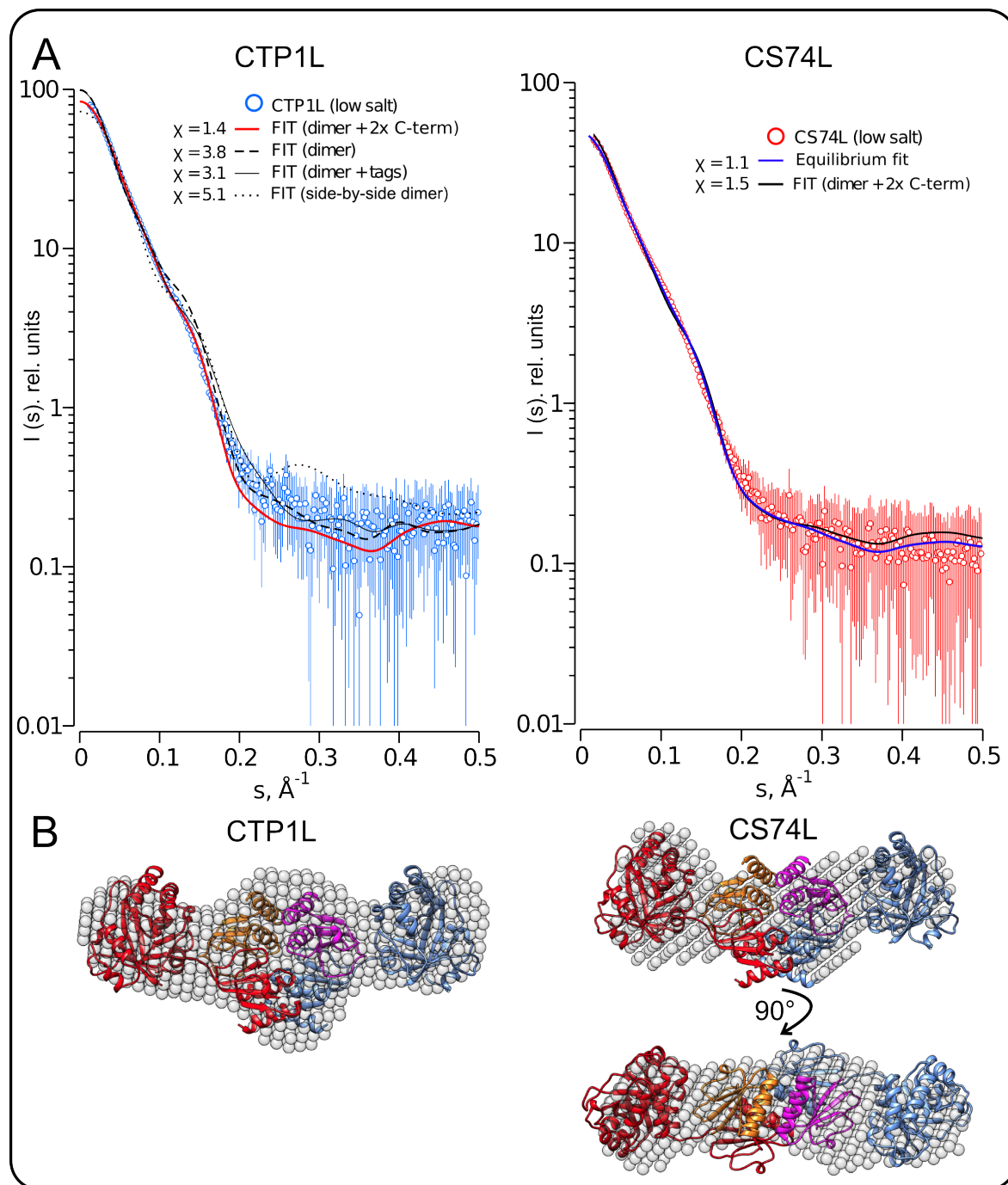
**Figure 17 Determination of the dimer configuration of CTP1L and CS74L in solution by SAXS A:** Overlay of the experimental scattering curves for CTP1L (blue) and CS74L (red) with the calculated scattering curves from different dimer configurations, as indicated in the legend, built as composite models from the crystal structure of CTP1L. Also shown for CS74L is the fit for the equilibrium model determined from OLIGOMER. **B:** Using SUPCOMB the crystal structure of CTP1L was superimposed with the best and average *ab initio* bead models reconstructed by DAMMIF for CTP1L and CS74L. Both envelopes have an elongated shape that accommodates the head-on dimer with the two C-terminal domains in the centre and the EADs at the exterior. In the centre of the head-on dimer the two cleaved C-terminal domains remain bound as seen in the crystal structure. For CTP1L the best fit was for the dimer + 2 C-terminal domains and the best fit for CS74L was the equilibrium model consisting of 70% dimer + 2 C-terminal domains, plus degradation products.

*A tetramer intermediate could facilitate the dimer switch of CTP1L*

For a single endolysin dimer to switch from the head-on to the side-by-side state, it would require a large conformational change to drastically rotate one of the N-terminal EADs by 180 degrees to place it adjacent to the other EAD. The likelihood of this happening is small and the switch between the two states most likely requires an intermediate step. Within the crystal structures of CD27L and CTP1L a tetramer can be observed between four individual C-terminal domains (Figure 8, A & B). This tetramer could represent the intermediate conformation as both the head-on and side-by-side dimer states are interacting (Figure 18, A). In CTP1L the central residue D215 appears to play a central role in connecting these two dimer states.
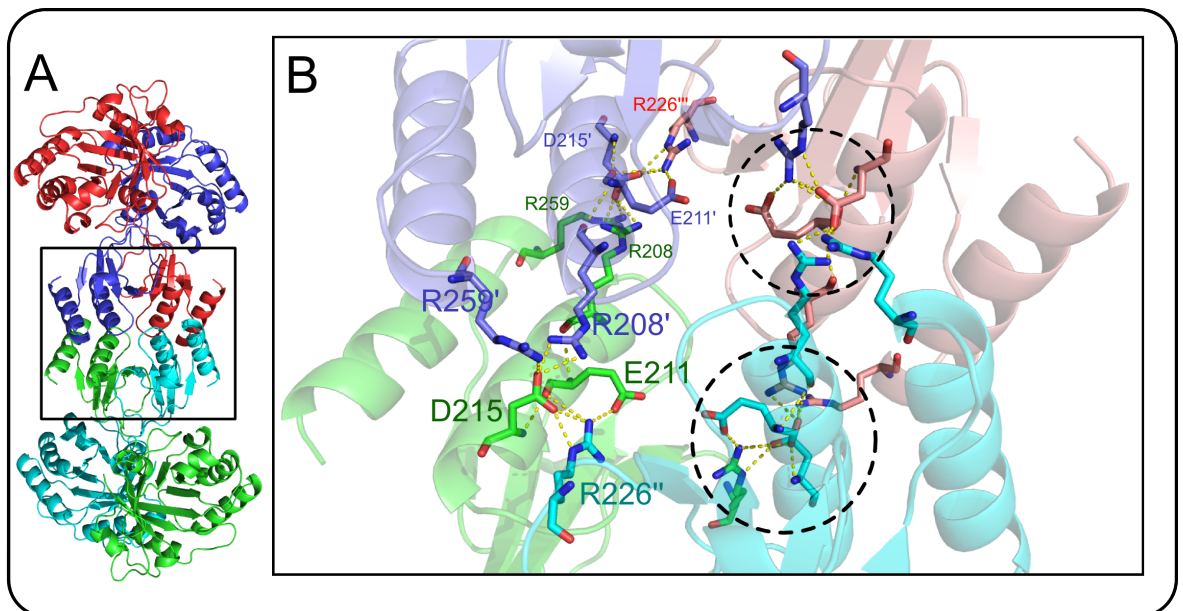


**Figure 18 The potential tetramer intermediate state of CTP1L. A:** The proposed tetrameric state involving both head-on and side-by-side interfaces, image generated in PYMOL by modeling the released EADs onto the cleaved domains. **B:** Salt bridging networks present between the C-terminal domains in the CTP1L crystal structure. The head-on dimer interface, stabilised by salt bridging between D215 and R259' and R208', between two C-terminal domains (green and blue), is attacked by the side chain of the R226'' residue after the formation of the side-by-side dimer (cyan and green). R226'' also interacts with E211. All the generated bonds are 2.8-3.0 Å in length, qualifying for salt bridges. The side-by-side conformation brings the two linkers into close proximity, where they interact and autocleavage is triggered to release the EADs.

The tetramer state of CTP1L consists of multiple salt bridging networks between the C-terminal domains (Figure 18, B). Within the head-on dimer, D215 interacts with R208' and R259' to form a network of electrostatic interactions (cyan and blue domains) at both ends of the head-on interface. In the side-by-side dimer, D215 and E211 interact with R226'' from the side-by-side domain (blue and green domains). Considering the head-on dimer is

the predominant state in solution, as observed by SAXS analysis, the interacting charges of the D215/R208'/R259' network is attacked by the incoming R226'', which also bonds with the E221 residue, as the C-terminal domains are forced into the side-by-side state. All these residues focus their charged side chains onto the D215 carboxyl group, with a high number of water molecules present that would help stabilise this tetramer state network. Potentially, the formation of the tetrameric intermediate forces the head-on dimer interfaces (cyan/blue and green/magenta) to break and the side-by-side dimer interface becomes the dominant interaction. The components of each dimer exchange and two side-by-side dimers are released (blue and green/cyan and magenta). The attack of the R226'' residue occurs at both ends of the head-on dimer interface so overall in the tetrameric state all four salt bridge networks are attacked by R226'' residues from opposing side-by-side domains (Figure 18, B, dotted circles). As the side-by-side dimer forms, the cleavage of the C-terminal domains is proposed to occur, as both cleavage sites come into close proximity.

As previously mentioned, the head-on network of R208/D215/R259 is highly conserved among CTP1L, CS74L and the 19 other aligned sequences. Sequence alignment also showed that R226 is present in CS74L and 9 of the 19 other sequences (Figure 9, B). Together this suggests that these two salt bridging networks could be a conserved feature for these endolysins. Visual inspection of the CD27L tetramer indicated a different tetrameric conformation to that of CTP1L and no such salt bridging networks could be established for CD27L. The only conserved residue is R259 in CTP1L with R257 in CD27L, which is conserved at the same position in the CD27L crystal structure (Figure 14, A).

### *Mutation within the cleavage site inhibits the proteolytic release of the C-terminal domain*

Proteolysis of the three endolysins occurs at the stem of the linker that connects the C-terminal domain to the N-terminal EAD. Proteolysis of CD27L occurs at methionine 186, for CTP1L at valine 195 and for CS74L at valine 185. To investigate if proteolysis was dependent on flexibility within the cleavage site, these three cleavage site residues were mutated to Proline (CD27L M186P, CTP1L V195P and CS74L V185P). The mutation to proline would strengthen the main chain and alter the mechanics of the linker at the hinge close to the C-terminal domain. Indeed for CD27L M186P proteolysis was totally inhibited as observed by SDS-PAGE (Figure 20, B). For CTP1L V195P and CS74L V185P,

proteolytic cleavage was significantly inhibited, however, when samples of these mutants were concentrated above ~ 3 mg.ml$^{-1}$ a protein fragment corresponding to the size of the C-terminal domain became detectable by SDS-PAGE, indicating that mutation to proline reduces but does not totally inhibit proteolysis (Figure 19, C & D). The oligomeric states of CD27L M186P and CTP1L V195P were analysed by RALS to check if the reduction of proteolytic activity affected dimerisation. CD27L M186P produced the same RALS profile as the side-by-side mutant C238R and eluted as a single broad peak (Appendix Figure 5). The calculated MW of 66 ± 6.5 kDa, matched the theoretical mass for the CD27L dimer (64 kDa).

The SEC-RALS profile of CTP1L V195P produced two peaks indicating the presence of two oligomeric species (Figure 19 B). Unfortunately, the two peaks at 10.2 ml and 12 ml merged together, as previously seen for the CTP1L T221C mutant, making MW estimations challenging. The MW was estimated at the start and middle of the first peak and gave MW estimations of 70 ± 4.3 and 51 ± 3 kDa respectively. Although these calculations were prone to fluctuations and the calculated values could only be used as an estimate they pointed towards the existence of a dimeric species for the first peak. Furthermore, the retention volume of this peak was equivalent to the retention volume for the wildtype dimer (10.16 ml) in the same conditions. The MWs correlated well through the second peak at 12 ml and allowed for a precise MW estimation of 34 ± 1 kDa, which matched the theoretical mass of the monomer (32.8 kDa). The retention volume and calculated MW of the second peak also matched the monomer observed for the CTP1L D215A head-on mutant.

The substitution of proline at the respective cleavage sites of CTP1L, CS74L and CD27L affected the intrinsic cleavage mechanism differently. Both CTP1L V195P and CS74L V185P were susceptible to proteolysis of their C-terminal domains, however, at a significantly reduced rate compared to their respective wildtype proteins. Whilst for CD27L M186P, C-terminal domain cleavage was totally inhibited. This suggested that there was a degree of plasticity of proteolysis within the CTP1L and CS74L cleavage sites, to allow the cleavage mechanism to occur even with the restricted flexibility of the proline incorporated into the scissile bond.

*Purification and crystallisation of CTP1L V195P*

CTP1L V195P was used for crystallisation trials to see if the reduction in proteolysis would yield full-length protein crystals. The merged peaks observed during SEC-RALS of CTP1L V195P could be resolved by changing the buffer conditions from HEPES 20 mM, NaCl 500 mM to just 50 mM TRIS (both solutions were at pH 7.4) during SEC (Figure 19, A). The first peak at 10.88 ml retention volume, matched with the single peak of wildtype CTP1L when used in the same buffer conditions, which was presumed to be the dimer peak. Therefore the second peak was suggested to be the monomeric species. Interestingly, crystals only appeared when protein from the first peak was crystallised separately and not with protein collected from the second peak. Crystals appeared after two weeks in the same mother liquor conditions as crystallisation of the wildtype CTP1L. An X-ray data set for a single crystal was collected to 2.1 Å. Using the C-terminal domain of CD27L as a search model, the structure of the C-terminal domain was solved by molecular replacement. There was no N-terminal domain present and the asymmetric unit contained a single molecule that had been truncated to proline 195. Data collection and refinement statistics can be found in Appendix table 4.

*Inhibition of proteolytic cleavage does not affect CD27L lytic activity*

To verify whether proteolytic cleavage of the C-terminal domain affected the activity of the endolysins, turbidity reduction assays were performed in collaboration with Melinda Mayer (Institute of Food Research, Norwich). Ni-NTA purified endolysins and their mutants were exogenously applied to cultures of their respective host cells and the drop in $OD_{600}$ was measured to understand the effect mutations had on the lytic activity of the endolysins.

No differences in lytic efficiency could be deducted between wildtype CD27L and any of the mutants tested. M186P that prevented proteolysis at the cleavage site, C238R that prohibited side-by-side dimer formation and W207A, W207R and Y22A that prevented head-on dimer formation, all had no effect on the lytic activity of the endolysin (Figure 20). This established that for CD27L, these mutants were enzymatically active, whether they cleaved the C-terminal domain or not. For CD27L it did not resolve whether autocleavage played a role in endolysin function. It has previously been shown that CD27L demonstrates the same lytic activity when it is applied to *C. difficile* cells as the full-length intact protein or as the truncated EAD (Mayer *et al.* 2011). Overall the lytic activity of CD27L appears to be insensitive to the cleavage mechanism of the C-terminal domain of

CD27L when tested exogenously. The function of CD27L may necessitate the cleavage of the C-terminal domain to become active during the inherent lytic mechanism within *C. difficile* cells. However, current analysis, by the external application of concentrated CD27L to cells for the turbidity assays, may be unreliable to measure this effect.
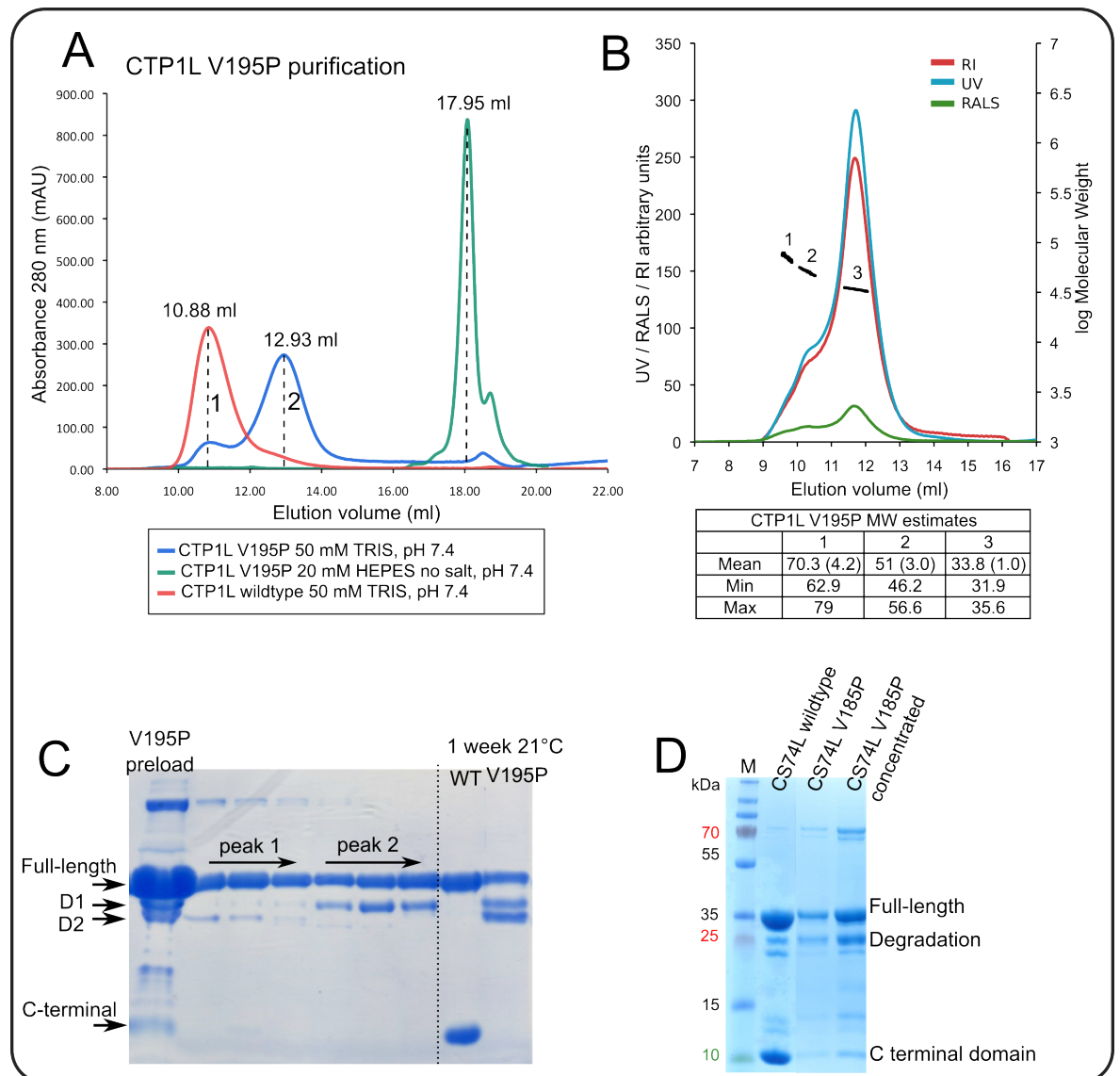


**Figure 19 Purification of V195P for crystallisation and in solution RALS measurements. A:** Comparison of SECs of Ni-NTA purified CTP1L V195P in 50 mM TRIS pH 7.4 (blue) or 20 mM HEPES pH 7.4 (green) and wildtype CTP1L in 50 mM TRIS pH 7.4 (red). For wildtype CTP1L and V195P, a shift in elution volume occurs from ~18 ml using HEPES buffer to lower retention volumes using TRIS buffer. V195P only crystallised using protein from elution peak 1 in 50 mM TRIS pH 7.4. **B:** UV/RI/RALS SEC traces and MW correlation estimates for CTP1L V195P. The protein eluted as two peaks merging together. The main peak gave a MW of 34 ± 1 kDa, close to the monomer MW. MW calculations of the smaller shoulder peak were influenced by the main peak, MWs calculated at two positions: 70 ± 4.2 kDa and 51 ± 3 kDa. **C:** SDS-PAGE analysis of V195P before SEC (preloaded) and samples collected across the two eluting peaks 1 and 2. Two degradation bands D1 and D2 were detected in different amounts between the two peaks. **D:** SDS-PAGE of CS74L wildtype (2mg.ml⁻¹) and CS74LV186P (1 mg.ml⁻¹ and 2 mg.ml⁻¹).
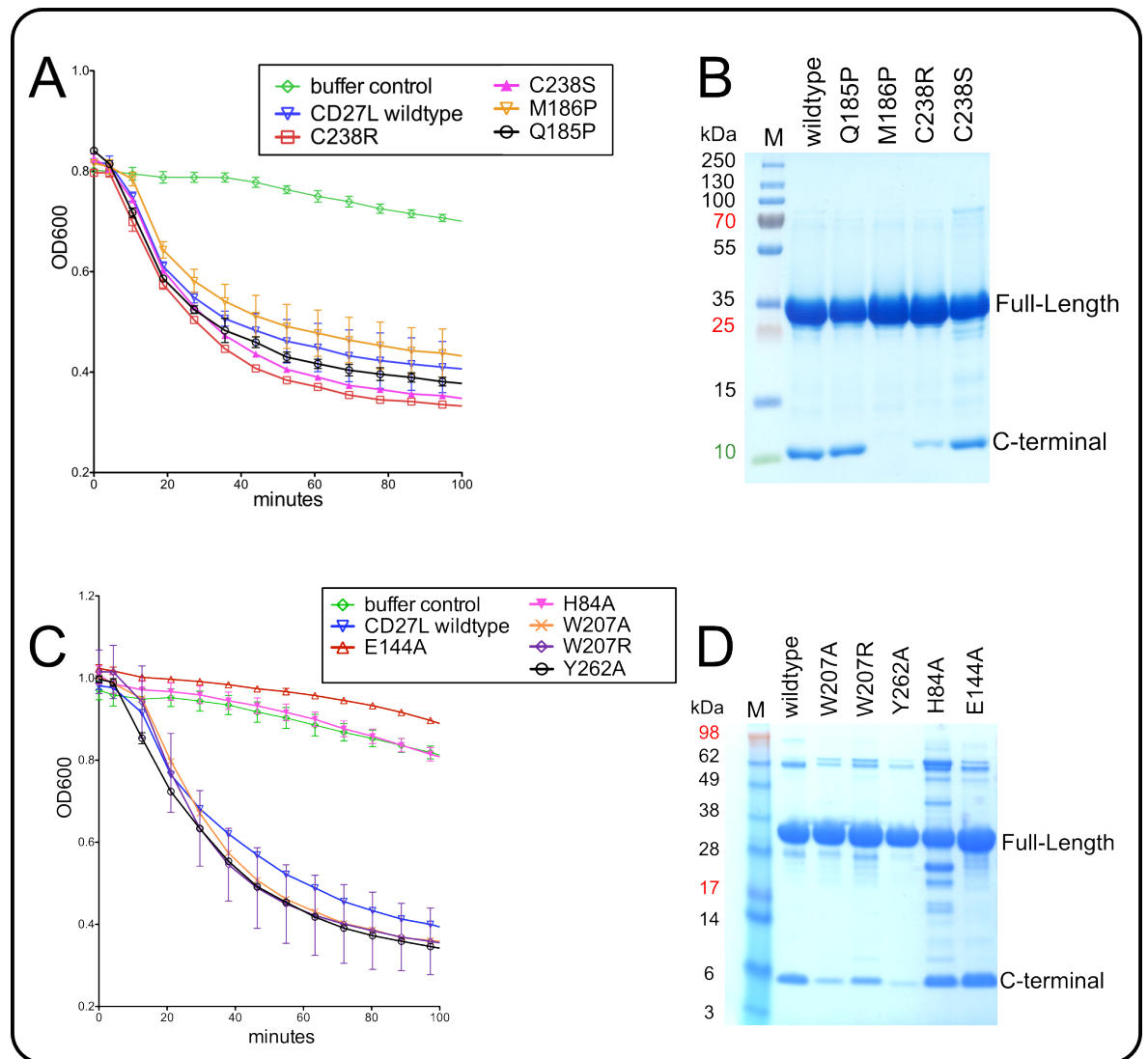
**Figure 20 Turbidity assays of CD27L and mutants on *C. difficile* cells. A & C:** Lysis assays with 10 µg of Ni-NTA purified CD27L and mutants on fresh *C. difficile* cells to determine if point mutations had an effect on the lytic activity of the CD27L endolysin. **B & D**: SDS-PAGE analysis of the mutants used for the mutant studies. 10 µl of 2mg.ml$^{-1}$ Ni-NTA purified protein was loaded to allow comparisons of the amount of autocleavage occurring. Mutations within the head-on dimerisation mode of CD27L W207A, W207R and Y262A and a single mutation C238R within the side-by-side dimer all showed a reduction in the proteolytic cleavage of the C-terminal domain as analysed by SDS-PAGE. However, they all had no effect on the lytic activity of CD27L, which remained at the same rate as the wildtype CD27L. Mutation M186P of the cleavage site totally inhibited proteolytic cleavage of the C-terminal domain but also had no effect on the lytic activity. *Figure 20 modified from the submitted article, Appendix page 170.*

## *Inhibition of proteolysis inhibited CTP1L lytic activity*

CTP1L was previously shown to only display activity as the full-length protein. Truncation of CTP1L to its individual glucosyl hydrolase EAD, resulted in a total inhibition of lytic activity against *C. tyrobutyricum* cells, indicating the requirement of the C-terminal domain during turbidity assay measurements (Mayer *et al.* 2010). Lysis by wildtype CTP1L showed a distinct drop in OD$_{600}$, reaching its maximum lytic effect within 15

minutes (Figure 21, A). Interestingly, the dimerisation mutants of CTP1L showed a drop in lytic efficiency that was proportional to the reduction in proteolysis. The side-by-side dimer mutants T221C and T221R, which totally lack the ability to cleave the C-terminal domain, produced no lytic activity. The cleavage site mutant V195P, which exhibited slight cleavage of the C-terminal domain, demonstrated a gradual lytic activity at a much lower rate when compared to wildtype CTP1L. Circular dichroism was used to show that lack of activity for these mutants was not due to unfolding. All circular dichroism spectra shared the same profile as wildtype CTP1L indicating that the secondary structure of these mutants was retained (Appendix Figure 7).
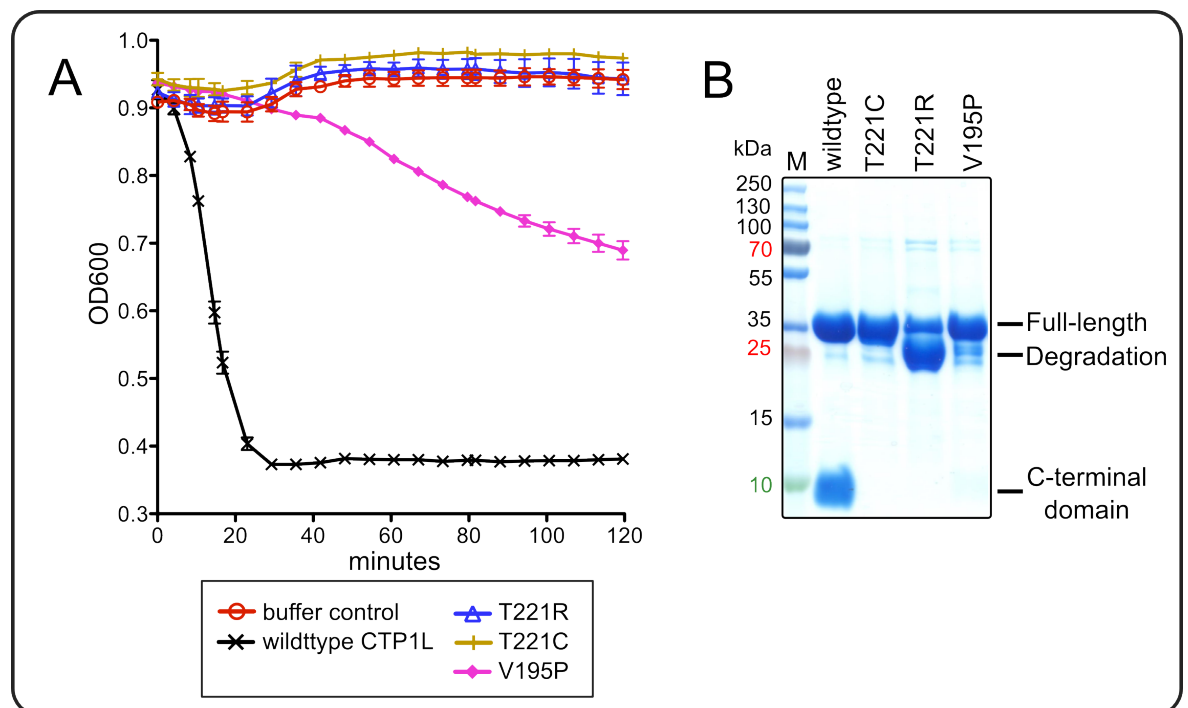


**Figure 21 Turbidity assays of CTP1L and mutants on *C. tyrobutyricum* cells**. **A**: Lysis assays with 10 µg of Ni-NTA purified CTP1L and mutants on pre-frozen *C. tyrobutyricum* cells to determine if point mutations had an effect on the lytic activity of the CTP1L endolysin. T221R and T221C totally inhibited the lytic activity of CTP1L and displayed the same profile as the buffer control. V195P dramatically reduced the lytic activity but did not inhibit lysis. Values are the means of results from duplicate assays ± standard deviations. **B:** SDS-PAGE analysis of mutants T221C, T2221R and V195P compared to wildtype CTP1L. All samples were loaded at 2 mg.ml$^{-1}$. *Figure 21 modified from the submitted article, Appendix page 170.*

For CTP1L, inhibition of the side-by-side dimer had the dual effect of inhibiting C-terminal proteolysis and inhibiting lytic activity (Figure 21, B). The lytic activity of the head-on dimer mutant CTP1L D215A still needs to be assessed, however, as apparent from the other tested CTP1L mutants with reduced C-terminal proteolysis, lytic activity would most likely also be inhibited.

*Mutagenesis to investigate linker residues of CTP1L involved in autocleavage*

Cleavage of the C-terminal domains appeared to be an intrinsic property of these three *Clostridial* endolysins. Inhibition of either of the two dimerisation modes caused a significant decrease in cleavage. CTP1L was reliant on autocleavage for activity, and the relative rate of lysis was dependent on the ability to cleave the C-terminal domain. Adversely, turbidity assays demonstrated that CD27L retained lytic activity whether dimerisation and the autocleavage mechanism was inhibited or not. Therefore, CTP1L was chosen as the model endolysin to investigate this cleavage mechanism. The asymmetric unit of CTP1L contains a full-length monomer and a cleaved C-terminal domain. When the missing EAD and linker are modelled onto the cleaved C-terminal domain the linkers and neighbouring residues of the two C-terminal domains begin to clash, opening up the possibility that these endolysins cleave themselves in an autoproteolytic process when the linkers come into close contact during side-by-side dimer formation.

Site-directed mutagenesis was used to test the involvement of various residues around the linker interface of CTP1L for their involvement in an intramolecular proteolytic mechanism. Mutants were assessed for their effect on proteolysis by SDS-PAGE analysis and turbidity assays to evaluate the effect certain mutants had when exogenously applied to *C. tyrobutyricum* cells.

*The N-terminal EAD of CTP1L and CD27L is not involved in autocleavage*

The catalytic domain of CTP1L belongs to the glucosyl hydrolase family 25 (Mayer *et al.* 2010). Two residues, an aspartate and a glutamate, are conserved in all proteins from this family and are vital for the catalytic activity of the enzyme (Fouche & Hash 1978). A single mutant of D92A and a double mutant of D92A_E94A of these catalytic residues in CTP1L were created to investigate if the autoproteolytic mechanism was dictated by the same active site as cell wall degradation. For D92A and D92A_E94A the C-terminal domain was detected in equal amounts to wildtype CTP1L by SDS-PAGE, indicating that the catalytic domain of CTP1L does not contribute to the autocleavage mechanism. However, as D92A and D92A_E94A totally inhibited the lytic activity of CTP1L against *C. tyrobutyricum* cells this provided further evidence that the active site belongs to the hydrolase 25 family (Figure 22, D). Introduction of mutants to two catalytic residues in the active site of the CD27L (H84A and E144A) also did not abolish proteolysis and also showed total inhibition of activity against *C. difficile* cells (Figure 20, C).

It is not surprising that the catalytic domains have no effect on autocleavage. The active site of the CTP1L EAD is positioned 36 Å from the cleavage site in the linker and also faces in the opposite direction to the linker. Intramolecular cleavage would require excessive conformational movement to bring the active site and the linker together. Intermolecular cleavage where the active site of one monomer could approach the linker of another monomer was ruled out as no such oligomerisation states could be observed in the crystal structures of full-length CTP1L or from SAXS analysis. Furthermore, the lytic activity of CTP1L is as a muramidase and CD27L as a peptidoglycan aminohydrolase (BLAST analysis also revealed that the CS74L EAD belongs to the same family as CD27L), so the likelihood that two different enzymatic types could perform the same autocleavage mechanism in this context is slim.

### *Mutagenesis of residues in close proximity to valine 195 in CTP1L*

The side-by-side dimer is the most likely conformation for an autoproteolytic mechanism to occur, as the two linkers are aligned parallel and come into close proximity providing the potential to interact. Protein self-processing reactions can be split into interdomain or intradomain autoproteolysis. The Ser/Lys cleavage dyad for LexA (Luo *et al.* 2001) and the Ser/Ser/Glu triad of BapA (Merz *et al.* 2012) are examples of intradomain proteolysis, contained within single domains that do not require oligomerisation for the donation of catalytic residues between domains. LexA and BapA are also examples of unconventional serine proteases, which do not involve the classical catalytic Ser/His/Asp triad. In fact, an array of catalytic triads, dyads or even single residues of serine or threonine can be used in the catalytic sites of proteases (Ekici *et al.* 2008). A combination of interdomain and intradomain auto-proteolytic cleavage has been described for the gp12 bacteriophage tail spike proteins whereby three Asp/Gln dyad protease sites are formed only when gp12 associates into a trimer. Trimerisation is essential for catalysis as separate monomers donate catalytic residues for interdomain proteolysis to occur (Xiang *et al.* 2009). For CD27L, CTP1L and CS74L dimerisation appears to be necessary for proteolysis, indicating a similar inter/intradomain mechanism reliant on the oligomeric state of the endolysins could exist.

For CTP1L there are many potential catalytic residues whose side chains could interact with the cleavage site V195, the majority of which were subsequently mutated: Y188A, N197A, C219A, D229A, S231A, N232A and K234A. The chemistry that the different side chains of these residues contribute in the context of proteolysis varies, as do the distances

of the residues to the cleavage site on the linker. In solution the endolysins exhibit a degree of flexibility, particularly as they form head-on or side-by-side dimers and have a flexible linker between the domains, allowing certain residues to form interactions with the cleavage site that are not visible in the rigid crystal structure. Surprisingly, all the mutants created around the cleavage site showed the same degree of C-terminal domain cleavage as wildtype CTP1L (Figure 22, B & C) and showed no difference in lytic activity as measured by turbidity assays, except for one mutant C219A. C219A showed a slight reduction in lytic activity (Figure 22, D). As C-terminal cleavage was not affected this could be due to instability of the mutant and still needs to be investigated.

*Autocleavage does not involve residues adjacent of the cleavage site*

Cis-autoproteolysis describes an autocleavage event initiated by a nucleophile attacking the carbonyl carbon of a neighbouring peptide bond (Buller *et al.* 2012). The nucleophilic attack causes a N→O acyl shift that converts the scissile peptide bond into an ester bond which is promptly hydrolysed, resulting in peptide bond autocleavage (Buller *et al.* 2012). The initial nucleophilic attack of the neighbouring residue is normally governed by a hydroxyl or thiol side chain (such as threonine, serine or cysteine) attacking the carbonyl carbon of the protein backbone. As a contrast, the cleavage site of CTP1L is flanked either side by glutamate residues, E194 and E196, whose side chains terminate with carboxyl groups. Single mutants of E194A and E196A and a double mutant E194A_E196A had no effect on autocleavage as the C-terminal band was present in equal quantities to the wildtype protein by SDS-PAGE. There was also no decrease in lytic activity for the single mutant E196A on *C. tyrobutyricum* cells and the lytic profile mirrored that of the wildtype CTP1L (Figure 22, C).
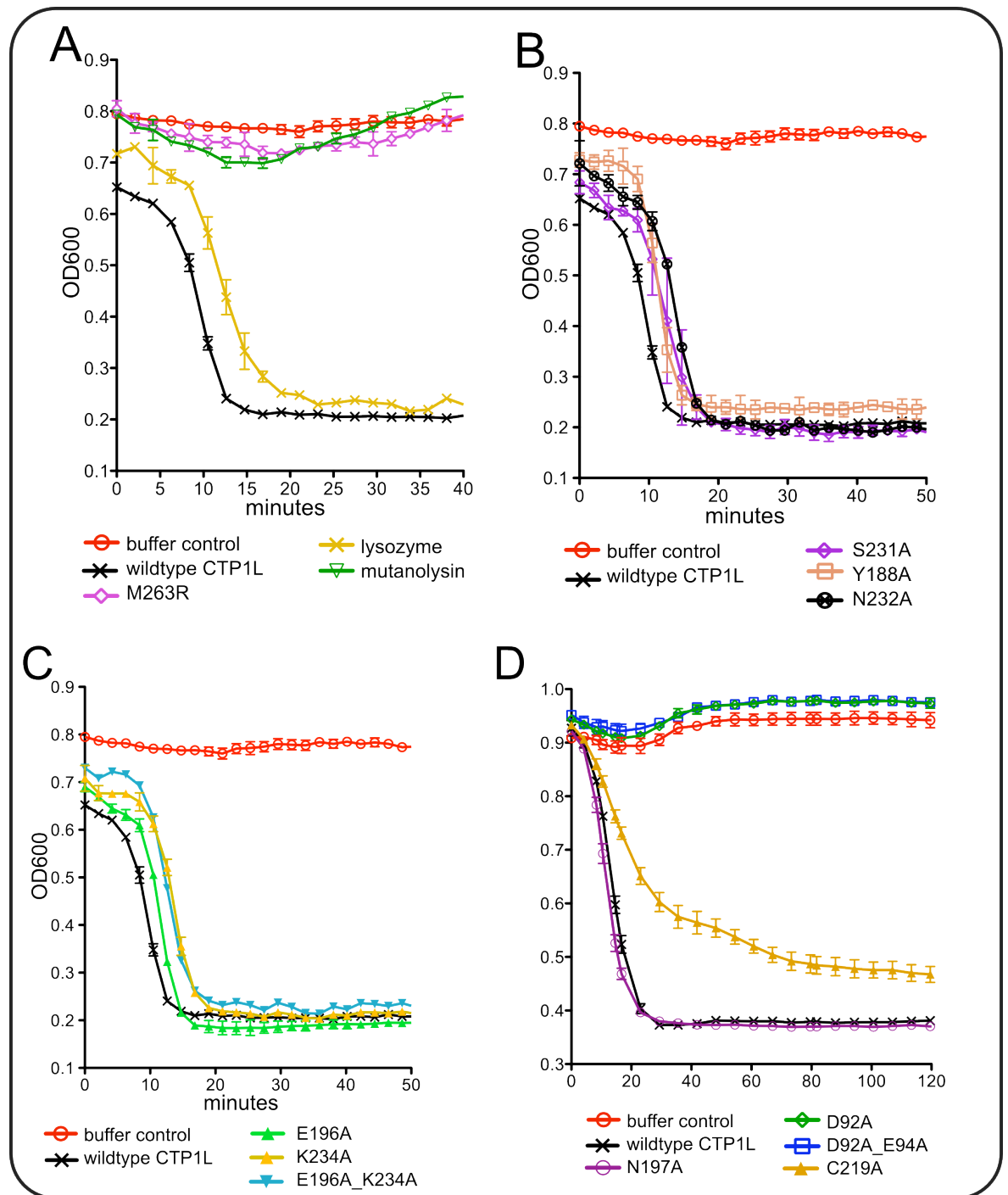
**Figure 22 Turbidity assays of CTP1L and mutants of residues in close proximity to the cleavage site**. **A**: Lysis assays with 10 µg of Ni-NTA purified CTP1L and mutants on pre-frozen *C. tyrobutyricum* cells to determine the lytic effect of mutants near the cleavage site. **A:** Mutation of M263R within the proposed cell wall binding site of the C-terminal domain inhibited lytic activity of the endolysin and displayed the same profile as the buffer control and mutanolysin (a lytic enzyme used as a negative control). Lysozyme was used as a positive control of lysis and gave a slightly reduced lytic activity when compared to wildtype CTP1L that presented full lysis within 15 minutes. **B, C & D:** All point mutations apart from M263R showed similar lytic rates to wildtype CTP1L except C219A, which has a reduced lytic activity and also a final higher OD600, indicating a lower number of total lysed cells. Values are the means of results from duplicate assays ± standard deviations.

*Proteolysis requires a linker but it does not have to be flexible*

By investigating the structures of other solved full-length endolysins it was noted that a common feature is a long flexible linker between the EAD and the C-terminal domain(s). PlyPSA has a 9 residue linker (PDB 1X0V) (Korndörfer *et al.* 2006), CPL-1 has a 12 residue linker (PDB 2IXU) (Pérez-Dorado *et al.* 2007) and Psm has a 7 residue linker (PDB 4KRT) (Tamai *et al.* 2014). Whilst these endolysins do not display a similar autocleavage mechanism for activation, the flexibility of the linker was tested to assess if it played a role in cleavage for CTP1L. The three residues upstream of the cleavage site V195 were mutated into prolines: E191P-D192P-E193P (CTP1L$^{PPP}$). As the linker between the EAD and C-terminal domain is relatively short at 6 residues, it was postulated that flexibility would be necessary to enable autocleavage activity. The addition of a polyproline linker adds structural rigidity. Surprisingly, CTP1L$^{PPP}$ had no inhibitory effect on autocleavage and the cleaved C-terminal domain was present after Ni-NTA purification and SDS-PAGE analysis (Figure 23, D preload lane). SEC of CTP1L$^{PPP}$ also gave a similar retention volume (9.93 ml) to the wildtype CTP1L (10.16 ml) with the same conditions and also had the same profile throughout. SDS-PAGE analysis of the elution samples across the peak showed the presence of full-length and a truncated C-terminal domain, the same as observed for the wildtype CTP1L (Figure 23, D).

To test if the length of the linker was necessary for proteolysis the same three residues that were mutated to proline in CTP1L$^{PPP}$ were deleted (ΔE191_ΔD192_ΔE193) to create CTP1L$^{ΔLinker}$, which reduced the length of the linker to just two residues, L190-G191, between the EAD and the V195 cleavage site. Interestingly, CTP1L$^{ΔLinker}$ totally inhibited proteolysis as detected by SDS-PAGE (Figure 23 C, Ni-NTA elutions). CTP1L$^{ΔLinker}$ was expressed in equivalent amounts to the wildtype CTP1L and was not prone to precipitation. Furthermore, the condensed linker did not affect the stability of the mutant as measured by circular dichroism (Appendix Figure 7). SEC-RALS was used to monitor if the inhibition of cleavage observed for CTP1L$^{ΔLinker}$ was connected to a change in oligomeric state. Surprisingly, CTP1L$^{ΔLinker}$ eluted at the same volume as the monomer species observed for the head-on mutant D215A. The calculated MW of 30 ± 2 kDa was also equivalent to the expected mass of the CTP1L monomer (32.4 kDa) (Figure 23, B).

Reducing the length of the linker to two residues was sufficient to inhibit the proteolytic cleavage of CTP1L, indicating an elongated linker is necessary for proteolysis. Combined with the fact proteolysis still occurred for the CTP1L$^{PPP}$, if the two linkers come together

in the side-by-side dimer they do so independently of the flexibility and chemistry of the linker.
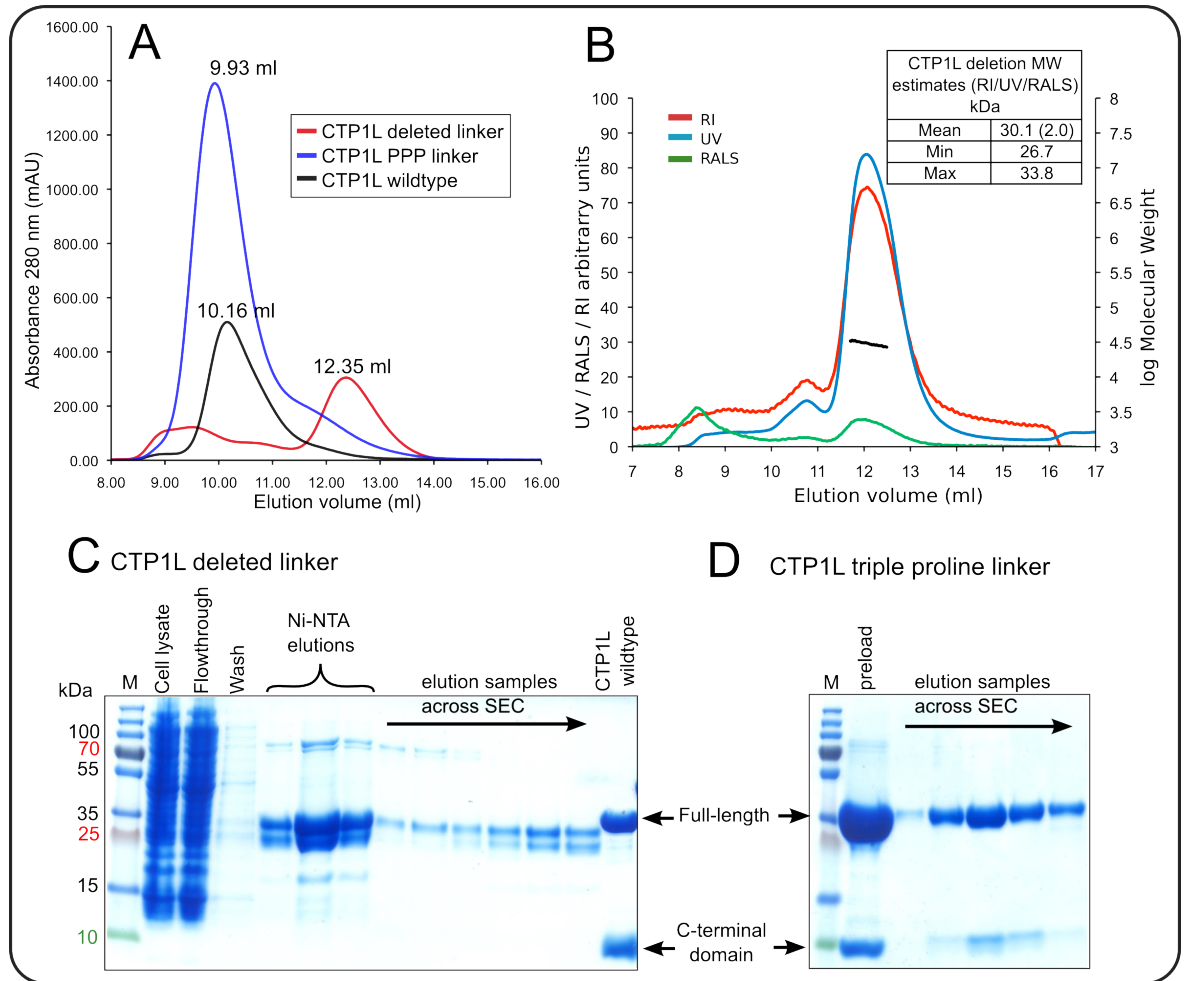


**Figure 23 CTP1L$^{\Delta Linker}$ and CTP1L$^{PPP}$ mutations within the CTP1L linker**. **A:** SEC profiles of CTP1L$^{\Delta Linker}$ and CTP1L$^{PPP}$ compared to wildtype CTP1L. **B:** UV/RI/RALS SEC traces and MW correlation estimates for CTP1L$^{\Delta Linker}$. The single main peak gave a calculated MW of 30 ± 2 kDa, indicating a monomer species in solution. **C:** SDS-PAGE analysis of the Ni-NTA purification of CTP1L$^{\Delta Linker}$ and samples taken from the SEC eluting peak at 12.35 ml. Compared to wildtype CTP1L (last lane) there is no detectable C-terminal domain. **D:** SDS-PAGE analysis of the preloaded SEC and elution profile of the CTP1L$^{PPP}$ peak at 10.16 ml, as for wildtype CTP1L the C-terminal domain is present throughout the peak.

# Summary of chapter two

In chapter two a novel autoproteolytic mechanism for the activation of the *Clostridial* endolysins CTP1L, CD27L and CS74L has been presented. Two distinct dimerisation modes were observed in the (previously solved) crystal structures of CTP1L and CD27L: the head-on and the side-by-side dimer states. These dimer states were validated in solution by the use of small angle X-ray scattering and right angle light scattering. For both endolysins a cleavage reaction was occurring on the linker between the EAD and the C-terminal domain during purification and crystallisation. Mutagenesis, to prevent either of the dimerisation states, had an inhibitory effect on the proteolytic mechanism, as a reduction in C-terminal domain release was detected by SDS-PAGE. For the CTP1L endolysin, the inhibition of proteolysis had an additional effect of inhibiting lytic activity against *C. tyrobutyricum* cells.

A model can therefore be proposed whereby the dimer switch and the proteolytic mechanism are part of an activation mechanism for these endolysins. The elongated head-on dimer exists as the dominant state before the moment of triggering, such as an environmental shock during holin lesion formation, which switches the endolysin into the side-by-side state. The side-by-side dimer state forces the linkers into close proximity to provide a potential configuration whereby an interdomain or intradomain autoproteolytic mechanism occurs. Unfortunately, current mutational analysis has not yet established common catalytic residues required for autoproteolysis.

Interestingly, a mass difference of approximately 40 Da was observed during MALDI-TOF mass spectrometric analysis on the cleaved C-terminal domains of CTP1L and CS74L. In the next chapter the mechanism for this mass difference is examined in detail and an attempt is made to derive its source. The hope is that identification of this mass shift could unravel details of the cleavage activation mechanism observed for these endolysins.

# Chapter 3: Results. Probing the autocleavage mechanism of endolysin activation

*Mass Spectrometry and tryptic digest analysis to probe the post-proteolytic modification*

The initial MALDI-TOF analysis of the cleaved C-terminal domains of CTP1L and CS74L indicated an approximate mass increase of 40 Da to the expected mass for a normal peptide hydrolysis of the C-terminal domains. The presence of a post-proteolytic modification to the C-terminal domains was therefore suggested. Below a mass to charge ratio (m/z) of 20000, the mass measurement accuracy from MALDI-TOF MS is quoted as ± 0.1% (Hillenkamp *et al.* 1991). The MW of the C-terminal domains is around 9000 Da and therefore the potential error would be ± 9 Da from the MALDI-TOF analysis. For higher resolution and higher mass-accuracy, liquid chromatography coupled to electrospray ionisation mass spectrometry (ESI-LC-MS) was used on freshly Ni-NTA purified samples of CTP1L, CS74L and CD27L to obtain the monoisotopic masses of the cleaved C-terminal domains .The MS analysis described in this chapter was performed by Stefan Leicht (EMBL, Heidelberg).

For CD27L, the calculated monoisotopic mass of 9579.9 Da for the cleaved C-terminal domain exactly matched the mass expected for a peptide hydrolysis between Q185 and M186, with the cleaved C-terminal domain retaining the methionine residue (Figure 24, A). Proteolytic cleavage at this bond is also observed in the crystal structure of the degraded full-length CD27L. For CTP1L the calculated monoisotopic mass was 9016.5 Da (Figure 24, B). For normal peptide hydrolysis between E194 and V195 at the cleavage site of CTP1L the mass would be 8984.5 Da. This indicated that post-proteolytic modification added 32 Da to the C-terminal domain. 32 Da is not a commonly recognised adduct to occur from ESI mass spectrometry. The same 32 Da mass increase was present for the C-terminal domain of CS74L and the experimental mass of 9061.5 was 32 Da higher than the theoretical mass of 9029.5 for proteolysis between K184 and V185 (Figure 24, C).

To further probe the 32 Da increase in mass for CTP1L and CS74L, the individual C-terminal domain bands were excised after SDS-PAGE and analysed by tryptic digestion. The trypsin-digested peptides were then analysed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Using the MASCOT server the MS/MS spectra were searched against a database containing over 3,500 proteins including the sequences of CTP1L and CS74L and directed to identify any common modifications to peptides such as acetylation, formylation and methionine oxidation. Numerous adducts were attempted to fit the mass shift of 32 Da. Ultimately, the 32 Da mass adduct was established to be from the covalent addition of a sulfhydryl group (R-SH) to the cleavage site at the N-terminus of the C-

terminal domain for both CTP1l and CS74L (Appendix Figure 6, B & D). The introduction of a sulfhydryl group adds 32 Da to the mass of the N-terminal valine (117.2 + 32 = 149.2 Da), generating the exact same mass as a methionine residue (149.2), hence the tryptic digest analysis indicated a modification of the terminal valine of the C-terminal domain to methionine or an isomer of methionine. As a control tryptic digestion on the SDS-PAGE bands corresponding to the full-length CTP1L and CS74L detected only valine at the cleavage sites (Appendix Figure 7, A & C), confirming that the cleavage site valines of the CTP1L and CS74L C-terminal domains are modified to a methionine-like residue only during proteolysis.
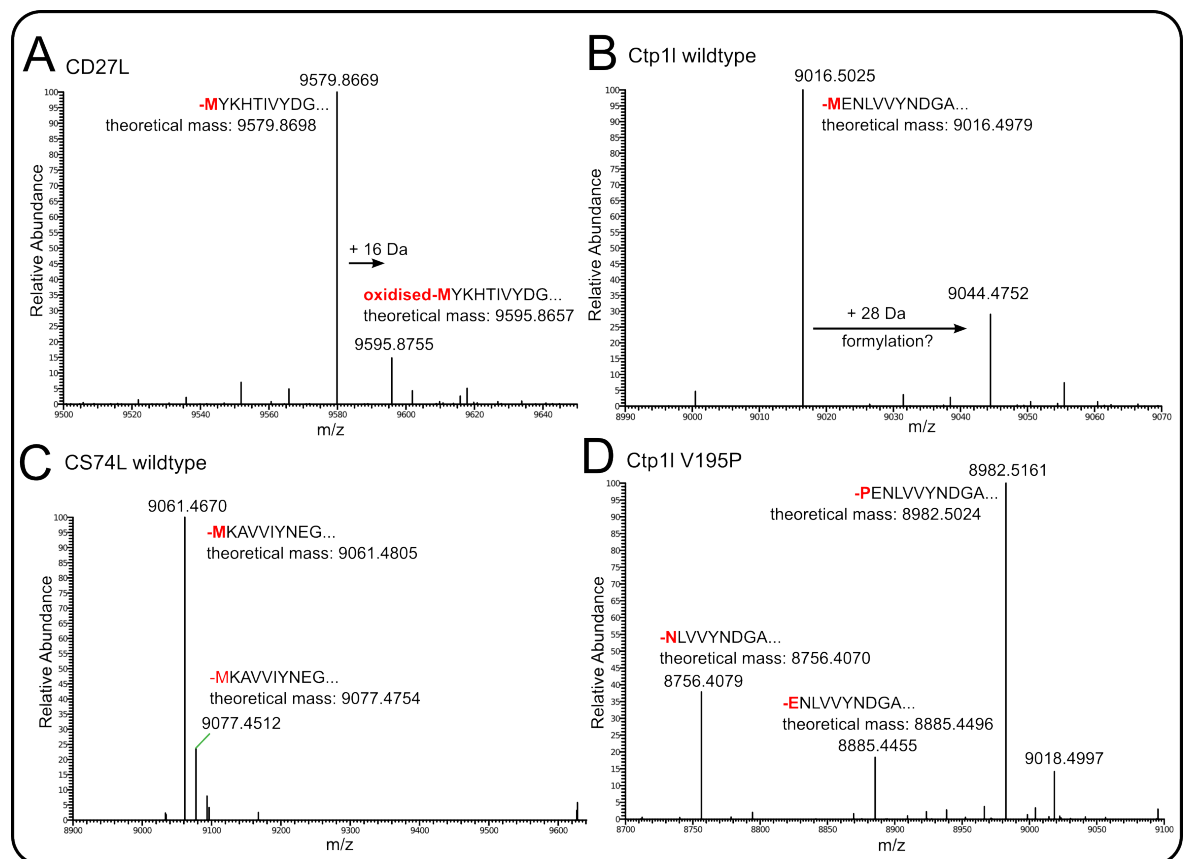


**Figure 24 Electrospray ionisation mass spectrometry on the cleaved C-terminal domains**. Liquid chromatography coupled to electrospray ionisation mass spectrometry (LC-MS) was used to accurately determine the full-length monoisotopic masses of the cleaved C-terminal domains for **A:** CD27L, **B:** CTP1L, **C:** CS74L and **D:** CTP1L V195P. For CD27L the mass corresponded to a peptide hydrolysis reaction between Q185-M186. For CTP1L and CS74L an additional mass of 32 Da was detected on their C-terminal domains, a methionine-like post-proteolytic modification for these proteins. CTP1L V195P was susceptible to multiple cleavage products.

*Proline incorporation at the cleavage site changes the kinetics of proteolysis*

Previously, it was shown that mutation of the cleavage site valine in CTP1l and CS74L to proline significantly reduced the rate of cleavage. CTP1L V195P was subsequently analysed by LC-MS to see if the same methionine-like modification could be detected with a proline residue at the cleavage site. Interestingly, a number of peptides were detected indicating cleavage was not just happening at one position (Figure 24, D). The most prominent species, at 8982.5 Da, corresponded to the cleavage product of peptide hydrolysis between E194 and P195. Monoisotopic masses of 8885.4 and 8756.4 Da were also detected and corresponded to the respective products of peptide hydrolysis after the cleavage site, between P195-E196 and E196-N195. The multiple products of peptide hydrolysis for CTP1L V195P were further confirmed by tryptic digest analysis (Figure 25, A). Additionally, the same methionine-like modification was detected at the N-terminus instead of proline and indicated that the same autocleavage mechanism observed for the wildtype CTP1L had also occurred. Tryptic digest analysis was also used to search for potential post-proteolytic modifications of CS74L V185P. Two proteolytic products were detected: the cleavage product of peptide hydrolysis between K184-P185 and the N-terminal methionine-like incorporation observed for wildtype CS74L (Figure 25, B).

For both CTP1L V195P and CS74L V185P, proline was still being modified to the methionine-like residue during cleavage, suggesting a similar mechanism of autocleavage to the wildtype endolysins was still possible. However, the observation of post-proteolytic products of peptide hydrolysis, without the methionine-like incorporation, suggested that incorporation of proline at the cleavage sites for these endolysins altered the mechanics of the autocleavage mechanism. Proline imposes restrictive constraints on the cleavage site, which could alter the kinetics of cleavage to allow normal peptide hydrolysis of the scissile bond between proline and its upstream residue. Furthermore, for CTP1L the peptide bonds one and two residues C-terminal of P195 were also susceptible to cleavage.

**Figure 25 MASCOT search results for tryptic digestion and LC-MS/MS analysis on CTP1L V195P and CS74L V185P. A: CTP1L V195P, B: CS74L V185P.** For both mutations methionine-like incorporation could be detected as the post-proteolytic modification on the N-terminus, indicating a 32 Da mass increase. Both searches also detected peptide hydrolysis of the cleavage site scissile bond as proline was retained as the N-terminal residue. For CTP1L the additional cleavage product of hydrolysis between E196-N195 was detected.

### Could hydrogen sulphide be the trigger for autocleavage?

Hydrogen sulphide ($H_2S$) was suggested as the causative agent of proteolysis and the additional mass of 32 Da to produce the methionine-like modification after autocleavage. $H_2S$ is present in high amounts within bacterial cells, particularly in anaerobic bacteria such as the *Clostridial* hosts for these phage endolysins (Fuchs & Bonde 1957). It was also recently shown that *C. difficile* has special hydrogen sulphide channels that actively remove $H_2S$ from the inside of the cell and pump it into the periplasm (Czyzewski & Wang 2012). In bacteria the majority of $H_2S$ biosynthesis is derived from the breakdown of L-cysteine or homocysteine by two enzymes: cystathionine β-synthase and cystathionine γ-lyase (Paul & Snyder 2012). $H_2S$ is also generated in small amounts within *E. coli* (Artman 1956), the cells used for endolysin expression throughout my studies. $H_2S$ has established roles in enzyme regulation, mostly by sulfydration of cysteine residues (Krishnan *et al.* 2011). It was hypothesised that during proteolysis the nucleophilic nature of the sulphur of $H_2S$ could interact with the amide of the cleavage site valine to generate a N-S bond, which would explain the addition mass of 32 Da.

Sodium hydrosulphide (NaHS) was used as a $H_2S$ donor, which, in solution dissociates to produce $H_2S$. NaHS was added to solutions containing the three endolysins in an array of environmental conditions (pH 6-9, NaCl 0 to 500 mM and at 21 or 37 ℃) to see if $H_2S$ could prompt a proteolytic reaction. SDS-PAGE was used to detect an increase in proteolysis by a decrease in full-length protein and an increase in C-terminal domain concentration. Furthermore, the released EAD domain should have become visible after cleavage. Unfortunately, it was promptly observed that the addition of $H_2S$ (10-100 mM) had little effect on the endolysins besides aggregating the samples in pH values above pH 8. The dissociation of NaHS to $H_2S$ and $HS^-$ ions increases with pH and therefore it was deduced that the aggregation of the three endolysins is from increased nucleophilic attack on the proteins by the increase $HS^-$ ionic content. Additionally, it was established that monitoring the initiation of autocleavage by detecting an increase in cleaved C-terminal domain by SDS-PAGE did not provide an accurate means for measurement, unless cleavage was close to 100%. Therefore, an experimental setup was developed for the quantification of cleavage upon the change of environmental conditions or from the addition of potential activators of autocleavage.

### *Testing for free sulfhydryl groups with Ellman's reagent*

Ellman's reagent (5,5'-dithiobis-(2-nitrobenzoic acid)), also known as DTNB, is a water-soluble compound for quantifying free sulfhydryl groups on proteins (Ellman 1959) . In solution the disulphide bond of DTNB breaks and $TNB^{2-}$ covalently attaches to free sulfhydryl groups, the other free $TNB^{2-}$ dianion exists as a yellow-coloured product with an absorbance at 412 nm that is used to quantify sulfhydryl incorporation using a spectrophotometer (Appendix Figure 8). Quantification of sulfhydryl groups was attempted for wildtype CTP1L, CS74L and the mutant CTP1L C219A as a control (the removal of the thiol group of C219A from the cleaved C-terminal domain would provide a quantifiable difference in absorbance readings). However, due to restricted access to the N-terminal domain or proteolytic heterogeneity between samples, calculations were variable and unreliable.

DTNB binding to sulfhydryl groups increases the mass of a protein by 198 Da, which can be measured by MS. DTNB was added to CTP1L and CS74L that had been denatured by 6 M urea. Denaturing the proteins ensured all potential thiols groups would be available to react with DTNB in solution. Samples were subjected to in-solution tryptic digestion and

full-length LC-MS/MS analysis. The successful incorporation of Ellman's reagent was detected for all C219 residues in wildtype CTP1L indicating the protocol worked, but no sulfhydryl-TNB mass shift could be detected at the N-terminus of the cleaved domain. Instead the 32 Da additional mass of the methionine-like residue was the only modification detected (Appendix Figure 9). As the N-terminal residue was not prone to DTNB modification this ruled out that the cleavage mechanism involved the incorporation of a terminal sulfhydryl group to the cleaved C-terminal domain by the attack of $H_2S$. However, the 32 Da additional mass could involve a different type of modification involving the addition of sulphur.

### *Measuring the effect of environmental conditions on endolysin activity*

In order to investigate whether a changing environment could affect the population of dimeric states and whether this could have an effect on autocleavage, an experiment was designed to quantify the release of the C-terminal domain for CTP1L. A construct of CTP1L was designed that incorporated a green fluorescent protein (GFP) tag onto the C-terminus of the protein (CTP1L-GFP). SDS-PAGE analysis demonstrated that the C-terminal domain-GFP product was still cleaved, indicating the autocleavage mechanism was still functional even with the GFP-tag attached (Figure 26, B, lane 2). CTP1L-GFP was then bound to Ni-NTA resin in a closed system with a constant flow of buffer over the resin (Figure 26, A). After cleavage, the C-terminal domain would be released into the constantly flowing solution. Samples were taken at certain time points of the eluting solution and the fluorescence signal of GFP measured using a spectrophotometer to quantify the amount of cleaved C-terminal domain released over time. After testing numerous conditions with different pH ranges, salt concentrations and additives it was discovered that pH 9 had an effect. Interestingly, the lytic activity of CTP1L against *C. tyrobutyricum* cells was previously shown to increase as the pH increased (Mayer *et al.* 2010).
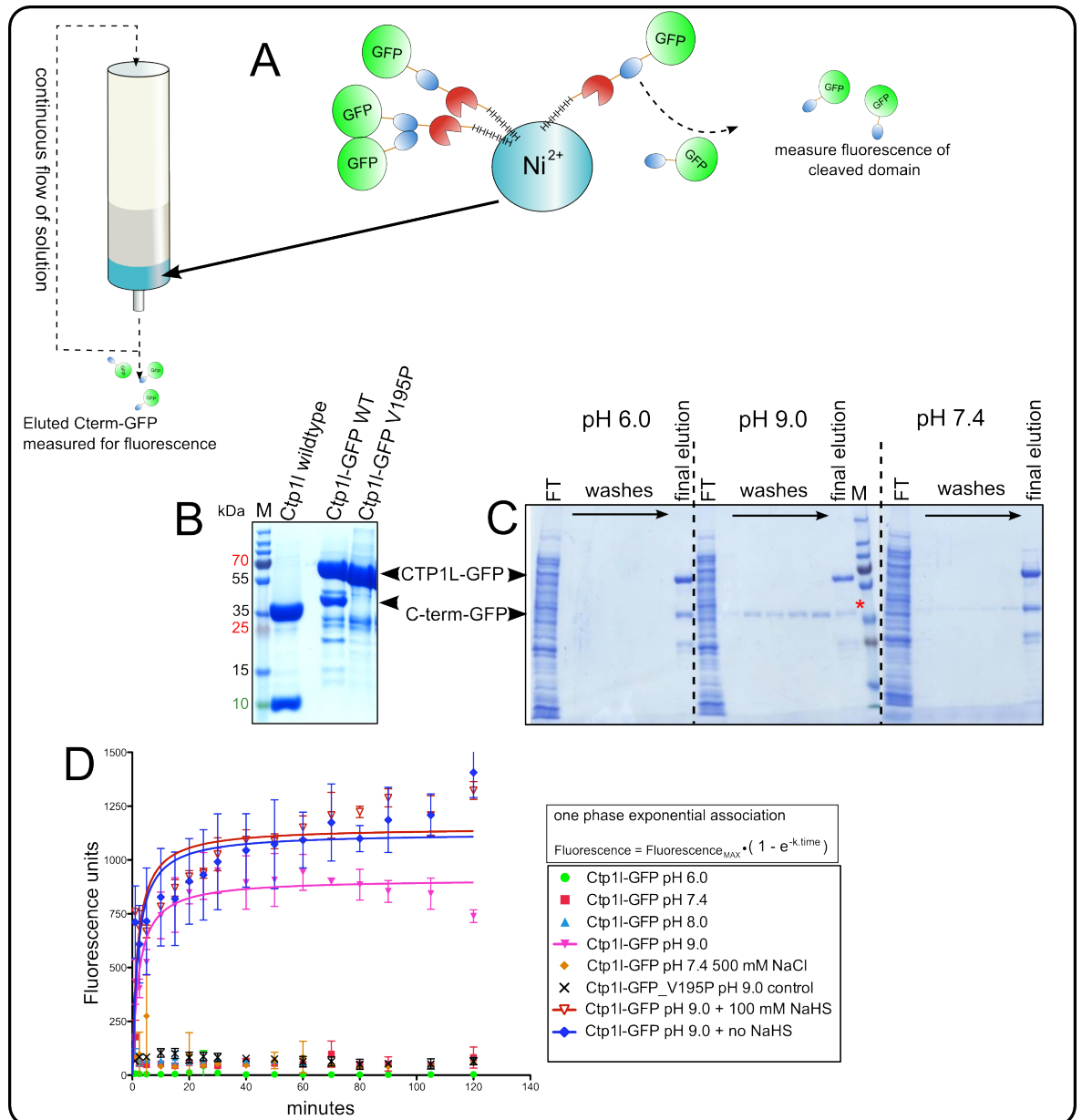
**Figure 26 Experimental setup to measure the release of the C-terminal domain by GFP fluorescence A:** The closed system used for measuring the effect of solution conditions on the rate of release of the C-terminal domain. The rate of release can be measured using the GFP tag attached to the C-terminal domain. **B:** SDS-PAGE analysis showing CTP1L-GFP can still cleave the C-terminal domain even with a GFP-tag attached. CTP1L-V195P-GFP inhibited the release of the C-terminal domain. **C:** SDS-PAGE detecting the release of the C-terminal domain-GFP product from CTP1L-GFP bound to Ni-NTA resin only occurs at pH 9.0 (red star). Ni-NTA beads were loaded with 2 mg of purified CTP1L-GFP and washed with the same volumes of wash buffer at different pHs. The bound protein was finally eluted from the resin for analysis (last lanes). **D:** Fluorescence vs. time is plotted for a range of different conditions tested to see if they released the C-terminal domain. The best fit to the data was by applying *one phase exponential association* kinetics (in legend). Although normally used to describe the association of a ligand and receptor whereby at each time interval more of the unoccupied binding sites of receptors become occupied until all receptors are bound. For this experimental process it was used to measure the release of the C-terminal domain from the attached state.

Within the first 20 minutes that the system was exposed to a pH 9 solution there was an increase in fluorescence, indicating the release of cleaved C-terminal domains (Figure 26, D). As the fluorescence becomes a flat line, all C-terminal domains with the potential to be released have done so within 40 minutes. Other conditions within a pH range 6 – 8 had no effect and fluorescence measurements were negligible. At the time of experimentation $H_2S$ was stipulated to be the source of the 32 Da mass shift, however, the addition of $H_2S$ did not increase fluorescence when added to a pH 9 solution. As a control, the cleavage site mutation V195P was introduced into CTP1L-GFP and SDS-PAGE indicated that the cleavage mechanism was inhibited (Figure 26, B lane 3). Furthermore, this construct also showed no fluorescence in the same pH 9 buffer conditions.

To test that the fluorescence detected was from the release of the C-terminal domain, samples were taken from a similar system as described in Figure 26, A except the solution was not recirculated over the beads but collected separately and analysed by SDS-PAGE (Figure 26, C). When CTP1L-GFP was washed at pH 6, no C-terminal domain-GFP product could be detected. An increase to pH 7.4 presented a small amount of cleaved product being released. Considering the fluorescence readings at pH 7.4 and 8 were always slightly above zero (Figure 26, D), the small amount of cleaved C-terminal domain-GFP could be due to the slow release of the C-terminal domain, at a reduced rate compared to the pH 9 solution. At pH 9 there is a larger release of the C-terminal domain-GFP product, which was continuously removed with multiple wash steps. The full-length protein does not elute at pH 9 and it can therefore be concluded that the fluorescence came solely from the cleaved C-terminal domain-GFP during the continuous flow experiment.

It was initially proposed that the increase to pH 9 could be causing a dimeric switch of the bound endolysins, initiating the autocleavage mechanism. However, the cleaved C-terminal domain that was detected was suggested to not be from the inititation of autocleavage, but from the dissociation of a C-terminal domain from its side-by-side interaction with a full-length monomer bound to the resin. The monomer plus a cleaved C-terminal domain we observed in the CTP1L asymmetric unit (Figure 8, A), was proposed as the most likely conformation that was initially attached to the column. SDS-PAGE indicated that even with successive pH 9 washes the full-length protein was still retained on the beads and could be eluted at the end of the wash steps (Figure 26, C). Interestingly, the final concentration of the C-terminal-GFP product was lower in the final elution step after the pH 9 washes, compared to the pH 6 and 7.4 washes (Figure 26, C, red star).

Indicating that at pH 9, the interaction between the cleaved C-terminal domain and the full-length monomer was destabilised, causing the C-terminal domain to dissociate.

Indirectly, this suggested that decreasing the pH could stabilise the side-by-side dimer, as at pH 6 no cleaved C-terminal domain could be detected by SDS-PAGE. Furthermore, no GFP fluorescence could be detected at pH 6 using the column flow experiment (Figure 26, D, green dots) and no cleaved C-terminal domain could be detected by SDS-PAGE from the successive pH 6 wash steps, suggesting that at pH 6 the interdomain interactions of the side-by-side dimer are the most stable.

***Mutation of the cleavage site valine to leucine was used to test for different sulphur modifications***

Both cleavage site valines of CTP1L and CS74L were mutated to leucine. Leucine is a branched chain amino acid, similar to valine except for an additional methyl group on the side chain. Therefore if leucine gained an additional mass of 32 Da on the amide group during proteolysis, the overall mass of the N-terminal leucine residue would increase to 163.2 Da, an identifiable difference using MS when compared to the methionine-like modification. It was previously showed that mutation of valine to proline at the cleavage site significantly reduces proteolysis, however, possibly due to the chemical similarity between valine and leucine, CTP1L V195L and CS74L V185L had the same proteolytic activity as wildtype proteins (Figure 29). The cleaved C-terminal domains were excised after SDS-PAGE and analysed by tryptic digestion and LC-MS/MS. Remarkably, the N-terminal residue was again recognised to have the mass of the methionine-like modification. The mass of leucine at the N-terminus and the 163.2 Da mass of the N-terminal residue expected if $H_2S$ was affixed to leucine could not be detected (Appendix Figure 11). The theory of $H_2S$ or a similar external mediator of proteolysis attaching a 32 Da mass to valine was dismissed, instead the above evidence suggested that valine is essentially switched with methionine during autocleavage. Instead of a sulfhydryl modification to produce an isomer of methionine, methionine itself was stipulated to be the modification, as a whole residue.

*Anomalous X-ray scattering detected methionine as the post-proteolytic modification in the CTP1L crystal*

Based on an X-ray data set collected at 6.5 keV on a crystal of the partially degraded CTP1L endolysin, a sulphur atom was verified at the N-terminus of the cleaved C-terminal domain within the asymmetric unit. At 6.5 keV it was possible to use the anomalous sulphur adsorption to pinpoint the locations of sulphur residues within the crystal structure of CTP1L. Using the phases from the refined crystal structure, a significant 5σ peak was identified in the phased anomalous difference map 3 Å away from the main chain nitrogen atom of V195 (Figure 27, A). The anomalous signals from the sulfhydryl groups of the two C219 residues are also identified. The modification of V195 for methionine also fitted well with the electron density in this region.
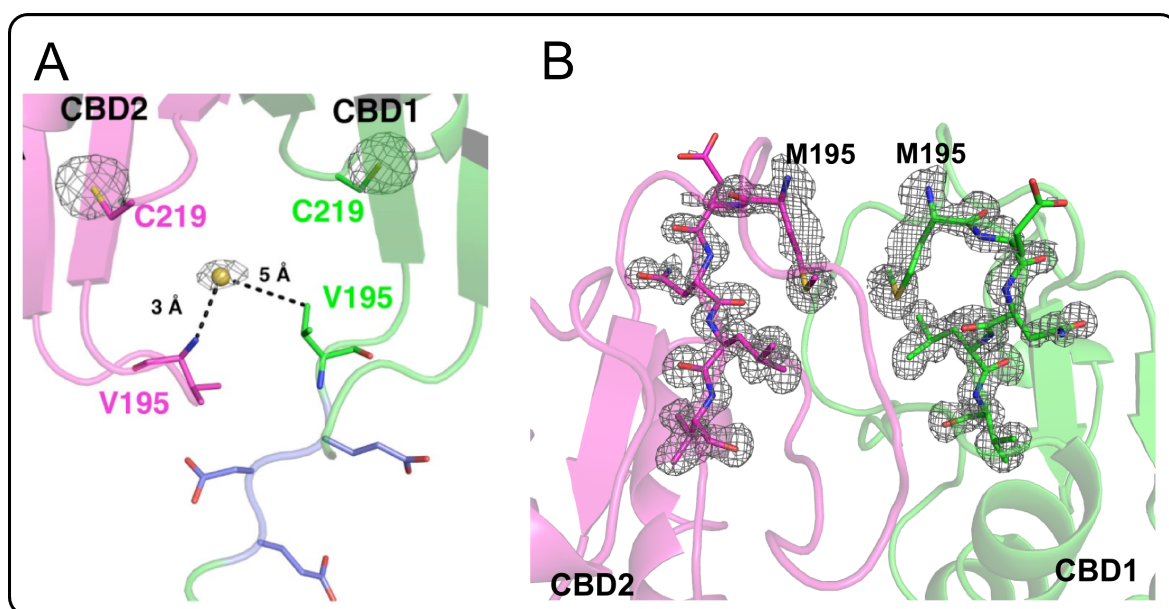


**Figure 27 Identification of methionine in the crystal structures of the partially and fully cleaved CTP1L**. **A:** The dimer interface in the partially cleaved structure of CTP1L between the two C-terminal domains shows the N-terminal V195 residue, the linker in blue and C219 in sticks. The peaks of the phased anomalous difference density map are shown as a mesh at a contour level of 4σ. A sulphur atom has been placed into the peak observed at the N-terminus and the distance between the nitrogen atom of V195 of the cleaved C-terminal domain (magenta) and the sulphur is shown, as well as the distance between the CG1 atom on the side chain of Val 195 of the full-length uncleaved CTP1L molecule (in green). **B:** Cartoon and stick diagram from the fully degraded 1.2 Ångstrom CTP1L crystal and Omit generated electron density map shown around the five N-terminal residues (V195-V199) contoured at 1.5σ, which shows the incorporation of methionine at both C-terminal domain cleavage sites.

The incorporation of methionine at the cleaved N-terminus was further confirmed by the 1.2 Å resolution crystal structure of the fully degraded CTP1L C-terminal domains, where the two methionine residues also come into close proximity (Figure 27, B). The

combination of structural information and mass spectrometric analysis supported the post-proteolytic incorporation of methionine at the cleavage sites of the C-terminal domains of CTP1L and CS74L. The incorporation of methionine as a fundamental step of a proteolytic mechanism has not been previously documented and could represent a novel autoproteolytic mechanism.

### *Selenomethionine could be incorporated as the post-proteolytic modification*

To test if methionine could truly be incorporated as a whole amino acid into the N-terminus of the C-terminal domain during proteolysis, CTP1L and CS74L were expressed in a reduced media containing the methionine analogue seleno-L-methionine. Using tryptic digestion combined with LC-MS/MS the mass difference of 47 Da between selenium and sulphur could be detected. Selenomethionine was detected at the N-terminus of the cleaved C-terminal domains for both CTP1L and CS74L (Figure 28). Replacement of seleno-methionine for methionine was not 100 % as we also detected ordinary methionine incorporation at the cleaved N-terminus. The expression protocol involved the use of a non-methionine synthetase auxotropic strain of *E. coli* (BL21-DE3) meaning residual amounts of methionine could still be generated within the cells. Furthermore, the initial preculture grown in lysogeny broth media would also contaminate the growth media with residual methionine. Nevertheless, selenomethionine replaced valine at the cleavage site of both CTP1L and CS74L C-terminal domains as the post-proteolytic modification.



**Figure 28 Incorporation of selenomethionine as the post-proteolytic modification of the CTP1L and CS74L C-terminal domains.** Sample MASCOT search results after tryptic digestion and LC-MS/MS for **A:** CTP1L and **B:** CS74L grown in methionine-reduced media supplemented with selenomethionine. Selenomethionine was detected as the N-terminal residue of the cleaved C-terminal domains for CTP1L and CS74L. Methionine was also detected as the N-terminal residue.

*Non-branch chained amino acids within the autocleavage sites of CTP1L and CS74L inhibit autocleavage*

It has previously been shown that mutation of the cleavage sites in CTP1L and CS74L to leucine had no effect on cleavage and methionine was the only post-proteolytic modification detected on their C-terminal domains (page 81). Furthermore, mutation to proline (CTP1L V195P, CS74L V185P) significantly reduced cleavage and produced multiple post-proteolytic products. Taken together, both mutations still indicated that the incorporation of methionine was possible after autocleavage, regardless of side-chain chemistry (Figure 25). In order to test if other amino acids, with a broader range of side-chain chemistries, would also be susceptible to cleavage and methionine incorporation, the cleavage site valines of CTP1L and CS74L were mutated to tyrosine, methionine, glycine or isoleucine. Methionine was incorporated to mimic the methionine at the cleavage site of CD27L (M186) to see if it could instigate an increase in autocleavage.

For CTP1L and CS74L, isoleucine and methionine incorporation had no effect and the amount of cleaved C-terminal domain was equal to valine or leucine incorporation at the cleavage site (Figure 29). Glycine incorporation showed a reduction in proteolysis for both CTP1L and CS74L, with the decrease in cleavage more apparent for CTP1L. Glycine has a side chain of a single hydrogen atom, although the reason why it showed a decrease is unknown. Small nonpolar residues may just not be favourable for cleavage and the presence of an alkyl group attached to the $C\alpha$ carbon may be required for efficient cleavage. Tyrosine incorporation totally inhibited cleavage in CTP1L as confirmed by SDS-PAGE and MS analysis, whereas in CS74L a decrease in autocleavage was detected, to the same degree as glycine incorporation. The inhibition of proteolysis observed when tyrosine was incorporated into CTP1L was most likely due to steric hindrance from the aromatic side chain of tyrosine as the two linkers come together during side-by-side dimer formation. Interestingly, for CS74L it just slightly decreased cleavage, suggesting the linker and the cleavage site in CS74L are more flexible, allowing larger side chains to interact.

The incorporation of glycine, isoleucine (or tyrosine for CS74L only) still produced the same N-terminal post-proteolytic modification of methionine incorporation as identified using tryptic digestion linked with LC-MS/MS (Appendix Figure 11). Even with the diverse array of side chains, the autocleavage mechanism observed for the wildtype endolysins was maintained. Interestingly, for these three mutations their respective peptide

hydrolysis products, from the cleavage of the scissile bond between the mutated cleavage site residue and the neighbouring residue, were also detected (Appendix Figure 11). As an example, for CTP1L V195G and CS74L V185G both terminal residues of glycine or methionine were detected by LC-MS/MS analysis on the cleaved C-terminal domain. The dynamics of cleavage must be distorted when these mutations are introduced, to allow two different cleavage mechanisms to give two different final products: peptide hydrolysis or methionine incorporation. Using the current tryptic digest and LC-MS/MS analysis it was not possible to quantify the ratio between methionine incorporation or peptide hydrolysis to infer which mutations provide a preference for the final modification. Interestingly, the cleavage site mutants that reduced proteolysis the most were the mutants that gave different cleavage products, indicating that methionine incorporation must be the favoured reaction, but a shift can occur that initiates peptide hydrolysis when the cleavage mechanism is hindered. This is best demonstrated for CTP1L V195P where multiple products of peptide hydrolysis around the cleave site were detected (Figure 25). Sequence analysis revealed that the cleavage sites for CTP1L, CS74L and the 19 analogous proteins are conserved for branched chain amino acids, with a preference for valine at this position. The other residues conserved at the cleavage site are leucine or methionine; the only residues introduced during this study that only detected methionine incorporation as the post-proteolytic modification and not peptide hydrolysis.

A final observation for the cleavage site mutants was the degradation product present after Ni-NTA purification of CS74L V185Y, V185G, V185P and also CTP1L V195P (Figure 19, C & D and Figure 29). The size of this degradation product was similar to the degradation product detected after Ni-NTA purification for mutants that interfered with dimerisation (CS74L T211R and CTP1L T221R, T221C) (Figure 15, E). Interestingly, the head-on dimer mutants (CTP1L D215A and CD27L W207A, W207R and Y262A) do not produce this degradation product. It is therefore assumed that these bands are due to degradation (similar to CS74L T211R) resulting from restrictions imposed on the side-by-side dimer or linker restrictions during autocleavage.
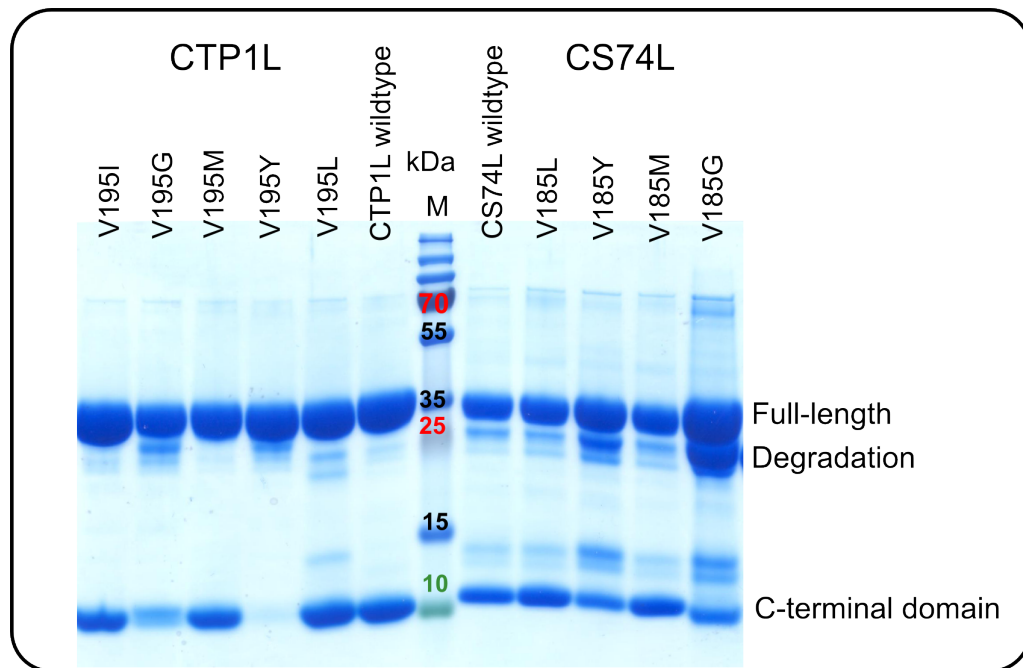
**Figure 29 SDS-PAGE of point mutations of the cleavage site valine of CTP1L and CS74L** All samples were loaded fresh after Ni-NTA at a concentration of 1.5 mg.ml$^{-1}$. The branch chain amino acids leucine and isoleucine, as well as methionine, showed a similar amount of cleavage as native valine at the cleavage sites of CTP1L and CS74L. Incorporation of glycine and tyrosine significantly affected proteolysis in both endolysins. For CS74L V185G and V185Y a degradation band just below the full-length protein was detected, a similar observation to the mutant CS74L T211R.

### Autocleavage is a continuous process throughout E. coli overexpression

It was not clear if proteolysis was happening at a particular stage of purification, such as during cellular lysis, or if it was a continuous event occurring throughout expression and purification. To monitor the proteolytic processing, a construct of CTP1L was designed which incorporated a C-terminal polyhistidine tag by cloning the construct into the pET21d vector. The sequential steps from *E. coli* induction to Ni-NTA elutions were monitored to find the moment at which autoproteolysis was initiated, by using immunoblotting to detect the polyhistidine tag. Immunoblotting provided higher sensitivity than SDS-PAGE and could differentiate between the C-terminal domain and other proteins of a similar MW. Interestingly, cleavage of the C-terminal domain could be detected within the cells two hours after induction (Figure 30, 2$^{nd}$ lane). In chapter two, the dynamic switch between dimer states was established as the trigger for autoproteolysis. Combined with the fact autoproteolysis occurred throughout *E. coli* expression, the endolysins must be able to undergo dynamic reconfiguration, between the two dimer states, within the cytosol. Furthermore, as the autocleavage mechanism occurred within the *E. coli* cell, an activator molecule or enzyme must be present in the cytosol for the incorporation of methionine as part of the cleavage mechanism. Interestingly, overexpression in *E. coli* was

shown to be responsible for the premature cleavage of gp5, a T4 phage tail lysozyme. Gp5 uses an internal cleavage mechanism that upon trimerisation activates and releases an N-terminal lysozyme domain from intermolecular restraints (Kanamaru *et al.* 2005). However, the cleaved lysozyme domain could also be detected after expression, unlike the potentially degraded EADs of the three clostridal endolysins presented here.

Theoretically, as the cytosolic concentration of the endolysins increases during overexpression in *E. coli,* the tendency to undergo oligomerisation also increases as the local concentration of the endolysins constitutively increases. This could prompt the premature activation and cleavage of the endolysins, which is why the C-terminal domain was detected throughout *E. coli* expression. Overexpression in *E. coli* has previously been attributed to the cytosolic activation of an autoproteolytic processing mechanism for the cysteine autoproteases NopT1 and NopT2 (Fotiadis *et al.* 2012).
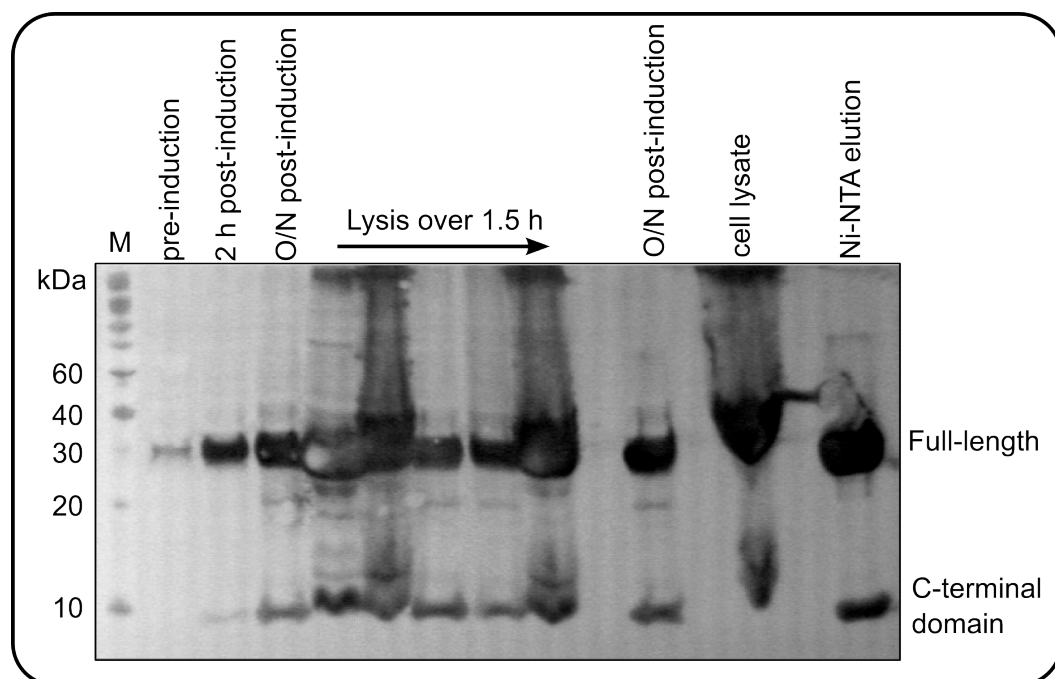


**Figure 30 Immunoblotting used for the detection of autocleavage during *E. coli* overexpression.** Using immunoblotting the polyhistidine tag containing C-terminal domain was detected throughout expression and purification. 2 hours post induction the C-terminal domain can be detected indicating autocleavage (second lane). Also, pre-induction indicated leaky expression of the full-length protein (first lane).

*Disappearance of the N-terminal EAD during purification*

The cleaved N-terminal EADs for all three endolysins could not be detected during purification by SDS-PAGE or immunoblotting. It was previously shown that the N-terminal domains of CD27L, CTP1L and CS74L could be expressed as truncated domains (Mayer *et al.* 2008; 2010; 2011) so theoretically they should also remain stable and folded during full-length purification. As the C-terminal domain is visible during expression within the *E. coli* cytosol it is presumed that the catalytic domain must be released during overexpression but does not remain stable within the contents of the *E. coli* cytosol and is degraded within the cell.

*Proteolysis is autonomously controlled by the C-terminal domains*

CD27L, CTP1L and CS74L have a modular organisation and separate the EAD and C-terminal domain with a linker. Previously, Mayer *et al.* showed that a chimera of the EAD of CS74L combined with the C-terminal domain of CD27L (CS-CD) was still active against *C. sporogenes* cells (Mayer *et al.* 2012) and that this chimera also demonstrated autocleavage of the C-terminal domain. An additional chimera has been produced with the EAD of CTP1L and the C-terminal domain of CD27L (CT-CD), which also demonstrated C-terminal cleavage, although this chimera was not active against cultures of C. *tyrobutyricum* or *C. difficile* cells (Appendix Figure 10). CS-CD and CT-CD present evidence that autoproteolysis occurs independently of the N-terminal domain.

To demonstrate that the autocleavage mechanism was entirely self-contained within the C-terminal domains, a chimera was produced for CTP1L that replaced the EAD with the non-related domain of small ubiquitin-related modifier 3 (SUMO3). The linker was kept the same except with two additional glycine residues incorporated to increase flexibility and deter potential steric clashing between SUMO3 and the C-terminal domain. Cleavage of the C-terminal domain from the SUMO3-CT chimera was low but detectable by SDS-PAGE and MALDI-TOF MS (Figure 31 A, B). Tryptic digest analysis detected the post-proteolytic methionine incorporation on the C-terminal domain indicating that the cleavage of the C-terminal domain was due to the same autocleavage mechanism (Figure 31, C). The side-by-side mutant T221R was introduced to the SUMO3-CT chimera, although T221R was highly unstable and showed a higher amount of general degradation. T221R totally inhibited cleavage as no band corresponding to the C-terminal domain could be detected by SDS-PAGE (Figure 31 A, middle lane) or MALDI-TOF MS. As previously mentioned, the N-terminal EAD is cleaved whilst CTP1L is sequestered in the *E. coli* cells but is absent after lysis and purification, suggesting the domain is degraded and cannot be

purified. It remains to be seen if the cleaved SUMO3 domain from the SUMO3-CT chimera is also degraded during *E. coli* expression. Isolation and characterisation of the C-terminus of the released SUMO3 would identify if the released domain is also modified during autocleavage and provide details to help elucidate the cleavage reaction.
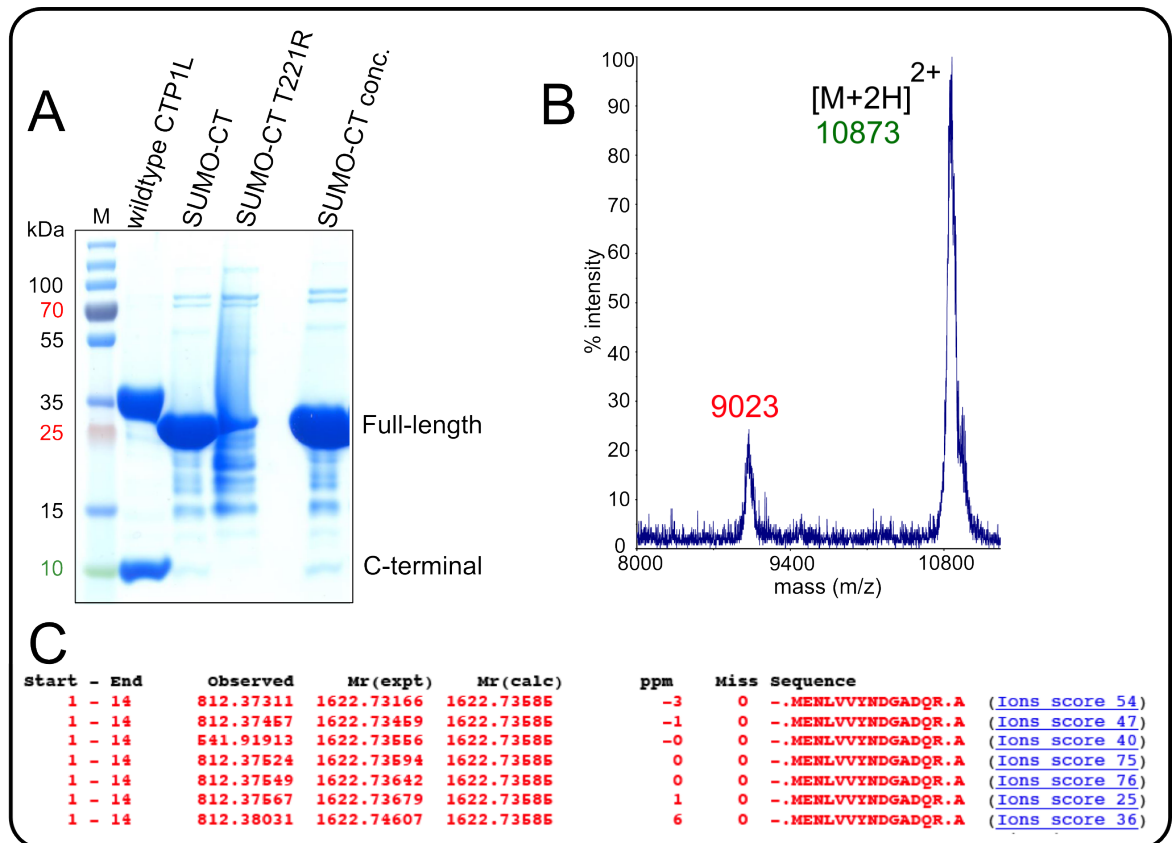


**Figure 31 Autocleavage and methionine incorporation still occurred after substitution of CTP1L EAD for SUMO3 A:** SDS-PAGE of SUMO3 linked to the C-terminal domain of CTP1L (SUMO-CT) showed a slight presence of the cleaved C-terminal domain. The mutation T221R that inhibits the side-by-side dimer was inserted into the chimera as a control and cleavage of the C-terminal domain was not detected. **B:** MALDI-TOF MS spectra of the SUMO-CT, which detected a peak at 9023 Da, a similar mass shift to the 9032 Da detected for the wildtype CTP1L C-terminal domain after cleavage. The double charged $[M+2H]^{2+}$ ion of the full-length SUMO-CT chimera (MW: 21.7 kDa) was also detected. **C:** Sample MASCOT search results after tryptic digest and LC-MS/MS for the SUMO-CT C-terminal domain that was excised after SDS-PAGE detected the same post-proteolytic modification of methionine on the C-terminal domain.

*HEK cell expression of CTP1L and CS74L*

To study the initiation of autoproteolysis it is necessary to begin with an inactive state of the endolysins to which an array of potential activator molecules can be added and the causative agent identified. As previously mentioned the endolysins are proteolytically active and cleaving themselves during *E. coli* expression (Figure 30) making identification of activator molecules difficult. Even with successive purification steps as the C-terminal domain remained bound to the full-length monomer. As proteolysis occurs during expression within *E. coli*, the activation molecule or enzyme would be constitutively present within the cytosol. Human embryonic kidney (HEK) 293 cells were investigated as a different expression model. HEK cytosolic components could be missing the elements required for initiation of proteolysis that was observed in *E. coli* cells. HEK cells also provide an alternative expression system that uses protein secretion. CTP1L and CS74L were cloned into a modified pXLGsec vector, which introduced an N-terminal signal peptide onto the endolysins. This allowed secretion of the endolysins during expression into the growth media, instead of being sequestered within the cytosol. Unfortunately, expression was only possible for CS74L and even then the endolysin was retained in the cell and not secreted. Sequestered proteins in the cytosol could be a sign that the protein is incorrectly folded and unfit for secretion and therefore unusable.

*Purification of the full-length endolysin alone for future experimentation*

Removal of the C-terminal domain was also attempted after *E. coli* expression. By performing the initial Ni-NTA purification under denaturing conditions any interdomain interactions, such as dimerisation between the C-terminal domains, would be removed. 6 M Guanidine hydrochloride (Gu.HCl) was added to the lysis buffer and 8 M urea was added for all subsequent washing and elution steps to ensure all proteins remained unfolded during purification. Proteins were eluted with a final pH 4. Using denatured purification it was possible to elute the single full-length endolysins without the additional C-terminal domain attached, as detected by SDS-PAGE (Figure 32 A, B, C). Purification under denaturing conditions inhibited the C-terminal domain from interacting with the full-length endolysins and hence was not co-purified. The eluted proteins were refolded using limited dilution and finally dialysed into 20 mM TRIS-HCl, pH 8 and concentrated to 1 mg.ml$^{-1}$. Concentrating the refolded endolysins did not initiate cleavage of the C-terminal domain, suggesting the cleavage mechanism could no longer occur until activation. As a preliminary test, 10 - 100 mM methionine and a sample of *E. coli* cell lysate were separately added to the concentrated refolded CTP1L and CS74L to see if autocleavage

could be induced. However, no cleaved C-terminal domain could be detected by SDS-PAGE, suggesting the cleavage mechanism could not be triggered (Figure 32 D). Possibly the refolding was not successful and the purified endolysin was not refolded correctly and cleavage could not be triggered, or, additional constituents are required to trigger activation. For instance during release from holin lesions, endolysin are exposed to an environmental shock, so spiking the refolded endolysins with a shift in pH or ionic conditions could be required for initiation of cleavage.

Interestingly, during refolding of both CTP1L and CS74L, a degradation band appeared just below the full-length protein band at an approximate mass of 24 kDa (Figure 32, D, red star). This degradation product corresponded in size to the degradation products observed for different mutants within the cleavage site or that inhibited side-by-side dimerisation, as previously mentioned (page 49). The observation of a similar degradation band for the refolded endolysins suggested that whilst the endolysins are still able to change conformation, the ability of the endolysins to perform the autocleavage mechanism is inhibited, potentially as the causative agent of proteolysis was not present, which caused the degradation. The refolded endolysins require future analysis, initially to clarify that the secondary structure is intact by using circular dichroism. The ability to purify the full-length endolysins alone provides a base for future work in the search for the activating molecule that triggers autocleavage.
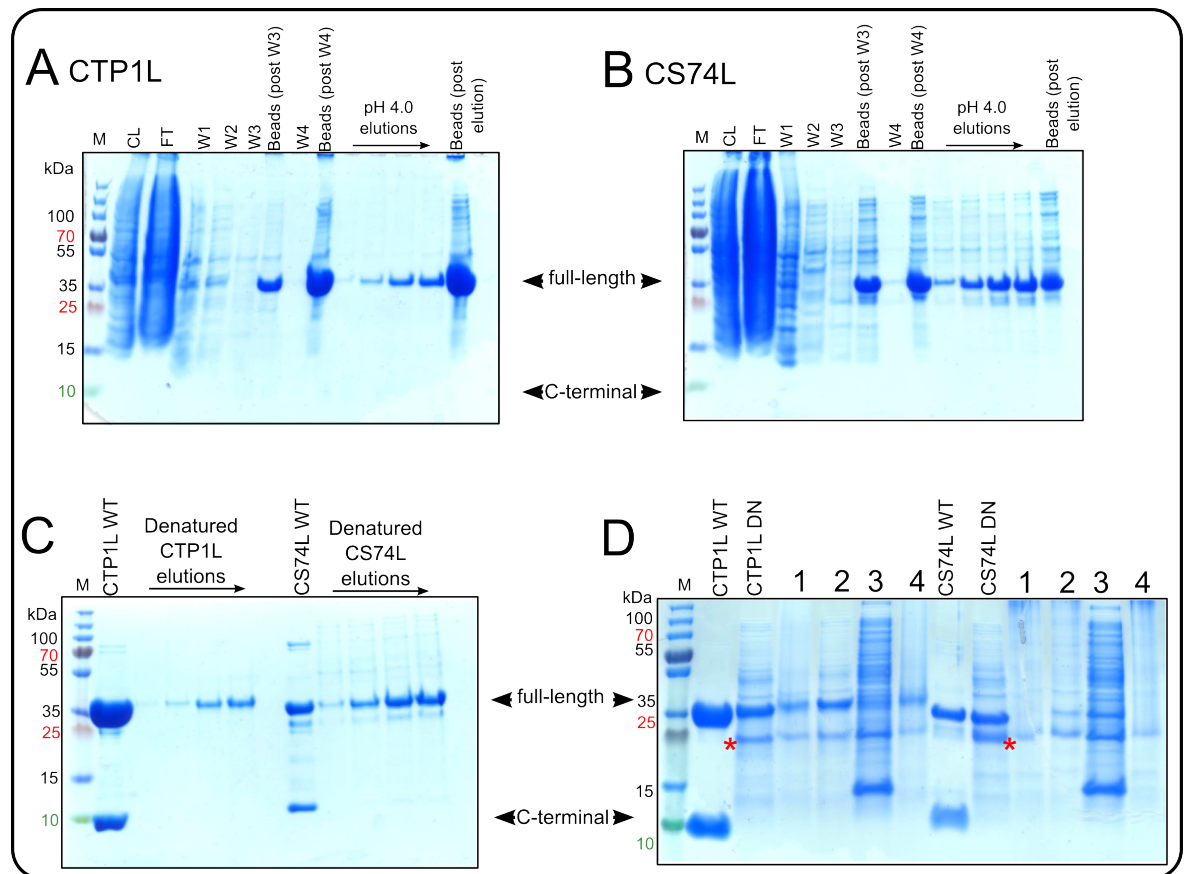
**Figure 32 Denatured Ni-NTA purification of full-length CTP1L and CS74L.**
**A (CTP1L), B (CS74L):** Sequential steps during denatured purification of the endolysins. CL: cell lysate in the presence of 6 M Gu.HCl, FT: flow through off Ni-NTA resin, W: wash steps using 8 M urea. Proteins bound to the resin were eluted at pH 4 in 8 M urea. Bead samples were taken to show only full-length protein was attached to the resin. A large amount of protein was still attached to the beads post elution, indicating improvements are required for this protocol. The C-terminal domain could not be detected and must be excluded in the flowthrough. **C:** Comparison of endolysins purified under native and denatured conditions. No C-terminal domain could be detected for the refolded CTP1L and CS74L after denatured Ni-NTA purification. **D:** Preliminary experiment to test the addition of 1: 10 mM L-methionine, 200 mM NaCl pH 7.4, 2: 10 mM L-methionine, 200 mM NaCl, pH9.0, 3: 6 µl *E. coli* cell lysate 1: 6 endolysin, 4: 100 mM L-methionine, 200 mM NaCl, pH7.4. Added to refolded and concentrated samples (1 mg.ml$^{-1}$) of refolded CTP1L or CS74L. WT: wildtype, DN: denatured, refolded and concentrated endolysins.

# Summary of chapter three

Chapter three explored the mechanism of autocleavage for the three endolysins. Numerous experiments were applied to test for causative agents that could attach a 32 Da mass to the cleaved C-terminal domain.  Eventually it was discovered that during the cleavage of the linker, the post-proteolytic modification of CTP1L and CS74L involved the substitution of methionine for valine at the cleavage site. Whilst for CD27L, the cleavage mechanism occurred by peptide hydrolysis between a methionine and glutamine residue. Modification of the CTP1L and CS74L cleavage site valine to a non-branched chain amino acid (glycine, tyrosine, proline) affected the amount of C-terminal domain cleavage and also a shift in proteolytic kinetics. As well as methionine incorporation, peptide hydrolysis was also observed within the cleavage site. Mutation of valine to the similar branch-chained amino acid leucine had no effect on proteolysis and methionine incorporation was the only post-proteolytic modification detected. The substitution of methionine for valine could be from another enzyme although no known enzymes could operate in this manner. This requires further investigation. It is interesting to note that the scissile bond broken in CD27L is between a methionine and glutamine and as methionine is incorporated in CTP1L and CS74L, suggests a role for methionine as a catalytic residue during autoproteolysis.

Proteolysis was demonstrated to occur after induction and during the entire *E. coli* expression. Overexpression within *E. coli* has previously been shown to induce oligomeric shifts for other proteins, which could be the reason for premature activation of the autoproteolytic mechanism for these endolysins. As the EAD could not be co-purified it was suggested that it is degraded after cleavage within the *E. coli* cell, or has yet to be detected. Interestingly, proteolysis was shown to be autonomous for the C-terminal domain as chimeras of the CD27L C-terminal domain still demonstrated autocleavage. Furthermore, a chimera of the unrelated SUMO3 domain linked to the C-terminal domain of CTP1L was also cleaved with the same autocleavage mechanism as the post-proteolytic methionine addition was detected by tryptic digestion for this chimera. Finally, it was shown that purification under denaturing conditions could be used to obtain the full-length endolysin alone, without the contamination of the cleaved C-terminal domain. Experimentation with the full-length endolysin alone will make the future study of the trigger mechanism of autoproteolysis less ambiguous.

# Chapter 4: Methodological development for the identification of endolysin cell wall ligands

*Introduction*

Cell wall binding domains recognise peptidoglycan and cell wall polymers, such as teichoic acid, with high affinity and specificity. Examples include the LysM repeat motif that bind peptidoglycan (Hu *et al.* 2010) and the Cpl-1 choline binding motifs (Garcia *et al.* 1988). Additionally, *Listeria* phage endolysins are documented to recognise different cell wall ligands. The cell wall binding domain of PlyP35 specifically recognises GlcNAc residues bound to teichoic acid (Eugster *et al.* 2011), whereas the *Listeria* phage endolysins Ply118, 511 and P40 specifically interact with the peptidoglycan backbone (Eugster & Loessner 2012). Specific ligands were confirmed using binding studies of fluorescently labelled cell wall binding domains against chemically treated cell walls (Hu *et al.* 2010) or genetically modified cell walls (Eugster *et al.* 2011; Eugster & Loessner 2012), where particular cell wall components were modified. For instance, removal of genes for GlcNAc decoration of teichoic acids inhibited PlyP35 binding, as measured by a decrease in GFP binding fluorescence (Eugster *et al.* 2011). Another reported method for identification of cell wall binding ligands is to measure the effect of cell wall modification on lytic ability. The *L. casei* phage endolysin Lc-Lys specifically targets the D-Asn interpeptide bridge of peptidoglycan. This was deducted by measuring the lytic ability of Lc-Lys against *L. lactis* cells where the peptidoglycan interpeptide cross-bridges had been mutated. Only peptidoglycan cross-bridges containing D-Asn could be bound and lysed (Regulski *et al.* 2013).

The parallel alpha helices α1 and α3 involved in the head-on dimerisation of the CTP1L and CD27L C-terminal domains are predicted to be the cell wall binding sites for these endolysins. Glycan binding proteins incorporate aromatic residues at their binding sites that are proposed to facilitate contact between the protein surface and glycans (Buist *et al.* 2008). The LysM motif binding domains, the *S. pneumoniae* choline-binding domains and other carbohydrate-binding domains reveal a shared YG motif within their binding sites (Turner *et al.* 2004). Attached to the alpha helices α1 and α3 of the CTP1L and CD27L C-terminal domains there are also a high number of aromatic residues that project into the cavity between the two helices that could enforce interactions with glycans. As previously described, the mutation of methionine to arginine (CTP1L M263R), whose side chain protrudes into the middle of the two alpha helices of the CTP1L C-terminal domain, continued to cleave the C-terminal (Figure 14, C), but inhibited the lytic activity of CTP1L (Figure 22, A). Mutation to arginine introduced a positive charge into this normally

hydrophobic region, which potentially modified the normal behaviour of the binding site and prevented attachment of the endolysin to the cell wall.

There is a limited amount of detailed cell wall binding domain epitopes so far identified. As more cell wall binding domain structures are being interpreted, more needs to be known regarding their interactions with cell wall ligands. Different peptidoglycan chemotypes are prone to species-specific modifications that could play a crucial role in endolysin recognition. For example, wall sugar patterns differ between *C. tyrobutyricum* and other butyric acid producing *Clostridia* (Schleifer & Kandler 1972) and binding may be susceptible to these particular modifications. A potential method to identify the specific epitopes that endolysins and other non-covalent binding proteins recognise is to use photo-activated cross-linkers to covalently attach ligands to proteins, which can then be analysed by mass spectrometry.

### *Methodological development: Photo-cross linking to locate the ligand for cell wall binding*

Unnatural amino acid incorporation of *p*-benzoyl-l-phenylalanine (pBpa) relies on an orthogonal tRNA synthetase and $tRNA_{CUA}$ pair for the *in vivo* incorporation of pBpa in response to the amber codon, TAG, at a specific site within a protein (Chin *et al.* 2002). Excitation of pBpa at 365 nm causes cross-linking of any C-H bonds within 10 Å of the carbonyl oxygen of pBpa (Farrell *et al.* 2005), thereby forming a covalent bond between the pBpa containing protein with possible binding partners. A proof of principal study was initially performed to see if it was possible to photo-cross-link peptidoglycan fragments to the C-terminal domain of CTP1L using pBpa incorporation. Figure 33, A describes the methodology in detail. The pBpa was incorporated at position Y212, situated on alpha helix α1 (Y212pBpa), and at position Y260, situated on alpha helix α3 (Y260pBpa), in different constructs of the truncated C-terminal domain of CTP1L. Incorporation was confirmed for both mutants by tryptic digestion followed by mass spectrometry (Figure 33, B).
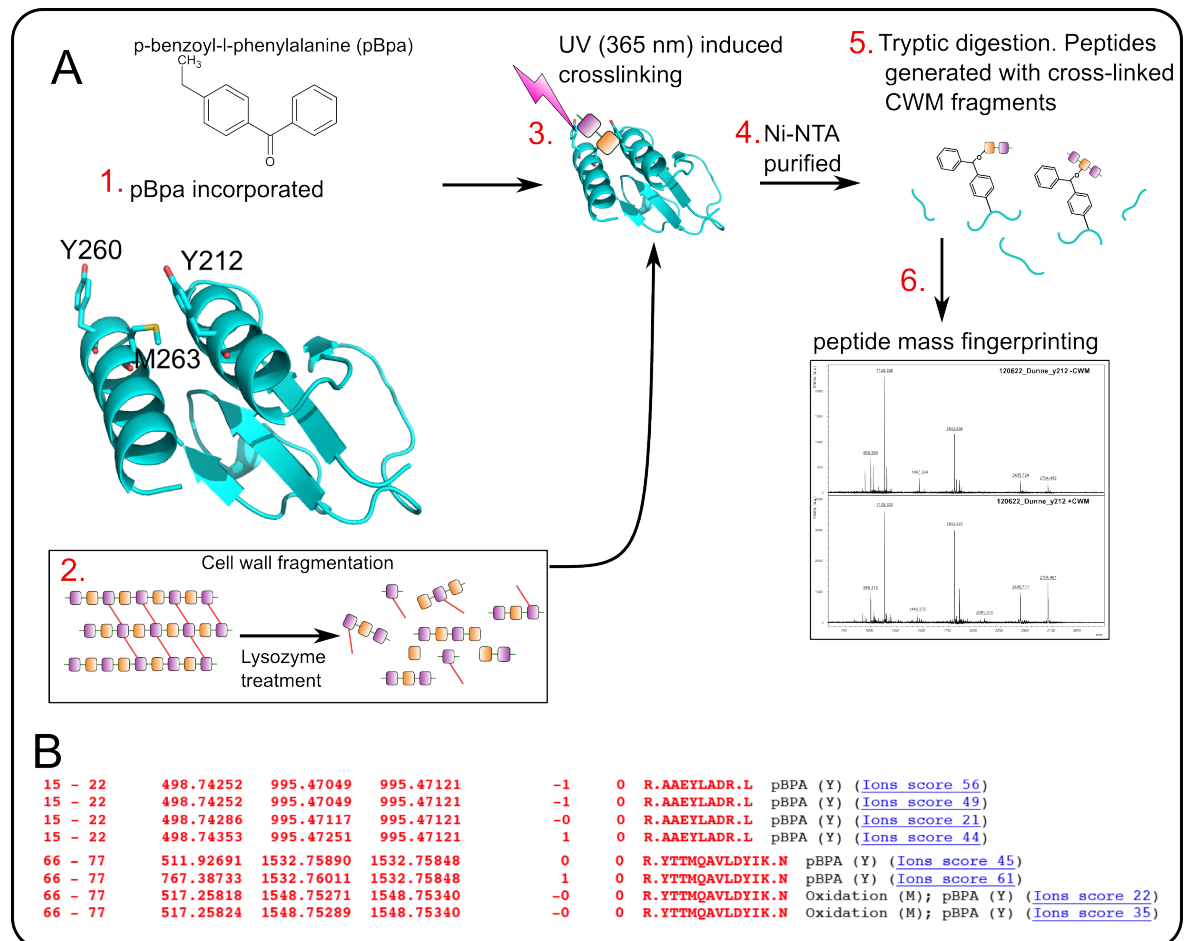
**Figure 33 Photo cross-linking methodology for ligand binding**. **A:** (1) The unnatural amino pBpa was incorporated at positions Y212 or Y260 of the truncated C-terminal domain of CTP1L. (2) *C. tyrobutyricum* CWM was purified and fragmented by lysozyme degradation. (3) The pBpa containing C-terminal domains were mixed with the *C. tyrobutyricum* CWM and irradiated with UV light (365 nm). (4) After sufficient incubation time the covalently linked complexes were Ni-NTA affinity purified to remove unbound CWM. (5) Samples were digested with trypsin. (6) Peptide mass fingerprinting was used to identify novel peptides that could have CWM fragments attached, for instance cross-linked peptidoglycan molecules. **B:** Example MASCOT search results for trypsin digestion linked with LC-MS/MS of Y212pBpa (top) andY260pBpa (bottom), demonstrated the successful identification of pBpa incorporation.

After detection of successful cross-linking, subsequent glycan analytical steps could be explored. As the location of pBpa is already known, position Y212 or Y260, cross-linked glycans could only be present on the respective tryptic digested peptides containing the unnatural amino acid. For the initial analysis MALDI-TOF-MS was used. All MS analysis was performed in collaboration with Dr. Diana Hildebrand at the University Medical Centre Hamburg-Eppendorf (UKE).

The native cell wall of *C. tyrobutyricum* is absent of teichoic and teichuronic acids, making peptidoglycan or peptidoglycan bound modifications the most likely binding sites for CTP1L (Bergère *et al.* 1986). Cell wall material (CWM) fractions were treated with

lysozyme in order to generate different sized fragments of peptidoglycan. CWM was also extracted using SDS to remove cell wall associated proteins and lipids. Lysozyme was used to fragment the peptidoglycan whilst maintaining features such as the peptide linker between MurNAc subunits. Peptidoglycan lengths can vary and digestion with lysozyme into smaller fragments makes detection and analysis via mass spectrometry easier, however, it is not possible to generate all peptidoglycan fragments of uniform size by lysozyme treatment. It was expected that after lysozyme treatment numerous fragments of the same peptidoglycan conformation would be generated.

Primarily, the intention was to detect differences in the MALDI-TOF-MS spectra between UV-irradiated samples with or without CWM present. In the sample without CWM only signals derived from tryptic peptides should be present. When cell wall material was incubated with the C-terminal domain the presence of any novel signal in the MS-spectra of the sample with CWM present could (in comparison to the signal pattern of the sample without CWM) indicate the presence of peptides containing an additional mass corresponding to an attached peptidoglycan fragment. Such putative peptide-peptidoglycan-derived signals could then be further analysed by MS/MS analysis to solve the structure of the complex and identify the attached binding partner. Unfortunately, no differences could be detected between spectra with or without the different CWM preparations for both the Y212 and Y260 pBpa-incorporated sites. The only notable observation after UV-irradiation of Y212pBpa was a dipeptide formed by pBpa-cross-linking between two peptides from opposing C-terminal domains. The cross-linked dipeptide was verified by Electrospray Ionisation Quadrupole Time-Of-Flight (ESI-Q-TOF)-MS/MS analysis (Appendix Figure 12) and showed that cross-linking was occurring to CTP1L whilst establishing the head-on dimer state. An excess of CWM was used compared to the pBpa containing C-terminal domains to try and increase the chance of interactions with CWM. It is therefore possible that no CWM fragments were covalently linked to the C-terminal domains and apart from the self-cross-linking to form covalent dimers, no other products could be formed.

Another explanation could be due to the highly heterogeneous nature and complexity of the CWM fragment-fraction regarding its molecular diversity. As a result "binding products" of the CWM fragments to the CTP1L are probably not detectable due to too low concentrations of each of the cross-linking products. In that case signal intensity of the cross-linking products will be too low and will be suppressed by the regular tryptic

peptides. For future perspective, specific glycopeptide enrichment steps, such as hydrophilic interaction liquid chromatography (HILIC) could also be used to increase the concentration of cross-linked products to increase signal intensities (Kolarich *et al.* 2012). Consideration is also necessary for the digestion/preparation to increase fragment homogeneity of the CWM, to increase detection of analogous cross-linked peptides and signal intensity.

### *The wildtype C-terminal domain can cross-link without pBpa incorporation*

A surprising result was observed with the wildtype CTP1L C-terminal domain, which self-cross-linked to form covalent dimers (of a similar MW to the cross-linked dimers present for Y212pBpa and Y260pBpa) upon UV-irradiation only in the presence of the *C. tyrobutyricum* CWM (Figure 34, A & C). The wildtype C-terminal domain does not contain the pBpa chemical cross-linker suggesting there must be a molecule present within the CWM extract that induced cross-linking after UV excitation. UV-induced cross-linking does not always require the introduction of a cross-linker. For instance, irradiation of protein-nucleic acid complexes with shortwave UV (254 nm) generates purine and pyrimidine free radicals that form covalent bonds between nucleic acids and proteins within close proximity (Chodosh 2001). Furthermore, monosaccharides can be synthesised to become photoactivatable by incorporating diazirine groups, which can be used to covalently interact with carbohydrate binding proteins (Tanaka & Kohler 2008).

As a control the wildtype C-terminal domain was mixed with *E. coli* CWM, prepared using the same protocol as the *C. tyrobutyricum* CWM. *E. coli* peptidoglycan consists of the same A1 gamma type peptidoglycan backbone as *C. tyrobutyricum* (Schleifer & Kandler 1972) and provided a suitable control to demonstrate that the non-pBpa cross-linking was specific for *C. tyrobutyricum* CWM. An additional control involved *C. tyrobutyricum* CWM that had been extensively treated with an excess of benzonase nuclease to degrade contaminating DNA and RNA fragments still retained in the CWM. Interestingly, after UV exposure the higher MW band corresponding to the dimer was only present when the C-terminal domain was mixed with the *C. tyrobutyricum* CWM, but not with the *E. coli* CWM, as monitored by western blot analysis (Figure 34, C). This further suggested that there is a specific molecule present in *C. tyrobutyricum* CWM that can induce cross-linking. After SDS-PAGE the dimer band was analysed by tryptic digestion linked with MS analysis, but unfortunately no additional peaks could be detected when compared to a control analysis of the C-terminal monomer without UV-irradiation. The signals from the

cross-linked peptides could be hidden and require enrichment before analysis. Full-length MS analysis needs to be performed on the cross-linked wildtype dimer to identify if an adduct gets incorporated into the dimer during UV irradiation. An additional mass besides the dimer mass would correspond to the mass of the cross-linked adduct. Identification of a potential cross-linking molecule could help identify the binding ligand for CTP1L and the reason why UV induced cross-linking was successful without photo-cross-linker incorporation.

### *Specific cross-linking of the CTP1L head-on dimer in solution*

Whilst cross-linking of the C-terminal domain to CWM remained unresolved, the specific cross-linking and dimer formation confirmed the existence of the head-on dimer for CTP1L in solution. After UV-irradiation both the Y212pBpa and the Y260pBpa mutants produced additional bands after SDS-PAGE analysis corresponding to dimer and tetramer formation (Figure 34, A). SDS-PAGE analysis on the unexposed Y212pBpa and the wild-type C-terminal domain do not present any cross-linked material. The higher molecular weight band was excised, digested with trypsin and analysed by MS to confirm that it contained only the C-terminal domain of CTP1L, due to the specificity of pBpa cross-linking. This verified the head-on dimer formation for CTP1L in solution.

The amber stop codon at position Y212 was also introduced into the full-length CTP1L endolysin. Y212 was the most efficient site for pBpa incorporation and gave the highest final yield of protein after purification when compared to incorporation at Y260. Full-length CTP1L was cross-linked upon exposure to UV light and analysed by SDS-PAGE. A mixture of species was observed, consisting of full-length CTP1L dimers, dimers of the C-terminal domains alone and a species with a MW that corresponded to one full length CTP1L cross-linked to a C-terminal domain fragment (Figure 34, B). The oligomerisation states of the CTP1L fragments gained by cross-linking support the interpretation of the size exclusion chromatography, SAXS and light scattering experiments performed for CTP1L and show that the head-on dimer is the significant species in solution.
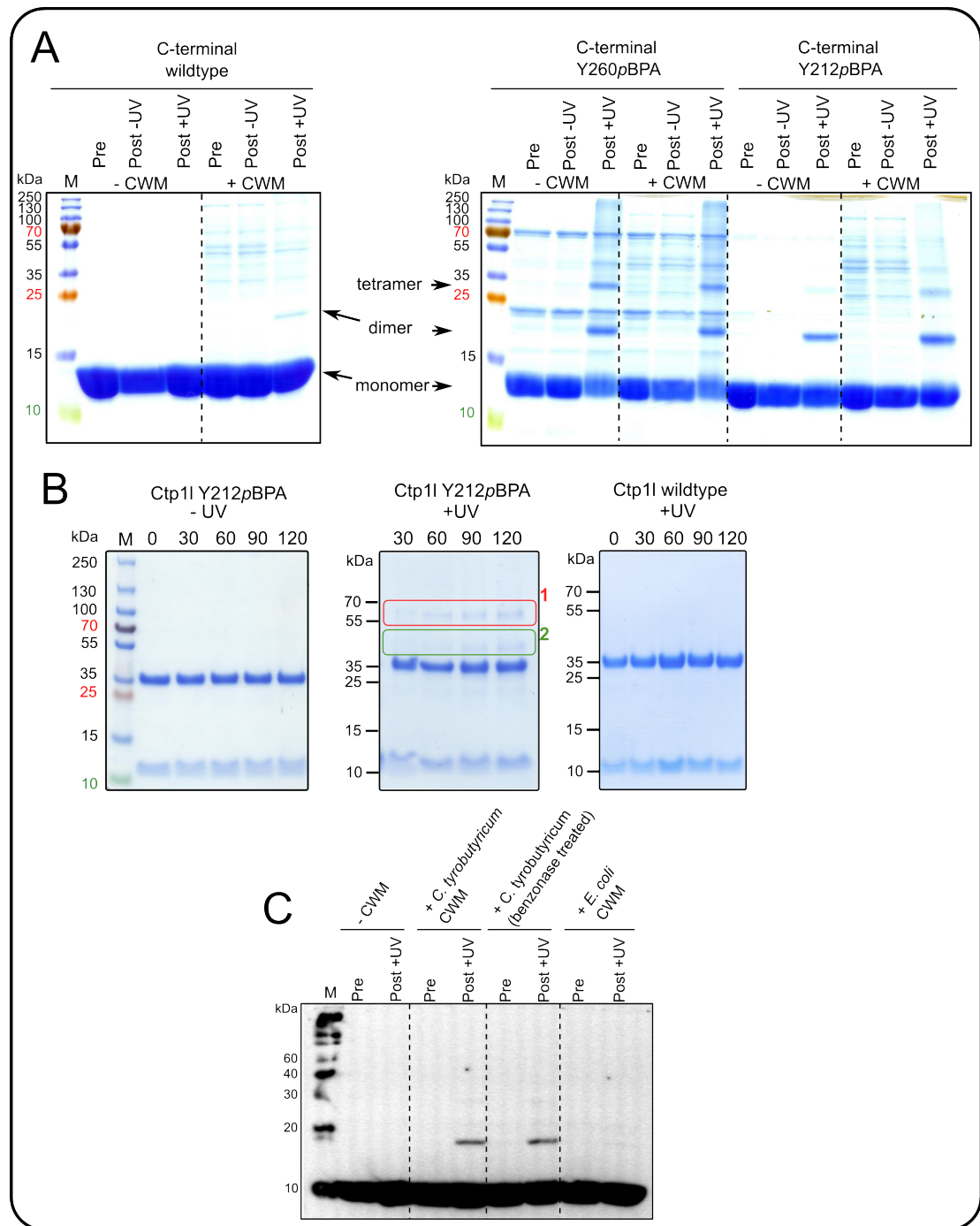
**Figure 34 Photo-cross-linking of CTP1L oligomers. A.** SDS-PAGE of the C-terminal domain of CTP1L with or without the light sensitive cross-linker pBpa incorporated with *C. tyrobutyricum* CWM. Dimer and tetramer states are observed for Y212pBpa and Y260pBpa after UV exposure. For the wildtype C-terminal domain, dimers could be cross-linked only in the presence of *C. tyrobutyricum* CWM. **B.** SDS-PAGE of the full-length endolysin with the Y212pBpa cross-linker mutant, detected the presence of a full-length dimer (red box) and an oligomer consisting of one C-terminal domain and a full-length CTP1L molecule (green box). **C:** Immunoblotting detected dimerisation of wildtype C-terminal domain only in the presence of *C. tyrobutyricum* CWM. Controls without *C. tyrobutyricum* CWM or with *E. coli* CWM did not detect any cross-linked dimer. *Figures A and B are reproduced from the submitted article, Appendix page 170.*

# Summary of chapter four

There remains limited information regarding the exact binding sites of endolysins within the cell walls of their hosts during lysis. In this chapter a potential method was described for the analysis of cell wall binding sites. As an example, photo-cross-linking was attempted with the C-terminal domain of CTP1L and its potential ligand within a mixture of *C. tyrobutyricum* cell wall fragments. Unfortunately, no positive hits could be detected for cross-linking of the C-terminal domain and the methodology requires future development. Major drawbacks for the method include the lack of control on heterogeneity of CWM fragmentation after lysozyme treatment and prior knowledge required of the cell wall binding domain structure in order to choose appropriate residues to modify to pBpa and increase the effectiveness of cross-linking. As a control, it would be interesting to attempt cross-linking with a different endolysin:ligand complex where the ligand identity is already known. A surprising discovery during experimentation was that the wildtype C-terminal domain could self-cross-link to form dimers when mixed with *C. tyrobutyricum* CWM. Identification of potential compounds that induce cross-linking remains a future goal and could provide insight into the binding ligand of CTP1L. Besides this novel observation, pBpa dimer photo-cross-linking further confirms that the head-on oligomerisation state exists for CTP1L in solution.

# Chapter 5: Discussion

*A novel activation mechanism for Clostridial bacteriophage endolysins*

Endolysin and holin are essential for host cell lysis by bacteriophage at the end of the lytic cycle. For the classic holin-endolysin system, endolysins are sequestered within the cytosol until holin lesion formation effects there release in an active state to begin peptidoglycan degradation. The molecular mechanisms underlying the activation of the cytosol-sequestered endolysins are unknown, but the sudden change in environment, such as changes in ionic strength, pH and reductive capacity, as they pass into the extracellular environment could act as the trigger. The depolarisation of the membrane and disruption of periplasmic electrochemical gradients from holin lesion formation, has previously been shown to activate other classes of endolysins, including the pre-secreted endolysin Lys44 (Nascimento *et al.* 2008) and the SAR-endolysins Lyz[P1], R[21] and Lyz[103] (Xu *et al.* 2004; Xu *et al.* 2005; Kuty *et al.* 2010).

Within this thesis, the activation of the three *Clostridial* endolysins CTP1L, CD27L and CS74L is proposed to occur via a novel autocleavage mechanism. Two distinct dimerisation modes were observed in the crystal structures of CTP1L and CD27L, an elongated head-on dimer and a side-by-side dimer. These dimer states were validated in solution by the use of small angle X-ray scattering and right angle light scattering. The head-on dimer was further confirmed by cross-linking experiments using the photoactivatable cross-linker, pBpa. Indirectly, it is shown that the conformational rearrangement of these endolysins from the head-on to the side-by-side dimer acts as the trigger for an autocleavage mechanism that activates these endolysins. The introduction of point mutations to prevent either of the dimerisation states had an inhibitory effect on the autoproteolytic mechanism, as they reduced the amount of detected C-terminal domain after purification (Figures 13-15). Furthermore, inhibition of either dimer state also inhibited the lytic activity of CTP1L (Figure 21).

The switch from the head-on to the side-by-side dimer is proposed to occur through a tetrameric intermediate, as observed in the crystals structures of CTP1L and CD27L, formed by the C-terminal domains (Figures 8 & 18). After holin lesion formation, the sudden change in ionic conditions and drop in pH is proposed to bring together two head-on dimers, which interact to form a tetrameric intermediate. Tetrameric formation concurrently forms the side-by-side interaction as well. It was demonstrated using a GFP-tag attached to the C-terminal domain, that the interdomain interactions of the side-by-side dimer could be stabilised at pH 6, whereas an increase to pH 9 interrupted with side-by-

side interactions and increased the release of the C-terminal domain (Figure 26). For CTP1L, as the tetramer state forms, a salt bridging network contributing to the head-on dimer interface is disrupted by the positive charge of an incoming R226 side chain of the C-terminal domain from the newly formed side-by-side state (Figure 18). This could provide the trigger to necessitate the switch between the two dimer states. The tetrameric intermediate subsequently dissociates and two side-by-side dimers are proposed to be released. A model for this trigger-release mechanism, via the tetrameric intermediate, is presented in Figure 35, A. The residues involved in both salt-bridging networks show partial conservation between CTP1l, CS74L and the 19 similar lysin sequences, suggesting this activation mechanism could be present in other lysins.

The switch between the two dimer states can be described in similar terms to a Monod-Wyman-Changeux (MWC) mechanism of allosteric regulation (Changeux 2012). The head-on dimer represents the inactive 'tensed state', as this is the most prominent state in solution, whereas the side-by-side dimer represents the active 'relaxed state' as this is the conformation that triggers autocleavage and the release of the N-terminal EADs. Essentially, the quaternary structure transition from tensed to relaxed state, is achieved by the stabilisation of side-by-side interactions, after the environmental shift due to the periplasmic entry of the endolysins. Allosteric regulation by interdomain interactions of the C-terminal domain on the EAD was previously hypothesised (Low *et al.* 2005) and later disregarded (Low *et al.* 2011). However, interdomain interactions were recently demonstrated using surface plasmon resonance for the mycobacterium phage lysin A, whereby interactions between catalytic and cell wall binding domains were proposed to inactivate the lysin (Pohane *et al.* 2014). For the *Clostridial* endolysins studied here, allosteric regulation is proposed to be solely from restraining the EAD by the linker and not by interdomain interactions. Autocleavage of the linker releases constraints on the EAD and increases the potential action radius of the EAD as a single domain. As previously proposed (Low *et al.* 2011), the small globular size of the EAD would allow further penetration within the bacterial cell wall compared to the full-length endolysin, to more efficiently exert its lytic activity. Interestingly, truncation of CD27L to the EAD alone showed an elevated lytic activity compared to the full-length endolysin (Mayer *et al.* 2011). Although truncation to the EAD alone does not increase the activity of all endolyins (Sanz *et al.* (1992); Porter *et al.* (2007); Sass & Bierbaum (2007)), including CTP1L (Mayer *et al.* 2010).
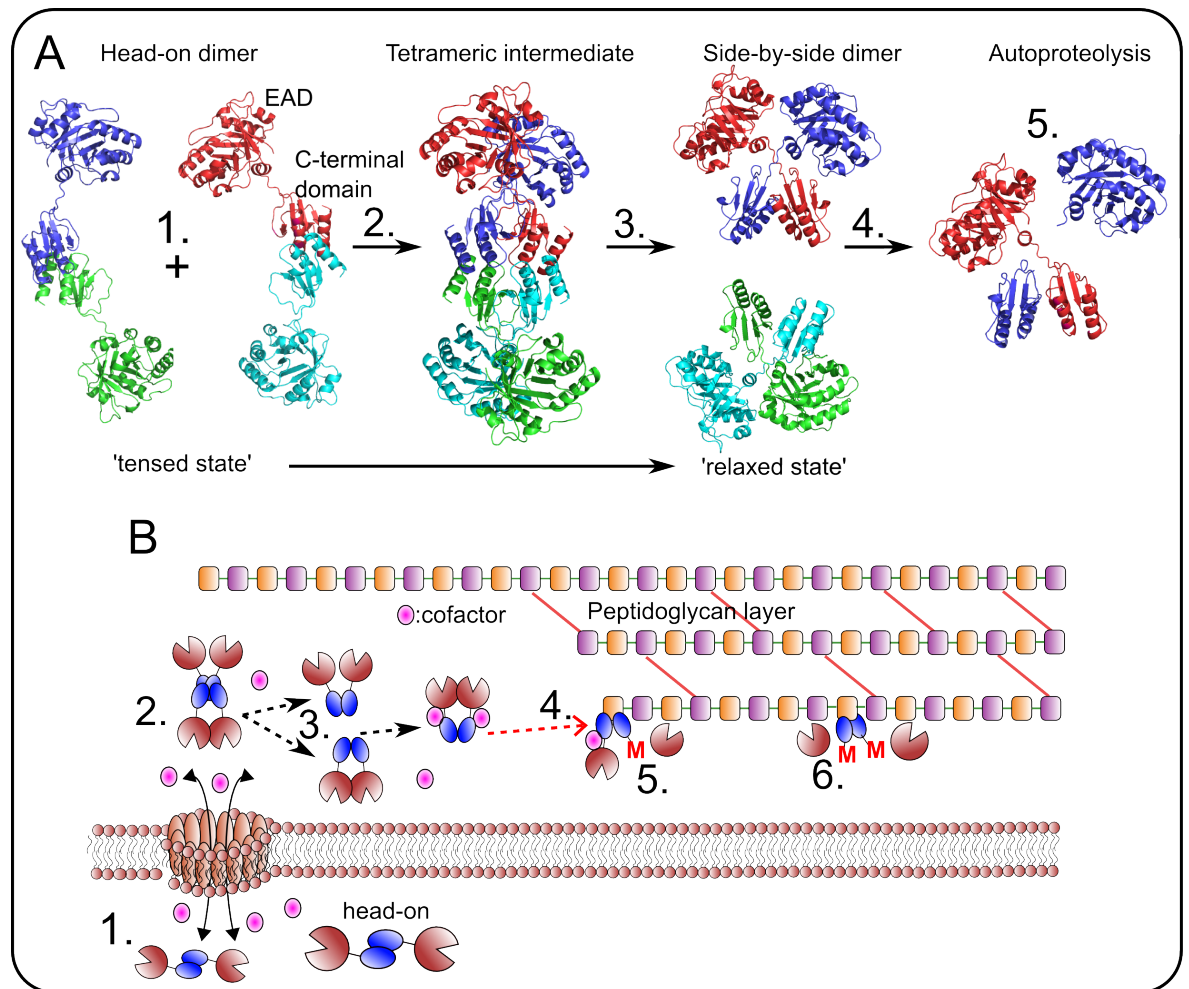
**Figure 35 A proposed model for the trigger-release mechanism of activation for the** *Clostridial* **endolysins.** Using CTP1L as the model endolysin. **A:** Structural model of the endolysin activation, generated using PYMOL. **B:** Cartoon schematic of endolysin activation during lysis. **1:** Within the cytosol the endolysins form head-on dimers, the most stable 'tensed state' of the endolysins as measured by right angle light scattering, SAXS and cross-linking. **2:** Upon holin lesion formation, the cytosol and periplasm are no longer separated and the head-on dimers are exposed to a sudden change in environmental conditions as they pass into the cell wall. This environmental shock is proposed to trigger the formation of the tetrameric intermediate directed through interactions between the C-terminal domains. **3:** Conformational restraints prompt the tetrameric state to dissipate and two side-by-side dimers are released. **4:** In the side-by-side 'relaxed state' the proposed cell wall binding sites of the α1 and α3 helices are exposed and target the endolysin to specific epitopes within the cell wall. The side-by-side dimer also forces the autocleavage sites into close proximity and activates the autocleavage mechanism. The autocleavage mechanism is hypothesised to involve a constitutively present cofactor (pink sphere), either a methionine containing molecule or an enzyme, which contributes to the post-proteolytic methionine modification (red M) onto the cleaved C-terminal domain. **5:** The release of the N-terminal EAD as a single domain increases the action radius of the EAD during peptidoglycan degradation. **6:** As observed for the fully degraded crystal forms of CD27L and CTP1L, the second EAD of the side-by-side dimer is also released from the constraints of the C-terminal domain to become active as well.

Oligomerisation influences the lytic activity of lysins from other bacteria and bacteriophage species. For instance, the pneumococcal autolysin LytA and pneumococcal phage endolysins PAL and Cpl-1 dimerise upon binding their cell wall ligand choline in solution (Romero *et al.* 2007; Varea *et al.* 2004; Buey *et al.* 2007). For LytA, the conformation of the active dimer closely resembles the side-by-side dimer presented here and removal of the residues involved in dimer formation forced LytA to remain as a monomer, which significantly reduced its lytic activity (Fernández-Tornero *et al.* 2002). For Cpl-1 the dimer state was stabilised by engineering a disulphide bond into the C-terminal domains, which increased lytic activity (Resch *et al.* 2011). Although activation of these lysins does not involve an autocleavage mechanism, dimerisation has been shown to play a crucial role in their activation and efficiency as lytic enzymes. Additionally, a T4 tail lysozyme, gp5, uses an internal cleavage mechanism that upon trimerisation activates and releases an N-terminal lysozyme domain from intermolecular restraints. Similar to CTP1L, inhibition of this cleavage mechanism inactivated the enzymatic activity of gp5 (Kanamaru *et al.* 2005). Furthermore, overexpression in *E. coli* was responsible for the premature cleavage of the gp5 protein during expression, as the full-length gp5 and its cleaved lysozyme domain were purified separately after lysis. This is analogous to the autocleavage observed during expression and purification of CD27L, CTP1L and CS74L (Figures 10 & 30).

Chimeras generated between the EADs of CS74L (Mayer *et al.* 2012) or CTP1L with the C-terminal domain of CD27L still demonstrated autocleavage of the C-terminal domain to the same degree as wildtype CD27L (Appendix Figure 10). Additionally, the post-proteolytic modification of methionine onto the C-terminal domain of CTP1L was detected even when the EAD was replaced with the unrelated protein SUMO3 (Figure 31). This demonstrated that autoproteolysis, by incorporation of methionine, was autonomously controlled by the C-terminal domains, irrespective of the N-terminal domains providing the N-terminal domains do not interfere with the oligomeric rearrangements of the C-terminal domains during activation.

For CTP1L the length of the linker, but not linker flexibility or side chain chemistry, is essential for efficient autocleavage. Mutation of half the linker to proline residues (CTP1L$^{PPP}$) did not affect autocleavage, whereas deletion of half the linker (CTP1L$^{\Delta Linker}$) totally inhibited autocleavage (Figure 23). Inhibition of autocleavage by the CTP1L$^{\Delta Linker}$

construct is potentially from the steric clashing of the N-terminal EADs in the side-by-side dimer state due to the shorter linker.

Autoproteolysis is proposed to occur as the stems of the linkers are forced into close proximity during side-by-side formation. The repulsion expected from the two linkers interacting and the two EADS coming into close proximity would be compensated by the stable formation of the side-by-side interface. As the stems of the linkers approach, conformational strain may drive distortion of peptide bonds at the cleavage sites and help destabilise the scissile bond for autocleavage, similar to the activation of cis-autoproteolytic mechanisms (Buller *et al.* 2012). Residues in close proximity to the cleavage sites were proposed to be involved in either intradomain (within the linker itself) or interdomain (from exchange between the two linkers) autocleavage. Bacteriophages have been shown to combine oligomerisation and interdomain and intradomain autocleavage for the initial infection of the bacterial cell (Xiang *et al.* 2009). Unfortunately, current mutational analysis of residues located around the cleavage site of CTP1L has not yet established common catalytic residues involved in autocleavage.

For CD27L, the autoproteolytic activity could be switched off with a point mutation (M186P) at the cleavage site. For the CTP1L endolysin, the inhibition of proteolysis had the additional effect of inhibiting lytic activity against *C. tyrobutyricum* cells (Figure 21). Furthermore, for CTP1L the introduction of a point mutation (V195P) at the cleavage site demonstrated that the rate of lytic activity was proportional to the degree of autocleavage, whereas for CD27L inhibition of autocleavage had no effect on lytic activity (Figure 20). In the context of "lysin-mediated lysis from without" (Abedon 2011), the lytic activity of CD27L was insensitive to autocleavage. However, autocleavage is expected to play a vital role for the activation of CD27L during the normal phage lytic cycle, when the endolysins act from the inside of the host. The activation mechanism would therefore have more importance compared to its external application.

During the cleavage of the linker, the post-proteolytic modification of CTP1L and CS74L involved the replacement of methionine for valine at the cleavage site, whilst for CD27L, the cleavage mechanism occurred by peptide hydrolysis between a methionine and glutamine residue (Figure 24). The presence of methionine at the scissile bond of CD27L and the incorporation of methionine after the CTP1L and CS74L suggests a role for the methionine side chain during the two different autocleavage events. Methionine is semi-conserved at this position in lysins with a similar domain to CD27L. The presence of

methionine within autoproteolytic sites of other nonrelated proteins has also been documented but not investigated (Fotiadis *et al.* 2012). Extensive experimentation could not yet identify a cofactor, either a methionine containing molecule or an enzyme, for the incorporation of methionine in CTP1L and CS74L and requires further investigation. The observation of autocleavage during overexpression in *E. coli* suggests the cofactor may be constitutively present on both sides of the membrane or it is released along with the endolysins during lesion formation. Figure 35, B is a hypothesised schematic for the autocleavage mechanism in the context of lysis from within the *Clostridial* host cell, whereby the side-by-side dimer formation and the potential cofactor initiate the autocleavage mechanism.

Crystallisation of CTP1L presented two products, either the partially degraded (a single EAD released) or fully degraded (both EADs released) conformation. Anomalous sulphur scattering on both crystal structures further confirmed that all cleaved C-terminal domains were cleaved through the post-proteolytic incorporation of methionine (Figure 27). The presence of methionine on all the cleaved C-terminal domains of the fully degraded crystal suggests that autocleavage is likely a two-step process. After the first EAD is released (to give the partially degraded crystal form) the incorporation of methionine at the cleavage site could act as a catalytic residue for the second cleavage event.

Modification of the CTP1L and CS74L cleavage site valine to a non-branched chain amino acid (glycine, tyrosine or proline) not only decreased the amount of C-terminal domain cleavage but also shifted the mechanics of autocleavage, as peptide hydrolysis of the scissile bond was observed as well as methionine incorporation (Figure 29). Steric hindrance of catalytic residues or hindering the accessibility of an approaching cofactor for methionine addition are possible suggestions that promoted the less effective peptide hydrolysis reaction for CTP1L and CS74L. Mutation of valine to the similar branch-chained amino acid leucine had no effect on proteolysis and methionine incorporation was the only post-proteolytic modification. Interestingly, for CS74L V185G and V185Y a degradation band appeared around 24 kDa, similar to the protein band that CS74L T211R eventually degraded to (Figure 15, E). Together this suggested that restrictions imposed on the side-by-side dimer or from linker restrictions during autocleavage could affect endolysin stability, leading to internal degradation.

The α1 and α3 helices of the C-terminal domain are suggested to not just play a role in head-on dimerisation but also cell wall binding. Introducing a positive charge into the proposed cell wall binding site of CTP1L (M263R) had no effect on autoproteolysis but totally inhibited the lytic activity of CTP1L. Interestingly, in the head-on dimer state, helices α1 and α3 are buried within each other, upon the switch and release to the side-by-side state the helices point in opposite directions and become exposed to the external environment. In conjunction with autocleavage activation, the switch to the side-by-side state would enable the α1 and α3 helices of the C-terminal domains to bind specific epitopes within the cell wall to increase lytic efficiency. Although the methodology requires future development, pBpa was incorporated into the proposed cell wall binding site of CTP1L to try and cross-link the domain with *C. tyrobutyricum* cell wall fragments for potential ligand identification by mass spectrometry.

### *Summary and future prospects*

A novel autocleavage mechanism has been demonstrated as an intrinsic property of activation for the three *Clostridial* endolysins CD27L, CTP1L and CS74L. Autocleavage is proposed to liberate the EAD from covalent restraints to increase the lytic activity of the EAD. After holin lesion formation and cell wall entry, the C-terminal domain is hypothesised to oligomerise and facilitate the release of the EAD by an autocleavage mechanism for more effective digestion. The C-terminal domain could also function as a cell wall binding domain, to facilitate the localisation of the endolysins to the cell wall, however, this requires further investigation to establish this additional role. The mechanism of catalytic domain release, presented in this thesis, could be applied for the production of recombinant endolysins with higher potency. The engineering of controlled C-terminal domain oligomerisation for endolysins may lead to novel fusion constructs with higher specificity and activity.

# Chapter 6: Material and Methods

Contributions to the various experiments and analytical techniques from collaborating authors have been specified. Parts of the material and methods have been replicated from the submitted article "The CD27L and CTP1L endolysins targeting *Clostridia* contain a built-in trigger and release factor", Appendix page 170.

# Molecular biology methods

### *Site-directed mutagenesis*

Site-specific mutants of CD27L, CTP1L and CS74L were created by polymerase chain reaction (PCR) site-directed mutagenesis following the Quikchange method (Stratagene). Plasmids pET15b-*cd27l/ctp1l/cs74l* were used as template DNA (provided by M. Mayer (IFR)). All PCR reactions were performed using Phusion High-Fidelity DNA Polymerase (NEB). A 50 μl total volume contained 10μl 5x GC Phusion Buffer for GC-rich templates, 10 mM of dNTPS, 1.5 μM of complementary forward and reverse primers for each mutation, 0.2μM template DNA and 0.7 μL of Phusion polymerase. For site-directed mutagenesis PCR cycle 2 was used, which has a longer elongation time to copy the entire plasmid (Table 3). Amplified products were digested by *Dpn*I at 37ºC overnight (NEB) to digest template DNA, before transformation into competent *E. coli* DH5α cells (Invitrogen). Creating the dual mutants D92A_E94A and E196A_K234A required two-step whole plasmid mutagenesis using pET15b-*ctp1l_D92A* and pET15b-*ctp1l_E196A* respectively as template DNA.

| PCR steps | PCR cycle 1 | PCR cycle 2 |
|---|---|---|
| 1.Initial denaturation | 95 ºC 60 seconds | 95 ºC 60 seconds |
| 2. Denaturation | 95 ºC 35 seconds | 95 ºC 35 seconds |
| 3. Annealing* | 58 ºC 35 seconds | 58 ºC 35 seconds |
| 4. Elongation | 72 ºC 45 seconds | 72 ºC 15 minutes |
| 5. Final elongation | 72 ºC 5 minutes | 72 ºC 18 minutes |

**Table 3 Cycling steps used for PCR**. Steps 2 to 4 were cycled 30 times for PCR cycle 1 and for 18 times in PCR cycle 2. *Annealing temperature was calculated at 5-10ºC lower than the melting temperature (Tm) for complementary primer pairs or for the lowest Tm of all primers in the reaction.

***Overlap extension Polymerase Chain Reaction***

Overlap extension PCR was used to delete linker residues (E192-E194) to generate CTP1L$^{\Delta Linker}$ and to mutate the same three residues to proline to generate CTP1L$^{PPP}$. Overlap extension PCR involved combining the products of two separate PCR events (PCR1 and PCR2), which generated complimentary ends containing the required mutation, into a final PCR cycle (PCR3) where the complimentary ends are joined together to complete the construct (Appendix Figure). PCR1, 2 & 3 were performed using PCR cycle 1 (Table 3). For both CTP1L$^{\Delta Linker}$ and CTP1L$^{PPP}$ the reaction mixtures for PCR1 and PCR2 are the same as described for site-directed mutagenesis. PCR1 used primers 1 and 2 and PCR2 used primers 3 and 4. Primer 1 contained an N-terminal *Nde*I site (5'- GGAATTC<u>CATATG</u>AAGAAAATAGCAGACA -3') and primer 4 contained a C-terminal *Bam*HI site (5'- CGC<u>GGATCC</u>CTATTTTAAATTTTTAATGTA-3'). Primer 2 and 3 overlapped and contained the required mutation for insertion into the linker. CTP1L$^{\Delta Linker}$ primer 2 (5'- TAAATTTTCCACCCCCTTAATA-3'), primer 3 (5'- TATTAAGGGGGTGGAAAATTTAGTA-3' ). CTP1L$^{PPP}$ primer 2 (5'- TTTCCAC<u>AGGAGGAGG</u>CCCCTTAATATATTTTA-3') primer 3 (5'- TAAGGGG<u>CCTCCTCCT</u>GTGGAAAATTTAGTAGT-3').


PCR1 & 2 were purified by agarose gel electrophoresis and extracted using the QIAquick gel extraction kit protocol (Qiagen). This removed contaminating primers 2 and 3 and the original template DNA. For PCR3 the template DNA consisted of 4 µl each of PCR1 and PCR2 products combined with primers 1 and 4. The final PCR3 product was purified using the QIAquick PCR purification protocol (Qiagen), digested with restriction enzymes *Nde*I and *Bam*HI according to the manufacturer's instructions (NEB), purified by agarose gel electrophoresis, extracted also using the QIAquick gel extraction kit and finally ligated back into the pET15b vector which had also been digested with *Nde*I and *Bam*HI and purified.


***Cloning of the CTP1L (EAD) – CD27L (C-terminal domain) chimera (CT-CD)***

Using the same protocol of overlap extension PCR as described above. Both domains were amplified seperatley with overlapping ends, before being joined in the final PCR step. PCR1 template: pET15b-*ctp1l* with primer 1 containing an *Nde*I site (5'- GGAATTC<u>CATATG</u>ATGAAGAAAATAGCAGACATAA-3') and primer 2 which overlapped with the beginning of the CD27L C-terminal sequence (5'- TAACTCCCTCATTATTCTTAATATATTTTATAAATTCAT-3'). PCR2 template:

pET15b-cd27l with primer 3 which overlapped with the end of the CTP1L EAD (5'-TAAAATATATTAAGAATAATGAGGGAGTTAAACA-3') and primer 4 containing a *Bam*HI site (5'-CGC**GGATCC**CTATCTATTAATAAAATCCAAT-3').

### Cloning for the GFP-autocleavage Ni-NTA experiment

The *ctp1l* gene was amplified from pET15b-*ctp1l* using a forward primer containing an *Nco*I site (5'-CATG**CCATGG**ATAAGAAAATAGCAGACATA -3') and a reverse primer containing a *Bam*HI site (5'-TAGC**GGATCC**TTTTAAATTTTTAATGTAA-3') with PCR cycle 1. The PCR product was digested and ligated into the pETM11-SUMO3GFP plasmid (EMBL, Heidelberg), replacing the SUMO tag with the *ctp1l* gene. The control construct CTP1L_V195P-GFP was generated by site-directed mutagenesis as described above.

### Cloning of the CTP1L C-terminal domain with SUMO3

The chimera of SUMO3 linked to the CTP1L C-terminal domain (SUMO-CT) was generated by amplifying gene ctp1l from pET15b-*ctp1l* using a forward primer with *Bam*HI site (5'-CGC**GGATCC**GGTGGTGGTGAAGTGGAAAATTTAGTAGT-3') and a reverse primer with *Xho*I site (5'- CGG**CTCGAG**CTATTTTAAATTTTTAATGTAAT-3'), with PCR cycle 1. The PCR product was digested and ligated between the *Bam*HI and *Xho*I sites of pETM11-SUMO3GFP. The GFP gene was consequently replaced with the CTP1L C-terminal domain. The point mutation T221R was introduced into SUMO-CT by site-directed mutagenesis as described above.

### Cloning of CTP1L with a C-terminal polyhistidine tag

The gene encoding CTP1L was amplified from pET15b-*ctp1l* using a forward primer containing *Nco*I (5'-CATG**CCATGG**ATATGAAGAAAATAGCAGACA-3') and *Xho*I containing reverse primer (5'-GTAGC**CTCGAG**TTTTAAATTTTTAATGTAATC-3') using PCR cycle 1. The PCR product was digested and ligated into the pET21d vector. The CTP1L construct now contained a polyhisitidine tag attached to the C-terminal domain for detecting when C-terminal cleavage occured during expression using immunoblotting.

### DNA ligation

Inserts were ligated into their respective vectors with a 5:1 molar ratio using T4 ligase (NEB) according to manufacturer's instructions. The ligated reaction mixture was transformed into DH5α cells and plated on selective LB-agar plates.

*Agarose gel electrophoresis*

DNA fragments were mixed with 6 x DNA loading dye (Fermentas) and loaded onto agaose gels containing 0.8% (w/v) agarose dissolved in Tris-acetate-EDTA  (TAE) buffer (40 mM TRIS pH 8.5, 1 mM EDTA, 20 mM Acetic acid). A drop of ethidium bromide (0.025%) was added to each gel. Electrophoresis was performed at 50 V for 70 minutes in TAE buffer. Gels were visualised under UV light and specific bands excised using a scalpel. DNA samples were recovered using the QIAquick gel extraction kit (Qiagen), following the manufacturer's instructions.

*Transformation of E.coli competent cells*

10-50 ng of plasmid DNA or 5 μl of *dpn1* digested DNA/ ligation mixture were added directly to frozen competent cells and incubated for 30 minutes on ice. DH5α were used for DNA plasmid preparation and BL21 (DE3) or BL21 arabinose-induced (AI) cells were used as expression strains. Cells were mixed by gentle inversion. The plasmid DNA was then transformed into the cells by heat shocking at exactly 42°C for 35 seconds in a heat block before quick reinsertion into ice for 2 minutes. 80 μl SOC medium was added to the cells prior to incubation at 37°C for 1 hour. Cells were spread on selective LB-agar and incubated overnight at 37°C. All steps involving exposure of the transformation mixture to the external environment were performed in sterile conditions using a Bunsen burner to limit contact with contaminating organisms.

*Plasmid preparation and sequencing*

Plasmid DNA was obtained from transformed DH5α cells using the QIAprep® Spin Miniprep Kit (Qiagen). Plasmid DNA was quantified and checked for quality by measuring absorbance at 260 and 280 nm. A 260:280 nm ratio of approximately 1.8 guaranteed "pure" DNA. Plasmid DNA was sequenced by MWG Eurofins to check for successful mutation using T7 forward and T7 termination primers.

# Protein biochemical methods

*Standard protein expression*

Transformed BL21 (DE3) or (AI) bacterial colonies were picked to inoculate a 50 ml preculture of selective Lysogeny broth (LB) media and shaken on a rotary shaker overnight at 37°C at 200 rpm. 5-10 ml of preculture was used to inoculate 1L of selective LB media at 37°C and shaken on a rotary shaker at 200 rpm, until an $OD_{600}$ of ~ 0.6. Cultures were cooled to 21°C and then induced with 1 mM isopropyl-β-D-thio-galactopyranoside (IPTG) for overnight expression at 21°C at 200 rpm. All steps involving exposure of the expression cultures to the external environment were performed in sterile conditions using a Bunsen burner. Protein expressing cells were harvested by centrifugation (5000 g, 30 min) and the supernatant discarded. Pelleted cells were transferred to a 50 ml falcon tube and stored at -20°C.

*Selenomethionine incorporation of CTP1L and CS74L*

Colonies of BL21 (DE3) cells transformed with pET15b-*ctp1l or cs74l* were picked to inoculate a 50 ml preculture of selective LB media and shaken overnight at 37°C at 200 rpm. 4 ml of preculture was used to inoculate 500 ml of selective SelenoMethionine Medium base plus nutrient mix (Molecular Dimensions). 1 ml of a Selenomethionine stock solution (25mg.ml$^{-1}$) was added to each culture. Cultures were incubated at 37°C at 200 rpm, until the $OD_{600}$ reached over 1.0. Cultures were cooled to 21°C and induced with 1 mM IPTG for overnight expression at 21°C at 200 rpm. Protein expressing cells were harvested by centrifugation (5000 g, 30 min) and the supernatant discarded. Pelleted cells were transferred to a 50 ml falcon tube and stored at -80°C. Selenomethionine is highly toxic so all steps involving open cultures and addition of selenomethionine was performed under a fume hood.

*Human Embryonic Kidney (HEK) cell expression*

*Cloning for HEK cell expression*

A modified mammalian expression vector, pXLGsec (created by Nina Krueger, EMBL Hamburg), incorporates an N-terminal pregnancy specific glycoprotein 1 (PSG1) secretion peptide for secretion and a C-terminal polyhistidine tag for immunoblotting. Genes

encoding CTP1L and CS74L were amplified, using PCR cycle 1, from pET15b-*ctp1l* and pET15b-*cs74l* respectively using *Kpn*I containing forward primers (CTP1L: 5'- CGG<u>GGTACC</u>ATGAAG AAAATAGCAGACATAA-3', CS74L: 5'- CGG<u>GGTACC</u>ATGAAGATAGGTATTGATAT-3') and *Sac*I containing reverse primers (CTP1L: 5'- C<u>GAGCTC</u>TTTTAAATTTTTAATGTAA-3', CS74L:5'- C<u>GAGCTC</u>CTTACCTCCGTTGATGAAGTTT-3').

Both genes were ligated into pXLGsec, transformed into *E. coli* DH5α and sequenced. Small scale HEK cell expression required 4μg of plasmid DNA per well so 200μg stocks were prepared using the Plasmid *Plus* Midi Kit (Qiagen) according to the manufacturer's instructions.

### *HEK cell expression*

*Expression steps performed by N. Krueger, biophysical analysis by M. Dunne.*

HEK293T cells were maintained at 37°C in a Wheaton incubator in a 5 % $CO^2$ atmosphere. The cells were grown adherently in 30 ml Dulbecco's modified eagle medium (DMEM), containing 2 mM L-glutamine, 1x non-essential amino acids and 10% foetal calf serum (FCS) in polysterol cell culture flasks with a surface area of $175cm^2$ (T175). Cells were cultured upon reaching 80-90 % confluency before being passaged. For passaging the cells, the culture media was aspirated and the cells were washed with 10 ml PBS to remove residual FCS. To remove the cells from the flask surface 3 ml 1x Trypsin-EDTA were added and the cells were incubated at room temperature for 5-10 minutes before neutralisation with 7 ml DMEM media. For small-scale expression one fifth of the cells from a T175 flask were resuspended in 12 ml of DMEM containing 10% FCS. Two ml of cell suspension were added then to each well of a 6-well plate 24 hours prior to transfection. The plates were cultured under humidified conditions to prevent dehydration of the media.

Transient gene expression was performed by chemical transfection using the branched cationic polymer Polyethylenimine 25 kDa (PEI). PEI (1 mg.ml$^{-1}$, pH 7) was incubated at a ratio of 1:2 with the plasmid of choice for 10 minutes in one quarter (500 μl per well) of DMEM without FCS of the final expression volume for complex formation. The culture medium was changed to DMEM containing 2% FCS for small-scale expression. Expression samples were taken 48 hours after transfection by centrifugation at 100 g, 4°C. Samples of the supernatant and of the cells were analysed by SDS-PAGE and Western blot to determine expression levels and if the endolysin had been secreted.

### *Native nickel affinity protein purification*

Pellets were resuspended in lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl, 1% Triton X-100, 10 mM Imidazole, 1 mg.ml$^{-1}$ Lysozyme, 25 U.ml$^{-1}$ Benzonase nuclease) and mixed for 60-90 minutes at 4°C. The lysed cell extract was centrifuged (27,000 g, 40 minutes) and the supernatant containing His-tagged endolysin was passed over a column containing 1 ml Ni-NTA resin. The resin was then washed with 25 ml wash buffer (20 mM Tris pH 8.0, 500 mM NaCl, 20 mM Imidazole). Protein was finally eluted in 4 x 1 ml steps using elution buffer (50 mM Tris pH 8.0, 150 mM NaCl, 200 mM Imidazole). Protein samples were then further processed depending on the next analytical steps.

### *Denatured nickel affinity purification*

Denatured Ni-NTA was used to purify the full-length endolysins without also purifying the cleaved C-terminal domain. The protocol was modified from the Ni-NTA Purification System handbook (Invitrogen (2006)). Pellets were resuspended in denaturing lysis buffer (20 mM Sodium Phosphate pH7.8, 500 mM NaCl, 6 M Guanidine Hydrochloride), lysed by sonication and centrifuged (27000 g, 40 minutes). The supernatant was passed over a column containing 1 ml Ni-NTA resin. The resin was washed 2 x 10 ml with denaturing binding buffer (20 mM Sodium Phosphate pH 7.8, 500 mM NaCl, 8 M Urea). Next, 2 x 10 ml of denaturing binding buffer at pH 6.0 was washed over the beads followed by 2 x 10 ml of the same buffer adjusted to pH 5.3. Denatured protein was finally eluted with 4 x 1 ml elution buffer (20 mM Sodium Phosphate pH 4.0, 500 mM NaCl, 8 M Urea) and monitored for protein content by measuring the 280 nm absorbance. Eluted fractions containing protein were pooled for refolding. Denatured protein was slowly adding drop-wise with a pipette into 50 ml refolding buffer (50 mM Tris pH8.0, 1mM EDTA, 1M L-Arginine, 1 mM Glutathione (reduced), 0.8 mM Gluthathione (oxidised)) at 4°C in a glass beaker stirred at a high speed. After all protein was added the solution was left for 3 minutes spinning at high speed, before 1 hour of stirring at slow speed also at 4°C. Protein solution was concentrated and dialysed against 25mM TRIS pH 8.0, ± 250 mM NaCl. Refolded protein was then concentrated to 1 mg.ml$^{-1}$ for further experimentation.

## SDS-PAGE analysis

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) (Laemmli 1970) was used to estimate relative molecular weights of proteins. 10-15% acrylamide Bis-TRIS gels were prepared with an upper stacking gel to improve resolution. Protein samples were mixed 4:1 with NuPage LDS Sample 4x buffer containing 10 % β-mercaptoethanol to dissociate disulphide bonds. Samples were heated for 5 minutes at 75 °C and subjected to 10-15% SDS polyacrylamide gel electrophoresis with the Mini-Protean Tetra Electrophoresis System (Qiagen). For Coomassie Blue staining, the SDS-PAGE gel was incubated respectively in Coomassie Blue staining solution (0.125% Coomassie Blue, 45% ethanol, 10% acetic acid), destaining solution (40% ethanol, 10% acetic acid) and drying solution (2% glycerol, 20% ethanol). SDS-PAGE was used to confirm protein expression, purity and to identify C-terminal domain cleavage. The marker used was PageRuler Prestained Protein Ladder (Thermo).

## Western Blot (immunoblot) analysis

Western blot analysis was used to detect specific His-tagged proteins or protein fragments. Roti®-Mark 10-150 (Roth) his-tagged marker was used to define regular molecular weights. Protein samples were separated by size using SDS-PAGE and then transferred onto a nitrocellulose membrane in a Mini Trans-Blot® System (BioRad) for 1 hour at 100 V in transfer buffer (192 mM glycine, 25 mM TRIS-base, 20% Ethanol). After protein transfer, the nitrocellulose membrane was washed in phosphate buffer saline- Tween 0.1% (PBS-T) (10 mM sodium phosphate, 137 mM NaCl, 2.7 mM KCl, 0.1% Tween-20) 3 x 5 minutes and blocked with 5 % milk solution for 2 hours at room temperature (RT), to prevent non-specific binding of antibodies to the membrane. The milk solution was removed and the membrane washed 3 x 5 minutes in PBS-T before incubation at 4°C overnight with the monoclonal mouse Anti-His antibody (Qiagen) diluted 1:2000 in PBS-T with 0.05% milk solution. To remove non-bound primary antibody, the membrane was washed 3 x 5 minutes in PBS-T. The membrane was then incubated for 2 hours at RT with the secondary antibody goat anti-mouse-horse radish peroxidase (HRP) (Thermo) diluted 1:5000 in PBS-T with 0.05% milk solution, followed by 3 x 10 minutes washing. For detection, excess wash buffer was removed and the membrane was incubated for 1-5 minutes in 10 ml SuperSignal West Pico Chemiluminescent Substrate (Thermo). Imaging was performed using the BioRad ChemiDoc System.

### Ellman's reagent to test for N-terminal sulfhydryl on C-terminal domain

A 0.1 M stock solution of Ellman's reagent (5,5'-dithiobis-(2-nitrobenzoic acid)) (DTNB) (Thermo scientific) was prepared in 20 mM TRIS pH 7.4. 1 μl of DTNB stock was added to 50 μl aliquots of CTP1L, CS74L and CTP1L C219A. To ensure all free sulfhydryl groups could be accessed, Gu.HCl was also used at a final concentration of 4 M to protein aliquots before addition of DTNB. According to manufacturer's instructions for sulfhydryl quantification, solutions were mixed and incubated at room temperature for 15 minutes. Absorbance was measured at 412nm and visual inspection indicated reaction was successful as it turned an orange colour. Samples after DTNB treatment were analysed by SDS-PAGE gel, the C-terminal domain bands were excised and monitored by tryptic digestion and LC-MS/MS analysis. After excess Ellman's reagent had been dialysed out, samples were also in solution tryptic digested before LC-MS/MS analysis and analysed by full-length LC-MS.

### Experimentation by the addition of potential causative agents of autocleavage

Ni-NTA purified CD27L, CS74L and CTP1L were dialysed into 10 mM TRIS pH 7.4, concentrated between 1-2 mg.ml$^{-1}$ and aliquots (50-150 μl) generated to which potential causative agents were added. 1-10 mM of L-Methionine, Dithiothreitol, β-mercaptoethanol, Dimethyl sulfoxide, Glutathionine (oxidised and reduced) or Hydrogen sulfide (H2S) were all tested by adding 5-15 μl of 100 mM stocks of each additive to separate aliquots of the endolysins. The 100 mM stock solution of H$_2$S was generated by adding sodium hydrosulfide (NaHS) into degassed water. NaHS dissociates to form HS$^-$ and H$_2$S in solution. Gaseous H$_2$S can be toxic so all experiments with NaHS were performed under the fume hood. pH 6-9 were tested with different additives and salt concentrations by spiking the aliquots with 15 μl from 1 M buffer stocks of 1 M MES pH 6, 1 M TRIS pH 7, 8 and 9. NaCl concentrations 50, 100, 250 and 500 mM were tested by spiking aliquots of the endolysins from a 5 M stock of NaCl.

### Protein purification for crystallography

Proteins were purified for crystallisation by size exclusion chromatography using an Aekta liquid chromatography system (Amersham Biosciences) and S75 10/300 GL (Tricorn) column (GE Healthcare) in 20 mM HEPES, pH 7.4. For CTP1L V195P the size exclusion buffer was 50 mM TRIS pH 7.4. Protein purity was verified by SDS-PAGE. Proteins were concentrated between 5-10 mg.ml$^{-1}$ as measured by UV absorption at 280 nm.

Crystallisation trials of wildtype CTP1L and CD27L had respectively previously been performed by R. Meijers and V. Garefalaki. Protein crystals were obtained by vapour diffusion in a hanging drop steup using Limbro Plates (Hampton Research). Crystals for degraded CD27L, that ultimately only contained the C-terminal domain, were obtained from a mother liquor containing 10 – 20 % PEG 4000 and 20 mM Tris pH 8.0. Crystals of the construct containing the C-terminal domain of CD27L and an N-terminal His tag were obtained from a mother liquor of 10 % PEG 20K and 20 mM Tris pH 8.0. Protein crystals, for CTP1L full length protein with the C-terminal domain attached, were obtained from a mother liquor containing 20 mM TRIS pH 8.0 and 6 % PEG 8000. For the CTP1L V195P mutant, crystals were obtained using the same mother liquor as described for wildtype CTP1L; the time for crystals to appear took 2 weeks compared to 24 hours. CTP1L T221C and CD27L M186P, C238R were submitted for crystallisation trials by hanging drop in the high-throughput facility for Sample Preparation and Characterisation (SPC) (EMBL, Hamburg). Unfortunately, no positive crystal hits could be observed.

# Protein biophysical methods

### Data collection for CTP1L

*Performed and written by R. Meijers (EMBL, Hamburg)*

Crystals of full length CTP1L and the cleaved cell wall binding domain were harvested a few days after they occurred and flash frozen in liquid nitrogen. X-ray diffraction data were collected on the PROXIMA I beamline at the Soleil Synchrotron at 100 Kelvin using a ADSC-315 CCD detector at an X-ray wavelength of 0.98 Ångstrom. Data were processed with XDS (Kabsch 2010) and SCALA (Evans 2006). For the crystal structure of full length CTP1L, a single crystal diffracted to 1.9 Ångstrom. Molecular replacement was performed in PHASER (McCoy *et al.* 2007) with a hybrid model built from the catalytic domains of 1JFX (Cellosyl) and 2NW0 (PlyB). Only one monomer was found (Z score of 9.3) in the asymmetric unit, although the Matthews' coefficient calculation predicted the presence of two molecules (Mc=2.14) rather than one monomer (Mc=4.28). The structure of the cell wall binding domain was solved using a density modification and automatic building protocol. The hybrid model from molecular replacement solution was submitted to rigid body refinement in Refmac5 (Murshudov *et al.* 2011). The resulting phases were used to extend the molecular envelope by density modification with PARROT (Cowtan 2012). The resulting map was used for automatic building by Buccaneer (Cowtan 2012).

This was not sufficient to extend the model, and it was necessary to insert a cycle of Arpwarp (Cohen *et al.* 2008) density modification using the molecular replacement mode where individual atoms are removed and added based on the electron density. After four megacycles of Parrot/Buccaneer/Arpwarp, the contours of two cell wall binding domains became visible. After manual inspection with Coot (Emsley *et al.* 2010) and the addition of TLS parameters for the individual domains, the structure was refined with Refmac5 to an R factor of 16.3 % (Rfree = 19.8 %). The stereochemistry of the model was verified with Molprobity (Chen *et al.* 2010) and contained 97.7 % of the residues within the favoured region of the Ramachandran plot, and no residues in disallowed regions. A single crystal of the digested cell wall binding domain of CTP1L diffracted to a resolution of 1.2 Ångstrom, and the structure was solved by molecular replacement. Refinement with Refmac5 combined with manual model building with Coot resulted in an R factor of 15 % (Rfree = 18.9 %). The N-terminal cleavage site was clearly visible, and an acetyl group was built to the N-terminus of Val 195. The stereochemistry for the model is such that all residues are within the favoured region of the Ramachandran plot, according to Molprobity. A phased anomalous difference map was calculated using an X-ray data set collected at 6.5 keV to a resolution of 2.5 Ångstrom. A model of the refined CTP1L full-length crystal structure, which has almost identical cell dimensions, was used to obtain phases. Using reflections up to a resolution of 3.3 Ångstrom, PHASER was used to identify the sulphur sites with a cut-off Z value of 5.0. In this way, a site with Z value 5.3 was obtained at 2.9 Ångstrom from the N-terminus of residue Val 195 of the orphaned cell wall binding domain. The refined site had occupancy of 0.2 and a B factor of 28 $Å^2$, which was comparable to the sites for sulphurs at methionine and cysteine positions, which had occupancies between 0.2 and 0.5.

A phased anomalous difference map was calculated using an X-ray data set collected at 6.5 keV to a resolution of 2.5 Ångstrom. A model of the refined CTP1L full-length crystal structure, which has almost identical cell dimensions, was used to obtain phases. Using reflections up to a resolution of 3.3 Ångstrom, PHASER was used to identify the sulphur sites with a cut-off Z value of 5.0. In this way, a site with Z value 5.3 was obtained at 2.9 Ångstrom from the N-terminus of residue Val 195 of the orphaned cell wall binding domain. The refined site had an occupancy of 0.2 and a B factor of 28 $Å^2$, which was comparable to the sites for sulphurs at methionine and cysteine positions, which had occupancies between 0.2 and 0.5.

### *Crystal Structure determination of CD27L*

*Performed by R. Meijers (EMBL, Hamburg) and reproduced from the submitted article Appendix page 170.*

The C-terminal domain of CD27L was first solved by single-wavelength anomalous dispersion using a mercury derivative (Table I). Crystals of the CD27L C-terminal domain construct alone with an N-terminal His tag were soaked in a cryo-protecting solution containing 15 % PEG 20K, 100 mM Tris pH8, 10 % (v/v) glycerol and the derivative 1 mM of Ethyl-mercury phosphate for a few minutes prior to freezing. A data set was collected on the X12 beamline at EMBL Hamburg, which is equipped with a MAR225 CCD detector. The crystal diffracted to a resolution of 3.5 Å, and the space group was $P2_1$. All the X-ray data were indexed, merged and scaled with DENZO and Scalepack (Otwinowski *et al.* 2003). The crystal contained eight copies of the C-terminal domain in the asymmetric unit, and 8 mercury sites were identified with SHELXD (Schneider & Sheldrick 2002). Density modification was performed with PARROT, and an initial model was built with BUCCANEER (Cowtan 2012). This model was used in PHASER (McCoy *et al.* 2007) to further improve the experimental phases and to find 5 additional mercury sites after several iterations. A native X-ray data set was collected on PROXIMA I at the Soleil Synchrotron (Gif-sur-Yvette, France), using a Q315 CCD detector from ADSC. The crystal diffracted to 2.3 Å and belonged to space group $P2_12_12_1$. The initial model was then used in molecular replacement using MOLREP (Vagin & Teplyakov 2010) to identify the contents of the crystals grown from initial full length CD27L. It was determined that these crystals contained six copies of the C-terminal domain in the asymmetric unit. The structure was refined with Refmac5 (Murshudov *et al.* 2011) to an R factor of 19.8 % (Rfree = 25.6 %). The stereochemistry of the model contained 98.2 % of the residues within the favored areas of the Ramachandran plot according to Molprobity (Chen *et al.* 2010), and no residues in the disallowed regions.

### *Crystal Structure determination of the C-terminal domain of CTP1L mutant V195P*

*Performed with R. Meijers (EMBL, Hamburg) and reproduced from the submitted article Appendix page 170.*

A native X-ray data set was collected on the EMBL beamline P14 at the PETRA3 synchrotron (Hamburg, Germany) using a MAR225 CCD detector. Although the crystal probably diffracted to at least 1.5 Ångstrom resolution, we were only able to collect usable data to a resolution of 2.1 Ångstrom due to a limited detector geometry. The crystal

diffraction also suffered from ice rings, limiting the completeness of the data to 92 %. Nevertheless, it was straightforward to solve the structure of the C-terminal domain of CTP1L by molecular replacement with MOLREP (Vagin & Teplyakov 2010) using the C-terminal domain of CD27L as a search model, since there is only one copy of the molecule in the asymmetric unit. The structure was refined with Refmac5 to an R factor of 17.2 % (Rfree = 26.4 %), and the electron density is of good quality. The stereochemistry of the refined model contained 98.8 % of the residues within the favored areas of the Ramachandran plot according to Molprobity (Chen *et al.* 2010), and no residues in the disallowed regions.

### *MALDI-TOF analysis*

*Performed by the SPC facility (EMBL, Hamburg).* Protein samples were dissolved in 1:1 or 1:2 ratio with Trifluoroacetic acid (TFA)(0.1% v/v TFA: $H_2O$) and 0.5 μl was spotted onto a MALDI plate. 0.5 μl of a matrix solution α-Cyano-4-hydroxycinnamic acid in Acetonitrile (ACN):H2O:TFA (50:50:0.1) was mixed with each sample and left to crystallise. After crystallisation of the matrix:sample solution the samples were analysed with a Voyager DE-STR MALDI- TOF (Applied Biosystems). Spectra for full-length MW determination were obtained between 8000 to 50,000 m/z.

### *Intact Protein Sample Analysis by LC-MS*

*Performed by S. Leicht (EMBL, Heidelberg) with parts reproduced from the submitted article Appendix page 170.* Fresh Ni-NTA proteins samples were dialysed overnight into 20 mM TRIS, pH 7.4 and concentrated to 2 mg.ml$^{-1}$ and sent for analysis at the proteomics core facility (PCF) (EMBL) by Stefan Leicht. Samples were acidified using 0.1% formic acid solution and transferred to vials prior to LC-MS analysis. Desalting and protein separation were carried out using an UltiMate 3000 RSLCnano system (Thermo Scientific) fitted with a trapping (Acclaim PepMap 100 $C_{18}$, 3 μm, 75 μm × 20 mm) and an analytical column (Acclaim PepMap RSLC $C_{18}$, 2 μm, 75 μm × 150 mm). The column was maintained at constant temperature (35 °C) throughout. The outlet of the column was coupled directly to a Q Exactive mass spectrometer (Thermo Scientific).

Solvent A was water, 0.1% formic acid and solvent B was acetonitrile, 0.1% formic acid. The samples (around 4 ng) were loaded onto the trapping column and desalted for 5 minutes at a flow rate of 6 µl/min, 100% A. The proteins were then eluted from the column

with a constant flow of 0.3 µl/min. During the elution step, the percentage of solvent B increased in a linear fashion from 4% to 85% in 20 minutes. Data were acquired in positive continuum mode, over a mass range 400-3000 *m/z* and a resolution of 70,000. Spectra across the protein chromatographic peak(s) were summed and intact mass was calculated using the Xtract algorithm (Thermo Scientific) to give the zero charge deconvoluted molecular weight.

### *Sample preparation for tryptic digest mass spectrometry*

After SDS-PAGE, protein bands were excised and transferred into an Eppendorf tube (1.5 ml) for analysis by Stefan Leicht (EMBL, Heidelberg). To stop keratin contamination and cross contamination of samples, all steps were performed under sterile conditions with separate scalpels used for each protein band excised.

### *Tryptic digestion linked with LC-MS/MS*

*Performed by S. Leicht (EMBL, Heidelberg) and reproduced from the submitted article Appendix page 170.* The tryptic in-gel-digest was performed based on the protocol from Rosenfeld (Rosenfeld *et al.* 1992). Modified procine trypsin, sequencing grade, from Promega (Madison, WI, USA), was used. The collected peptide samples were dried and dissolved in 3% acetonitrile and 0.1% formic acid. Peptides were separated using an UltiMate 3000 RSLCnano system (Thermo Scientific) fitted with a trapping (Acclaim PepMap 100 $C_{18}$, 3 µm, 75 µm × 20 mm) and an analytical column (Acclaim PepMap RSLC $C_{18}$, 2 µm, 75 µm × 150 mm). The column was maintained at constant temperature (35 °C) throughout. The outlet of the column was coupled directly to a Q Exactive mass spectrometer (Thermo Scientific) via a PicoTip emitter (FS360-20-10-D, coating 1P-4P, New Objective).

Peptides were eluted through the analytical column at a constant flow of 0.3 µl per min and total runtime was 30 min. Full scan mass spectrometry spectra with mass range 300–2,000 mass-to-charge ratio (*m/z*) were acquired with a resolution of 70,000. The most intense ions (up to 15) from the full-scan mass spectrometry were selected for fragmentation. MS2 was carried out with a resolution of 17,500 with a fixed first mass at 100 m/z and normalised collision energy of 25.

The raw data were processed using MaxQuant (Cox & Mann 2008) and MS/MS spectra were searched using the MASCOT search engine (version 2.2.07, Matrix Science) against

a self-created database containing around 3,300 entries. Enzyme specificity was set to trypsin, and a maximum of three missed cleavages were allowed. Carbamidomethylation was set as fixed modification while methionine oxidation and acetylation on K and N-term were included as variable modifications. The search was performed with an initial mass tolerance of 10 ppm for the precursor ion and 0.1 Da for the MS/MS spectra and a minimum ion score of 20 was applied.

### *SEC-RALS/RI/UV molecular weight determination*

*Performed with the assistance of Cy Jeffries (EMBL, Hamburg) with parts reproduced from the submitted article Appendix page 170.* Molecular weight estimates were evaluated using size-exclusion chromatography in combination with right-angle light scattering (RALS), refractive index (RI) and UV ($l_{280\ nm}$) measurements (Malvern Instruments Viscotek, RALS/RI/UV 305 TDA detector equipped with a 670 nm laser diode). All measurements were performed at room temperature. 50-75 µl samples were separately injected at concentrations between 6-8 mg/ml$^{-1}$ (with the exception of CD27L W207A: 1.2 and 3.7 mg.ml$^{-1}$) onto a GE-Healthcare Tricorn S75 10/300 GL column equilibrated in 20 mM HEPES pH7.4, 500 mM NaCl at a flow rate of 0.4 ml.min$^{-1}$. The molecular weight (*MW*) of each species eluting from the SEC column were assessed using concentration (*c*) measurements derived from base-line corrected RI or UV measurements in combination with base-line corrected RALS intensities calibrated against a bovine serum albumin narrow (monomeric) standard (RALS = $c(dn/dc)^2.MW.k_{RALS}$; RI = $c(dn/dc)k_{RI}$ and; UV = $cek_{UV}$, where dn/dc is the refractive index increment of unmodified protein, 0.185 mL.g$^{-1}$, $k_{RI}$, $k_{UV}$ and $k_{RALS}$ are the TDA instrument calibration constants relative to a BSA and *e* the $l_{280\ nm}$ $E_{0.1\%}$ extinction coefficient of each protein in mg.mL$^{-1}$). The *MW* correlations across the selected range of each CD27L elution peak and the final *MW* estimates quoted in the text were calculated using OmniSEC Software (Malvern Instruments).

### *SAXS data collection and shape determination*

*Performed with H. Mertens (EMBL, Hamburg) reproduced from the submitted article Appendix page 170.* Synchrotron radiation X-ray scattering data were collected on the EMBL X33 and P12 beamlines of the storage rings DORIS III and PETRA III (DESY, Hamburg), respectively, using PILATUS 1M and 2M pixel detectors (DECTRIS, Switzerland). For the wild-type CD27L, CTP1L and CS74L data were acquired at X33, with 8 frames of 15 s exposure time collected. Samples were measured in a temperature-

controlled cell at 10°C. CD27L was measured in a buffer of 20 mM HEPES 150 mM NaCl pH 7.4. For CTP1L and CS74L samples were measured in 20 mM HEPES ± 500 mM NaCl pH 7.4. Protein concentrations were 0.9 – 4.0 mg/mL for all three samples. The sample-to-detector distance was 2.7 m, covering a range of momentum transfer $0.01 \leq s \geq 0.6$ Å$^{-1}$ ($s = 4\pi \sin\theta / \lambda$, where $2\theta$ is the scattering angle, and $\lambda = 1.54$ Å is the X-ray wavelength). For the C238R mutant data were acquired at P12, with 20 frames of 0.05 s exposure time collected. C238R was measured while flowing through a temperature-controlled capillary at 10°C in 20 mM Tris buffer, 500 mM NaCl pH 7.4 at protein concentrations of 1.0 – 8.5 mg/mL. The sample-to-detector distance was 3.1 m, covering a range of momentum transfer $0.008 \leq s \geq 0.458$ Å$^{-1}$ ($s = 4\pi \sin\theta / \lambda$, where $2\theta$ is the scattering angle, and $\lambda = 1.24$ Å is the X-ray wavelength). Based on comparison of successive frames, no detectable radiation damage was observed. Data from the detector were normalised to the transmitted beam intensity, averaged and the scattering of buffer solutions subtracted. The difference curves were scaled for solute concentration and the 1.0 mg.ml$^{-1}$ (low-$s$) and 8.4 mg.ml$^{-1}$ (high-$s$) data sets merged for modeling. All data manipulations were performed using PRIMUS (Konarev *et al.* 2003).

The forward scattering $I(0)$ and radius of gyration, $Rg$ were determined from Guinier analysis (Guinier n.d.), assuming that at very small angles ($s \leq 1.3/Rg$ ) the intensity is represented as $I(s)=I(0)\exp(-(sRg)^2/3)$). These parameters were also estimated from the full scattering curves using the indirect Fourier transform method implemented in the program GNOM (Semenyuk & Svergun 1991), along with the distance distribution function $p(r)$ and the maximum particle dimension $D_{max}$. Molecular masses (MMs) of solutes were estimated from SAXS data by comparing the extrapolated forward scattering with that of a reference solution of bovine serum albumin, and also from the hydrated-particle/Porod volume $V_p$, where molecular mass is estimated as 0.625 times $V_p$.

### *Ab initio Shape Determination and Molecular Modelling*

*Performed with H. Mertens (EMBL, Hamburg) reproduced from the submitted article Appendix page 170.* Low-resolution shape envelopes for all constructs were determined using the *ab initio* bead-modelling program DAMMIF (Franke & Svergun 2009), using both P1 and P2 symmetry for all constructs. The results of 10 independent DAMMIF runs were analysed using the program DAMAVER (Volkov & Svergun 2003) to identify the most representative/typical models. Modeling using P2 symmetry was only attempted

following the identification of excluded solvent volumes, $V_{ex}$ in models generated in P1 (slow mode) consistent with that expected for dimers (see Table 2).

Molecular modelling was conducted using, as rigid bodies and where appropriate, the crystal structures of the catalytic (PDB: 3QAY, (Mayer *et al.* 2011a) and the C-terminal domain of CD27L determined in this study. Rigid-body models were generated using the program CORAL (Petoukhov *et al.* 2012) and 10 independent runs assessed for convergence with DAMAVER. For CTP1L and CS74L the CTP1L crystal structure, in a head-on configuration with its symmetry mate, was used for modeling. The ddditional fitting of PDB files to the SAXS data was conducted using CRYSOL (Svergun *et al.* 1995).

### *Oligomeric equilibrium analysis*

*Performed with H. Mertens (EMBL, Hamburg) parts reproduced from the submitted article Appendix page 170.* Using the program OLIGOMER (Petoukhov *et al.* 2012), the SAXS data of the individual constructs were used to model potential multicomponent mixtures of species in solution. Form factors of input PDB files were calculated using the program FFMAKER (Petoukhov *et al.* 2012). Input PDBs for CD27L and C238R were generated based on the CORAL rigid body described above. Input PDBs for CTP1L and CS74L were based on the solved CTP1L crystal structure. Form factors were calculated for individual domains and substructures of the intact PDB files to represent products of autolysis and averaged. This averaging was performed, as the identity of the exact solution composition of these lysis products could not be established. Volume fractions corresponding to each component were determined by OLIGOMER using a non-negative least squares procedure.

### *Turbidity reduction assays*

*Perfomed by M. Mayer (IFR, Norwich). I performed the turbidity assays referred to in Figure 22 with the assistance of M. Mayer.* Cells of *C. difficile* strain NCTC 11204 or *C. tyrobutyricum* strain NCIMB 9582 were grown, harvested and assayed as described previously Mayer et al 2008 and 2010. Lysis assays were performed on freshly harvested cells in 300 µl volumes with 10 µg or 30 µg Ni-NTA-purified protein or elution buffer (control). Results are the mean of duplicate assays ± standard deviation. For controls Lysozyme (1000U) and Mutanolysin (50U) were used.

*Measuring the effect of environmental conditions and additives on endolysin activity using GFP fluorescence*

500 µl of Ni-NTA agarose beads (Qiagen) contained in poly prep chromatography columns (Biorad) were pre-equilibrated in 20 mM Tris buffer, pH 7.4, before addition of 2 mg of CTP1L-GFP or CTP1L_V195P-GFP. Protein bound beads were washed with 1.5 ml of 20 mM Tris buffer, pH 7.4 to remove unbound protein. This was followed by addition of 5 ml of experimental buffer, differing in pH and NaCl concentration (50 mM Tris pH 9.0, 50 mM Tris pH 8.0, 50 mM Tris pH 7.4, 50 mM MES pH 6.0, 50 mM Tris pH 7.4 500 mM NaCl) with or without 100 mM NaHS. The flowthrough was collected in a 50 ml beaker and a continuous flow of buffer was created over the Ni-NTA beads via pipetting the flowthrough collected in the 50 ml beaker (bottom reservoir) back to the top reservoir of buffer solution (Figure 26). Beads were continuously washed with experimental buffer. Over 2 hours, at specified time points, 100 µl of the flow-through was obtained from the collection well for fluorescence measurements. GFP fluorescence was measured with a Tecan Infinite M1000 plate reader using 96-well flat bottom black microplates (Greiner) containing 100 µl sample volume. Mode of determination was Fluorescence Top Reading (50 flashes at 400 Hz, excitation: $395 \pm 5$ nm, emission: $510 \pm 5$ nm). The fluorescence produced from the autocleaved C-terminal domain-GFP product was plotted as a function of time for each condition. Using the software GraphPad Prism version 5.04, the kinetics toward equilibrium was fitted to one phase exponential association, as described in Figure 26, where Ymax represents the fluorescence reading at equilibrium, Y the fluorescence reading of cleaved cell wall binding domain-GFP at time X and K is the observed cleavage rate constant. All Y values are the means of duplicate samples ± standard deviations. The exponential equation could only be fitted for CTP1l-GFP wildtype. This equation normally describes the association kinetics between ligand and receptor but can also describe the dissociation of the C-terminal domain from the full-length monomer.

*CD spectropolarimetry measurements on CD27L and CTP1L endolysins and mutants*
*Reproduced from the submitted article Appendix page 170.*
Proteins were purified by Ni-NTA affinity purification as described above and by size exclusion chromatography using an Aekta liquid chromatography system (Amersham Biosciences) and S75 10/300 GL (Tricorn) column (GE Healthcare). Size exclusion buffer for CD27L and mutants was 20 mM HEPES, pH 7.4 with 500 mM NaCl and for CTP1L and mutants 20 mM HEPES, pH 7.4. Purified protein was eluted and dialysed against 20

mM HEPES pH 7.4 and concentration was measured by UV absorption at 280 nm. Spectra were recorded at 20°C on a Chirascan CD Spectrometer (Applied Photophysics), between 200 and 260 nm in a 0.1 cm cuvette with machine settings as follows: 1 nm bandwidth, 0.5-sec response, and 0.5-nm data pitch. Spectra were background-subtracted and converted into mean residue ellipticity. Each curve represents the mean of three separate measurements.

# Material and methods for the identification of endolysin cell wall ligands

***Sub cloning and expression of* p-*benzoyl-L-phenylalanine incorporated into CTP1L constructs***

*Parts reproduced from the submitted article Appendix page 170.*

CTP1L was amplified from *ctp1l*-pET15b using a forward primer with NcoI site (5'-CATGCCATGGATATGAAGAAAATAGCAGACA-3') and reverse primer with XhoI site (5'- GTAGCCTCGAGTTTTAAATTTTTAATGTAATC-3') and ligated into the pET21d vector. The *amber* codon (TAG) was previously incorporated at position Y212 or Y260 using site-directed mutagenesis to generate pET21d-*Y260TAG* and pET21d-*Y212TAG*. Sub cloning into pET21d introduced a C-terminal polyhistidine-tag, which selective purified only full-length proteins that have *p*Bpa incorporated. C-terminal constructs Y260TAG and Y212TAG were generated by amplifying the C-terminal domains of *Y212TAG*-pET21d and *Y260TAG*-pET21d and reinserting the constructs back into the pET21d vector to create pET21d-c-*termY260TAG* and pET21d-c-*termY212TAG*. As a control, the wild-type C-terminal domain was sub cloned into pET21d with no amber stop codon incorporated. Following the principal method outlined by Farrell *et al.* (Ian S. Farrell *et al.* 2005), the photo-activated amino acid *p*-benzoyl-L-phenylalanine (*p*Bpa) (BACHEM) was incorporated into the full-length CTP1L endolysin and the truncated C-terminal domain of CTP1L.

*E.coli* BL21 (AI) cells were transformed with pEVOL-*p*BPA (aminoacyl-tRNA synthetase/suppressor tRNA) and either of the plasmids encoding an *amber* codon containing construct, *Y212TAG*-pET21d, C-term*Y212TAG*-pET21d or *C-termY260TAG*-pET21d. Cells were grown in 500 ml Lysogeny broth (LB) media supplemented with 1 mM *p*Bpa in the presence of ampicillin and chloramphenicol. When an $OD_{600}$ of 0.6 had

been reached the cultures were induced with Arabinose (final concentration 0.02%) and expressed at 21°C overnight. Cells were harvested by centrifugation (5000 g, 30 min) and the supernatant discarded.

### *Photo-cross-linking full length CTP1L and the C-terminal domain of CTP1L without cell wall material present*

*Reproduced from the submitted article Appendix page 170.*

Full-length Y212*pBpa* was Ni-NTA purified as described above and dialysed into 25 mM TRIS, pH 7.4 and concentrated to 0.5 mg.ml$^{-1}$, as measured by UV absorption at 280 nm. A 1 ml aliquot of Y212*pBpa* was pipetted into a single well of a 24-well clear polystyrene plate, typically used for protein crystallography. Clear polystyrene allowed efficient passage of UV light at 365 nm. The lid was kept on to prevent sample evaporation, and placed inside an RPR-100 UV reactor equipped with 350-370 nm bulbs (Rayonet). The reactor was kept at 4°C with the combined cooling fan on. The samples were exposed to UV light for 30 minute intervals. During each interval the solutions were stirred by gentle pipetting and SDS-PAGE samples taken for each time point. As a control, samples were also taken every 30 minutes from a parallel sample of full-length Y212*p*BPA kept at 4°C separate from the reactor in the dark with no UV exposure. Cross-linking was analysed by SDS-PAGE by comparing the pre-UV and post-UV exposed samples.

Photo-cross-linking was also performed for the C-terminal domain constructs. Ni-NTA purified C-termY212*p*Bpa, C-termY260*p*Bpa and the wild-type C-terminal domain were dialysed into 25mM TRIS, pH 7.4 and concentrated to 2 mg.ml$^{-1}$ as measured by UV absorption at 280 nm. Following the same protocol as described above for the full-length Y212*p*Bpa, 500 μl aliquots of each protein were pipetted into separate wells of a 24 well clear polystyrene plate and exposed to UV light for a total of 120 minutes with stirring of the samples at 30 minute intervals. As a control SDS-PAGE samples were taken from parallel aliquots of each protein kept at 4°C in the dark with no UV exposure.

### *Photo-cross-linking of C-terminal domains in the presence of cell wall material*

C-termY212*p*Bpa, C-termY260*p*Bpa and wildtype C-term (25mM TRIS, pH 7.4) were concentrated to 1.5 mg.ml$^{-1}$. To 300 μl aliquots varying amounts of the different cell wall preparations (1 and 2) were added (3 – 30 μl) and mixed by gentle vortexing and pipetting

fro 30 minutes. Aliquots were then split into separate wells of a 24 well clear polystyrene plate and exposed to UV light for 90 minutes with stirring of the samples at 10 minute intervals. SDS-PAGE samples were taken before and after UV exposure. As a control SDS-PAGE samples were taken from parallel aliquots with or without CWM of each protein kept at 4°C in the dark with no UV exposure. Cross-linked protein-CWM complexes were affinity purified using Ni-NTA. A total of 100 ul Ni-NTA resin was used for each sample and followed the same standard protocol as described above. Affinity purification removed non-cross-linked CWM. Samples were then desalting into Tris buffer 20 mM, pH 7.4 to remove the imidazole and stored at -20°C for MS analysis.

### Cell wall growth and extraction

*Performed by M. Mayer (IFR)*. Crude *C.tyrobutyricum* cell wall material (CWM) preparation was modified from Pritchard *et al.* (2004). A 500 ml culture of BHI broth was inoculated with an overnight culture of *C. tyrobutyricum* strain NCIMB 9582 at 2% and grown to $OD_{600}$ 0.5. The culture was centrifuged for 20 minutes, 4000 g at 4°C, and the pellets resuspended in 3 ml ice-cold growth media. Resuspensions were then boiled at 100°C for ten minutes to inactivate autolysins and finally centrifuged for 7 minutes at 8000 g at 4°C for storage at -20°C.

### Cell wall material fragmentation

This formed the basic CWM that was then split between different preparative steps. (1) CWM was resuspended in TRIS pH 7 and extensively bead beaten for 5 x 10 second intervals with 5-10 minutes on ice in between. Beads were removed and suspension centrifuged for 5 minutes 5000 g at 4°C to remove unbroken cells. Supernatant transferred to fresh tube and centrifuged for 45 minutes at 27000 g at 4°C to recover cell walls (2) CWM was resuspended in PBS containing $1mg.ml^{-1}$ lysozyme and incubated with gentle shaking for 30 minutes at 37°C then centrifuged for 5 minutes 5000 g at 4°C to remove unbroken cells to separate the digested CWM, which was then stored at -20°C. Excess lysozyme would be removed during Ni-NTA purification of the potentially cross-linked C-terminal domain : CWM complexes after UV exposure. The final CWM solutions for all steps were composed of a semi translucent suspension.

*MALDI-TOF-MS analysis*

*Performed with the core facility for mass spectrometric protein analysis, Prof. Dr. H. Schlüter/Dr. D. Hildebrand (UKE).* Trypsin digested peptide samples were evaporated to complete dryness using a cooling trap (4-16K; Christ) equipped vacuum centrifuge (RCF10, Thermo Scientific and dissolved in 10µl 3% acetonitrile (ACN; Merck) and 0.1% formic acid (v/v) (Fluka, Sigma)/HPLC-grade Water (Merck). For MALDI-TOF-MS measurement 0.8µl of the peptide samples were pipetted onto the MALDI-target plate (Bruker, AnchorChip Standardtarget) and mixed with 0.8µl of the matrix 2,5-dihydroxybenzoic acid (DHB; 30mg/ml dissolved in ACN/0.1% (v/v), Trifluoroacetic acid (TFA) in HPLC-grade Water (ratio 1:1). After crystallisation of the matrix:sample solution the samples were analysed by MALDI-TOF-MS (Reflex IV, Bruker). Applied instrument settings were the following: Lasershots per single spectrum=100, number of summed up single spectra=100, Laser intensity 50-50%, Laser Frequency 9Hz, Ion deflection from 0-500 Da, Pulsed ion extraction time=400ns, Ion source 1 voltage=20kV, Ion source 2 voltage=16,35kV. Peptide fingerprinting was performed in a mass range between 500 to 5000 m/z. Spectra were evaluated with the Software Bruker Flex Control and Flex Analysis. Manually the signal m/z values between C-termY212$p$BPA and C-termY260$p$BPA ± CWM were compared. For identification of the cross-linked peptides by C-termY212$p$BPA, *in silico* cross-linked tryptic peptides were calculated and also searched for manually.

*ESI-Q-TOF-MS/MS-Analysis*

*Performed with the core facility for mass spectrometric protein analysis, Prof. Dr. H. Schlüter/Dr. D. Hildebrand (UKE).* ESI-Q-TOF analysis was performed on a Q-TOF-2 electrospray mass spectrometer (Waters). The experiments were carried out in the positive ion mode (ES (+)). Nanoflow capillaries were drawn and coated with gold (done in-house). The capillaries were loaded with a 2-µL sample of the desalted tryptic peptides, and low pressure nitrogen gas was used to initiate the flow through the capillary. The capillary tip was set to a potential of 0.63 kV, and the cone voltage was set to 34 V. The source temperature was 25 °C. For MS/MS experiments using collision-induced dissociation (CID) experiments, ions were selected within a precursor mass window of ± 1 Da in the quadrupole analyser and fragmented in the collision cell using a collision gas (Ar) and collision energies of 27 to 30 eV. The cycle time was about 1.1 s with a scan duration of 1 s. Raw data were acquired and analysed using the software, MassLynx 3.5 (Micromass, Manchester, UK). MS/MS spectra were automatically processed with the micromass

charge state de-encryption algorithm (MaxEnt-3) and signal anlaysis including aminosequence assignment was performed manually.

# Software used for thesis documentation

The crystal structure analysis of CD27L and CTP1L was performed using PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.)

The thesis was written in Microsoft Word and all referencing performed with Zotero (www.zotero.org). Chemical drawings were produced using ACD/ChemSketch (version 12.01). All Figures were created using the open source vector graphics editor Inkscape (www.inkscape.org).

# References

Abedon, Stephen T. 2011. "Lysis from without." *Bacteriophage* 1 (1): 46–49. doi:10.4161/bact.1.1.13980.

Ackermann, H-W. 2007. "5500 Phages Examined in the Electron Microscope." *Archives of Virology* 152 (2): 227–43. doi:10.1007/s00705-006-0849-1.

Artman, M. 1956. "The Production of Hydrogen Sulphide from Thiosulphate by Escherichia Coli." *Journal of General Microbiology* 14 (2): 315–22.

Bergère, J L, H Hayes, and J Commissaire. 1986. "Major Protein Components in the Cell Envelope of Clostridium Tyrobutyricum." *Annales de l'Institut Pasteur. Microbiology* 137B (3): 271–82.

Biswas, Raja, Raul E. Martinez, Nadine Göhring, Martin Schlag, Michaele Josten, Guoqing Xia, Florian Hegler, Cordula Gekeler, Anne-Kathrin Gleske, Friedrich Götz, Hans-Georg Sahl, Andreas Kappler, and Andreas Peschel. 2012. "Proton-Binding Capacity of Staphylococcus Aureus Wall Teichoic Acid and Its Role in Controlling Autolysin Activity." *PLoS ONE* 7 (7): e41415. doi:10.1371/journal.pone.0041415.

Bradbury, Mark, Paul Greenfield, David Midgley, Dongmei Li, Nai Tran-Dinh, Frank Vriesekoop, and Janelle L. Brown. 2012. "Draft Genome Sequence of Clostridium Sporogenes PA 3679, the Common Nontoxigenic Surrogate for Proteolytic Clostridium Botulinum." *Journal of Bacteriology* 194 (6): 1631–32. doi:10.1128/JB.06765-11.

Braun, V., and K. Hantke. 1977. "Bacterial Receptors for Phages and Colicins as Constituents of Specific Transport Systems." In *Microbial Interactions*, edited by J. L. Reissig, 99–137. Receptors and Recognition 3. Springer US.

Brennan, P J. 2003. "Structure, Function, and Biogenesis of the Cell Wall of Mycobacterium Tuberculosis." *Tuberculosis (Edinburgh, Scotland)* 83 (1-3): 91–97.

Briers, Yves, Liesbet M Peeters, Guido Volckaert, and Rob Lavigne. 2011. "The Lysis Cassette of Bacteriophage ?KMV Encodes a Signal-Arrest-Release Endolysin and a Pinholin." *Bacteriophage* 1 (1): 25–30. doi:10.4161/bact.1.1.14868.

Briers, Yves, Mathias Schmelcher, Martin J Loessner, Jelle Hendrix, Yves Engelborghs, Guido Volckaert, and Rob Lavigne. 2009. "The High-Affinity Peptidoglycan Binding Domain of Pseudomonas Phage Endolysin KZ144." *Biochemical and*

*Biophysical Research Communications* 383 (2): 187–91. doi:10.1016/j.bbrc.2009.03.161.

Briers, Yves, Guido Volckaert, Anneleen Cornelissen, Stijn Lagaert, Chris W Michiels, Kirsten Hertveldt, and Rob Lavigne. 2007. "Muralytic Activity and Modular Structure of the Endolysins of Pseudomonas Aeruginosa Bacteriophages phiKZ and EL." *Molecular Microbiology* 65 (5): 1334–44. doi:10.1111/j.1365-2958.2007.05870.x.

Buey, Rubén M, Begoña Monterroso, Margarita Menéndez, Greg Diakun, Pablo Chacón, Juan Antonio Hermoso, and J Fernando Díaz. 2007. "Insights into Molecular Plasticity of Choline Binding Proteins (pneumococcal Surface Proteins) by SAXS." *Journal of Molecular Biology* 365 (2): 411–24. doi:10.1016/j.jmb.2006.09.091.

Buist, Girbe, Anton Steen, Jan Kok, and Oscar P. Kuipers. 2008. "LysM, a Widely Distributed Protein Motif for Binding to (peptido)glycans." *Molecular Microbiology* 68 (4): 838–47. doi:10.1111/j.1365-2958.2008.06211.x.

Buller, Andrew R., Michael F. Freeman, Nathan T. Wright, Joel F. Schildbach, and Craig A. Townsend. 2012. "Insights into Cis-Autoproteolysis Reveal a Reactive State Formed through Conformational Rearrangement." *Proceedings of the National Academy of Sciences of the United States of America* 109 (7): 2308–13. doi:10.1073/pnas.1113633109.

Bustamante, Noemi, Nuria E. Campillo, Ernesto Garcia, Cristina Gallego, Benet Pera, Gregory P. Diakun, Jose Luis Saiz, Pedro Garcia, J. Fernando Diaz, and Margarita Menendez. 2010. "Cpl-7, a Lysozyme Encoded by a Pneumococcal Bacteriophage with a Novel Cell Wall-Binding Motif." *Journal of Biological Chemistry* 285 (43): 33184–96. doi:10.1074/jbc.M110.154559.

Calamita, Heather G, and R J Doyle. 2002. "Regulation of Autolysins in Teichuronic Acid-Containing Bacillus Subtilis Cells." *Molecular Microbiology* 44 (3): 601–6.

Canchaya, Carlos, Ghislain Fournous, Sandra Chibani-Chennoufi, Marie Lise Dillmann, and Harald Brüssow. 2003. "Phage as Agents of Lateral Gene Transfer." *Current Opinion in Microbiology* 6 (4): 417–24.

Casjens, Sherwood R. 2005. "Comparative Genomics and Evolution of the Tailed-Bacteriophages." *Current Opinion in Microbiology* 8 (4): 451–58. doi:10.1016/j.mib.2005.06.014.

Catalao, Maria Joao, Filipa Gil, Jose Moniz-Pereira, and Madalena Pimentel. 2010. "The Mycobacteriophage Ms6 Encodes a Chaperone-like Protein Involved in the Endolysin Delivery to the Peptidoglycan." *Molecular Microbiology* 77 (3): 672–86. doi:10.1111/j.1365-2958.2010.07239.x.

Catalão, Maria João, Filipa Gil, José Moniz-Pereira, Carlos São-José, and Madalena Pimentel. 2013. "Diversity in Bacterial Lysis Systems: Bacteriophages Show the Way." *FEMS Microbiology Reviews* 37 (4): 554–71. doi:10.1111/1574-6976.12006.

Changeux, Jean-Pierre. 2012. "Allostery and the Monod-Wyman-Changeux Model After 50 Years." *Annual Review of Biophysics* 41 (1): 103–33. doi:10.1146/annurev-biophys-050511-102222.

Chen, Vincent B, W Bryan Arendall 3rd, Jeffrey J Headd, Daniel A Keedy, Robert M Immormino, Gary J Kapral, Laura W Murray, Jane S Richardson, and David C Richardson. 2010. "MolProbity: All-Atom Structure Validation for Macromolecular Crystallography." *Acta Crystallographica. Section D, Biological Crystallography* 66 (Pt 1): 12–21. doi:10.1107/S0907444909042073.

Chin, Jason W., Andrew B. Martin, David S. King, Lei Wang, and Peter G. Schultz. 2002. "Addition of a Photocrosslinking Amino Acid to the Genetic Code of Escherichia Coli." *Proceedings of the National Academy of Sciences* 99 (17): 11020 –11024. doi:10.1073/pnas.172226299.

Chodosh, Lewis A. 2001. "UV Crosslinking of Proteins to Nucleic Acids." In *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc.

Cohen, Serge X, Marouane Ben Jelloul, Fei Long, Alexei Vagin, Puck Knipscheer, Joyce Lebbink, Titia K Sixma, Victor S Lamzin, Garib N Murshudov, and Anastassis Perrakis. 2008. "ARP/wARP and Molecular Replacement: The next Generation." *Acta Crystallographica. Section D, Biological Crystallography* 64 (Pt 1): 49–60. doi:10.1107/S0907444907047580.

Cowtan, Kevin. 2012. "Completion of Autobuilt Protein Models Using a Database of Protein Fragments." *Acta Crystallographica. Section D, Biological Crystallography* 68 (Pt 4): 328–35. doi:10.1107/S0907444911039655.

Cox, Jürgen, and Matthias Mann. 2008. "MaxQuant Enables High Peptide Identification Rates, Individualized P.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification." *Nature Biotechnology* 26 (12): 1367–72. doi:10.1038/nbt.1511.

Crooks, Gavin E, Gary Hon, John-Marc Chandonia, and Steven E Brenner. 2004. "WebLogo: A Sequence Logo Generator." *Genome Research* 14 (6): 1188–90. doi:10.1101/gr.849004.

Czyzewski, Bryan K, and Da-Neng Wang. 2012. "Identification and Characterization of a Bacterial Hydrosulphide Ion Channel." *Nature* 483 (7390): 494–97. doi:10.1038/nature10881.

D' Herelle, Félix. 1917. "An Invisible Microbe That Is Antagonistic to the Dysentery Bacillus." *Comptes Rendus Acad. Sci. Paris* 165: 373–75.

Damjanovic, Marlen, Arun S. Kharat, Alice Eberhardt, Alexander Tomasz, and Waldemar Vollmer. 2007. "The Essential tacF Gene Is Responsible for the Choline-Dependent Growth Phenotype of Streptococcus Pneumoniae." *Journal of Bacteriology* 189 (19): 7105–11. doi:10.1128/JB.00681-07.

Davies, Julian. 2006. "Where Have  All the Antibiotics Gone?" *The Canadian Journal of Infectious Diseases & Medical Microbiology* 17 (5): 287–90.

Deresinski, Stan. 2009. "Bacteriophage Therapy: Exploiting Smaller Fleas." *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 48 (8): 1096–1101. doi:10.1086/597405.

Dewey, Jill S., Christos G. Savva, Rebecca L. White, Stanislav Vitha, Andreas Holzenburg, and Ry Young. 2010. "Micron-Scale Holes Terminate the Phage Infection Cycle." *Proceedings of the National Academy of Sciences* 107 (5): 2219–23. doi:10.1073/pnas.0914030107.

Donovan, David M., and Juli Foster-Frey. 2008. "LambdaSa2 Prophage Endolysin Requires Cpl-7-Binding Domains and Amidase-5 Domain for Antimicrobial Lysis of Streptococci." *FEMS Microbiology Letters* 287 (1): 22–33. doi:10.1111/j.1574-6968.2008.01287.x.

Ekici, Ozlem Dogan, Mark Paetzel, and Ross E. Dalbey. 2008. "Unconventional Serine Proteases: Variations on the Catalytic Ser/His/Asp Triad Configuration." *Protein Science : A Publication of the Protein Society* 17 (12): 2023–37. doi:10.1110/ps.035436.108.

Ellman, George L. 1959. "Tissue Sulfhydryl Groups." *Archives of Biochemistry and Biophysics* 82 (1): 70–77. doi:10.1016/0003-9861(59)90090-6.

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

Eugster, Marcel R, Martina C Haug, Simona G Huwiler, and Martin J Loessner. 2011. "The Cell Wall Binding Domain of Listeria Bacteriophage Endolysin PlyP35 Recognizes Terminal GlcNAc Residues in Cell Wall Teichoic Acid." *Molecular Microbiology*, July. doi:10.1111/j.1365-2958.2011.07774.x.

Eugster, Marcel R, and Martin J Loessner. 2012. "Wall Teichoic Acids Restrict Access of Bacteriophage Endolysin Ply118, Ply511, and PlyP40 Cell Wall Binding Domains to the Listeria Monocytogenes Peptidoglycan." *Journal of Bacteriology* 194 (23): 6498–6506. doi:10.1128/JB.00808-12.

Evans, Philip. 2006. "Scaling and Assessment of Data Quality." *Acta Crystallographica. Section D, Biological Crystallography* 62 (Pt 1): 72–82. doi:10.1107/S0907444905036693.

Farrell, I. S, R. Toroney, J. L Hazen, R. A Mehl, and J. W Chin. 2005. "Photo-Cross-Linking Interacting Proteins with a Genetically Encoded Benzophenone." *Nature Methods* 2 (5): 377–84.

Farrell, Ian S., Rebecca Toroney, Jennifer L. Hazen, Ryan A. Mehl, and Jason W. Chin. 2005. "Photo-Cross-Linking Interacting Proteins with a Genetically Encoded Benzophenone." *Nature Methods* 2 (5): 377–84. doi:10.1038/nmeth0505-377.

Fenton, Mark, Paul Ross, Olivia McAuliffe, Jim O'Mahony, and Aidan Coffey. 2010. "Recombinant Bacteriophage Lysins as Antibacterials" 1 (1): 9–16. doi:10.4161/bbug.1.1.9818.

Fernández-Tornero, Carlos, Ernesto García, Rubens López, Guillermo Giménez-Gallego, and Antonio Romero. 2002. "Two New Crystal Forms of the Choline-Binding Domain of the Major Pneumococcal Autolysin: Insights into the Dynamics of the Active Homodimer." *Journal of Molecular Biology* 321 (1): 163–73.

Fischer, Samuel, Sophie Kittler, Günter Klein, and Gerhard Glünder. 2013. "Impact of a Single Phage and a Phage Cocktail Application in Broilers on Reduction of Campylobacter Jejuni and Development of Resistance." *PLoS ONE* 8 (10): e78543. doi:10.1371/journal.pone.0078543.

Fischetti, Vincent A. 2008. "Bacteriophage Lysins as Effective Antibacterials." *Current Opinion in Microbiology* 11 (5): 393–400. doi:10.1016/j.mib.2008.09.012.

Fotiadis, Christos T., Maria Dimou, Dimitrios G. Georgakopoulos, Panagiotis Katinakis, and Anastasia P. Tampakaki. 2012. "Functional Characterization of NopT1 and NopT2, Two Type III Effectors of Bradyrhizobium Japonicum." *FEMS Microbiology Letters* 327 (1): 66–77. doi:10.1111/j.1574-6968.2011.02466.x.

Fouche, P B, and J H Hash. 1978. "The N,O-Diacetylmuramidase of Chalaropsis Species. Identificaiton of Aspartyl and Glutamyl Residues in the Active Site." *The Journal of Biological Chemistry* 253 (19): 6787–93.

Franke, Daniel, and Dmitri I. Svergun. 2009. "DAMMIF, a Program for Rapid *Ab-Initio* Shape Determination in Small-Angle Scattering." *Journal of Applied Crystallography* 42 (2): 342–46. doi:10.1107/S0021889809000338.

Frias, M. J., J. Melo-Cristino, and M. Ramirez. 2009. "The Autolysin LytA Contributes to Efficient Bacteriophage Progeny Release in Streptococcus Pneumoniae." *Journal of Bacteriology* 191 (17): 5428–40. doi:10.1128/JB.00477-09.

Frias, Maria João, José Melo-Cristino, and Mário Ramirez. 2013. "Export of the Pneumococcal Phage SV1 Lysin Requires Choline-Containing Teichoic Acids and Is Holin-Independent." *Molecular Microbiology* 87 (2): 430–45. doi:10.1111/mmi.12108.

FUCHS, A R, and G J BONDE. 1957. "The Availability of Sulphur for Clostridium Perfringens and an Examination of Hydrogen Sulphide Production." *Journal of General Microbiology* 16 (2): 330–40.

Gaeng, Susanne, Siegfried Scherer, Horst Neve, and Martin J. Loessner. 2000. "Gene Cloning and Expression and Secretion of Listeria Monocytogenes Bacteriophage-Lytic Enzymes in Lactococcus Lactis." *Applied and Environmental Microbiology* 66 (7): 2951–58.

Ganguly, Jhuma, Lieh Y Low, Nazia Kamal, Elke Saile, L Scott Forsberg, Gerardo Gutierrez-Sanchez, Alex R Hoffmaster, Robert Liddington, Conrad P Quinn, Russell W Carlson, and Elmar L Kannenberg. 2013. "The Secondary Cell Wall Polysaccharide of Bacillus Anthracis Provides the Specific Binding Ligand for the C-Terminal Cell Wall-Binding Domain of Two Phage Endolysins, PlyL and PlyG." *Glycobiology* 23 (7): 820–32. doi:10.1093/glycob/cwt019.

Garcia, E, J L Garcia, P Garcia, A Arraras, J M Sanchez-Puelles, and R Lopez. 1988. "Molecular Evolution of Lytic Enzymes of Streptococcus Pneumoniae and Its Bacteriophages." *Proceedings of the National Academy of Sciences of the United States of America* 85 (3): 914–18.

Gerding, Dale N., Carlene A. Muto, and Robert C. Owens, Jr. 2008. "Measures to Control and Prevent *Clostridium Difficile* Infection." *Clinical Infectious Diseases* 46 (s1): S43–S49. doi:10.1086/521861.

Gil, Filipa, Maria João Catalão, José Moniz-Pereira, Paula Leandro, Michael McNeil, and Madalena Pimentel. 2008. "The Lytic Cassette of Mycobacteriophage Ms6 Encodes an Enzyme with Lipolytic Activity." *Microbiology (Reading, England)* 154 (Pt 5): 1364–71. doi:10.1099/mic.0.2007/014621-0.

Gil, Filipa, Anna E Grzegorzewicz, Maria João Catalão, João Vital, Michael R McNeil, and Madalena Pimentel. 2010. "Mycobacteriophage Ms6 LysB Specifically Targets the Outer Membrane of Mycobacterium Smegmatis." *Microbiology (Reading, England)* 156 (Pt 5): 1497–1504. doi:10.1099/mic.0.032821-0.

Gouet, Patrice, Xavier Robert, and Emmanuel Courcelle. 2003. "ESPript/ENDscript: Extracting and Rendering Sequence and 3D Information from Atomic Structures of Proteins." *Nucleic Acids Research* 31 (13): 3320–23.

Gründling, A, M D Manson, and R Young. 2001. "Holins Kill without Warning." *Proceedings of the National Academy of Sciences of the United States of America* 98 (16): 9348–52. doi:10.1073/pnas.151247598.

Guinier, André. 2013. "La Diffraction Des Rayons X Aux Tres Petits Angles: Applications a L'etude de Phenomenes Ultramicroscopiques."

Hatfull, Graham F, Deborah Jacobs-Sera, Jeffrey G Lawrence, Welkin H Pope, Daniel A Russell, Ching-Chung Ko, Rebecca J Weber, Manisha C Patel, Katherine L Germane, Robert H Edgar, Natasha N Hoyte, Charles A Bowman, Anthony T Tantoco, Elizabeth C Paladin, Marlana S Myers, Alexis L Smith, Molly S Grace, Thuy T Pham, Matthew B O'Brien, Amy M Vogelsberger, Andrew J Hryckowian, Jessica L Wynalek, Helen Donis-Keller, Matt W Bogel, Craig L Peebles, Steven G Cresawn, and Roger W Hendrix. 2010. "Comparative Genomic Analysis of 60 Mycobacteriophage Genomes: Genome Clustering, Gene Acquisition, and Gene Size." *Journal of Molecular Biology* 397 (1): 119–43. doi:10.1016/j.jmb.2010.01.011.

Hatheway, C L. 1990. "Toxigenic Clostridia." *Clinical Microbiology Reviews* 3 (1): 66–98.

Hermoso, Juan A., Begoña Monterroso, Armando Albert, Beatriz Galán, Oussama Ahrazem, Pedro García, Martín Martínez-Ripoll, José Luis García, and Margarita Menéndez. 2003. "Structural Basis for Selective Recognition of Pneumococcal Cell Wall by Modular Endolysin from Phage Cp-1." *Structure* 11 (10): 1239–49. doi:10.1016/j.str.2003.09.005.

Hillenkamp, Franz, Michael Karas, Ronald C. Beavis, and Brian T. Chait. 1991. "Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Biopolymers." *Analytical Chemistry* 63 (24): 1193A–1203A. doi:10.1021/ac00024a716.

Holm, Liisa, and Päivi Rosenström. 2010. "Dali Server: Conservation Mapping in 3D." *Nucleic Acids Research* 38 (Web Server issue): W545–549. doi:10.1093/nar/gkq366.

Horgan, Marianne, Gary O'Flynn, Jennifer Garry, Jakki Cooney, Aidan Coffey, Gerald F. Fitzgerald, R. Paul Ross, and Olivia McAuliffe. 2009. "Phage Lysin LysK Can Be Truncated to Its CHAP Domain and Retain Lytic Activity against Live Antibiotic-Resistant Staphylococci." *Applied and Environmental Microbiology* 75 (3): 872–74. doi:10.1128/AEM.01831-08.

Hu, Shumin, Jian Kong, Wentao Kong, Tingting Guo, and Mingjie Ji. 2010. "Characterization of a Novel LysM Domain from Lactobacillus Fermentum Bacteriophage Endolysin and Its Use as an Anchor to Display Heterologous Proteins on the Surfaces of Lactic Acid Bacteria." *Applied and Environmental Microbiology* 76 (8): 2410–18. doi:10.1128/AEM.01752-09.

Ince, Jennifer, and Alan McNally. 2009. "Development of Rapid, Automated Diagnostics for Infectious Disease: Advances and Challenges." *Expert Review of Medical Devices* 6 (6): 641–51. doi:10.1586/erd.09.46.

Johnson, Mark, Irena Zaretskaya, Yan Raytselis, Yuri Merezhuk, Scott McGinnis, and Thomas L. Madden. 2008. "NCBI BLAST: A Better Web Interface." *Nucleic Acids Research* 36. doi:10.1093/nar/gkn201.

Kabsch, Wolfgang. 2010. "XDS." *Acta Crystallographica. Section D, Biological Crystallography* 66 (Pt 2): 125–32. doi:10.1107/S0907444909047337.

Kanamaru, Shuji, Yasutaka Ishiwata, Toshiharu Suzuki, Michael G. Rossmann, and Fumio Arisaka. 2005. "Control of Bacteriophage T4 Tail Lysozyme Activity During the Infection Process." *Journal of Molecular Biology* 346 (4): 1013–20. doi:10.1016/j.jmb.2004.12.042.

Klijn, N., F. F. Nieuwenhof, J. D. Hoolwerf, C. B. van der Waals, and A. H. Weerkamp. 1995. "Identification of Clostridium Tyrobutyricum as the Causative Agent of Late Blowing in Cheese by Species-Specific PCR Amplification." *Applied and Environmental Microbiology* 61 (8): 2919–24.

Kolarich, Daniel, Pia H Jensen, Friedrich Altmann, and Nicolle H Packer. 2012. "Determination of Site-Specific Glycan Heterogeneity on Glycoproteins." *Nature Protocols* 7 (7): 1285–98. doi:10.1038/nprot.2012.062.

Konarev, Petr V., Vladimir V. Volkov, Anna V. Sokolova, Michel H. J. Koch, and Dmitri I. Svergun. 2003. "PRIMUS: A Windows PC-Based System for Small-Angle Scattering Data Analysis." *Journal of Applied Crystallography* 36 (5): 1277–82. doi:10.1107/S0021889803012779.

Korndörfer, Ingo P, Joseph Danzer, Mathias Schmelcher, Markus Zimmer, Arne Skerra, and Martin J Loessner. 2006. "The Crystal Structure of the Bacteriophage PSA

Endolysin Reveals a Unique Fold Responsible for Specific Recognition of Listeria Cell Walls." *Journal of Molecular Biology* 364 (4): 678–89. doi:10.1016/j.jmb.2006.08.069.

Kretzer, Jan W, Rainer Lehmann, Mathias Schmelcher, Manuel Banz, Kwang-Pyo Kim, Corinna Korn, and Martin J Loessner. 2007. "Use of High-Affinity Cell Wall-Binding Domains of Bacteriophage Endolysins for Immobilization and Separation of Bacterial Cells." *Applied and Environmental Microbiology* 73 (6): 1992–2000. doi:10.1128/AEM.02402-06.

Krishnan, Navasona, Cexiong Fu, Darryl J Pappin, and Nicholas K Tonks. 2011. "H2S-Induced Sulfhydration of the Phosphatase PTP1B and Its Role in the Endoplasmic Reticulum Stress Response." *Science Signaling* 4 (203): ra86. doi:10.1126/scisignal.2002329.

Krissinel, Evgeny, and Kim Henrick. 2007. "Inference of Macromolecular Assemblies from Crystalline State." *Journal of Molecular Biology* 372 (3): 774–97. doi:10.1016/j.jmb.2007.05.022.

Kutter, Elizabeth, and Alexander Sulakvelidze. 2004. *Bacteriophages: Biology and Applications*. CRC Press.

Kuty, Gabriel F., Min Xu, Douglas K. Struck, Elizabeth J. Summer, and Ry Young. 2010. "Regulation of a Phage Endolysin by Disulfide Caging." *Journal of Bacteriology* 192 (21): 5682–87. doi:10.1128/JB.00674-10.

Labrie, Simon J., Julie E. Samson, and Sylvain Moineau. 2010. "Bacteriophage Resistance Mechanisms." *Nature Reviews Microbiology* 8 (5): 317–27. doi:10.1038/nrmicro2315.

Laemmli, U. K. 1970. "Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4." *Nature* 227 (5259): 680–85. doi:10.1038/227680a0.

Le Bourhis, Anne-Gaëlle, Joël Doré, Jean-Philippe Carlier, Jean-François Chamba, Michel-Robert Popoff, and Jean-Luc Tholozan. 2007. "Contribution of C. Beijerinckii and C. Sporogenes in Association with C. Tyrobutyricum to the Butyric Fermentation in Emmental Type Cheese." *International Journal of Food Microbiology* 113 (2): 154–63. doi:10.1016/j.ijfoodmicro.2006.06.027.

Ledford, Heidi. 2009. "Engineered Viruses Fight Bacteria." *Nature News*, March. doi:10.1038/news.2009.131.

Lee, Byong H. 1996. *Fundamentals of Food Biotechnology*. John Wiley & Sons.

Loeffler, J. M., and V. A. Fischetti. 2003. "Synergistic Lethal Effect of a Combination of Phage Lytic Enzymes with Different Activities on Penicillin-Sensitive and - Resistant Streptococcus Pneumoniae Strains." *Antimicrobial Agents and Chemotherapy* 47 (1): 375–77. doi:10.1128/AAC.47.1.375-377.2003.

Loeffler, Jutta M, Svetolik Djurkovic, and Vincent A Fischetti. 2003. "Phage Lytic Enzyme Cpl-1 as a Novel Antimicrobial for Pneumococcal Bacteremia." *Infection and Immunity* 71 (11): 6199–6204.

Loeffler, Jutta M., Daniel Nelson, and Vincent A. Fischetti. 2001. "Rapid Killing of Streptococcus Pneumoniae with a Bacteriophage Cell Wall Hydrolase." *Science* 294 (5549): 2170–72. doi:10.1126/science.1066869.

Loessner, M J, G Wendlinger, and S Scherer. 1995. "Heterogeneous Endolysins in Listeria Monocytogenes Bacteriophages: A New Class of Enzymes and Evidence for Conserved Holin Genes within the Siphoviral Lysis Cassettes." *Molecular Microbiology* 16 (6): 1231–41.

Loessner, Martin J. 2005. "Bacteriophage Endolysins -- Current State of Research and Applications." *Current Opinion in Microbiology* 8 (4): 480–87. doi:10.1016/j.mib.2005.06.002.

Loessner, Martin J, Karl Kramer, Frank Ebel, and Siegfried Scherer. 2002. "C-Terminal Domains of Listeria Monocytogenes Bacteriophage Murein Hydrolases Determine Specific Recognition and High-Affinity Binding to Bacterial Cell Wall Carbohydrates." *Molecular Microbiology* 44 (2): 335–49.

López, Rubens, and Ernesto García. 2004. "Recent Trends on the Molecular Biology of Pneumococcal Capsules, Lytic Enzymes, and Bacteriophage." *FEMS Microbiology Reviews* 28 (5): 553–80. doi:10.1016/j.femsre.2004.05.002.

Low, Lieh Yoon, Chen Yang, Marta Perego, Andrei Osterman, and Robert Liddington. 2011. "Role of Net Charge on Catalytic Domain and Influence of Cell Wall Binding Domain on Bactericidal Activity, Specificity, and Host Range of Phage Lysins." *The Journal of Biological Chemistry* 286 (39): 34391–403. doi:10.1074/jbc.M111.244160.

Low, Lieh Yoon, Chen Yang, Marta Perego, Andrei Osterman, and Robert C Liddington. 2005. "Structure and Lytic Activity of a Bacillus Anthracis Prophage Endolysin." *The Journal of Biological Chemistry* 280 (42): 35433–39. doi:10.1074/jbc.M502723200.

Luo, Y, R A Pfuetzner, S Mosimann, M Paetzel, E A Frey, M Cherney, B Kim, J W Little, and N C Strynadka. 2001. "Crystal Structure of LexA: A Conformational Switch for Regulation of Self-Cleavage." *Cell* 106 (5): 585–94.

Mayer, Melinda J, Vasiliki Garefalaki, Rebecca Spoerl, Arjan Narbad, and Rob Meijers. 2011. "Structure-Based Modification of a Clostridium Difficile-Targeting Endolysin Affects Activity and Host Range." *Journal of Bacteriology* 193 (19): 5477–86. doi:10.1128/JB.00439-11.

Mayer, Melinda J, John Payne, Michael J Gasson, and Arjan Narbad. 2010. "Genomic Sequence and Characterization of the Virulent Bacteriophage phiCTP1 from Clostridium Tyrobutyricum and Heterologous Expression of Its Endolysin." *Applied and Environmental Microbiology* 76 (16): 5415–22. doi:10.1128/AEM.00989-10.

Mayer, Melinda J., Michael J. Gasson, and Arjan Narbad. 2012. "Genomic Sequence of Bacteriophage ATCC 8074-B1 and Activity of Its Endolysin and Engineered Variants against Clostridium Sporogenes." *Applied and Environmental Microbiology* 78 (10): 3685–92. doi:10.1128/AEM.07884-11.

Mayer, Melinda J., Arjan Narbad, and Michael J. Gasson. 2008. "Molecular Characterization of a Clostridium Difficile Bacteriophage and Its Cloned Biologically Active Endolysin." *J. Bacteriol.* 190 (20): 6734–40. doi:10.1128/JB.00686-08.

McCoy, Airlie J, Ralf W Grosse-Kunstleve, Paul D Adams, Martyn D Winn, Laurent C Storoni, and Randy J Read. 2007. "Phaser Crystallographic Software." *Journal of Applied Crystallography* 40 (Pt 4): 658–74. doi:10.1107/S0021889807021206.

McGowan, Sheena, Ashley M Buckle, Michael S Mitchell, James T Hoopes, D Travis Gallagher, Ryan D Heselpoth, Yang Shen, Cyril F Reboul, Ruby H P Law, Vincent A Fischetti, James C Whisstock, and Daniel C Nelson. 2012. "X-Ray Crystal Structure of the Streptococcal Specific Phage Lysin PlyC." *Proceedings of the National Academy of Sciences of the United States of America* 109 (31): 12752–57. doi:10.1073/pnas.1208424109.

Mellroth, Peter, Robert Daniels, Alice Eberhardt, Daniel Ronnlund, Hans Blom, Jerker Widengren, Staffan Normark, and Birgitta Henriques-Normark. 2012. "LytA, Major Autolysin of Streptococcus Pneumoniae, Requires Access to Nascent Peptidoglycan." *The Journal of Biological Chemistry* 287 (14): 11018–29. doi:10.1074/jbc.M111.318584.

Merz, Tobias, Tobias Heck, Birgit Geueke, Peer R. E. Mittl, Christophe Briand, Dieter Seebach, Hans-Peter E. Kohler, and Markus G. Grütter. 2012. "Autoproteolytic and

Catalytic Mechanisms for the B-Aminopeptidase BapA—A Member of the Ntn Hydrolase Family." *Structure* 20 (11): 1850–60. doi:10.1016/j.str.2012.07.017.

Mo, Kai-For, Xiuru Li, Huiqing Li, Lieh Yoon Low, Conrad P. Quinn, and Geert-Jan Boons. 2012. "Endolysins of Bacillus Anthracis Bacteriophages Recognize Unique Carbohydrate Epitopes of Vegetative Cell Wall Polysaccharides with High Affinity and Selectivity." *Journal of the American Chemical Society* 134 (37): 15556–62. doi:10.1021/ja3069962.

Murshudov, Garib N, Pavol Skubák, Andrey A Lebedev, Navraj S Pannu, Roberto A Steiner, Robert A Nicholls, Martyn D Winn, Fei Long, and Alexei A Vagin. 2011. "REFMAC5 for the Refinement of Macromolecular Crystal Structures." *Acta Crystallographica. Section D, Biological Crystallography* 67 (Pt 4): 355–67. doi:10.1107/S0907444911001314.

Nascimento, João Gil, Maria Carolina Guerreiro-Pereira, Sérgio Fernandes Costa, Carlos São-José, and Mário Almeida Santos. 2008. "Nisin-Triggered Activity of Lys44, the Secreted Endolysin from Oenococcus Oeni Phage fOg44." *Journal of Bacteriology* 190 (1): 457–61. doi:10.1128/JB.01195-07.

Nelson, Daniel C, Mathias Schmelcher, Lorena Rodriguez-Rubio, Jochen Klumpp, David G Pritchard, Shengli Dong, and David M Donovan. 2012. "Endolysins as Antimicrobials." *Advances in Virus Research* 83: 299–365. doi:10.1016/B978-0-12-394438-2.00007-4.

Nelson, Daniel, Raymond Schuch, Peter Chahales, Shiwei Zhu, and Vincent A. Fischetti. 2006. "PlyC: A Multimeric Bacteriophage Lysin." *Proceedings of the National Academy of Sciences* 103 (28): 10765–70. doi:10.1073/pnas.0604521103.

Nelson, Eric J., Jason B. Harris, J. Glenn Morris, Stephen B. Calderwood, and Andrew Camilli. 2009. "Cholera Transmission: The Host, Pathogen and Bacteriophage Dynamic." *Nature Reviews Microbiology* 7 (10): 693–702. doi:10.1038/nrmicro2204.

Neuhaus, Francis C., and James Baddiley. 2003. "A Continuum of Anionic Charge: Structures and Functions of D-Alanyl-Teichoic Acids in Gram-Positive Bacteria." *Microbiology and Molecular Biology Reviews* 67 (4): 686–723. doi:10.1128/MMBR.67.4.686-723.2003.

O'Flaherty, Sarah, R. Paul Ross, and Aidan Coffey. 2009. "Bacteriophage and Their Lysins for Elimination of Infectious Bacteria." *FEMS Microbiology Reviews* 33 (4): 801–19. doi:10.1111/j.1574-6976.2009.00176.x.

Otsuji, Nozomu, Mutsuo Sekiguchi, Teiji Iijima, and Yasuyuki Takagi. 1959. "Induction of Phage Formation in the Lysogenic Escherichia Coli K-12 by Mitomycin C." *Nature* 184 (4692): 1079–80. doi:10.1038/1841079b0.

Otwinowski, Zbyszek, Dominika Borek, Wladyslaw Majewski, and Wladek Minor. 2003. "Multiparametric Scaling of Diffraction Intensities." *Acta Crystallographica. Section A, Foundations of Crystallography* 59 (Pt 3): 228–34.

Pang, Ting, Tinya C. Fleming, Kit Pogliano, and Ry Young. 2013. "Visualization of Pinholin Lesions *in Vivo*." *Proceedings of the National Academy of Sciences of the United States of America* 110 (22): E2054–E2063. doi:10.1073/pnas.1222283110.

Parfitt, Tom. 2005. "Georgia: An Unlikely Stronghold for Bacteriophage Therapy." *The Lancet* 365 (9478): 2166–67.

Park, Taehyun, Douglas K Struck, John F Deaton, and Ry Young. 2006. "Topological Dynamics of Holins in Programmed Bacterial Lysis." *Proceedings of the National Academy of Sciences of the United States of America* 103 (52): 19713–18. doi:10.1073/pnas.0600943103.

Park, Taehyun, Douglas K. Struck, Chelsey A. Dankenbring, and Ry Young. 2007. "The Pinholin of Lambdoid Phage 21: Control of Lysis by Membrane Depolarization." *Journal of Bacteriology* 189 (24): 9135–39. doi:10.1128/JB.00847-07.

Parreira, Ricardo, Carlos São-José, Anabela Isidro, Susana Domingues, Graça Vieira, and Mário A. Santos. 1999. "Gene Organization in a Central DNA Fragment of Oenococcus Oeni Bacteriophage fOg44 Encoding Lytic, Integrative and Non-Essential Functions." *Gene* 226 (1): 83–93. doi:10.1016/S0378-1119(98)00554-X.

Paul, Bindu D., and Solomon H. Snyder. 2012. "H2S Signalling through Protein Sulfhydration and beyond." *Nature Reviews Molecular Cell Biology* 13 (8): 499–507. doi:10.1038/nrm3391.

Pérez-Dorado, Inmaculada, Nuria E Campillo, Begoña Monterroso, Dusan Hesek, Mijoon Lee, Juan A Páez, Pedro García, Martín Martínez-Ripoll, José L García, Shahriar Mobashery, Margarita Menéndez, and Juan A Hermoso. 2007a. "Elucidation of the Molecular Recognition of Bacterial Cell Wall by Modular Pneumococcal Phage Endolysin CPL-1." *The Journal of Biological Chemistry* 282 (34): 24990–99. doi:10.1074/jbc.M704317200.

Pérez-Dorado, Inmaculada, Nuria E. Campillo, Begoña Monterroso, Dusan Hesek, Mijoon Lee, Juan A. Páez, Pedro García, Martín Martínez-Ripoll, José L. García, Shahriar Mobashery, Margarita Menéndez, and Juan A. Hermoso. 2007b. "Elucidation of the Molecular Recognition of Bacterial Cell Wall by Modular Pneumococcal Phage

Endolysin CPL-1." *Journal of Biological Chemistry* 282 (34): 24990–99. doi:10.1074/jbc.M704317200.

Petoukhov, Maxim V., Daniel Franke, Alexander V. Shkumatov, Giancarlo Tria, Alexey G. Kikhney, Michal Gajda, Christian Gorba, Haydyn D. T. Mertens, Petr V. Konarev, and Dmitri I. Svergun. 2012a. "New Developments in the ATSAS Program Package for Small-Angle Scattering Data Analysis." *Journal of Applied Crystallography* 45 (2): 342–50. doi:10.1107/S0021889812007662. 2012. "New Developments in the *ATSAS* Program Package for Small-Angle Scattering Data Analysis." *Journal of Applied Crystallography* 45 (2): 342–50. doi:10.1107/S0021889812007662.

Pevsner, Jonathan. 2013. *Bioinformatics and Functional Genomics*. John Wiley & Sons.

Pohane, Amol Arunrao, Himanshu Joshi, and Vikas Jain. 2014. "Molecular Dissection of Phage Endolysin: An Interdomain Interaction Confers Host Specificity in Lysin A of Mycobacterium Phage D29." *Journal of Biological Chemistry*, March, jbc.M113.529594. doi:10.1074/jbc.M113.529594.

Porter, Corrine J, Raymond Schuch, Adam J Pelzek, Ashley M Buckle, Sheena McGowan, Matthew C J Wilce, Jamie Rossjohn, Ryann Russell, Daniel Nelson, Vincent A Fischetti, and James C Whisstock. 2007. "The 1.6 A Crystal Structure of the Catalytic Domain of PlyB, a Bacteriophage Lysin Active against Bacillus Anthracis." *Journal of Molecular Biology* 366 (2): 540–50. doi:10.1016/j.jmb.2006.11.056.

Pritchard, David G, Shengli Dong, John R Baker, and Jeffrey A Engler. 2004. "The Bifunctional Peptidoglycan Lysin of Streptococcus Agalactiae Bacteriophage B30." *Microbiology (Reading, England)* 150 (Pt 7): 2079–87. doi:10.1099/mic.0.27063-0.

Pritchard, David G, Shengli Dong, Marion C Kirk, Robert T Cartee, and John R Baker. 2007. "LambdaSa1 and LambdaSa2 Prophage Lysins of Streptococcus Agalactiae." *Applied and Environmental Microbiology* 73 (22): 7150–54. doi:10.1128/AEM.01783-07.

Rakhuba, D V, E I Kolomiets, E Szwajcer Dey, and G I Novik. 2010. "Bacteriophage Receptors, Mechanisms of Phage Adsorption and Penetration into Host Cell." *Polish Journal of Microbiology / Polskie Towarzystwo Mikrobiologów = The Polish Society of Microbiologists* 59 (3): 145–55.

Ramanculov, E., and R. Young. 2001. "Functional Analysis of the Phage T4 Holin in a Λ Context." *Molecular Genetics and Genomics* 265 (2): 345–53. doi:10.1007/s004380000422.

Rau, A., T. Hogg, R. Marquardt, and R. Hilgenfeld. 2001. "A New Lysozyme Fold. Crystal Structure of the Muramidase from Streptomyces Coelicolor at 1.65 A Resolution." *J.Biol.Chem.* 276 (June): 31994–99. doi:10.1074/jbc.M102591200.

Rea, M. C., D. Alemayehu, R. P. Ross, and C. Hill. 2013. "Gut Solutions to a Gut Problem: Bacteriocins, Probiotics and Bacteriophage for Control of Clostridium Difficile Infection." *Journal of Medical Microbiology* 62 (Pt_9): 1369–78. doi:10.1099/jmm.0.058933-0.

Regulski, Krzysztof, Pascal Courtin, Saulius Kulakauskas, and Marie-Pierre Chapot-Chartier. 2013. "A Novel Type of Peptidoglycan-Binding Domain Highly Specific for Amidated D-Asp Cross-Bridge, Identified in Lactobacillus Casei Bacteriophage Endolysins." *Journal of Biological Chemistry*, June, jbc.M112.446344. doi:10.1074/jbc.M112.446344.

Resch, Gregory, Philippe Moreillon, and Vincent A Fischetti. 2011. "A Stable Phage Lysin (Cpl-1) Dimer with Increased Antipneumococcal Activity and Decreased Plasma Clearance." *International Journal of Antimicrobial Agents* 38 (6): 516–21. doi:10.1016/j.ijantimicag.2011.08.009.

Rodríguez-Rubio, Lorena, Dolores Gutiérrez, Beatriz Martínez, Ana Rodríguez, and Pilar García. 2012. "Lytic Activity of LysH5 Endolysin Secreted by Lactococcus Lactis Using the Secretion Signal Sequence of Bacteriocin Lcn972." *Applied and Environmental Microbiology* 78 (9): 3469–72. doi:10.1128/AEM.00018-12.

Romero, Patricia, Rubens López, and Ernesto García. 2007. "Key Role of Amino Acid Residues in the Dimerization and Catalytic Activation of the Autolysin LytA, an Important Virulence Factor in Streptococcus Pneumoniae." *Journal of Biological Chemistry* 282 (24): 17729–37. doi:10.1074/jbc.M611795200.

Rosenfeld, J, J Capdevielle, J C Guillemot, and P Ferrara. 1992. "In-Gel Digestion of Proteins for Internal Sequence Analysis after One- or Two-Dimensional Gel Electrophoresis." *Analytical Biochemistry* 203 (1): 173–79.

Ruhr, E, and H G Sahl. 1985. "Mode of Action of the Peptide Antibiotic Nisin and Influence on the Membrane Potential of Whole Cells and on Cytoplasmic and Artificial Membrane Vesicles." *Antimicrobial Agents and Chemotherapy* 27 (5): 841–45.

Sao-Jose, Carlos, Ricardo Parreira, Graca Vieira, and Mario A. Santos. 2000. "The N-Terminal Region of the Oenococcus Oeni Bacteriophage fOg44 Lysin Behaves as a Bona Fide Signal Peptide in Escherichia Coli  and as a Cis-Inhibitory Element, Preventing Lytic  Activity on Oenococcal Cells." *Journal of Bacteriology* 182 (20): 5823–31.

Schleifer, K H, and O Kandler. 1972. "Peptidoglycan Types of Bacterial Cell Walls and Their Taxonomic Implications." *Bacteriological Reviews* 36 (4): 407–77.

Schmelcher, Mathias, David M Donovan, and Martin J Loessner. 2012. "Bacteriophage Endolysins as Novel Antimicrobials." *Future Microbiology* 7 (10): 1147–71. doi:10.2217/fmb.12.97.

Schmelcher, Mathias, and Martin J Loessner. 2014. "Application of Bacteriophages for Detection of Foodborne Pathogens." *Bacteriophage* 4 (1). doi:10.4161/bact.28137.

Schmelcher, Mathias, Tatiana Shabarova, Marcel R. Eugster, Fritz Eichenseher, Vincent S. Tchang, Manuel Banz, and Martin J. Loessner. 2010. "Rapid Multiplex Detection and Differentiation of Listeria Cells by Use of Fluorescent Phage Endolysin Cell Wall Binding Domains." *Applied and Environmental Microbiology* 76 (17): 5745–56. doi:10.1128/AEM.00801-10.

Schmelcher, Mathias, Vincent S Tchang, and Martin J Loessner. 2011. "Domain Shuffling and Module Engineering of Listeria Phage Endolysins for Enhanced Lytic Activity and Binding Affinity." *Microbial Biotechnology*, April. doi:10.1111/j.1751-7915.2011.00263.x.

Schneider, Thomas R, and George M Sheldrick. 2002. "Substructure Solution with SHELXD." *Acta Crystallographica. Section D, Biological Crystallography* 58 (Pt 10 Pt 2): 1772–79.

Schuch, Raymond, Daniel Nelson, and Vincent A. Fischetti. 2002. "A Bacteriolytic Agent That Detects and Kills Bacillus Anthracis." *Nature* 418 (6900): 884–89. doi:10.1038/nature01026.

Semenyuk, A. V., and D. I. Svergun. 1991. "GNOM – a Program Package for Small-Angle Scattering Data Processing." *Journal of Applied Crystallography* 24 (5): 537–40. doi:10.1107/S002188989100081X.

Shen, Yang, Thomas Köller, Bernd Kreikemeyer, and Daniel C Nelson. 2013. "Rapid Degradation of Streptococcus Pyogenes Biofilms by PlyC, a Bacteriophage-Encoded Endolysin." *The Journal of Antimicrobial Chemotherapy* 68 (8): 1818–24. doi:10.1093/jac/dkt104.

Smith, H. Williams, and M. B. Huggins. 1982. "Successful Treatment of Experimental Escherichia Coli Infections in Mice Using Phage: Its General Superiority over Antibiotics." *Journal of General Microbiology* 128 (2): 307–18. doi:10.1099/00221287-128-2-307.

Summers, W C. 1993. "Cholera and Plague in India: The Bacteriophage Inquiry of 1927-1936." *Journal of the History of Medicine and Allied Sciences* 48 (3): 275–301.

Sun, Qingan, Gabriel F Kuty, Arulandu Arockiasamy, Min Xu, Ry Young, and James C Sacchettini. 2009. "Regulation of a Muralytic Enzyme by Dynamic Membrane Topology." *Nat Struct Mol Biol* 16 (11): 1192–94. doi:10.1038/nsmb.1681.

Svergun, D., C. Barberato, and M. H. J. Koch. 1995. "CRYSOL – a Program to Evaluate X-Ray Solution Scattering of Biological Macromolecules from Atomic Coordinates." *Journal of Applied Crystallography* 28 (6): 768–73. doi:10.1107/S0021889895007047.

Tamai, Eiji, Hiromi Yoshida, Hiroshi Sekiya, Hirofumi Nariya, Shigeru Miyata, Akinobu Okabe, Tomomi Kuwahara, Jun Maki, and Shigehiro Kamitori. 2014. "X-Ray Structure of a Novel Endolysin Encoded by Episomal Phage phiSM101 of Clostridium Perfringens." *Molecular Microbiology*, n/a–n/a. doi:10.1111/mmi.12559.

Tanaka, Yoshihito, and Jennifer J. Kohler. 2008. "Photoactivatable Crosslinking Sugars for Capturing Glycoprotein Interactions." *Journal of the American Chemical Society* 130 (11): 3278–79. doi:10.1021/ja7109772.

Turner, Mark S, Louise M Hafner, Terry Walsh, and Philip M Giffard. 2004. "Identification and Characterization of the Novel LysM Domain-Containing Surface Protein Sep from Lactobacillus Fermentum BR11 and Its Use as a Peptide Fusion Partner in Lactobacillus and Lactococcus." *Applied and Environmental Microbiology* 70 (6): 3673–80. doi:10.1128/AEM.70.6.3673-3680.2004.

Twort, F.W. 1915. "An Investigation on the Nature of Ultra-Microscopic Viruses ." *The Lancet* 186 (4814): 1241–43. doi:10.1016/S0140-6736(01)20383-3.

Vagin, Alexei, and Alexei Teplyakov. 2010. "Molecular Replacement with MOLREP." *Acta Crystallographica. Section D, Biological Crystallography* 66 (Pt 1): 22–25. doi:10.1107/S0907444909042589.

Varea, Julio, Begoña Monterroso, José L. Sáiz, Consuelo López-Zumel, José L. García, José Laynez, Pedro García, and Margarita Menéndez. 2004. "Structural and Thermodynamic Characterization of Pal, a Phage Natural Chimeric Lysin Active against Pneumococci." *Journal of Biological Chemistry* 279 (42): 43697–707. doi:10.1074/jbc.M407067200.

Volkov, Vladimir V., and Dmitri I. Svergun. 2003. "Uniqueness of *Ab Initio* Shape Determination in Small-Angle Scattering." *Journal of Applied Crystallography* 36 (3): 860–64. doi:10.1107/S0021889803000268.

Vollmer, Waldemar, and Stephen J. Seligman. 2010. "Architecture of Peptidoglycan: More Data and More Models." *Trends in Microbiology* 18 (2): 59–66. doi:10.1016/j.tim.2009.12.004.

Wang, I N, D L Smith, and R Young. 2000. "Holins: The Protein Clocks of Bacteriophage Infections." *Annual Review of Microbiology* 54: 799–825. doi:10.1146/annurev.micro.54.1.799.

Wang, Ing-Nang, John Deaton, and Ry Young. 2003. "Sizing the Holin Lesion with an Endolysin-Beta-Galactosidase Fusion." *Journal of Bacteriology* 185 (3): 779–87.

Wecke, J, M Perego, and W Fischer. 1996. "D-Alanine Deprivation of Bacillus Subtilis Teichoic Acids Is without Effect on Cell Growth and Morphology but Affects the Autolytic Activity." *Microbial Drug Resistance (Larchmont, N.Y.)* 2 (1): 123–29.

Weidenmaier, Christopher, and Andreas Peschel. 2008. "Teichoic Acids and Related Cell-Wall Glycopolymers in Gram-Positive Physiology and Host Interactions." *Nature Reviews Microbiology* 6 (4): 276–87. doi:10.1038/nrmicro1861.

White, Rebecca, Shinobu Chiba, Ting Pang, Jill S. Dewey, Christos G. Savva, Andreas Holzenburg, Kit Pogliano, and Ry Young. 2010. "Holin Triggering in Real Time." *Proceedings of the National Academy of Sciences*, December. doi:10.1073/pnas.1011921108.

Whitman, W B, D C Coleman, and W J Wiebe. 1998. "Prokaryotes: The Unseen Majority." *Proceedings of the National Academy of Sciences of the United States of America* 95 (12): 6578–83.

Witzenrath, Martin, Bernd Schmeck, Jan M Doehn, Thomas Tschernig, Janine Zahlten, Jutta M Loeffler, Maren Zemlin, Holger Müller, Birgitt Gutbier, Hartwig Schütte, Stefan Hippenstiel, Vincent A Fischetti, Norbert Suttorp, and Simone Rosseau. 2009. "Systemic Use of the Endolysin Cpl-1 Rescues Mice with Fatal Pneumococcal Pneumonia." *Critical Care Medicine* 37 (2): 642–49. doi:10.1097/CCM.0b013e31819586a6.

Wlodarska, M., B. Willing, K. M. Keeney, A. Menendez, K. S. Bergstrom, N. Gill, S. L. Russell, B. A. Vallance, and B. B. Finlay. 2011. "Antibiotic Treatment Alters the Colonic Mucus Layer and Predisposes the Host to Exacerbated Citrobacter Rodentium-Induced Colitis." *Infection and Immunity* 79 (4): 1536–45. doi:10.1128/IAI.01104-10.

Xiang, Ye, Petr G. Leiman, Long Li, Shelley Grimes, Dwight L. Anderson, and Michael G. Rossmann. 2009. "Crystallographic Insights Into the Autocatalytic Assembly

Mechanism of a Bacteriophage Tail Spike." *Molecular Cell* 34 (3): 375–86. doi:10.1016/j.molcel.2009.04.009.

Xu, M., D. K Struck, J. Deaton, I. N Wang, and R. Young. 2004. "A Signal-Arrest-Release Sequence Mediates Export and Control of the Phage P1 Endolysin." *Proceedings of the National Academy of Sciences of the United States of America* 101 (17): 6415.

Xu, Min, Arockiasamy Arulandu, Douglas K Struck, Stephanie Swanson, James C Sacchettini, and Ry Young. 2005. "Disulfide Isomerization after Membrane Release of Its SAR Domain Activates P1 Lysozyme." *Science (New York, N.Y.)* 307 (5706): 113–17. doi:10.1126/science.1105143.

Yang, Hang, Yun Zhang, Junping Yu, Yanling Huang, Xian-En Zhang, and Hongping Wei. 2014. "Novel Chimeric Lysin with High-Level Antimicrobial Activity against Methicillin-Resistant Staphylococcus Aureus *in Vitro* and *in Vivo*." *Antimicrobial Agents and Chemotherapy* 58 (1): 536–42. doi:10.1128/AAC.01793-13.

Young, Ry. 2002. "Bacteriophage Holins: Deadly Diversity." *Journal of Molecular Microbiology and Biotechnology* 4 (1): 21–36.

Young, Ry, Ing-Nang Wang, and William D. Roof. 2000. "Phages Will out: Strategies of Host Cell Lysis." *Trends in Microbiology* 8 (3): 120–28. doi:16/S0966-842X(00)01705-4.

Ziedaite, Gabija, Rimantas Daugelavicius, Jaana K. H. Bamford, and Dennis H. Bamford. 2005. "The Holin Protein of Bacteriophage PRD1 Forms a Pore for Small-Molecule and Endolysin Translocation." *Journal of Bacteriology* 187 (15): 5397–5405. doi:10.1128/JB.187.15.5397-5405.2005.

# Appendix

## CTP1L amino acid sequence

```
        10         20         30         40         50         60
MKKIADISNL NGNVDVKLLF NLGYIGIIAK ASEGGTFVDK YYKQNYTNTK AQGKITGAYH

        70         80         90        100        110        120
FANFSTIAKA QQEANFFLNC IAGTTPDFVV LDLEQQCTGD ITDACLAFLN IVAKKFKCVV

       130        140        150        160        170        180
YCNSSFIKEH LNSKICAYPL WIANYGVATP AFTLWTKYAM WQFTEKGQVS GISGYIDFSY

       190        200        210        220        230        240
ITDEFIKYIK GEDEVENLVV YNDGADQRAA EYLADRLACP TINNARKFDY SNVKNVYAVG

       250        260        270
GNKEQYTSYL TTLIAGSTRY TTMQAVLDYI KNLK
```

## Table of different mutations and their effect in CTP1L

| Ctp1L Mutation | Position | C-terminal Cleavage | Methionine incorporation? | Lytic activity |
|---|---|---|---|---|
| Wild-type | | +++ | + | +++ |
| D92A | Catalytic domain active site | - | | - |
| D92A_E94A | | - | | - |
| M263R | Cell wall binding region | +++ | | - |
| T221C | Side-by-side dimer interface | - | | - |
| T221R | | - | | - |
| D215A | Head-on dimer interface | - | | N/A |
| V195P | Linker between Enzymatically active domain and C-terminal domain | + | + | + |
| E194A_E196A | | +++ | | N/A |
| CTP1LΔLinker | | +++ | | N/A |
| CTP1LPPP | | +++ | | N/A |
| C219A | Residues proximal to cleavage site, hypothesised to engage in autocleavage mechanism | +++ | | +++ |
| N197A | | +++ | | +++ |
| N232A | | +++ | | +++ |
| Y188A | | +++ | | +++ |
| S231A | | +++ | N/A | +++ |
| E194A | | +++ | | N/A |
| E196A | | +++ | | +++ |
| K234A | | ++ | | +++ |
| E196A_K234A | | +++ | | +++ |
| V195G | Mutations are all made to the V195 cleavage residue. | ++ | + | |
| V195L | | + | + | |
| V195Y | | --- | + | N/A |
| V195I | | +++ | + | |
| V195M | | + | + | |

**Appendix Figure 1**: CTP1L amino acid sequence. Enzymatically active domain (red), linker (green) with cleavage site residue underlined and C-terminal domain (blue). Summary table of all tested CTP1L mutants. +++ maximum effect similar to wildtype, + minimal effect, - total inhibition.

## CS74L amino acid sequence

```
        10         20         30         40         50         60
MKIGIDMGHT LSGADYGVVG LRPESVLTRE VGTKVIYKLQ KLGHVVVNCT VDKASSVSES

        70         80         90        100        110        120
LYTRYYRANQ ANVDLFISIH FNATPGGTGT EVYTYAGRQL GEATRIRQEF KSLGLRDRGT

       130        140        150        160        170        180
KDGSGLAVIR NTKAKAMLVE CCFCDNPNDM KLYNSESFSN AIVKGITGKL PNGESGNNNQ

       190        200        210        220        230        240
GGNKVKAVVI YNEGADRRGA EYLADYLNCP TISNSRTFDY SCVEHVYAVG GKKEQYTKYL

       250        260
KTLLSGANRY DTMQQILNFI NGGK
```

Table of different mutations and their effect in CS74L

| CS74L Mutation | Position | C-terminal Cleavage | Methionine incorporation? |
|---|---|---|---|
| Wild-type | | +++ | + |
| V185P | | + | + |
| V185Y | | + | + |
| V185G | Mutations are all made to the | ++ | + |
| V185I | V185 cleavage residue. | +++ | + |
| V185L | | ++ | + |
| V185M | | ++ | + |
| T211R | Side-by-side dimer interface | - | N/A |

**Appendix Figure 2:** CS74L amino acid sequence. Enzymatically active domain (red), estimated linker (green) with cleavage site residue underlined and C-terminal domain (blue). and summary table of all tested CS74L mutants. +++ maximum effect similar to wildtype, + minimal effect, - total inhibition.

**CD27L amino acid sequence**

```
         10         20         30         40         50         60
MKICITVGHS ILKSGACTSA DGVVNEYQYN KSLAPVLADT FRKEGHKVDV IICPEKQFKT

         70         80         90        100        110        120
KNEEKSYKIP RVNSGGYDLL IELHLNASNG QGKGSEVLYY SNKGLEYATR ICDKLGTVFK

        130        140        150        160        170        180
NRGAKLDKRL YILNSSKPTA VLIESFFCDN KEDYDKAKKL GHEGIAKLIV EGVLNKNINN

        190        200        210        220        230        240
EGVKQMYKHT IVYDGEVDKI SATVVGWGYN DGKILICDIK DYVPGQTQNL YVVGGGACEK

        250        260        270
ISSITKEKFI MIKGNDRFDT LYKALDFINR
```

Table of different mutations and their effect in CD27L

| CD27L Mutation | Position | C-terminal Cleavage | Lytic activity |
|---|---|---|---|
| Wild-type | | +++ | +++ |
| H84A | Catalytic domain active site | +++ | - |
| E144A | | +++ | - |
| W207A | | + | +++ |
| W207R | Head-on dimer interface | + | +++ |
| Y262A | | + | +++ |
| C238R | Side-by-side dimer interface | + | +++ |
| C238S | | +++ | +++ |
| M186P | Linker between EAD and C-terminal domain | - | +++ |
| Q185P | | +++ | +++ |

**Appendix Figure 3:** CD27L amino acid sequence. Enzymatically active domain (red), linker (green) with cleavage site residue underlined and C-terminal domain (blue). Summary table of all tested CD27L mutants. +++ maximum effect similar to wildtype, + minimal effect, - total inhibition.
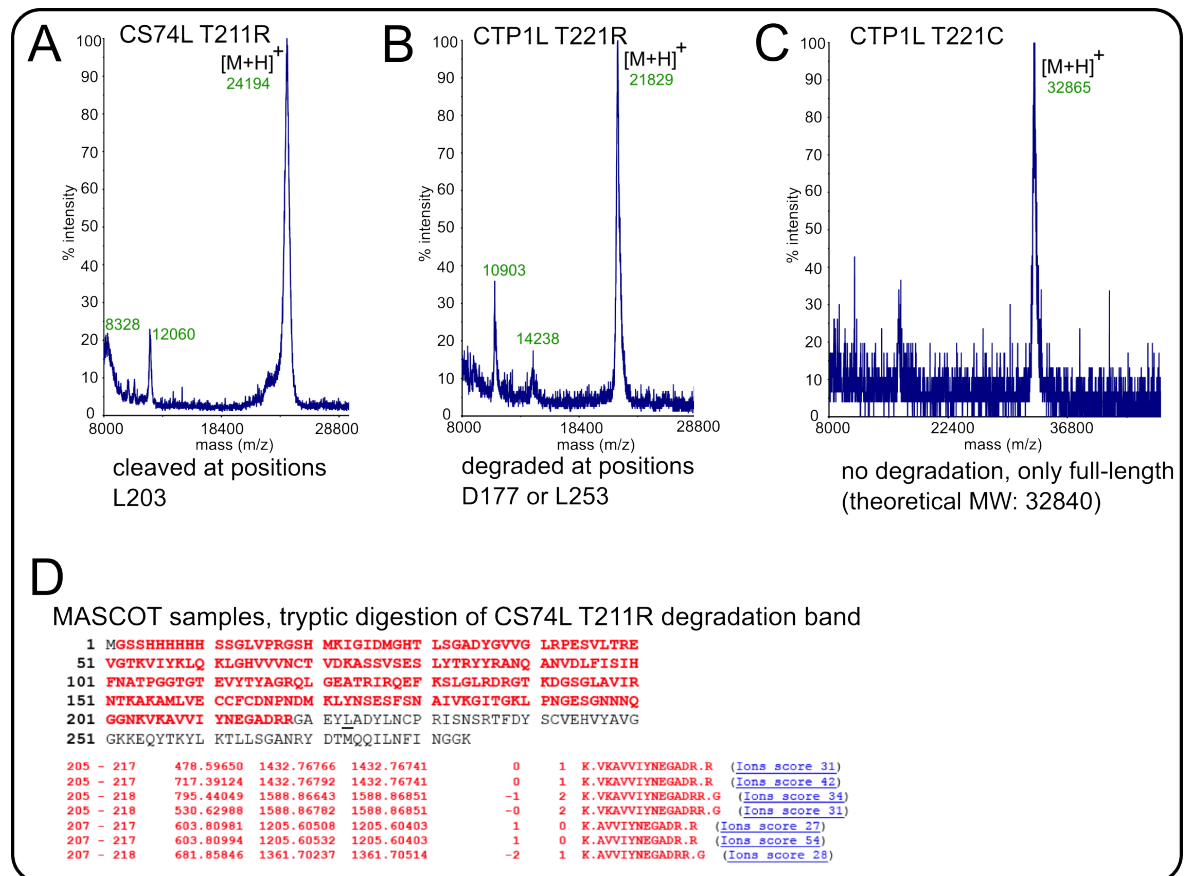
**Appendix Figure 4 MALDI-TOF analyses of CTP1L T221R, T221C and CS74L T211R A:** CS74L T211R fully degraded to a product of m/z value 24 kDa, which was most likely cleavge within the C-terminal domain at position L203. **B:** CTP1L T221R degradation product was approximately 21.8 kDa and could occur from degradation at position D177, within the EAD or at position L253 within the C-terminal domain. Tryptic digest analysis was always contaminated with a small amount of full-length protein, meaning deciding the correct location was complicated. **C:** TP1L T221C did not degrade and only the full-length protein could be detected by MS. **D:** MASCOT search result sample for CS74L T211R, peptides including the degradation site of L203 or C-terminal there on, could not be detected indicating cleavage was most likely at the position L203 within the C-terminal domain.
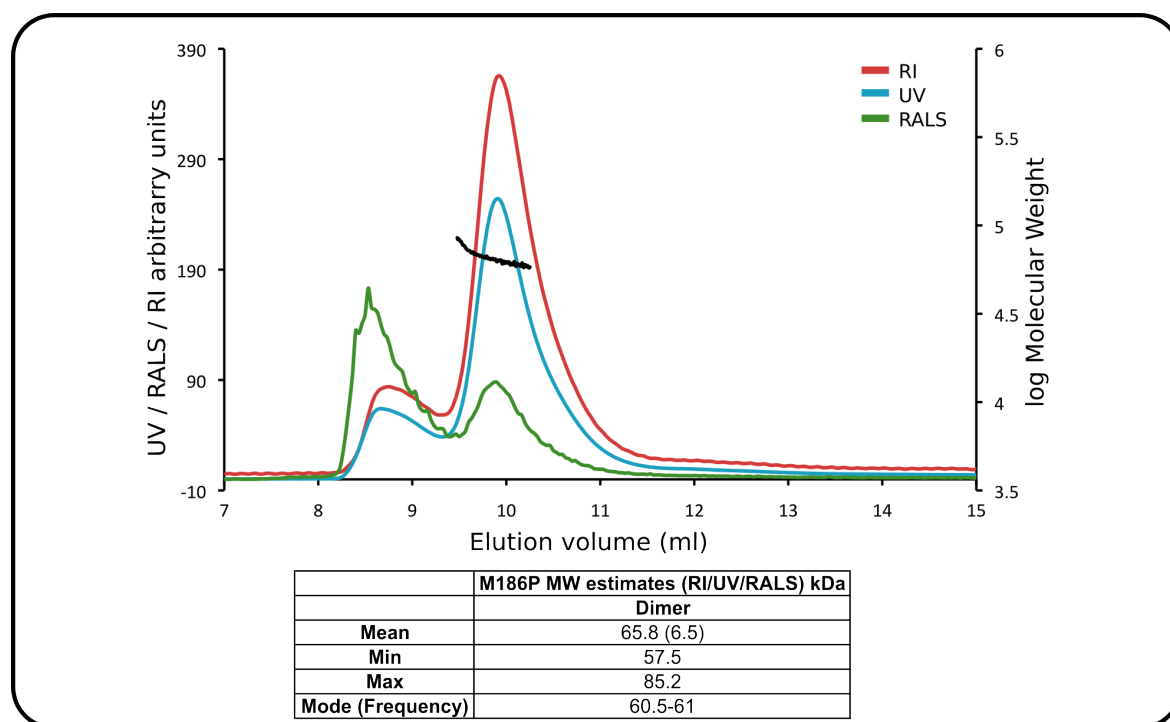
**Appendix Figure 5 UV/RI/RALS size exclusion chromatography traces and MW correlation for CD27L M186P.** The main peak was calculated to be the dimer of CD27L M186P with a MW of 65.8 ± 6.5 kDa. (theoretical MW of dimer: 64.6 kDa).

**A**    100% sequence coverage of full-length CTP1L monomer

```
  1  MGSSHHHHHH SSGLVPRGSH MKKIADISNL NGNVDVKLLF NLGYIGIIAK
 51  ASEGGTFVDK YYKQNYTNTK AQGKITGAYH FANFSTIAKA QQEANFFLNC
101  IAGTTPDFVV LDLEQQCTGD ITDACLAFLN IVAKKFKCVV YCNSSFIKEH
151  LNSKICAYPL WIANYGVATP AFTLWTKYAM WQFTEKGQVS GISGYIDFSY
201  ITDEFIKYIK GEDEVENLVV YNDGADQRAA EYLADRLACP TINNARKFDY
251  SNVKNVYAVG GNKEQYTSYL TTLIAGSTRY TTMQAVLDYI KNLK
```

sample tryptic peptides showing native valine incorporation at site 195

```
211 - 228   1011.45569  2020.89682  2020.89739   -0   0  K.GEDEVENLVVYNDGADQR.A  (Ions score 54)
211 - 228   1011.45569  2020.89682  2020.89739   -0   0  K.GEDEVENLVVYNDGADQR.A  (Ions score 91)
211 - 228   1011.45569  2020.89682  2020.89739   -0   0  K.GEDEVENLVVYNDGADQR.A  (Ions score 99)
211 - 228    674.63965  2020.89712  2020.89739   -0   0  K.GEDEVENLVVYNDGADQR.A  (Ions score 22)
211 - 228    674.63965  2020.89712  2020.89739   -0   0  K.GEDEVENLVVYNDGADQR.A  (Ions score 47)
```

**B**    100% sequence coverage of cleaved C-terminal domain of CTP1L

```
  1  MENLVVYNDG ADQRAAEYLA DRLACPTINN ARKFDYSNVK NVYAVGGNKE
 51  QYTSYLTTLI AGSTRYTTMQ AVLDYIKNLK
```

sample tryptic peptides showing methionine incorporation at site 195 of cleaved CTP1L C-terminal domain

```
 1 - 14    541.91937  1622.73629  1622.73585    0   0  -.MENLVVYNDGADQR.A  (Ions score 38)
 1 - 14   1623.74365  1622.73638  1622.73585    0   0  -.MENLVVYNDGADQR.A  (Ions score 26)
 1 - 14    820.37164  1638.72873  1638.73077   -1   0  -.MENLVVYNDGADQR.A  Oxidation (M) (Ions score 27)
 1 - 14    820.37164  1638.72873  1638.73077   -1   0  -.MENLVVYNDGADQR.A  Oxidation (M) (Ions score 68)
 1 - 14    820.37164  1638.72873  1638.73077   -1   0  -.MENLVVYNDGADQR.A  Oxidation (M) (Ions score 81)
```

**C**    100% sequence coverage of full-length CS74L monomer

```
  1  MGSSHHHHHH SSGLVPRGSH MKIGIDMGHT LSGADYGVVG LRPESVLTRE
 51  VGTKVIYKLQ KLGHVVVNCT VDKASSVSES LYTRYYRANQ ANVDLFISIH
101  FNATPGGTGT EVYTYAGRQL GEATRIRQEF KSLGLRDRGT KDGSGLAVIR
151  NTKAKAMLVE CCFCDNPNDM KLYNSESFSN AIVKGITGKL PNGESGNNNQ
201  GGNKVKAVVI YNEGADRRGA EYLADYLNCP TISNSRTFDY SCVEHVYAVG
251  GKKEQYTKYL KTLLSGANRY DTMQQILNFI NGGK
```

sample tryptic peptides showing native valine incorporation at site 185

```
205 - 217   717.39081  1432.76706  1432.76741   -0   1  K.VKAVVIYNEGADR.R  (Ions score 67)
205 - 217   478.59647  1432.76757  1432.76741    0   1  K.VKAVVIYNEGADR.R  (Ions score 37)
205 - 218   795.43903  1588.86350  1588.86851   -3   2  K.VKAVVIYNEGADRR.G  (Ions score 35)
205 - 218   530.62848  1588.86361  1588.86851   -3   2  K.VKAVVIYNEGADRR.G  (Ions score 36)
205 - 218   398.22327  1588.86396  1588.86851   -3   2  K.VKAVVIYNEGADRR.G  (Ions score 20)
205 - 218   530.62982  1588.86764  1588.86851   -1   2  K.VKAVVIYNEGADRR.G  (Ions score 27)
```

**D**    100% sequence coverage of cleaved C-terminal domain of CS74L

```
  1  MKAVVIYNEG ADRRGAEYLA DYLNCPTISN SRTFDYSCVE HVYAVGGKKE
 51  QYTKYLKTLL SGANRYDTMQ QILNFINGGK
```

sample tryptic peptides showing methionine incorporation at site 185 of cleaved CS74L C-terminal domain

```
Start - End   Observed    Mr(expt)    Mr(calc)   ppm  Miss Sequence
  1 - 13    489.25372  1464.73934  1464.73947   -0   1  -.MKAVVIYNEGADR.R  (Ions score 40)
  1 - 13    741.37402  1480.73349  1480.73439   -1   1  -.MKAVVIYNEGADR.R  Oxidation (M) (Ions score 53)
  1 - 13    494.58542  1480.73443  1480.73439    0   1  -.MKAVVIYNEGADR.R  Oxidation (M) (Ions score 52)
  1 - 13    494.58542  1480.73443  1480.73439    0   1  -.MKAVVIYNEGADR.R  Oxidation (M) (Ions score 42)
  1 - 14    541.28674  1620.83840  1620.84058   -1   2  -.MKAVVIYNEGADRR.G  (Ions score 55)
  1 - 14    811.42719  1620.83982  1620.84058   -0   2  -.MKAVVIYNEGADRR.G  (Ions score 89)
```

**Appendix Figure 6 Sample MASCOT search results after tryptic digestion linked with LC-MS/MS analysis for CTP1L, CS74L and their respective cleaved C-terminal domains. A:** CTP1L full-length endolysin. **B:** CTP1L C-terminal domain. **C:** CS74L CS74L full-length endolysin **D:** CS74L C-terminal domain. For both full-length endolysins only valine was detected at the cleavage site, whilst methionine incorporation is only detected after cleavage as the post-protoelytic modification of the C-terminal domain.
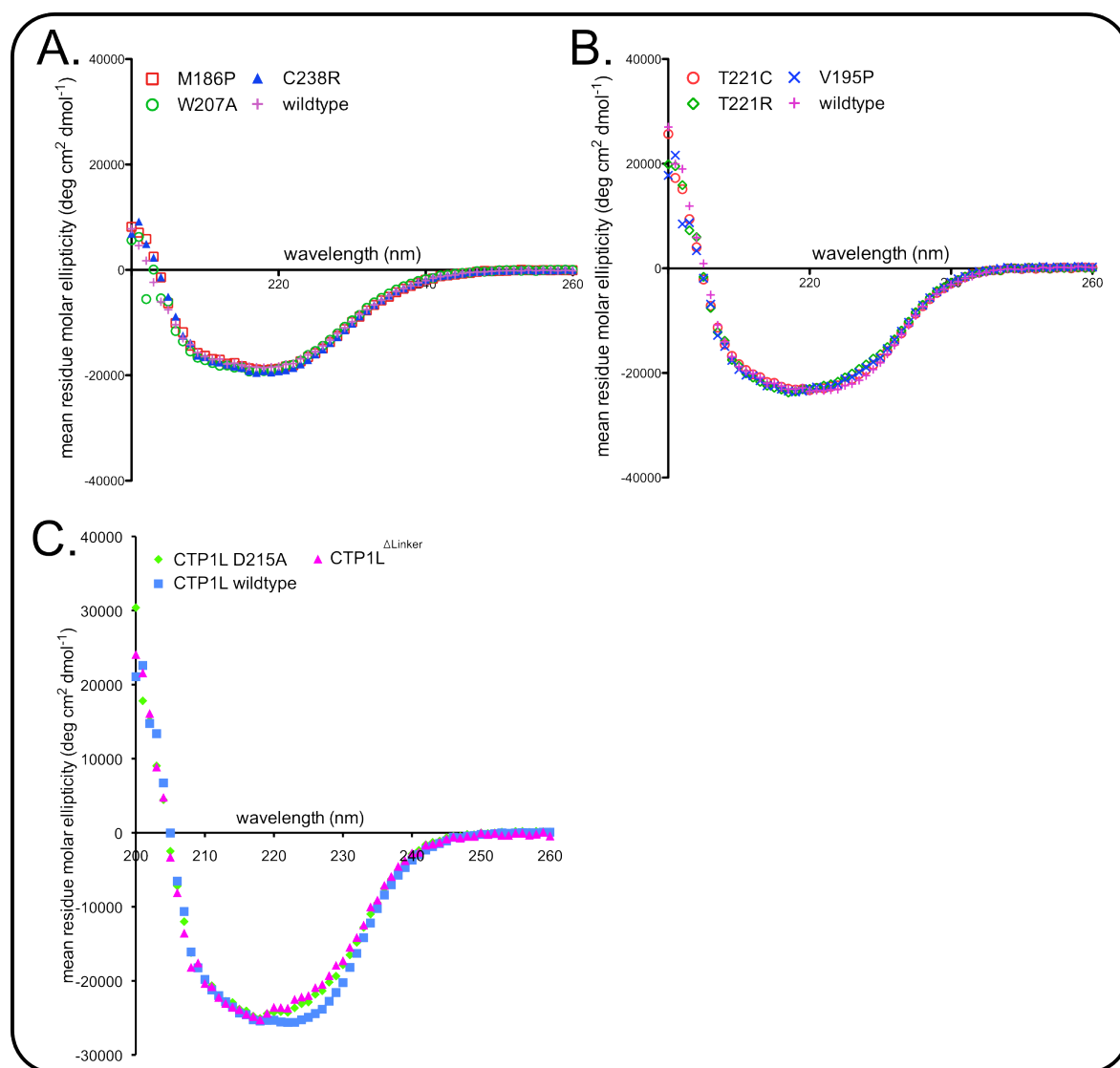
**Appendix Figure 7 Circular dichroism measurements (mean residue molar ellipticity) to show that the integrity of secondary structure has been retained**. **A:** CD27L wild-type and the mutants that prevent autocleavage (M186P, W207A and C238R). **B:** CTP1L wild-type and the mutants that prevent autocleavage (V195P, T221C and T221R). **C:** CTP1L D215A and CTP1L$^{\Delta Linker}$. All the profiles follow a similar shape to the circular dichroism measurements of their respective wildtype proteins.
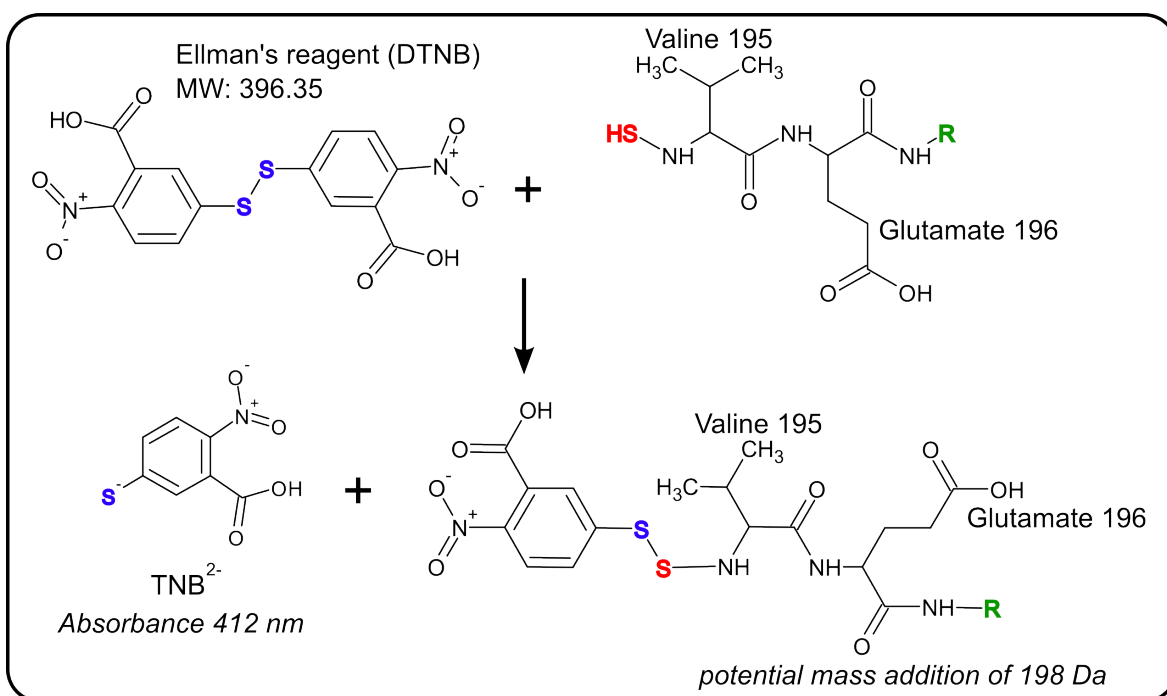
**Appendix Figure 8 The use of Ellman's reagent to test for sulfhydryl addition.** The addition of Ellman's reagent, (5,5'-dithiobis-(2-nitrobenzoic acid)) (DTNB), was used to test if during proteolysis the N-terminal modification of valine involved the addition of a sulfhydryl group. The disulphide bond in DTNB is broken by free sulfuhydryl groups, such as cysteine, in solution. The proposed addition of a sulfhydryl group was proposed to occur on the free amide of valine to produce an N-S bond, which would still cleave DTNB and attach a TNB to the N-terminus of the cleaved domain, providing a mass addition that could be measured by LC-MS.
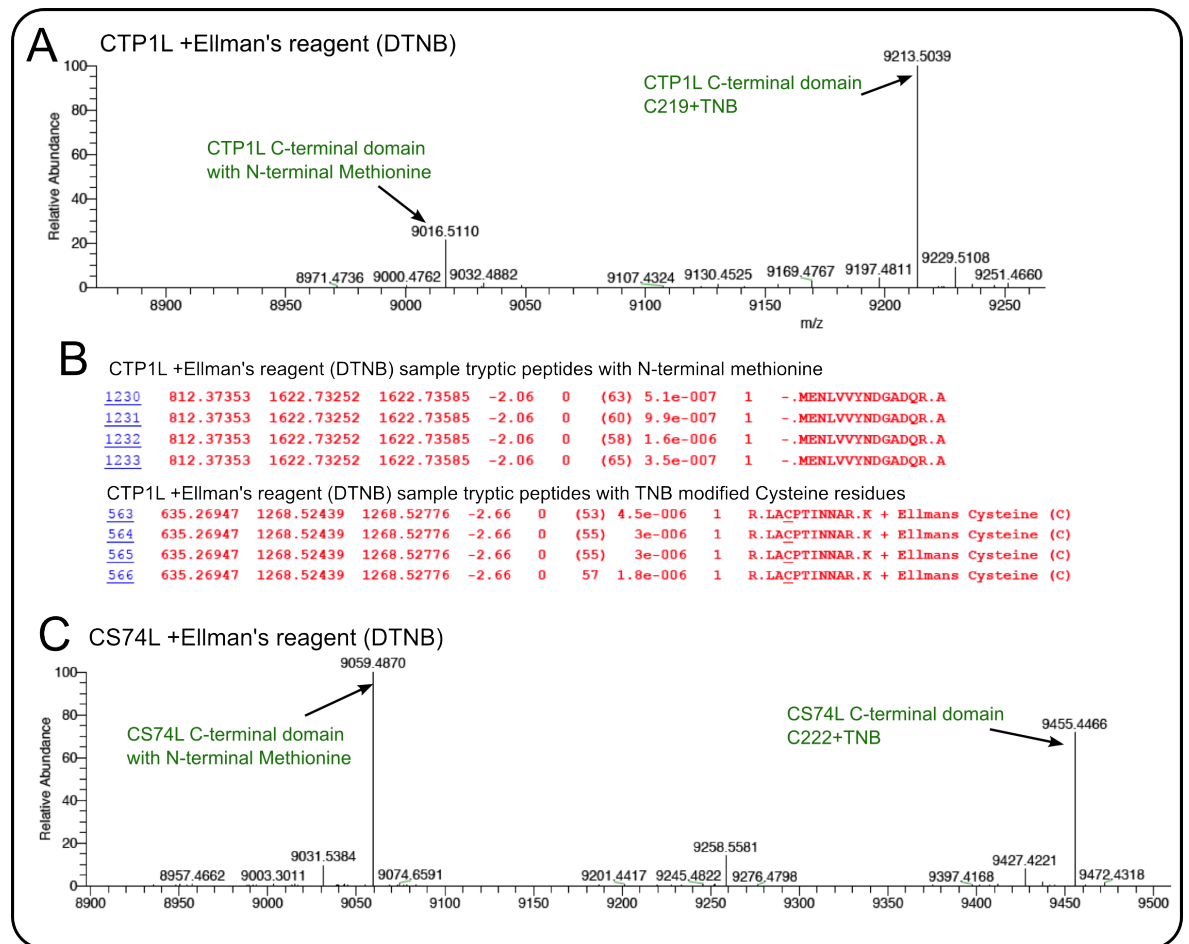
**Appendix Figure 9 Electrospray ionisation mass spectrometry on the cleaved C-terminal domains after addition of Ellman's reagent**. **A:** (CTP1L), **B:** (CS74L) Liquid chromatography coupled to electrospray ionisation mass spectrometry (LC-MS) was used to accurately determine the full-length mass of the cleaved C-terminal domains after the addition of Ellman's reagent to detect any sulfhydryl groups attached to the protein. For both wildtypes, the C-terminal domain contains a single cysteine (CTP1L C219, CS74L C222) which was the only modification detected, indicating no sulfhydryl group was introduced during autocleavage. C: Tryptic digest linked with LC-MS/MS on CTP1L confirmed methionine was still the post-proteolytic product and that TNB was covalently attached to C219.
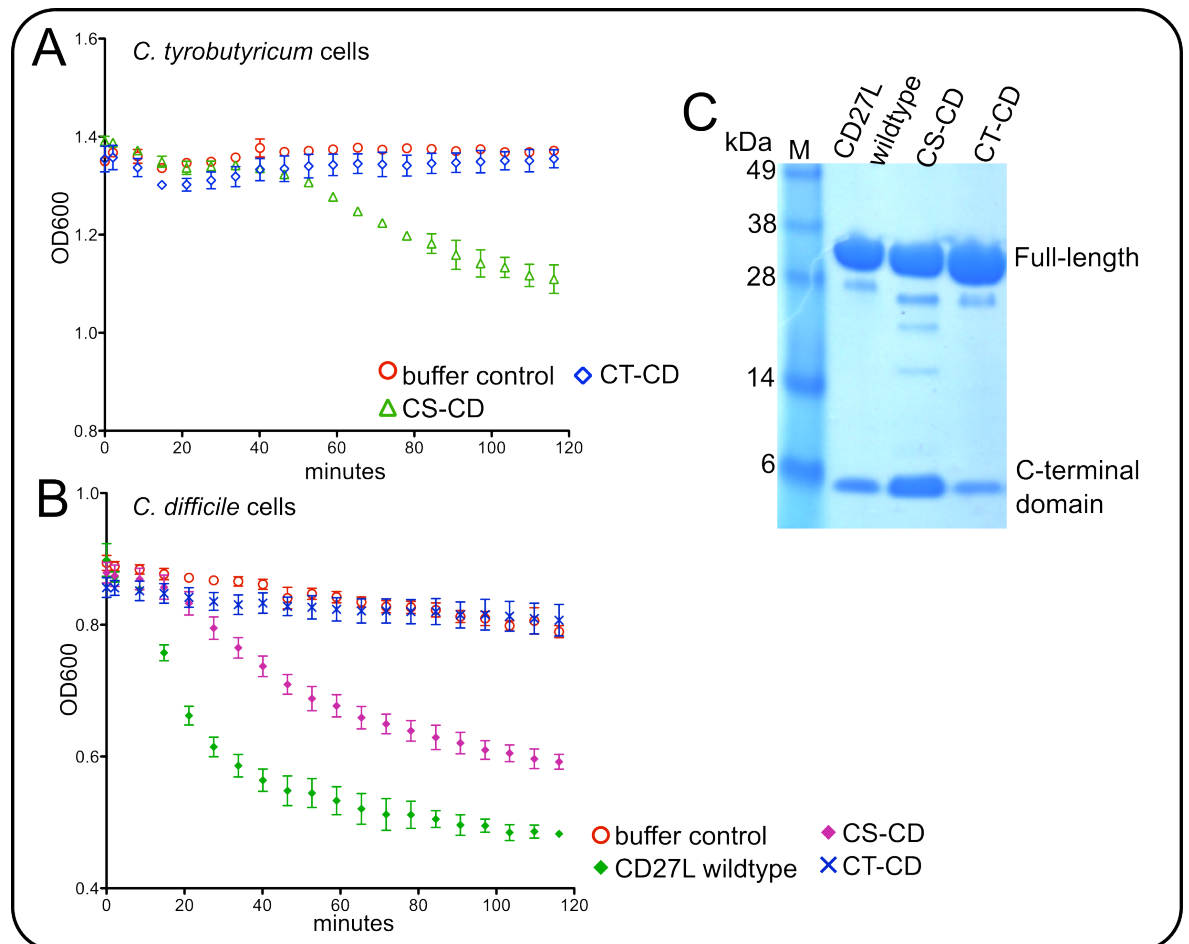
**Appendix Figure 10 Turbidity assays and SDS-PAGE analysis on the chimeras CD-CD and CT-CD. A:** Lysis assays with 10 µg of Ni-NTA purified CTP1L and mutants on *C. tyrobutyricum* cells to determine the lytic effect of the chimeras. CS-CD showed a moderate lytic activity against the cells, CT-CD was inactive. B: Lysis assays with 10 µg of Ni-NTA purified CTP1L and mutants on *C. difficile* cells. CD27L wildtype was used as a positiv control, CS-CD showed moderate activity whilst CT-CD was totally inactive. C: SDS-PAGE of the chimeras, comparing the amount of cleaved CD27L C-terminal domain compared to wildtype CD27L. All samples were loaded fresh after Ni-NTA purification at 2 mg.ml$^{-1}$.

**A** Sample tryptic peptides detected for CTP1LV195G cleaved C-terminal domain with terminal methionine

```
1200   812.37781  1622.74106  1622.73585   3.21   0  (28)  0.0015    1   -.MENLVVYNDGADQR.A
1217   820.37152  1638.72849  1638.73077  -1.39   0  (57)  2.2e-006  1   -.MENLVVYNDGADQR.A + Oxidation (M)
1218   820.37152  1638.72849  1638.73077  -1.39   0  (83)  5.2e-009  1   -.MENLVVYNDGADQR.A + Oxidation (M)
```

Sample tryptic peptides detected for CTP1LV195G cleaved C-terminal domain with terminal glycine

```
1 - 14   775.36414  1548.71372  1548.71684         -2   0  -.GENLVVYNDGADQR.A  (Ions score 58)
1 - 14   775.36414  1548.71372  1548.71684         -2   0  -.GENLVVYNDGADQR.A  (Ions score 50)
1 - 14   775.36414  1548.71372  1548.71684         -2   0  -.GENLVVYNDGADQR.A  (Ions score 49)
1 - 14   775.36414  1548.71372  1548.71684         -2   0  -.GENLVVYNDGADQR.A  (Ions score 38)
```

Sample tryptic peptides detected for CTP1LV195I cleaved C-terminal domain with terminal methionine

```
835   812.37396  1622.73337  1622.73585  -1.53   0  (72)  5.7e-008  1   -.MENLVVYNDGADQR.A
836   812.37396  1622.73337  1622.73585  -1.53   0  (65)  3.1e-007  1   -.MENLVVYNDGADQR.A
837   812.37396  1622.73337  1622.73585  -1.53   0  (89)  1.2e-009  1   -.MENLVVYNDGADQR.A
```

Sample tryptic peptides detected for CTP1LV195L cleaved C-terminal domain with terminal methionine

```
1330   541.91827  1622.73299  1622.73585  -1.76   0  (60)  1e-006    1   -.MENLVVYNDGADQR.A
1331   541.91827  1622.73299  1622.73585  -1.76   0  (51)  7.4e-006  1   -.MENLVVYNDGADQR.A
1332   541.91827  1622.73299  1622.73585  -1.76   0  (42)  6.4e-005  1   -.MENLVVYNDGADQR.A
```

Sample tryptic peptides detected for CTP1LV195M cleaved C-terminal domain with terminal methionine

```
830   812.37457  1622.73459  1622.73585  -0.78   0  (61)  8e-007    1   -.MENLVVYNDGADQR.A
831   812.37457  1622.73459  1622.73585  -0.78   0  (82)  6.3e-009  1   -.MENLVVYNDGADQR.A
832   812.37457  1622.73459  1622.73585  -0.78   0  (59)  1.4e-006  1   -.MENLVVYNDGADQR.A
```

**B** Sample tryptic peptides detected for CS74LV185G cleaved C-terminal domain with terminal methionine

```
Start - End   Observed   Mr(expt)    Mr(calc)   ppm  Miss Sequence
   1 - 13   733.37671  1464.73887  1464.73947   -0   1  -.MKAVVIYNEGADR.R  (Ions score 70)
   1 - 13   741.37097  1480.72739  1480.73439   -5   1  -.MKAVVIYNEGADR.R  Oxidation (M) (Ions score 79)
   1 - 13   494.58322  1480.72784  1480.73439   -4   1  -.MKAVVIYNEGADR.R  Oxidation (M) (Ions score 21)
   1 - 13   741.37592  1480.73728  1480.73439    2   1  -.MKAVVIYNEGADR.R  Oxidation (M) (Ions score 44)
```

Sample tryptic peptides detected for CS74LV185G cleaved C-terminal domain with terminal glycine

```
205 - 217   696.36713  1390.71970  1390.72046   -1   1  K.GKAVVIYNEGADR.R  (Ions score 30)
205 - 217   696.36713  1390.71970  1390.72046   -1   1  K.GKAVVIYNEGADR.R  (Ions score 40)
205 - 217   696.36713  1390.71970  1390.72046   -1   1  K.GKAVVIYNEGADR.R  (Ions score 28)
205 - 217   696.36713  1390.71970  1390.72046   -1   1  K.GKAVVIYNEGADR.R  (Ions score 30)
```

Sample tryptic peptides detected for CS74LV185I cleaved C-terminal domain with terminal methionine

```
Start - End   Observed   Mr(expt)    Mr(calc)   ppm  Miss Sequence
   1 - 13   489.25397  1464.74007  1464.73947    0   1  -.MKAVVIYNEGADR.R  (Ions score 44)
   1 - 13   741.37280  1480.73105  1480.73439   -2   1  -.MKAVVIYNEGADR.R  Oxidation (M) (Ions score 67)
   1 - 13   741.37280  1480.73105  1480.73439   -2   1  -.MKAVVIYNEGADR.R  Oxidation (M) (Ions score 30)
   1 - 13   741.37280  1480.73105  1480.73439   -2   1  -.MKAVVIYNEGADR.R  Oxidation (M) (Ions score 53)
```

Sample tryptic peptides detected for CS74LV185I cleaved C-terminal domain with terminal isoleucine

```
1 - 13   724.39654  1446.77854  1446.78305   -3   1  -.IKAVVIYNEGADR.R   (Ions score 29)
1 - 13   483.26825  1446.78292  1446.78305   -0   1  -.IKAVVIYNEGADR.R   (Ions score 41)
1 - 14   802.44934  1602.88413  1602.88416   -0   2  -.IKAVVIYNEGADRR.G  (Ions score 30)
1 - 14   401.72858  1602.88520  1602.88416    1   2  -.IKAVVIYNEGADRR.G  (Ions score 27)
```

Sample tryptic peptides detected for CS74LV185L cleaved C-terminal domain with terminal methionine

```
1 - 13   733.37671  1464.73887  1464.73947   -0   1  -.MKAVVIYNEGADR.R  (Ions score 54)
1 - 13   733.37671  1464.73887  1464.73947   -0   1  -.MKAVVIYNEGADR.R  (Ions score 21)
1 - 13   489.25381  1464.73962  1464.73947    0   1  -.MKAVVIYNEGADR.R  (Ions score 30)
1 - 13   741.37244  1480.73032  1480.73439   -3   1  -.MKAVVIYNEGADR.R  Oxidation (M) (Ions score 57)
```

Sample tryptic peptides detected for CS74LV185M cleaved C-terminal domain with terminal methionine

```
Start - End   Observed   Mr(expt)    Mr(calc)   ppm  Miss Sequence
   1 - 13   733.37683  1464.73911  1464.73947   -0   1  -.MKAVVIYNEGADR.R  (Ions score 57)
   1 - 13   489.25378  1464.73952  1464.73947    0   1  -.MKAVVIYNEGADR.R  (Ions score 39)
   1 - 13   489.25378  1464.73952  1464.73947    0   1  -.MKAVVIYNEGADR.R  (Ions score 29)
   1 - 13   741.37231  1480.73008  1480.73439   -3   1  -.MKAVVIYNEGADR.R  Oxidation (M) (Ions score 63)
```

Sample tryptic peptides detected for CS74LV185Y cleaved C-terminal domain with terminal tyrosine

```
205 - 217   749.38800  1496.76145  1496.76231   -1   1  K.YKAVVIYNEGADR.R  (Ions score 53)
205 - 217   749.38800  1496.76145  1496.76231   -1   1  K.YKAVVIYNEGADR.R  (Ions score 36)
205 - 217   749.38800  1496.76145  1496.76231   -1   1  K.YKAVVIYNEGADR.R  (Ions score 47)
205 - 217   749.38849  1496.76243  1496.76231    0   1  K.YKAVVIYNEGADR.R  (Ions score 31)
```

Sample tryptic peptides detected for CS74LV185Y cleaved C-terminal domain with terminal methionine

```
1 - 13   733.37707  1464.73960  1464.73947    0   1  -.MKAVVIYNEGADR.R  (Ions score 66)
1 - 13   489.25397  1464.74007  1464.73947    0   1  -.MKAVVIYNEGADR.R  (Ions score 40)
1 - 13   741.37183  1480.72910  1480.73439   -4   1  -.MKAVVIYNEGADR.R  Oxidation (M) (Ions score 71)
1 - 13   494.58371  1480.72930  1480.73439   -3   1  -.MKAVVIYNEGADR.R  Oxidation (M) (Ions score 54)
```

**Appendix Figure 11 Sample MASCOT search results of the cleaved C-terminal domain, for different point mutations made to the cleavage sites of CTP1L and CS74L A:** CTP1L V195G, V195I, V195L, V195M. **B:** CS74L V185G, V185I, V185L, V185M, V185Y. Post-protrolytic methionine incorporation was observed for all mutants as well as peptide hydrolysis for CTP1L V195G, CS74L V185G, V185I and V185Y
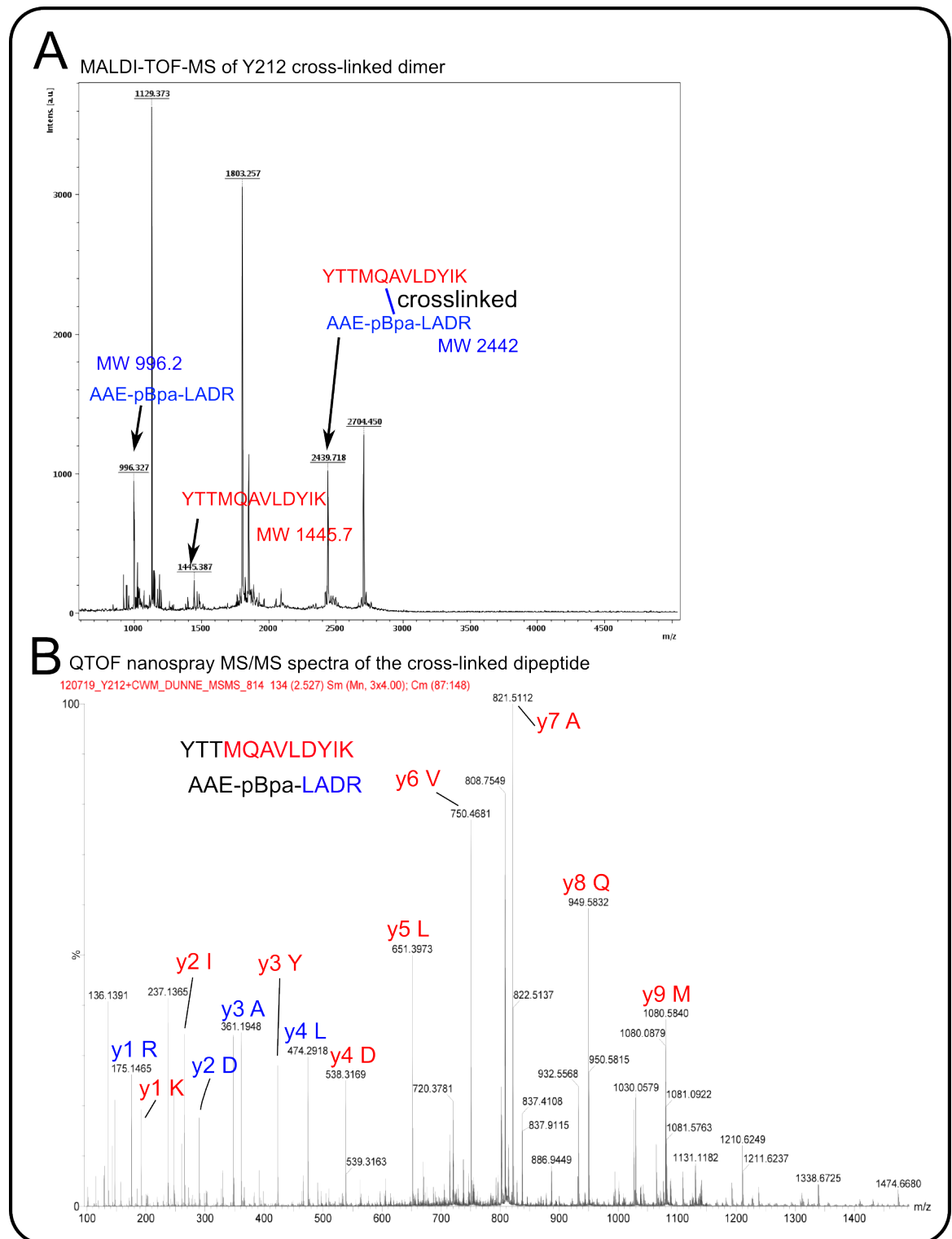
**Appendix Figure 12 Y212-cross-linking between opposing C-terminal domains within the head-on dimer state. A:** MALDI-TOF-MS peptide analysis of in solution cross-linked Y212*p*BPA after tryptic digestion. Beside the signal for the pBPA-containing peptide "AAE-pBPA-LARD" (m/z=996.2) a peak with m/z = 2439.7, probably corresponding to a dipeptide of the highlighted peptides was observed. **B:** ESI-nanospray Q-TOF-MS/MS-sprectrum the Y212pBPA-dipeptide indicated cross-linking between the two peptides via the pBpa molecule. For identification the triply charged precursor ion (m/z=814) of the putative dipeptide was isolated and fragmented using CID resulting in the shown MS/MS-spectrum. The partial amino acid sequence of the dipeptide was deduced by assigning the corresponding y-fragmention-series.

| | CD27L Proteolytic fragment | CTP1L V195P mutant (proteolytic fragment)# | ΔN-CD27L Hg derivative |
|---|---|---|---|
| **Data collection** | | | |
| Space group | P2$_1$2$_1$2$_1$ | I222 | P2$_1$ |
| Cell dimensions | | | |
|   *a*, *b*, *c* (Å) | 75.3, 82.1, 83.8 | 44.9, 48.8, 77.2 | 63.1, 84.7, 65.3 |
|   a, b, g (°) | 90.0, 90.0, 90.0 | 90.0, 90.0, 90.0 | 90.0, 92.0, 90.0 |
| Wavelength (Å) | 0.97 | 1.223 | 0.998 |
| Resolution (Å) | 20 – 2.24 (2.37-2.24) | 20 – 2.10 (2.17 – 2.10) | 20 – 3.5 (3.66 – 3.50) |
| $R_{sym}$ or $R_{merge}$ | 12.8 (59.8) | 2.5 (4.4) | 23.8 (55.8) |
| $I$ / s$I$ | 7.0 (2.0) | 48.9 (29.0) | 4.7 (2.6) |
| Completeness (%) | 97.9 (93.8) | 92 (55.5) | 99.9 (99.9) |
| Redundancy | 2.7 (2.6) | 5.7 (4.4) | 4.8 (4.8) |
| | | | |
| **Refinement** | | | |
| Resolution (Å) | 30 – 2.24 | 20 – 2.10 | |
| No. reflections | 24189 | 4489 | |
| $R_{work}$ / $R_{free}$ | 20.2/24.6 | 17.2 (26.4) | |
| No. atoms | | | |
|   Protein | 3978 | 645 | |
|   Ligand/ion | n/a | n/a | |
|   Water | 495 | 90 | |
| *B*-factors | | | |
|   Protein | 45 | 22 | |
|   Ligand/ion | n/a | n/a | |
|   Water | 57 | 32 | |
| R.m.s. deviations | | | |
|   Bond lengths (Å) | 0.01 | 0.01 | |
|   Bond angles (°) | 1.3 | | |

**Appendix Table 4 Data collection and refinement statistics set one**
*Values in parentheses are for highest-resolution shell.
# The data collection was affected by ice rings and a limited detector geometry
*Table reproduced from the submitted article Appendix page 170.*

|  | CTP1L full length | CTP1L CBD degraded |
| --- | --- | --- |
| **Data collection** | | |
| Space group | P41212 | I222 |
| Cell dimensions | | |
| $a, b, c$ (Å) | 136.20, 136.20, 56.46 | 45.07, 48.74, 77.36 |
| $\alpha, \beta, \gamma$ (°) | 90.0, 90.0, 90.0 | 90.0, 90.0, 90.0 |
| Resolution (Å) | 30 – 1.90 (2.00 – 1.90) * | 20 – 1.18 (1.24 – 1.18) |
| $R_{sym}$ or $R_{merge}$ | 14.8 (75.4) | 7.3 (83.4) |
| $I / \sigma I$ | 9.6 (2.6) | 12.5 (2.2) |
| Completeness (%) | 100.0 (100.0) | 99.1 (98.0) |
| Redundancy | 8.2 (8.0) | 6.2 (5.0) |
| | | |
| **Refinement** | | |
| Resolution (Å) | 30 – 1.90 | 20 - 1.18 |
| No. reflections | 40260 | 26983 |
| $R_{work} / R_{free}$ | 16.3 (19.8) | 15.0/18.9 |
| No. atoms | | |
| Protein | 2567 | 554 |
| Ligand/ion | 38 | n/a |
| Water | 749 | 322 |
| $B$-factors | | |
| Protein | 36 | 16 |
| Ligand/ion | 52 | n/a |
| Water | 47 | 38 |
| R.m.s. deviations | | |
| Bond lengths (Å) | 0.03 | 0.03 |
| Bond angles (°) | 1.9 | 2.5 |

**Appendix Table 5 Data collection and refinement statistics set two**
*Values in parentheses are for highest-resolution shell.
*Table provided by R. Meijers*

| Data collection parameters | CD27L (wild-type) | CD27L (C238R) |
|---|---|---|
| Instrument | EMBL X33 beam line (DORIS-III, DESY, Hamburg) | EMBL P12 beam line (PETRA-III, DESY, Hamburg) |
| Beam geometry | 2.0 x 0.6 mm$^2$ | 0.2 x 0.12 mm$^2$ |
| Wavelength (Å) | 1.54 | 1.24 |
| $s$ range (Å$^{-1}$)[a] | 0.01-0.6 | 0.01-0.46 |
| Exposure time (s) | 8 x 15 | 1 (20 x 0.05 s) |
| Concentration range (mg/mL) | 0.9-4.0 | 1.0-8.5 |
| Temperature (K) | 283 | 283 |
| **Structural parameters[b]** | | |
| $I(0)$ (relative) [from $p(r)$] | 44 ± 2 | 3653 ± 14 |
| $R_g$ (Å) [from $p(r)$] | 33 ± 1 | 43 ± 2 |
| $I(0)$ (cm$^{-1}$) (from Guinier) | 45.6 ± 0.5 | 3664 ± 14 |
| $R_g$ (Å) (from Guinier) | 33 ± 1 | 42 ± 1 |
| $D_{max}$ (Å) | 106 | 147 |
| Porod volume estimate (Å$^3$) | 72151 ± 10000 | 91690 ± 10000 |
| Excluded volume estimate (Å$^3$) | 94000 ± 10000 | 123000 ± 10000 |
| Dry volume calculated from sequence (Å$^3$) | 39121/78219 (mon/dim) | |
| **Molecular-mass determination** | | |
| $I(0)$ (cm$^{-1}$) BSA (66,000 Da) | 71.4 ± 0.4 | 3791 ± 10 |
| Molecular mass $M_r$ (Da) [from $I(0)$] | 42150 ± 5000 | 63780 ± 5000 |
| Molecular mass $M_r$ (Da) [from Porod volume ($V_p/1.6$)] | 45094 ± 5000 | 57306 ± 5000 |
| Molecular mass $M_r$ (Da) [from excluded volume ($V_{ex}/2$)] | 47000 ± 5000 | 61500 ± 5000 |
| Calculated monomeric $M_r$ from sequence | ~32335 | |
| **Software employed** | | |
| Primary data reduction | RADAVER | |
| Data processing | PRIMUS/Qt | |
| *Ab initio* analysis | DAMMIF | |
| Validation and averaging | DAMAVER | |
| Rigid-body modeling | CORAL | |
| Equilibrium analysis | OLIGOMER | |
| Computation of model intensities | CRYSOL | |
| 3D graphics representations | PyMOL, UCSF Chimera | |

**Appendix Table 5 SAXS Data collection and derived parameters for CD27L.** Abbreviations: $M_r$: molecular mass; $R_g$: radius of gyration; $D_{max}$: maximal particle dimension; $V_p$: Porod volume; $V_{ex}$: Particle excluded volume. *Table reproduced from the submitted article Appendix page 170.*

# The CD27L and CTP1L endolysins targeting *Clostridia* contain a built-in trigger and release factor

Matthew Dunne[1], Haydyn D. T. Mertens[1], Vasiliki Garefalaki[1#], Cy M. Jeffries[1], Andrew Thompson[2], Edward Lemke[3], Dmitri I. Svergun[1], Melinda Mayer[4]\*, Arjan Narbad[4], Rob Meijers[1]\*

**Affiliations:**

[1]European Molecular Biology Laboratory (EMBL), Notkestrasse 85, 22587, Hamburg, Germany.

[2]Synchrotron Soleil, L'Orme des Merisiers, BP 48, Saint Aubin, 91192 Gif sur Yvette, France.

[3]European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, 69117 Heidelberg, Germany.

[4]Institute of Food Research, Colney NR4 7UA, Norwich, United Kingdom.

\*Correspondence to: Rob Meijers (r.meijers@embl-hamburg.de) and Melinda Mayer (Melinda.Mayer@ifr.ac.uk)

#Current address: Democritus University of Thrace, Department of Molecular Biology and Genetics, Dragana Campus, Building 10, Alexandroupolis, Greece

*Author contributions*

My contribution to this publication included: construct mutagenesis; protein purification for biophysical characterisation and turbidity assays; crystallisation of CTP1L V195P; SAXS setup and measurements; circular dichroism measurements; pBpa photo-cross-linking and editing of the manuscript. R. Meijers performed the crystal structure determination of CTP1L V195P and CD27L and was the principal editor of the manuscript, H. Mertens performed the SAXS analysis, V. Garefalaki purified the wildtype CD27L for SAXS analysis, C. Jeffries assisted with RALS measurements and their analysis and M. Mayer performed the turbidity assays.

*Abstract*

The bacteriophage ΦCD27 is capable of lysing *Clostridium difficile*, a pathogenic bacterium that is a major cause for nosocomial infection. A recombinant CD27L endolysin lyses *C. difficile in vitro*, and represents a promising alternative as a bactericide. To better understand the lysis mechanism, we have determined the crystal structure of an autoproteolytic fragment of the CD27L endolysin. The structure covers the C-terminal domain of the endolysin, and represents a novel fold that is identified in a number of lysins that target *Clostridia* bacteria. The structure indicates endolysin cleavage occurs at the stem of the linker connecting the catalytic domain with the C-terminal domain. We also solved the crystal structure of a slow cleaving mutant of the CTP1L endolysin C-terminal domain. Two distinct dimerisation modes are observed in the crystal structures for both endolysins, despite a sequence identity of only 22 % between the domains. The dimers are validated to be present for the full-length protein in solution by right angle light scattering, small angle X-ray scattering and cross-linking experiments using the cross-linking amino acid p-benzoyl-L-phenylalanine (pBpa). Mutagenesis on residues contributing to the dimer interfaces indicates that there is a link between the dimerisation modes and the autocleavage mechanism. We show that for the CTP1L endolysin, there is a reduction in lysis efficiency that is proportional to the cleavage efficiency. We propose a model for endolysin triggering, where the extended dimer prevails in the cytosol, and a switch to the side-by-side dimer occurs when the endolysin passes through holin lesions into the exterior environment. This leads to the release of the catalytic portion of the endolysin, enabling the efficient digestion of the bacterial cell wall.

*Author Summary*

*Clostridium difficile* infection is a common cause of hospital-acquired diarrhea, following broad-spectrum antibiotic treatment particularly in elderly patients. Bacteriophage therapy could provide an alternative treatment, but a better understanding of the viral components that lyse the bacterial cell is necessary. Here, we report on the activation of the endolysin from the bacteriophage ΦCD27 that is capable of lysing *C. difficile*. X-ray crystallography was used to determine the structure of an autoproteolytic fragment of the endolysin, and the molecular interactions in the crystal indicate the existence of several oligomeric states. Based on the structure, we introduced mutations that affect the autolytic cleavage of the enzymatic portion of the endolysin, and we show that the oligomeric state has an effect on the cleavage mechanism. To see whether this is a general mechanism for endolysins targeting *Clostridia* bacteria, we solved the crystal structure for another endolysin that targets *C. tyrobutyricum*, and show that a similar mechanism is at work. We propose that the cleavage and oligomerisation are linked, and they provide the endolysin with a trigger and release mechanism that leads to activation. The identification of a trigger and release factor may not only be relevant to *Clostridia* endolysins, but could be an important factor in the triggering of many bacteriophage endolysins. A fuller understanding of this activation mechanism will help in the design of recombinant endolysins or bacteriophages with a more efficient therapeutic potential.

## Introduction

The increasing emergence of antibiotic resistance has led to a renaissance in the use of bacteriophage therapy as an alternative to eradicate pathogenic bacteria [1]. These bacterial viruses are potentially effective bactericides, with the additional advantage that they only affect a small portion of the human microbiome, in contrast to the broad spectrum antibiotics in use [2,3]. Many antibiotics have an effect on a large portion of the microbiome, leading to a shift in bacterial populations after treatment. A striking example is the emergence of *Clostridium difficile* as a causative agent of antibiotic-associated diarrhea. *C. difficile* is resistant to many of the antibiotics used in hospitals, and it colonizes the gut after antibiotic treatment [4]. In search of an alternative treatment, a bacteriophage named ΦCD27 was isolated from a strain of *C. difficile* [5]. The genome of the ΦCD27 phage revealed the presence of a canonical holin/endolysin system. Endolysins are produced by many double stranded DNA bacteriophages to effect the release of new virions from an infected cell by degrading the bacterial cell wall [6]. The recombinant endolysin CD27L was shown to lyse *C. difficile in vitro* [5]. We have also shown that the N-terminal domain of CD27L consisting of a zinc dependent N-acetylmuramoyl-L-alanine amidase alone is effective in lysis, and that the host range of the endolysin can be affected by a mutation in the substrate binding pocket [7].

Bacteriophages co-evolve with their bacterial hosts, and the continuous adaptation of the phage may limit its lethality. Many bacteriophages isolated from the host environment are therefore not efficient in the rapid eradication of pathogenic hosts, as is the case with ΦCD27. The potential to engineer more potent bacteriophages requires knowledge of the most important components of the lysis machinery [8]. Cell lysis is tightly regulated by the phage which only triggers cell lysis once it has finished the production of new viral particles inside the cell [9]. The endolysin is sequestered in the cytoplasm until it can penetrate the peptidoglycan layer following the formation of lesions in the cell membrane that are created by holin, another phage encoded protein [10]. Endolysins typically consist of a peptidoglycan hydrolase domain and a C-terminal domain that is often termed as a cell wall binding domain. The efficient use of endolysins as bactericides is limited by a poor understanding in most systems of the mechanisms that relate catalytic activity to the role of the C-terminal domain [8]. Many recombinantly produced endolysins can lyse a population of bacteria efficiently only after the protein has been incubated or converted with cell wall material from the host [11,12]. For some endolysins, the catalytic domain expressed in isolation is more effective than the full-length protein [7,13] and for other endolysins, the

catalytic domain alone shows little or no lytic activity at all [12,14]. For a pinholin-dependent phage, endolysin activation was shown to depend on disulphide isomerisation that triggers cleavage of the enzyme from the bacterial membrane [15]. For the highly efficient endolysin PlyC active against streptococcal species, it was found that two catalytic components are tethered in a non-covalent way to eight components of the cell wall binding domain [16]. However, for the classical endolysin/holin system, it is not clear how the endolysins are activated. Here, we present the crystal structures of autoproteolytic fragments of the CD27L and CTP1L endolysins, covering the C-terminal domain. Structure-based mutagenesis allowed us to manipulate autolytic cleavage, and we show that the rate of cleavage is proportional to lysis efficiency for the CTP1L endolysin.

## *Results*

### The C-terminal domain of CD27L adopts a novel protein fold

When crystallization trials for full length CD27L endolysin were set up, crystals appeared overnight from freshly purified protein. Any delay in the purification or crystal tray setup would prevent crystallization, and the crystals dissolved after three weeks. An X-ray data set to 2.3 Ångstrom was collected from a fresh crystal, and it was realized that the crystal most likely contained the C-terminal portion of the endolysin, because molecular replacement with the previously determined crystal structure of the catalytic domain [7] was not successful. To determine the structure, the C-terminal portion of CD27L was also independently cloned, expressed and purified. This N-terminal deletion construct (ΔN-CD27L) was crystallized, and the crystal structure was determined by single wavelength anomalous diffraction (SAD) using a mercury derivative (See Table I for details). The structure was used as a model to solve the structure of the full-length CD27L crystals by molecular replacement. It was found that the "full length" construct had been proteolyzed and the crystal contained six copies of the C-terminal portion of CD27L alone. The refined structure shows clear electron density for all six monomers of the C-terminal domain with a Matthews coefficient of 2.3, and there is no space in the crystal lattice for an additional N-terminal domain.

The C-terminal portion of CD27L consists of a platform of four parallel beta strands, flanked by an alpha helix with two additional alpha helices mounted on top (Figure 1A,B). The N-terminus contains beta strand b1 at the center of the beta sheet, connected to alpha helix a1. This is followed by beta strand b2 at the outer side of the sheet that is connected through an extended loop including a single $3^{10}$ helical turn (h1) to beta strand b3 at the center of the beta sheet. The a2 alpha helix connects beta strands b3 and b4 and the fold

ends with an alpha helix a3 at the C-terminus of the protein. A DALI search of the PDB for domains with a similar fold did not identify a structure with significant similarity [17]. A BLAST search was done with the sequence covering the proteolytic fragment to identify other proteins that may have a similar domain, and 14 unique sequences were found with an E value <0.01. All these proteins are lysins that target *Clostridia* species, and the sequence variation is too large to identify residues that define the fold (Figure 2A).

**Proteolytic processing of full length recombinant endolysin CD27L**

Expression of the full-length CD27L endolysin was hampered by severe and continuous proteolysis that could not be diminished by protease inhibitors. An SDS-PAGE gel of freshly purified material typically showed a protein band for the full length protein and a second band with a molecular weight that corresponds to the C-terminal domain (Figure 3A, B). Proteolytic products isolated from an SDS-PAGE gel of CD27L were analyzed by mass spectrometry following tryptic digestion, and this confirmed that the fragments were the intact N-terminal catalytic domain and the C-terminal domain respectively. The proteolytic fragment covering the C-terminal domain was also observed in liquid chromatography coupled to an electrospray mass spectrometry system, and the N-terminal residue was identified as methionine M186. (Figure 3C).

These observations are not unprecedented, and similar proteolytic processes can be uncovered from studies on other unrelated endolysins. For instance, crystallization of several endolysins was achieved only after a substantial incubation period [11,18], or the individual domains had to be cloned and crystallized separately due to the degradation of the full-length protein [13]. By investigating the structures of full-length endolysins that underwent these treatments (PDB codes 1XOV [18] and 2IXU [19]) we observed that the linker between the domains is always extended and exposed to the solvent. In addition, the catalytic domain and the C-terminal domain are expressed as separate components in PlyC, the most efficient endolysin isolated to date [16]. This raised the possibility that the autoproteolytic cleavage of the catalytic domain in CD27L has a functional role.

**Mutagenesis at the cleavage site prevents endolysin cleavage**

In an attempt to find the residues involved in the cleavage of the endolysin, we investigated the N-terminus of the proteolytic fragment of CD27L. The catalytic domain precedes the C-terminal domain, and when the two crystal structures are concatenated, there is a seven residue linker between the domains (Figure 1C). The autolytic fragment of

the C-terminal domain starts at the end of the linker at methionine M186, which is still integrated in the C-terminal domain. Among the six copies of the C-terminal domain, there are no consistent contacts between M186 and other residues within the C-terminal domain. The methionine side chain only forms a hydrogen bond with the main chain nitrogen of threonine T227 in two out of six molecules.

Since there is no clear candidate among the adjacent residues to be involved in protein cleavage, we decided to mutate methionine M186 to a proline. The M186P mutant will strengthen the main chain at the cleavage point, and would alter the mechanics of the linker at the hinge close to the C-terminal domain. Indeed, the M186P mutant abolishes the cleavage of the endolysin as observed by SDS-PAGE (Figure 3A). In addition, we mutated the amino acid that precedes the methionine (glutamine Q185) to a proline. This residue forms part of the linker and is fully exposed to the solvent. In this case, endolysin cleavage was not affected. This indicates methionine M186 is critical in the cleavage process, and it forms an integral part of the C-terminal domain that is not accessible for external proteolytic cleavage.

**Structure of the C-terminal domain of the CTP1L endolysin mutant V195P**

Another previously characterized phage endolysin that targets Clostridia is CTP1L, which lyses *C. tyrobutyricum* [14]. This endolysin also contains a C-terminal domain that is approximately 80 residues long, but the sequence identity with the C-terminal domain of CD27L is low (22%). SDS-PAGE analysis confirmed that purified CTP1L wild type endolysin undergoes cleavage of the C-terminal domain (Figure 3D). We then transferred the critical mutation that affect CD27L cleavage, involving the stem of the linker of the C-terminal domain of CTP1L (V195P). The SDS-PAGE analysis of purified recombinant CTP1L shows a reduction in cleavage for the V195P mutant (Figure 3D).

We attempted to crystallize the CTP1L V195P mutant to see if this slowly cleaving mutant would yield crystals of the full length protein. After 2 weeks, crystals appeared, an X-ray data set was collected to 2.1 Ångstrom and the structure of the C-terminal domain was solved by molecular replacement using the C-terminal domain of CD27L as a search model. As with CD27L, there was no N-terminal domain present in the crystal lattice. The C-terminal domain is truncated at Pro195, and there is only one molecule present in the asymmetric unit. The fold of the C-terminal domain of CTP1L is very similar to that of CD27L, except that the second alpha helix a2 is deleted, and the alpha helix a3 is extended

in CTP1L (Figure 2B). A superimposition of the two domains based on secondary structure elements using Coot [20] gives an RMSD of 1.5 Å for 75 aligned residues, even though the domains have a low sequence identity (Figure 1D). A BLAST search for other proteins that align to the C-terminal domain of CTP1L reveal a separate set of amino acid sequences of lysins targeting *Clostridia* (Figure 2C). It is not possible to come up with conserved amino acids that define the fold. The only conserved residues are an aspartate on helix a1 (Asp 206 in CTP1L and Asp198 in CD27L), a threonine on helix a3 (Thr 262 in CTP1L and Thr 261 in CD27L), and an arginine (Arg 259 in CTP1L and Arg 258 in CD27L). The conserved aspartate/threonine form a hydrogen bond through a water molecule in both structures, connecting the outer alpha helices, but this is not sufficient to keep the fold together.

**Two dimerization modes suggestive of endolysin activation**

The proteolytic fragments of CD27L form a mixture of dimers within the crystal lattice. All six molecules are engaged in one common dimerization mode, where the alpha helices a1 and a3 from one molecule stack on their symmetry mate from a second molecule. The a1 and a3 helices run parallel and in the same direction, forming a platform with a concave surface (Figure 4A). The dimerization is such that the N-termini of both monomers are pointing away from the dimer interface, and we term this dimerization mode a 'head-on' dimer. The buried surface area is between 1200 and 1300 Å$^2$ for the three head-on dimers found in the asymmetric unit, as determined by the PISA server [21]. The docking for the three head-on dimers observed in the crystal lattice is very similar, and superimposition of the Ca atoms with LSQKAB [22] using both protomers gave RMSDs of 0.71 Å and 0.84 Å respectively. There is a 2-fold symmetry axis running perpendicular to the parallel alpha helices, with a hydrophobic core at the center consisting of residues valine V204, leucine L261 and leucine L265. Further along the rim, there are additional aromatic residues (tryptophan W207, phenylalanine F258 and tyrosine Y262) whose symmetry mates are involved in dimerization. The strong hydrophobic component, combined with the stacking of aromatic rings indicates this is a stable dimerization mode. The head-on dimer is also present in the crystal lattice of the C-terminal domain of CTP1L. In fact, it is possible to superimpose the whole dimer unit based on secondary structure elements in Coot, with an RMSD for the Ca backbone of 2.1 A for 146 residues out of a total of 160 (Figure 4B). None of the residues in the head-on dimer face is conserved between CTP1L and CD27L. To test the significance of the head-on dimer, we performed mutagenesis on two of the aromatic residues involved in the dimer interface (W207A/W207R and Y262A). These

mutants had a surprising effect on the autolytic cleavage, since a decrease was observed in the cleavage product present on an SDS-PAGE gel (Figure 3B). This mutant is situated at the opposite site of the linker that connects the C-terminal domain to the catalytic domain.

An alternative dimerization mode is found among the six C-terminal domains present in the crystal structure of the proteolytic fragment of CD27L endolysin, between two molecules that are each involved in separate head-on dimers as well (Figure 4C). The a2 helices of the opposing monomers stack against each other and the buried surface area is 1216 $\text{Å}^2$, similar to the values found for the head-on dimer. The side chains of cysteine C238 of the symmetry mates face each other at the center of this dimer, with a sulphur-sulphur distance of 3.4 Å (Figure 4D). This distance is too large to qualify for a covalent bond. The cysteine is in close proximity to methionine M251 (Figure 4D), with a sulphur-sulphur distance of 3.7 Å (4.1 Å for the symmetry mate). Moreover, lysine K253 forms a hydrogen bond between the NZ atom and the sulphur with a distance of 2.8 Å (3.1 Å for the symmetry mate). Together, M251 and L253 seem to destabilize the formation of the disulphide bond. Although the closely related phage endolysins contain cysteine C238 (Figure 2A), the surrounding residues seem to vary.

To test the significance of this side-by-side dimer, we mutated cysteine C238 to an arginine (C238R), eliminating a potential disulphide bond and introducing opposing positive charges. This mutant had a similar effect on the autolytic cleavage as the head-on dimer mutant (W207A), showing a significant reduction in the production of the proteolytic product (Figure 3A). The cleavage site M186 is approximately 20 Ångstrom away from cysteine C238, indicating that disruption of both dimer interfaces have an effect on the autocleavage of the endolysin.

There is a similar side-by-side dimer present in the crystal lattice of the CTP1L C-terminal fragment, but the domains point in opposite directions compared to the same dimer observed in CD27L. There is no cysteine present in the CTP1L side-by-side dimer interface, and the residue that is closest positioned is threonine T221 (Figure 2B). We mutated threonine T221 to a cysteine as well as an arginine, to see if we could emulate the effects observed for the C238R mutant in CD27L. We observed that both the T221C and the T221R mutant reduce autocleavage to almost undetectable levels, and that these mutants have a stronger effect than the V195P mutant (Figure 3D). This provides further support for the role of an oligomeric switch in the autocleavage of these endolysins.

**Endolysin oligomerization in solution**

To determine the low resolution shape and the oligomeric state of CD27L endolysin in solution, small-angle X-ray scattering (SAXS) experiments were conducted using freshly purified material (Table 2). We used the crystal structures of the catalytic domain of CD27L (PDB code 3QAY) and the crystal structure of the C-terminal domain presented here to make a composite model of the full length CD27L endolysin using the structure of the intact PlyPSA amidase (PDB code 1XOV) to place the two domains. This model was employed to test the presence in solution of the two dimeric states of the C-terminal domain observed in the crystal structure. The molecular mass of the solute of wild type full length CD27L, estimated from the forward scattering intensity was $42\pm4$ kDa, significantly lower than expected for a 64 kDa dimer and indicative of a possible equilibrium of the dimers with dissociation products. SAXS curves calculated from both the head-on and side-by-side dimers using CRYSOL [23] produced poor fits (discrepancy $\chi=1.8$ and 4.0) to the experimental data from the wide type protein (Figure 5A). For the C238R mutant, however, the experimental data fit a scattering curve calculated from the head-on dimer configuration of the composite model (discrepancy $\chi=1.0$) (Figure 5A). The side-by-side dimer is not compatible with the scattering curve for this mutant ($\chi=3.3$). This is an indication that this mutation has driven the equilibrium of the oligomeric states towards the head-on dimer. The distance distribution function $p(r)$ of the C238R mutant (Figure 5A, insert) displays two distinct peaks, the one at larger distance (about 70 Å) matching the distance between the centers of the catalytic domains in the head-on dimer. The $p(r)$ function of the wild type lacks this feature and displays a smaller maximum size, again suggesting an equilibrium of dimers and dissociation products.

Low resolution shape reconstruction from the SAXS data for the wild-type and C238R mutant yields compact and extended structures, respectively. These models represent an average of the conformations of all particles present in solution. The volumes of the models constructed from the wild-type ($V_{ex} = 94000 \pm 10000$ Å$^3$) and C238R ($V_{ex} = 123000 \pm 10000$ Å$^3$) data are consistent with that of a mixture and of a dimeric CD27 structure, respectively (Table 2). The extended shape reconstructed for the C238R mutant overlays well with the head-on dimer model, providing a good low resolution representation of the solution structure (Figure 5C).

We then performed gel exclusion chromatography coupled with right-angle light scattering and refractive index/UV measurements to assess the molecular weights of each endolysin species, comparing the separation profiles of the wild type with the C238R and M186P mutants. The elution profile of the wild type endolysin is rather complicated (Figure 6A). We interpret the peak with a molecular weight mass of 68±4 kDa as predominantly containing endolysin dimers (expected *MW* is 64 kDa), and the peak with a molecular weight mass of 33±7 as a CD27L monomer. In between, there is a peak at 43±2 kDa molecular weight that we interpret as a mixture of CBD-cleaved monomer in complex with full length protein (expected *MW*, 42 kDa). We conclude that the wild-type protein exists in different oligomeric states in solution that are affected by autoproteolytic cleavage. The C238R mutation produces an elution profile with a single peak corresponding to the MW of a dimer (61±3 kDa) (Figure 6B). The M186P mutant also appears to exist predominantly as a dimer (MW, 63±3 kDa Figure 6C), but it has a tendency to form aggregates.

It is interesting to note that both mutations force the endolysin to adopt a dimeric state, even though only one mutation, M186P, is incorporated directly at the autoproteolytic cleavage site. The results of the size exclusion analysis suggest that both the integrity of the internal cleavage site combined with how endolysin self-associates are key factors that dictate the final auto-cleavage event. M186P abolishes cleavage, as indicated by the disappearance of the intermediate 43 kDa species as well as the monomer peak from the elution profile. Abolishing side-by-side dimer formation via the introduction of a C238R mutation produces a dimeric state that is less-prone to aggregation. In addition, autoproteolytic cleavage has ceased, and the sole presence of the head-on dimer leads to an elution profile with a single dimer peak. Consequently, auto-proteolytic cleavage appears to be a spatially controlled *trans* event that occurs between endolysin monomers but only when these monomers associate to adopt the appropriate dimeric – or oligomeric – state.

To further investigate oligomerisation states and potential degradation of the CD27L samples in solution, the experimental data was analyzed in terms of possible mixtures using OLIGOMER [24]. The extended head-on and compact side-by-side dimer models and their individual domains were used to generate form-factor files for a fitting procedure, where volume fractions of each component present were determined that minimize the discrepancy between the theoretical scattering of the components and the experimental data (Table 3). The contribution from the potential degraded products including the lysed

side-by-side dimer with a missing catalytic domain, dimers of C-terminal domains and the individual domains were pooled together as an additional component. The C238R data is described exclusively by the extended head-on dimer component scattering. The head-on dimer is also the dominant component in solution for wild-type, but the other components show noticeable contributions providing the best description of the wild-type data. This result further explains the low apparent molecular mass determined from the wild-type SAXS data, and also why the individual structures show such poor fits to the wild-type data (Figure 5A).

**Specific cross-linking confirms existence of head-on dimer in the CTP1L endolysin**

To independently verify oligomerization of the CTP1L endolysin, we cloned the C-terminal domain alone and expressed it in *E. coli*. We introduced an amber stop codon at position Y212 which sits an alpha helix a1 (Y212pBpa). We also introduced an amber stop codon at position Y260, which is situated on alpha helix a3 (Y260pBpa). Both alpha helix a1 and a3 are involved in head-on dimerization (Figure 2B and 4A, B). We then expressed both amber mutants in the presence of the cross-linkable amino acid p-benzoyl-L-phenylalanine (pBpa) together with a pBpa specific tRNA and a tRNA synthetase that are capable of placing the pBpa at the position of the amber stop codon. In this way, a light sensitive cross-linker is introduced with a specific radius of interaction of approximately 10 Ångstrom [25]. The incorporation of the unnatural amino acid was confirmed for both mutants by tryptic digest, followed by mass spectrometry. We show that upon exposure to UV light, both the Y212pBpa and the Y260pBpa mutants show an additional band on an SDS-PAGE gel at double the molecular weight of the C-terminal domain alone (Figure 7A), whereas the unexposed and the wild-type protein do not show any cross-linked material. The band with elevated molecular weight was treated with trypsin and analyzed by mass spectrometry and it was confirmed that it contained the C-terminal domain of CTP1L. Since the pBpa cross-linking is quite specific, we conclude that the head-on dimer is also formed by the C-terminal domain of CTP1L in solution. Finally, we introduced an amber stop codon in the full length CTP1L endolysin replacing Y212, which was most efficient in cross-linking the C-terminal domain alone (Figure 7B). The full length CTP1L protein is cross-linked only upon exposure to UV light and forms a mixture of full length CTP1L dimers, dimers of the C-terminal domain alone as well as a species that based on the molecular weight consists of one full length CTP1L and a C-terminal domain fragment. All bands were analyzed by tryptic digest followed by peptide fingerprinting with mass spectrometry and this confirmed the bands contained the C-terminal domain of CTP1L.

The oligomerization states of the CTP1L fragments observed by cross-linking reinforce the interpretation of the size exclusion chromatography and light scattering experiments done on the CD27L endolysin, showing that both the cleavage and the oligomerization occur in both endolysins.

**Inhibition of autoproteolytic cleavage inactivates the related endolysin CTP1L**

To verify whether the autoproteolytic cleavage affects the activity of the endolysins containing the C-terminal domain, we performed cell lysis on *C. difficile* cultures with recombinant CD27L wild type and mutants using turbidity reduction assays (Figure 7C, D). We observed no difference in lysis efficiency between the wild type protein and mutants which prevent/reduce cleavage either at the cleavage site (M186P) or by affecting the side-by side (C238R and C238S) or head-on dimers (W207A, W207R and Y262A). This establishes at the least that these mutants are enzymatically active, but it does not resolve whether autocleavage plays a role in endolysin function. Introduction of mutations at the catalytic site (H84A and E144A) did abolish lytic activity (Figure 7D) but did not affect cleavage (Figure 3A). It was previously established that CD27L applied externally to *C difficile* cells shows the same efficiency for full length protein as well as for a construct that contains the enzymatic domain alone [7]. Therefore, the lysis assay is insensitive to the trigger and release function of the C-terminal domain for CD27L.

However, CTP1L is only active as an intact, full length protein, and the enzymatic domain alone does not lyse *C. tyrobutyricum* cultures [14]. Lysis of *C. tyrobutyricum* cultures by wild-type CTP1L is robust (Figure 7D), leading to a drop in optical density (OD) at 600 nm. The mutants show a drop in lysis efficiency that is proportional to the reduction in autocleavage. The V195P mutant is still somewhat active, whereas the T221R and T221C mutants show no lysis at all. We verified that these mutants are similar in secondary structure to the wild-typeWe conclude that in the context of an externally applied recombinant endolysin, CTP1L depends on the autoproteolytic cleavage of its C-terminal domain to lyse *C. tyrobutyricum*.

*Discussion*

Bacteriophages release endolysins at the end of the phage life cycle to lyse the host bacterial cell following a well-timed trigger mechanism [9]. The molecular mechanisms underlying such a trigger are unknown, but it is thought that endolysins are activated after the formation of holin lesions in the bacterial cell membrane [10]. When the endolysins

pass from the cytosol to the extra-cellular environment, they will undergo a substantial change in environment and this may activate the endolysin.

The crystal structure suggests that CD27L exists in two distinct dimeric states. We show indirectly that one of the dimeric states is associated with an autocleavage mechanism, because a mutation (C238R) in the side-by-side dimer interface reduces autocleavage. Endolysin dimerization has been shown for other bacteriophage species, and the dimerization seems to influence endolysin activity. The pneumococcal autolysin LytA does dimerize into a conformation resembling the side-by-side dimer presented here, and it was suggested that the dimer conformation may contribute to its activity [26]. The CPL-1 phage endolysin that targets *Streptococcus pneumoniae* was engineered to stabilize the (what we call) side-by-side dimerization mode, and this led to a ten fold increase in its activity [27].

We propose that the head-on dimer is more prevalent in the cytosol, whereas the side-by-side dimer is more prevalent in the reductive environment of the extra-cellular space (Figure 8). There is an equilibrium between the dimer states, and the transition between the states is facilitated by a tetrameric intermediate (as observed in the crystal structures of the C-terminal portion of CD27L, Figure 4E). The oligomeric switch can be described in terms of an Monod-Wyman-Changeux mechanism [28], with a 'tensed' state that represents the inactive endolysin and a 'relaxed' state that represents the active endolysin. We propose that the 'tensed' state is related to the head-on dimer, because autocleavage is reduced when this dimer is most prominent. The 'relaxed' state is the side-by-side dimer, which promotes autocleavage and the release of the catalytic domain from the C-terminal domain.

Autocleavage increases the action radius of the catalytic module, and as previously suggested [29], the small globular size of this enzyme may allow it to further penetrate the bacterial cell wall which may act as a sieve. Bacteriophages have been shown to use a mechanism of autocleavage and oligomerization when entering the bacterial cell wall upon infection [30]. Some bacterial toxins are activated upon autocleavage [31,32]. We have not been able to identify residues that catalyze the cleavage of the catalytic domain, but we managed to switch the cleavage off with a point mutation (M186P) at the hinge of the linker. The presence of a methionine at this position for CD27L seems to be of significance, as can be seen from the sequence alignment between lysins with a similar domain (Figure 2A). According to the sequence alignment presented in Figure 2, all lysins that have a cysteine present at position 238, also have a methionine at the start of the

domain. It is interesting to note in this respect that a chimera between the catalytic domain of CS74L and the C-terminal domain of CD27L (CS74L$_{1-177}$-CD27L$_{180-270}$) also cleaves off its C-terminal domain [33]. The C-terminal domain could therefore be involved in autonomous self-cleavage, but this needs to be further investigated.

At this stage, we can only speculate about the role of the side-by-side dimer in the autocleavage mechanism. We believe that this dimerization mode will affect the conformation of the linker that connects the two domains, possibly bringing two linkers within close proximity. The methionine may be involved in cis- (within the linker itself) or trans (in an exchange between the two linkers) autoproteolysis, such as is observed for other bacterial enzymes that undergo maturation [34]. This would represent a new form of protein splicing, involving two copies of the endolysin, rather than a single autonomous splicing unit such as is observed in inteins [35]. We are in the process of further investigating this splicing mechanism.

We have shown that autocleavage is an intrinsic property of two endolysins targeting Clostridia, and we believe that this mechanism occurs in other endolysin systems as well. The most potent lysin identified to date (PlyC) consists of two components that are expressed independently [16]. Structural characterization revealed that one component provides dual catalytic activity, whereas the other component is an octomeric cell wall binding unit. The lack of a covalent link between the enzymatic portion and the cell wall binding domain is probably key for its increased potency. We therefore believe that the engineered clustering of endolysins through a controlled oligomerization of the C-terminal domains may lead to more efficient enzymes with high specificity. This opens new opportunities to produce recombinant phage or endolysins that can lyse specifically pathogenic bacteria without affecting the microbiome overall.

*Materials and Methods*

**Protein expression, purification and crystallization**

The nucleotide sequence of the full-length endolysins CD27L and CTP1L mutant V195P, as well as the C-terminal domain CD27L$_{180-270}$ were inserted in pET15b, containing an N-terminal His tag and a thrombin cleavage site as described previously. Constructs were expressed in *E. coli* BL21(DE3) grown in Lysogeny broth (LB) media until an OD$_{600}$ ~ 0.6 was induced with 1 mM isopropyl-β-D-thio-galactopyranoside for overnight expression at

21°C. Protein expressing cells were harvested by centrifugation (5000 g, 30 min) and the supernatant discarded. Pelleted cells were lysed chemically in lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl, 1% Triton X-100, 10 mM Imidazole, 1 mg/ml Lysozyme, 25 U/ml Benzonase nuclease) for 30 min at 4°C. Lysed cell extract was centrifuged (27000 g, 40 min) and supernatant containing His-tagged endolysin purified by nickel-nitrilotriacetic acid (Ni-NTA) purification (Qiagen). Protein was eluted in a final elution buffer of 50 mM Tris pH 8.0, 150 mM NaCl, 200 mM Imidazole. Proteins were purified for crystallization by size exclusion chromatography using an Aekta liquid chromatography system (Amersham Biosciences) and S75 10/300 GL (Tricorn) column (GE Healthcare) in 20 mM HEPES, pH 7.4. The protein was concentrated to 10 mg/mL as measured by UV absorption at 280 nm. Protein crystals for degraded CD27L, that ultimately only contained the C-terminal domain, were obtained from a mother liquor containing 10 – 20 % PEG 4000 and 20 mM Tris pH 8.0. Crystals of the construct containing the C-terminal domain of CD27L and an N-terminal His tag were obtained from a mother liquor of 10 % PEG 20K and 20 mM Tris pH 8.0. For the CTP1L V195P mutant, crystals were obtained from a mother liquor containing 20mM TRIS pH8.0 and 6 % PEG 8000.

**Crystal Structure determination of CD27L**

The C-terminal domain of CD27L was first solved by single-wavelength anomalous dispersion using a mercury derivative (Table I). Crystals of the CD27L C-terminal domain construct alone with an N-terminal His tag were soaked in a cryo-protecting solution containing 15 % PEG 20K, 100 mM Tris pH8, 10 % (v/v) glycerol and the derivative 1 mM of Ethyl-mercury phosphate for a few minutes prior to freezing. A data set was collected on the X12 beamline at EMBL Hamburg, which is equipped with a MAR225 CCD detector. The crystal diffracted to a resolution of 3.5 Å, and the space group was $P2_1$. All the X-ray data were indexed, merged and scaled with DENZO and Scalepack [36]. The crystal contained eight copies of the C-terminal domain in the asymmetric unit, and 8 mercury sites were identified with SHELXD [37]. Density modification was performed with PARROT, and an initial model was built with BUCCANEER [38]. This model was used in PHASER [39] to further improve the experimental phases and to find 5 additional mercury sites after several iterations.

A native X-ray data set was collected on PROXIMA I at the Soleil Synchrotron (Gif-sur-Yvette, France), using a Q315 CCD detector from ADSC. The crystal diffracted to 2.3 Å and belonged to space group $P2_12_12_1$. The initial model was then used in molecular replacement using MOLREP [40] to identify the contents of the crystals grown from initial full length CD27L. It was determined that these crystals contained six copies of the C-terminal domain in the asymmetric unit. The structure was refined with Refmac5 [41] to an R factor of 19.8 % (Rfree = 25.6 %). The stereochemistry of the model contained 98.2 % of the residues within the favored areas of the Ramachandran plot according to Molprobity [42], and no residues in the disallowed regions.

**Crystal Structure determination of the C-terminal domain of CTP1L mutant V195P**

A native X-ray data set was collected on the EMBL beamline P14 at the PETRA3 synchrotron (Hamburg, Germany) using a MAR225 CCD detector. Although the crystal probably diffracted to at least 1.5 Ångstrom resolution, we were only able to collect usable data to a resolution of 2.1 Ångstrom due to a limited detector geometry. The crystal diffraction also suffered from ice rings, limiting the completeness of the data to 92 %. Nevertheless, it was straightforward to solve the structure of the C-terminal domain of CTP1L by molecular replacement with MOLREP [40] using the C-terminal domain of CD27L as a search model, since there is only one copy of the molecule in the asymmetric unit. The structure was refined with Refmac5 to an R factor of 17.2 % (Rfree = 26.4 %), and the electron density is of good quality. The stereochemistry of the refined model contained 98.8 % of the residues within the favored areas of the Ramachandran plot according to Molprobity [42], and no residues in the disallowed regions.

**Mutagenesis of CD27L and CTP1L mutants**

The mutants of CD27L and CTP1L were created by PCR site-directed mutagenesis following the Quikchange method (Stratagene). Plasmid pET15b-*cd27l* (Mayer 2008, 2010) was used as template DNA. Complementary primer pairs for each mutation were used for whole plasmid mutagenesis PCR performed using Phusion polymerase (NEB). Complementary primers can be found in Supplementary Table S1. Template DNA was digested by *Dpn*I (NEB) before transformation into competent *E. coli* DH5α (Invitrogen). Plasmid DNA was obtained by Miniprep (Qiagen) for sequence confirmation. Mutants were expressed and purified using the same method as wild-type CD27L.

**SDS-PAGE analysis**

Samples of purified CD27L wild-type and mutants were mixed with reducing Laemmli buffer, heated for 5 minutes at 75 °C and subjected to 15% SDS polyacrylamide gel electrophoresis. For Coomassie Blue staining, the SDS-PAGE gel was incubated respectively in Coomassie Blue staining solution (0.125% Coomassie Blue, 45% ethanol, 10% acetic acid), destaining solution (40% ethanol, 10% acetic acid) and drying solution (2% glycerol, 20% ethanol).

**Intact Protein Sample Analysis by LC-MS**

Protein samples (around 2 mg/mL) were acidified using 1% formic acid solution and transferred to vials prior to LC-MS analysis. Desalting and protein separation were carried out using an Acquity UPLC system (Waters) fitted with a $C_4$ column (2.1 mm x 15 cm, 5 μm particle size). The column was maintained at constant temperature (40 °C) throughout. The outlet of the column was coupled directly to a Q-Tof II mass spectrometer (Waters) using the standard ESI source in positive ion mode.

Solvent A was water, 0.1 % formic acid and solvent B was acetonitrile, 0.1 % formic acid. The samples (between 1 and 20 μL) were loaded onto the column and desalted for 5 minutes at a flow rate of 0.2 mL/min, 4% B. The proteins were then eluted from the column with a constant flow of 0.2 mL/min. During the elution step, the percentage of solvent B increased in a linear fashion from 5 % to 25 % in 1 minute, then increased to 80 % in a further 11 minutes. On the Q-Tof, a spray voltage of 3.5 kV was applied, with a cone voltage of 35 V and extraction cone at 10 V. A collision energy of 8 eV was used, with Argon in the collision cell. The desolvation temperature was set at 320°C, with a source temperature of 120°C. Data were acquired in continuum mode, over a mass range 500-3500 *m/z* with a scan time of 0.5 s and interscan delay of 0.1 s. Data were externally calibrated against a reference standard of intact myoglobin, acquired immediately after sample data acquisition. Spectra across the protein chromatographic peak(s) were summed and intact mass was calculated using the MaxEnt1 maximum entropy algorithm (Waters/Micromass) to give the zero charge deconvoluted molecular weight.

**SEC-RALS/RI/UV molecular weight determination.**

Molecular weight estimates of each CD27L variant (CD27L wild-type, CD27L C239R and CD27L M186P) were evaluated using size-exclusion chromatography in combination with right-angle light scattering (RALS), refractive index (RI) and UV ($l_{280\ nm}$) measurements

(Malvern Instruments Viscotek, RALS/RI/UV 305 TDA detector equipped with a 670 nm laser diode). All measurements were performed at room temperature. Samples were separately injected at their respective concentrations (75 mL at 6.37, 7.15 and 6.17 mg.mL$^{-1}$) onto a GE-Healthcare Tricorn S75 10/300 GL column equilibrated in 20 mM HEPES pH7.4, 500 mM NaCl at a flow rate of 0.4 mL.min$^{-1}$. The molecular weight (*MW*) of each species eluting from the SEC column were assessed using concentration (*c*) measurements derived from base-line corrected RI or UV measurements in combination with base-line corrected RALS intensities calibrated against a bovine serum albumin narrow (monomeric) standard (RALS = $c$(dn/dc)$^2$.$MW$.k$_{RALS}$; RI = $c$(dn/dc)k$_{RI}$ and; UV = $ce$k$_{UV}$, where dn/dc is the refractive index increment of unmodified protein, 0.185 mL.g$^{-1}$, k$_{RI}$, k$_{UV}$ and k$_{RALS}$ are the TDA instrument calibration constants relative to a BSA and *e* the l$_{280\ nm}$ E$_{0.1\%}$ extinction coefficient of each protein in mg.mL$^{-1}$). The *MW* correlations across the selected range of each CD27L elution peak and the final *MW* estimates quoted in the text were calculated using OmniSEC Software (Malvern Instruments). The molecular weight ranges, frequencies, modes and averages for each CD27L variant are summarised in Table S2 and Figure S1.


**SAXS data collection and shape determination**

Synchrotron radiation X-ray scattering data were collected on the EMBL X33 and P12 beamlines of the storage rings DORIS III and PETRA III (DESY, Hamburg), respectively, using PILATUS 1M and 2M pixel detectors (DECTRIS, Switzerland). For the wild-type CD27L data were acquired at X33, with 8 frames of 15 s exposure time collected. Samples were measured in a temperature controlled cell at 10°C in 20 mM HEPES buffer, 150 mM NaCl pH 7.4 at protein concentrations of 0.9 – 4.0 mg/mL. The sample-to-detector distance was 2.7 m, covering a range of momentum transfer $0.01 \leq s \geq 0.6$ Å$^{-1}$ ($s = 4\pi \sin\theta /\lambda$, where $2\theta$ is the scattering angle, and $\lambda = 1.54$ Å is the X-ray wavelength). For the C238R mutant data were acquired at P12, with 20 frames of 0.05 s exposure time collected. Solutions were measured while flowing through a temperature-controlled capillary at 10°C in 20 mM Tris buffer, 500 mM NaCl pH 7.4 at protein concentrations of 1.0 – 8.5 mg/mL. The sample-to-detector distance was 3.1 m, covering a range of momentum transfer $0.008 \leq s \geq 0.458$ Å$^{-1}$ ($s = 4\pi \sin\theta /\lambda$, where $2\theta$ is the scattering angle, and $\lambda = 1.24$ Å is the X-ray wavelength). Based on comparison of successive frames, no detectable radiation damage was observed. Data from the detector were normalised to the transmitted beam intensity, averaged and the scattering of buffer solutions subtracted. The difference curves were scaled for solute concentration and the 1.0 mg/mL (low-*s*) and 8.4 mg/mL (high-*s*)

data sets merged for modeling. All data manipulations were performed using PRIMUS [43].

The forward scattering $I(0)$ and radius of gyration, $Rg$ were determined from Guinier analysis [44], assuming that at very small angles ($s \leq 1.3/Rg$ ) the intensity is represented as $I(s)=I(0)exp(-(sRg)^2/3))$. These parameters were also estimated from the full scattering curves using the indirect Fourier transform method implemented in the program GNOM [45], along with the distance distribution function $p(r)$ and the maximum particle dimension $D_{max}$. Molecular masses (MMs) of solutes were estimated from SAXS data by comparing the extrapolated forward scattering with that of a reference solution of bovine serum albumin, and also from the hydrated-particle/Porod volume $V_p$, where molecular mass is estimated as 0.625 times $V_p$.

### *Ab initio* Shape Determination and Molecular Modelling

Low-resolution shape envelopes for all constructs were determined using the *ab initio* bead-modelling program DAMMIF [46], using both P1 and P2 symmetry. The results of 10 independent DAMMIF runs were analyzed using the program DAMAVER [47] to identify the most representative/typical models. Modeling using P2 symmetry was only attempted following the identification of excluded solvent volumes, $V_{ex}$ in models generated in P1 (slow mode) consistent with that expected for dimers (see Table 2).

Molecular modelling was conducted using, as rigid bodies and where appropriate, the crystal structures of the catalytic and the C-terminal domains of CD27L determined in this study. Rigid-body models were generated using the program CORAL [24] and 10 independent runs assessed for convergence with DAMAVER. Additional fitting of PDB files to the SAXS data was conducted using CRYSOL [23].

### Oligomeric equilibrium analysis

Using the program OLIGOMER [24], the SAXS data for both wild-type CD27L and the C238R mutant was used to model potential multicomponent mixtures of species in solution. Form factors of input PDB files were calculated using the program FFMAKER [24]. Form factors were also calculated for individual domains and substructures of the intact PDB files to represent products of autolysis and averaged. This averaging was performed as the identity of the exact solution composition of these lysis products could not be established. Volume fractions corresponding to each component (eg. extended

dimer, compact dimer and "degraded components") were determined by OLIGOMER utilising a non-negative least squares procedure.

## Subcloning and expression of *p*-benzoyl-L-phenylalanine incorporated into CTP1L

Following the principal method outlined by Farrell *et al.* [48], the photo-activated amino acid *p*-benzoyl-L-phenylalanine (*p*BPA) (BACHEM) was incorporated into the full-length CTP1L endolysin and the truncated C-terminal domain of CTP1L. CTP1L was amplified from *ctp1l*-pET15b and inserted into the pET21d vector to create pET21d. The *amber* codon (TAG) was incorporated at position Y212 or Y260 following the Quikchange method of PCR site-directed mutagenesis to generate *Y260TAG*-pET21d and *Y212TAG*-pET21d. Sub-cloning into pET21d introduced a C-terminal hexa histidine-tag to the construct, which permits selective Ni-NTA purification of full-length proteins that have only incorporated *p*BPA. C-terminal constructs Y260TAG and Y212TAG were generated by amplifying the C-terminal domains of *Y212TAG*-pET21d and *Y260TAG*-pET21d between positions V195 and K274, and the PCR products were inserted back into the pET21d vector to create *C-termY260TAG*-pET21d and *C-termY212TAG*-pET21d. As a control, the wild-type C-terminal domain was sub cloned into pET21d with no amber stop codon incorporated. *E.coli* BL21(AI) cells were transformed with pEVOL-*p*BPA (aminoacyl-tRNA synthetase/suppressor tRNA) and one of the plasmids encoding an *amber* codon containing construct: *Y212TAG*-pET21d, C-term*Y212TAG*-pET21d or *C-termY260TAG*-pET21d. Cells were grown in 500 ml Lysogeny broth (LB) media supplemented with 1 mM *p*BPA in the presence of ampicillin and chloramphenicol. When an $OD_{600}$ of 0.6 had been reached the cultures were induced with Arabinose (final concentration 0.02%) and expressed at 21°C overnight. Cells were harvested by centrifugation (5000 g, 30 min) and the supernatant discarded.

## Photo-cross-linking full length CTP1L and the C-terminal domain of CTP1L

Y212*pBPA* was Ni-NTA purified as described above and dialyzed into 25mM TRIS, pH 7.4. The protein was concentrated to 0.5 mg.ml$^{-1}$, as measured by UV absorption at 280 nm. A 1 ml aliquot of Y212*pBPA* was pipetted into a single well of a 24-well clear polystyrene plate, typically used for protein crystallography. The lid was kept on to prevent sample evaporation, and placed inside an RPR-100 UV reactor equipped with 350-nm bulbs (Rayonet). The reactor was kept at 4 °C with the cooling fan on. Proteins were exposed to UV light for 30-minute intervals during which the solutions were stirred by gentle pipetting and samples taken for each time point for SDS-PAGE analysis. As a

control, samples were also taken every 30 minutes from a parallel sample of Y212*p*BPA that was kept at 4°C in the dark with no UV exposure. Cross-linking was analyzed by SDS-PAGE by comparison of the pre-UV and post-UV exposed samples. The same photo-cross-linking experiment was performed for the C-terminal domain constructs. Ni-NTA purified C-termY212*p*BPA, C-termY260*p*BPA and wild-type were dialyzed into 25mM TRIS, pH 7.4 and concentrated to 2 mg.ml$^{-1}$ as measured by UV absorption at 280 nm. Following the same protocol as described above for the full-length Y212*p*BPA, 500 μl aliquots of each protein were pipetted into separate wells of a 24 well clear polystyrene plate and exposed to UV light for 120 minutes with stirring of the samples at 30 minute intervals. As a control SDS-PAGE samples were taken from parallel aliquots of each protein kept at 4°C in the dark with no UV exposure.

**Turbidity reduction assays**

Lysis assays were performed on fresh cells of *C. difficile* NCTC 11204 and *C. tyrobutyricum* NCIMB 9582. Cells were cultured, harvested and resuspended in PBS pH 7.3 as described previously [5, 14]. Lysis assays were performed on freshly harvested cells in 300 μl volumes with 10 μg Ni-NTA-purified protein or elution buffer. Results are the mean of duplicate assays +/- standard deviation.

**Acknowledgements**

*References*

1.    Henry M, Debarbieux L (2012) Tools from viruses: bacteriophage successes and beyond. Virology 434: 151–161. doi:10.1016/j.virol.2012.09.017.

2.    Rea MC, Alemayehu D, Ross RP, Hill C (2013) Gut solutions to a gut problem: bacteriocins, probiotics and bacteriophage for control of Clostridium difficile infection. J Med Microbiol 62: 1369–1378. doi:10.1099/jmm.0.058933-0.

3.      Meader E, Mayer MJ, Steverding D, Carding SR, Narbad A (2013) Evaluation of bacteriophage therapy to control Clostridium difficile and toxin production in an *in vitro* human colon model system. Anaerobe 22: 25–30. doi:10.1016/j.anaerobe.2013.05.001.

4.      Venugopal AA, Johnson S (2012) Current state of Clostridium difficile treatment options. Clin Infect Dis Off Publ Infect Dis Soc Am 55 Suppl 2: S71–76. doi:10.1093/cid/cis355.

5.      Mayer MJ, Narbad A, Gasson MJ (2008) Molecular characterization of a Clostridium difficile bacteriophage and its cloned biologically active endolysin. J Bacteriol 190: 6734–6740. doi:10.1128/JB.00686-08.

6.      Loessner MJ (2005) Bacteriophage endolysins--current state of research and applications. Curr Opin Microbiol 8: 480–487. doi:10.1016/j.mib.2005.06.002.

7.      Mayer MJ, Garefalaki V, Spoerl R, Narbad A, Meijers R (2011) Structure-based modification of a Clostridium difficile-targeting endolysin affects activity and host range. J Bacteriol 193: 5477–5486. doi:10.1128/JB.00439-11.

8.      Hermoso JA, García JL, García P (2007) Taking aim on bacterial pathogens: from phage therapy to enzybiotics. Curr Opin Microbiol 10: 461–472. doi:10.1016/j.mib.2007.08.002.

9.      White R, Chiba S, Pang T, Dewey JS, Savva CG, et al. (2011) Holin triggering in real time. Proc Natl Acad Sci U S A 108: 798–803. doi:10.1073/pnas.1011921108.

10.     Dewey JS, Savva CG, White RL, Vitha S, Holzenburg A, et al. (2010) Micron-scale holes terminate the phage infection cycle. Proc Natl Acad Sci U S A 107: 2219–2223. doi:10.1073/pnas.0914030107.

11.     Hermoso JA, Monterroso B, Albert A, Galán B, Ahrazem O, et al. (2003) Structural basis for selective recognition of pneumococcal cell wall by modular endolysin from phage Cp-1. Struct Lond Engl 1993 11: 1239–1249.

12.     Porter CJ, Schuch R, Pelzek AJ, Buckle AM, McGowan S, et al. (2007) The 1.6 A crystal structure of the catalytic domain of PlyB, a bacteriophage lysin active against Bacillus anthracis. J Mol Biol 366: 540–550. doi:10.1016/j.jmb.2006.11.056.

13.     Low LY, Yang C, Perego M, Osterman A, Liddington RC (2005) Structure and lytic activity of a Bacillus anthracis prophage endolysin. J Biol Chem 280: 35433–35439. doi:10.1074/jbc.M502723200.

14.     Mayer MJ, Payne J, Gasson MJ, Narbad A (2010) Genomic sequence and characterization of the virulent bacteriophage phiCTP1 from Clostridium tyrobutyricum and heterologous expression of its endolysin. Appl Environ Microbiol 76: 5415–5422. doi:10.1128/AEM.00989-10.

15.     Xu M, Arulandu A, Struck DK, Swanson S, Sacchettini JC, et al. (2005) Disulfide isomerization after membrane release of its SAR domain activates P1 lysozyme. Science 307: 113–117. doi:10.1126/science.1105143.

16.     McGowan S, Buckle AM, Mitchell MS, Hoopes JT, Gallagher DT, et al. (2012) X-ray crystal structure of the streptococcal specific phage lysin PlyC. Proc Natl Acad Sci. Available: http://www.pnas.org/content/early/2012/07/17/1208424109. Accessed 24 January 2013.

17.     Holm L, Sander C (1999) Protein folds and families: sequence and structure alignments. Nucleic Acids Res 27: 244–247.

18. Korndörfer IP, Danzer J, Schmelcher M, Zimmer M, Skerra A, et al. (2006) The crystal structure of the bacteriophage PSA endolysin reveals a unique fold responsible for specific recognition of Listeria cell walls. J Mol Biol 364: 678–689. doi:10.1016/j.jmb.2006.08.069.

19. Pérez-Dorado I, Campillo NE, Monterroso B, Hesek D, Lee M, et al. (2007) Elucidation of the molecular recognition of bacterial cell wall by modular pneumococcal phage endolysin CPL-1. J Biol Chem 282: 24990–24999. doi:10.1074/jbc.M704317200.

20. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66: 486–501. doi:10.1107/S0907444910007493.

21. Krissinel E (2011) Macromolecular complexes in crystals and solutions. Acta Crystallogr D Biol Crystallogr 67: 376–385. doi:10.1107/S0907444911007232.

22. Kabsch W (1978) A discussion of the solution for the best rotation to relate two sets of vectors. Acta Crystallogr Sect A 34: 827–828. doi:10.1107/S0567739478001680.

23. Svergun D, Barberato C, Koch MHJ (1995) CRYSOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. J Appl Crystallogr 28: 768–773. doi:10.1107/S0021889895007047.

24. Petoukhov MV, Franke D, Shkumatov AV, Tria G, Kikhney AG, et al. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. J Appl Crystallogr 45: 342–350. doi:10.1107/S0021889812007662.

25. Hino N, Okazaki Y, Kobayashi T, Hayashi A, Sakamoto K, et al. (2005) Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid. Nat Methods 2: 201–206. doi:10.1038/nmeth739.

26. Fernández-Tornero C, García E, López R, Giménez-Gallego G, Romero A (2002) Two new crystal forms of the choline-binding domain of the major pneumococcal autolysin: insights into the dynamics of the active homodimer. J Mol Biol 321: 163–173.

27. Resch G, Moreillon P, Fischetti VA (2011) A stable phage lysin (Cpl-1) dimer with increased antipneumococcal activity and decreased plasma clearance. Int J Antimicrob Agents 38: 516–521. doi:10.1016/j.ijantimicag.2011.08.009.

28. Changeux J-P (2011) Allostery and the Monod-Wyman-Changeux Model After 50 Years. Annu Rev Biophys. Available: http://www.ncbi.nlm.nih.gov/pubmed/22224598. Accessed 3 April 2012.

29. Low LY, Yang C, Perego M, Osterman A, Liddington R (2011) Role of net charge on catalytic domain and influence of cell wall binding domain on bactericidal activity, specificity, and host range of phage lysins. J Biol Chem 286: 34391–34403. doi:10.1074/jbc.M111.244160.

30. Xiang Y, Leiman PG, Li L, Grimes S, Anderson DL, et al. (2009) Crystallographic insights into the autocatalytic assembly mechanism of a bacteriophage tail spike. Mol Cell 34: 375–386. doi:10.1016/j.molcel.2009.04.009.

31. Egerer M, Giesemann T, Jank T, Satchell KJF, Aktories K (2007) Auto-catalytic cleavage of Clostridium difficile toxins A and B depends on cysteine protease activity. J Biol Chem 282: 25314–25321. doi:10.1074/jbc.M703062200.

32. Reineke J, Tenzer S, Rupnik M, Koschinski A, Hasselmayer O, et al. (2007) Autocatalytic cleavage of Clostridium difficile toxin B. Nature 446: 415–419. doi:10.1038/nature05622.

33.    Mayer MJ, Gasson MJ, Narbad A (2012) Genomic sequence of bacteriophage ATCC 8074-B1 and activity of its endolysin and engineered variants against Clostridium sporogenes. Appl Environ Microbiol 78: 3685–3692. doi:10.1128/AEM.07884-11.

34.    Buller AR, Freeman MF, Wright NT, Schildbach JF, Townsend CA (2012) Insights into cis-autoproteolysis reveal a reactive state formed through conformational rearrangement. Proc Natl Acad Sci U S A 109: 2308–2313. doi:10.1073/pnas.1113633109.

35.    Paulus H (2000) Protein splicing and related forms of protein autoprocessing. Annu Rev Biochem 69: 447–496. doi:10.1146/annurev.biochem.69.1.447.

36.    Otwinowski Z, Borek D, Majewski W, Minor W (2003) Multiparametric scaling of diffraction intensities. Acta Crystallogr A 59: 228–234.

37.    Schneider TR, Sheldrick GM (2002) Substructure solution with SHELXD. Acta Crystallogr D Biol Crystallogr 58: 1772–1779.

38.    Cowtan K (2012) Completion of autobuilt protein models using a database of protein fragments. Acta Crystallogr D Biol Crystallogr 68: 328–335. doi:10.1107/S0907444911039655.

39.    McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, et al. (2007) Phaser crystallographic software. J Appl Crystallogr 40: 658–674. doi:10.1107/S0021889807021206.

40.    Vagin A, Teplyakov A (2010) Molecular replacement with MOLREP. Acta Crystallogr D Biol Crystallogr 66: 22–25. doi:10.1107/S0907444909042589.

41.    Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, et al. (2011) REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr 67: 355–367. doi:10.1107/S0907444911001314.

42.    Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66: 12–21. doi:10.1107/S0907444909042073.

43.    Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J Appl Crystallogr 36: 1277–1282. doi:10.1107/S0021889803012779.

44.    Guinier A (n.d.) La diffraction des rayons X aux tres petits angles: applications a l'etude de phenomenes ultramicroscopiques. Available: http://publikationen.ub.uni-frankfurt.de/frontdoor/index/index/docId/15232. Accessed 19 August 2013.

45.    Semenyuk AV, Svergun DI (1991) GNOM – a program package for small-angle scattering data processing. J Appl Crystallogr 24: 537–540. doi:10.1107/S002188989100081X.

46.    Franke D, Svergun DI (2009) DAMMIF, a program for rapid *ab-initio* shape determination in small-angle scattering. J Appl Crystallogr 42: 342–346. doi:10.1107/S0021889809000338.

47.    Volkov VV, Svergun DI (2003) Uniqueness of *ab initio* shape determination in small-angle scattering. J Appl Crystallogr 36: 860–864. doi:10.1107/S0021889803000268.

48.    Farrell IS, Toroney R, Hazen JL, Mehl RA, Chin JW (2005) Photo-cross-linking interacting proteins with a genetically encoded benzophenone. Nat Methods 2: 377–384. doi:10.1038/nmeth0505-377.

49. Gouet P, Robert X, Courcelle E (2003) ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res 31: 3320–3323. doi:10.1093/nar/gkg556.
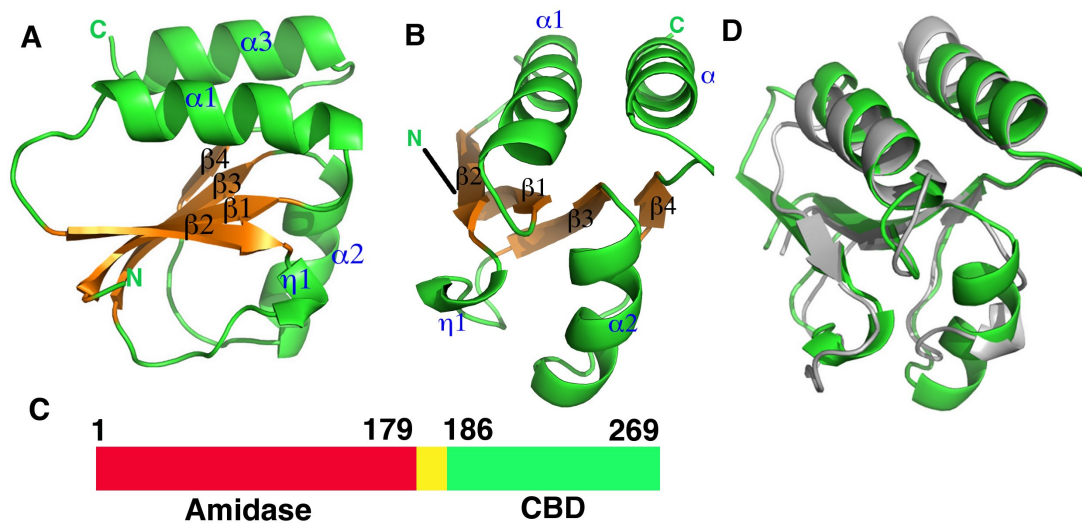
*Figures*



**Figure 1: Overall structures of the proteolytic fragments of the CD27L and CTP1L endolysins reveal a novel fold for the C-terminal domain.** A. Ribbon diagram of the C-terminal domain of endolysin CD27L. The beta sheet at the core of the domain is colored in gold and the secondary structure elements have been labeled. B. View of the C-terminal domain at a 90 degree angle from the view presented in A. C. Linear map of the domain organization of CD27L, with the enzymatic portion (Amidase) in red, the linker in yellow and the C-terminal domain in green. D. Superposition of the ribbon diagrams of the C-terminal domains of (gray) and CD27L (green). Molecular graphics were produced with Pymol (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.).
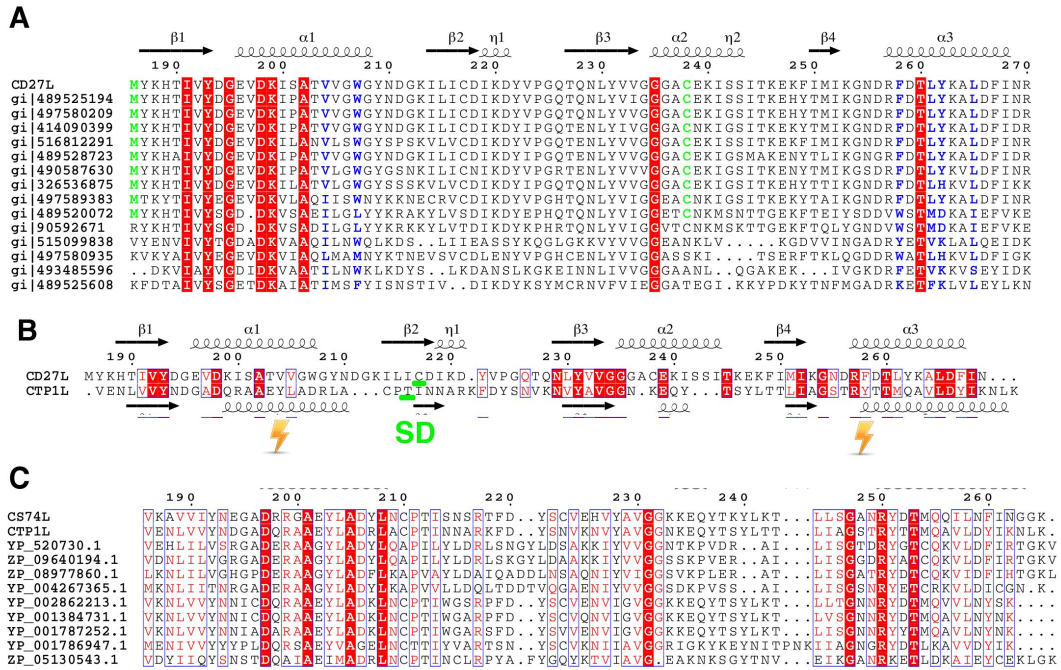
**Figure 2: Sequence alignment of the C-terminal domains of CD27L and CTP1L showing this fold is prevalent among lysins that target *Clostridia*.** A. Sequence alignment of the C-terminal domain of CD27L and other sequences with a significant BLAST score (E < 0.01) produced with ESPRIPT [49]. Conserved residues are colored red. The secondary structure of the C-terminal domain of CD27L is depicted with arrows for beta strands and curls for alpha helices. Hydrophobic residues that contribute to the head-on dimer interface are colored blue, and the cysteine residue involved in the side-by-side dimer formation is colored green. B. Structure-based sequence alignment of the C-terminal domain of CD27L with the C-terminal domain of CTP1L. Residues where the cross-linking amino acid pBpa is inserted are marked by a blitz, and the area involved in the side-by-side dimer is marked SD. C. Sequence alignment of the C-terminal domain of CTP1L with other sequences that give a significant BLAST score.
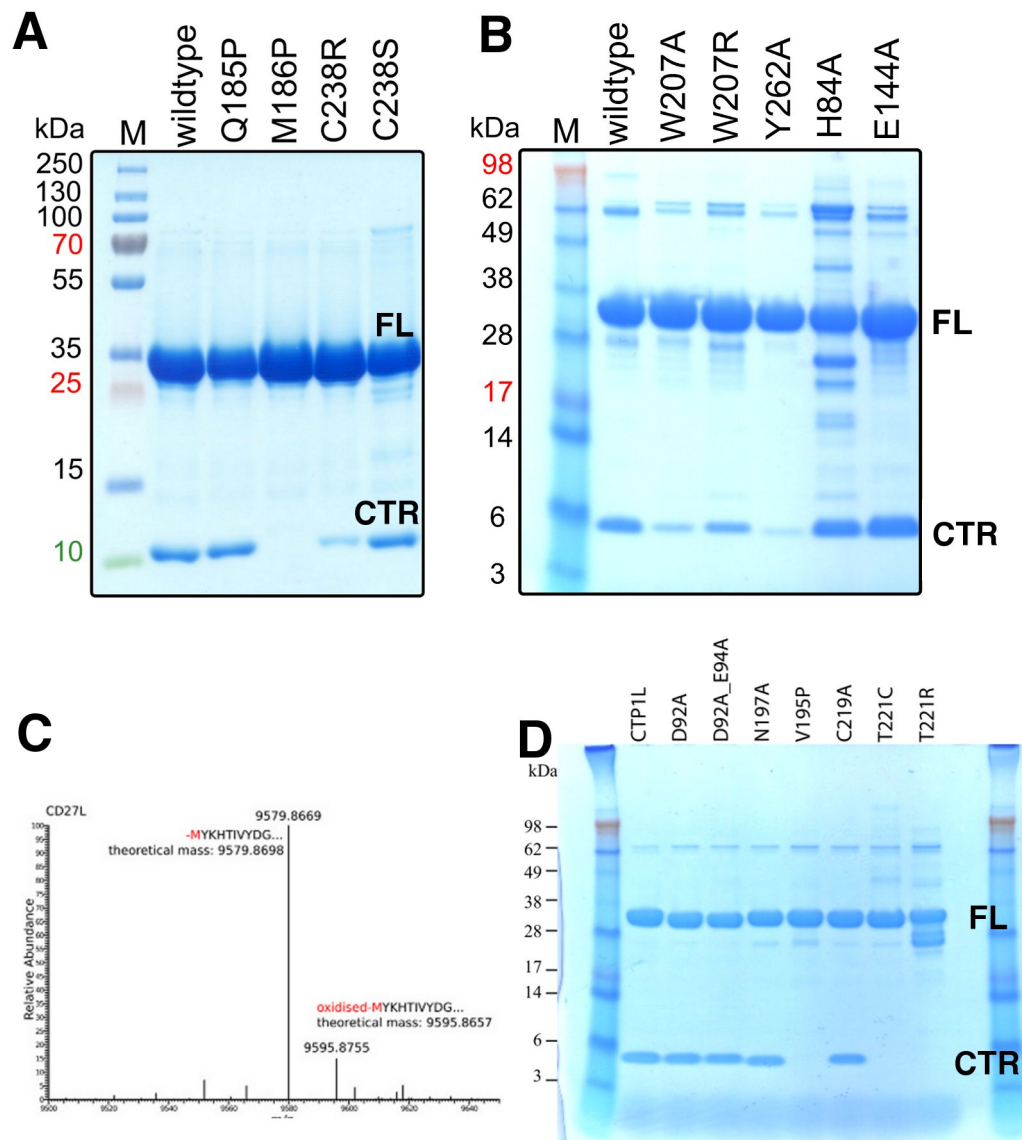
**Figure 3: The cleavage of the CD27L and CTP1L endolysins occur at a specific site and is affected by mutagenesis.** A, B. SDS-PAGE gel of equal amounts of purified samples of CD27L wild-type and mutants, showing the full length (FL) protein and the 10 kDa band that corresponds to the C-terminal domain (CTR). C. LC-MS spectrum of the eluted fraction corresponding to C-terminal domain of CD27L, showing that the proteolysis occurs at M186. D. SDS-PAGE gel of equal amounts of purified wild-type and mutants. The mutants D92A and D92A_E94A that neutralize the catalytic residues of the catalytic domain do not affect autocleavage, whereas V195P, T221C and T221R inhibit autocleavage.
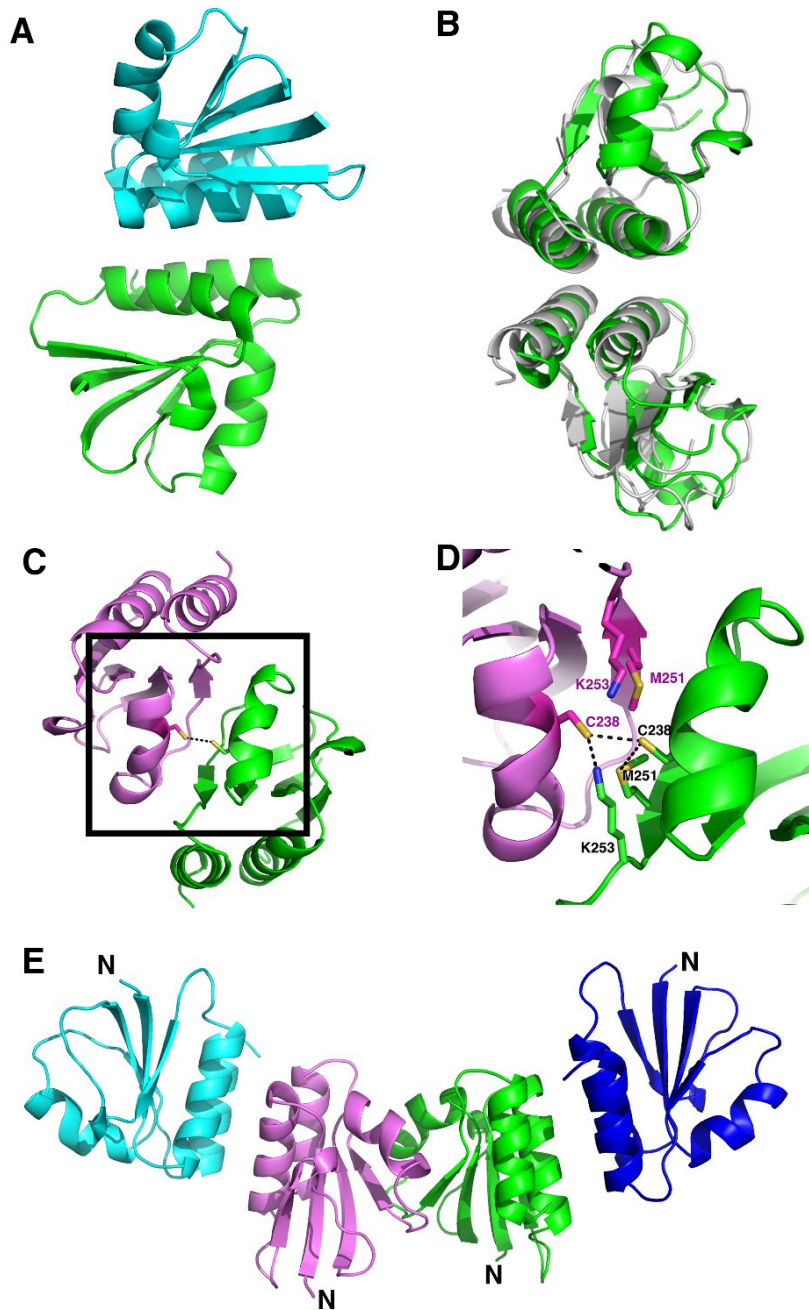
**Figure 4: Overview of the oligomerization modes observed in the crystal structure of the proteolytic fragment of endolysin CD27L.** A. Ribbon diagram of the head-on dimer configuration with one monomer coloured green and the other blue. B. The head-on dimer configuration rotated 90 degrees along the horizontal axis, with the head-on dimer of CTP1L superimposed in gray. C. Ribbon diagram of the side-by-side dimer, with one monomer in green and the other in pink. The side chain of cysteine C238 is shown as ball and stick for both monomers. D. Close-up of the side-by-side dimer interface, showing the interactions between cysteine C238 and symmetry mates of lysine K251 and methionine M253 shown in sticks. E. Tetrameric assembly observed in the crystal structure, with the side-by-side dimer in the center, where the N-terminus points in the same direction for both monomers, and two head-on dimers at the periphery.
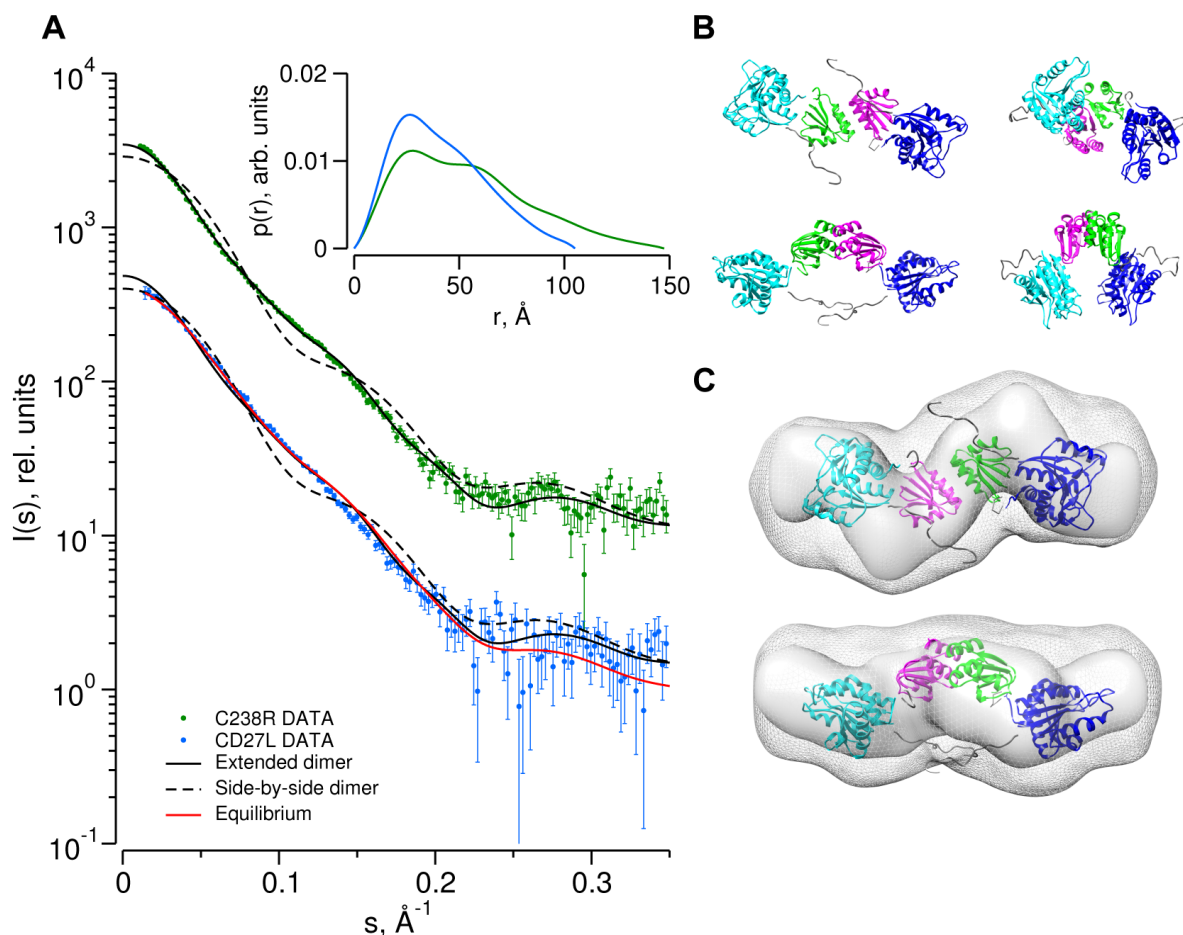
**Figure 5: Determination of the dimer configuration of wild-type CD27L and mutant C238R in solution by SAXS.** A. Overlay of the experimental scattering curves for wild-type CD27L (blue circles) and the C238R mutant (green circles) with the calculated scattering curves from a head-on-dimer/extended configuration (solid black line) and a side-by-side/compact dimer configuration (broken black line), built as a composite model from the crystal structures of the catalytic domain of CD27L (PDB code 3QAY) and the structure of the C-terminal domain presented in the paper, using the full length endolysin with a very similar catalytic domain PlyPSA amidase (PDB code 1XOV) to compose the full length structure. Missing regions of the structure (n-terminal histidine tag and interdomain linker residues) were refined against the SAXS data, keeping the domains fixed using the program CORAL. Also shown is the fit of the equilibrium model determined from OLIGOMER (solid red line) to the wild-type data. The inset shows the corresponding real-space distance-distribution functions $p(r)$ determined by indirect Fourier transformation. B. Cartoon representations of the head-on dimer (left) and side-by-side dimer (right) configurations of CD27L C. Overlay of the reconstituted domains of CD27L C238R refined by rigid body modeling (cartoon) with the best and average SAXS bead models in surface and mesh presentation, respectively. At the center of the dimer sit two C-terminal domains in a head-on configuration. The catalytic domains are at the exterior, and the N-terminal His tag and inter-domain linker are shown as grey spheres. Models rotated by 90° are shown below each corresponding reference structure.
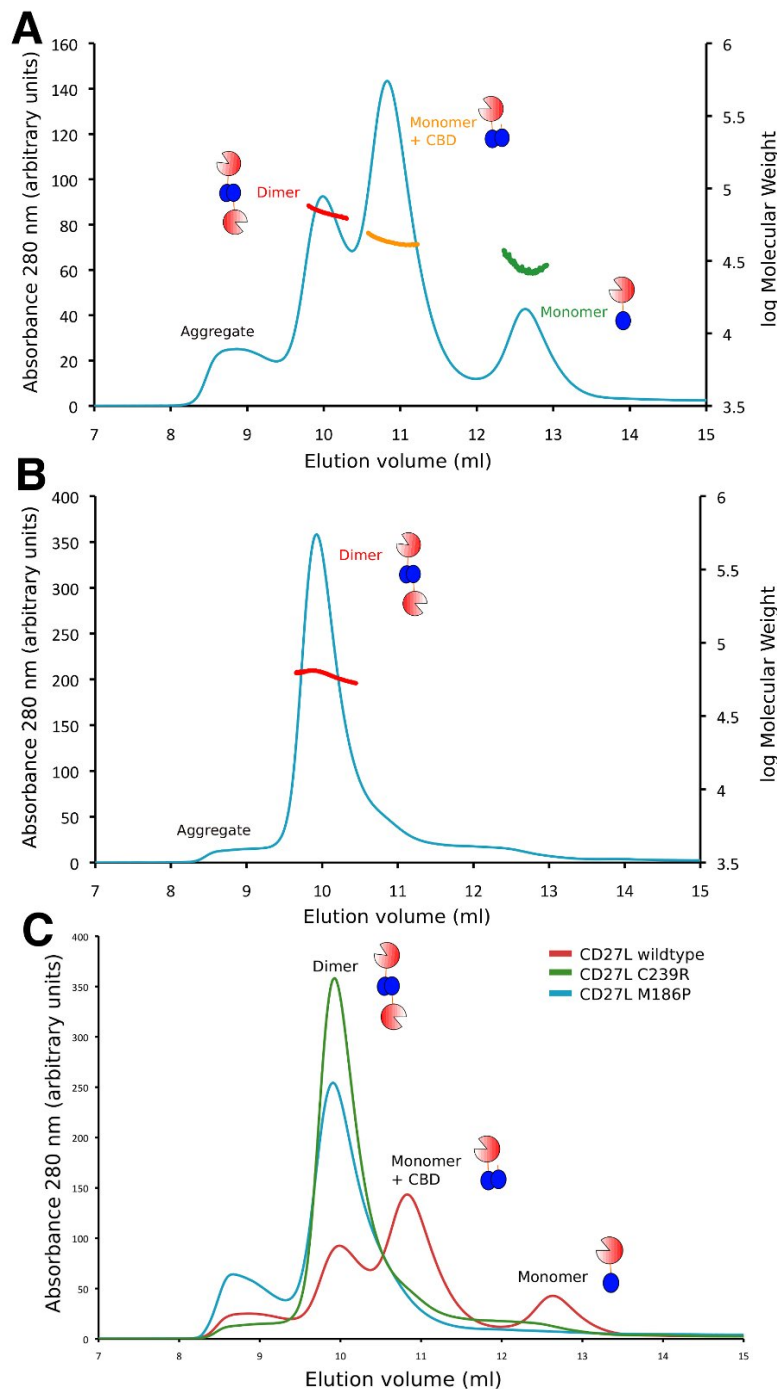
**Figure 6: Oligomerization of CD27L in solution for wild-type and the M186P and C238R mutants.** A. Size exclusion chromatography elution profile of the UV trace at 280 nm combined with static light scattering of the wild-type CD27L. There are three peaks that correspond to the dimer (68±4 kDa), a mixture of monomer and cleaved C-terminal domain (43±2 kDa), and the monomer (33±7 kDa). B. Elution profile of the CD27L C238R mutant, showing a single peak with a dimeric species (62 kDa). C. Superposition of the elution profiles of the wild-type and the C238R and M186P mutants, showing the relative peak heights at equivalent amounts of sample loaded, with increased aggregation for the M186P mutant.
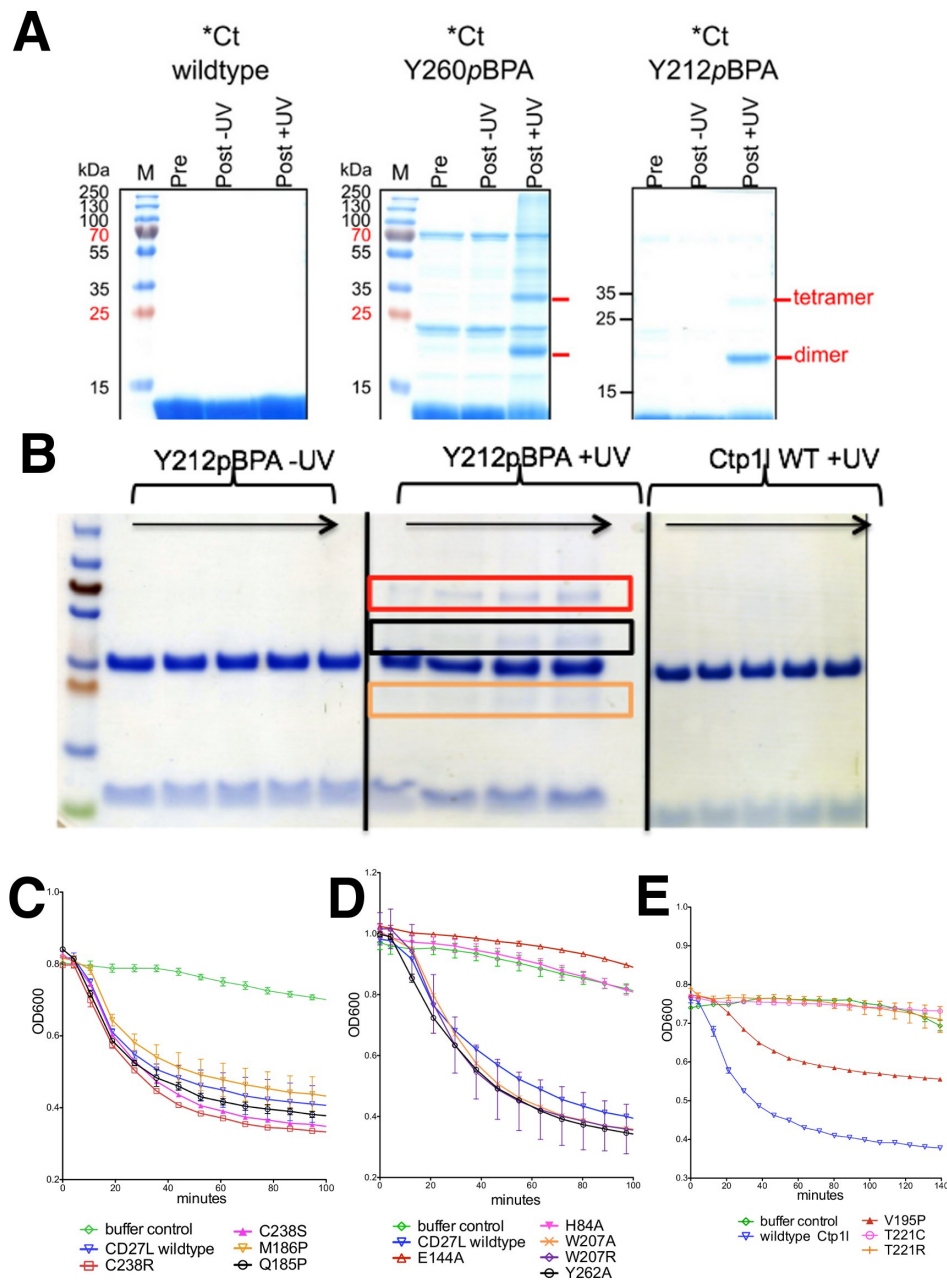
**Figure 7: Cross-linking of oligomers and lysis activity of endolysins CD27L and CTP1L.** A. SDS-PAGE of the C-terminal domain of CTP1L with or without the light sensitive cross-linker pBpa incorporated. B. SDS-PAGE of the full-length endolysin with the Y212pBpa cross-linker mutant, showing the presence of 1) a full-length dimer, 2) a C-terminal domain dimer and 3) an oligomer consisting of one C-terminal domain and a full-length CTP1L endolysin molecule. C. Lysis assays of 10 mg of recombinant CD27L applied to a culture of *C. difficile* showing that the autocleavage mutants do not affect lysis *in vitro* D. Lysis assays on CD27L showing that amidase active site mutants are not active. E. Lysis assays of CTP1L on *C. tyrobutyricum* cells showing the effect of mutants that reduce autocleavage (V195P, T221C and T221R).
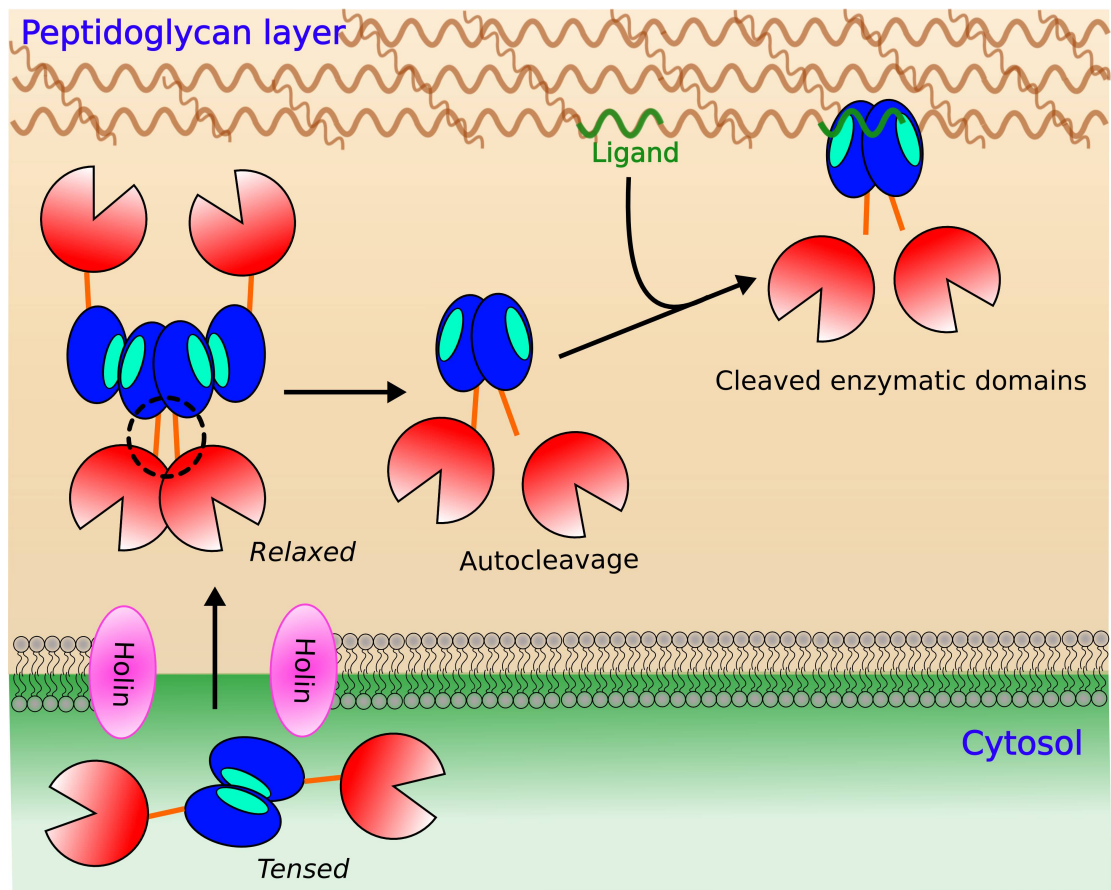
**Figure 8: Schematic representation of the proposed trigger and release mechanism of endolysin CD27L.** The endolysin is sequestered in the cytoplasm in a 'tensed' inactive state that corresponds to the head-on dimer, until holin lesions are formed. The endolysin can pass through the lesions to the extra-cellular space, where it will convert to a side-by-side dimer mode through a tetrameric intermediate. This triggers the release of the enzymatic domains, which may be further accelerated when the c-terminal domain binds to a ligand on the peptidoglycan layer. The release of the enzymatic domain expedites the digestion of the bacterial cell wall.

**Table I  Data collection and refinement statistics**

| | CD27L Proteolytic fragment | CTP1L V195P mutant (proteolytic fragment)# | ΔN-CD27L Hg derivative |
|---|---|---|---|
| **Data collection** | | | |
| Space group | $P2_12_12_1$ | I222 | $P2_1$ |
| Cell dimensions | | | |
| $a$, $b$, $c$ (Å) | 75.3, 82.1, 83.8 | 44.9, 48.8, 77.2 | 63.1, 84.7, 65.3 |
| a, b, g  (°) | 90.0, 90.0, 90.0 | 90.0, 90.0, 90.0 | 90.0, 92.0, 90.0 |
| Wavelength (Å) | 0.970 | 1.223 | 0.998 |
| Resolution (Å) | 20 – 2.24 (2.37-2.24) | 20 – 2.10 (2.17 – 2.10) | 20 – 3.5 (3.66 – 3.50) |
| $R_{sym}$ or $R_{merge}$ | 12.8 (59.8) | 2.5 (4.4) | 23.8 (55.8) |
| $I$ / s$I$ | 7.0 (2.0) | 48.9 (29.0) | 4.7 (2.6) |
| Completeness (%) | 97.9 (93.8) | 92 (55.5) | 99.9 (99.9) |
| Redundancy | 2.7 (2.6) | 5.7 (4.4) | 4.8 (4.8) |
| **Refinement** | | | |
| Resolution (Å) | 30 – 2.24 | 20 – 2.10 | |
| No. reflections | 24189 | 4489 | |
| $R_{work}$ / $R_{free}$ | 20.2/24.6 | 17.2 (26.4) | |
| No. atoms | | | |
| Protein | 3978 | 645 | |
| Ligand/ion | n/a | n/a | |
| Water | 495 | 90 | |
| $B$-factors | | | |
| Protein | 45 | 22 | |
| Ligand/ion | n/a | n/a | |
| Water | 57 | 32 | |
| R.m.s. deviations | | | |
| Bond lengths (Å) | 0.01 | 0.01 | |
| Bond angles (°) | 1.3 | | |

. *Values in parentheses are for highest-resolution shell.
# The data collection was affected by ice rings and a limited detector geometry

**Table 2 SAXS Data collection and derived parameters for CD27L.** Abbreviations: $M_r$: molecular mass; $R_g$: radius of gyration; $D_{max}$: maximal particle dimension; $V_p$: Porod volume; $V_{ex}$: Particle excluded volume.
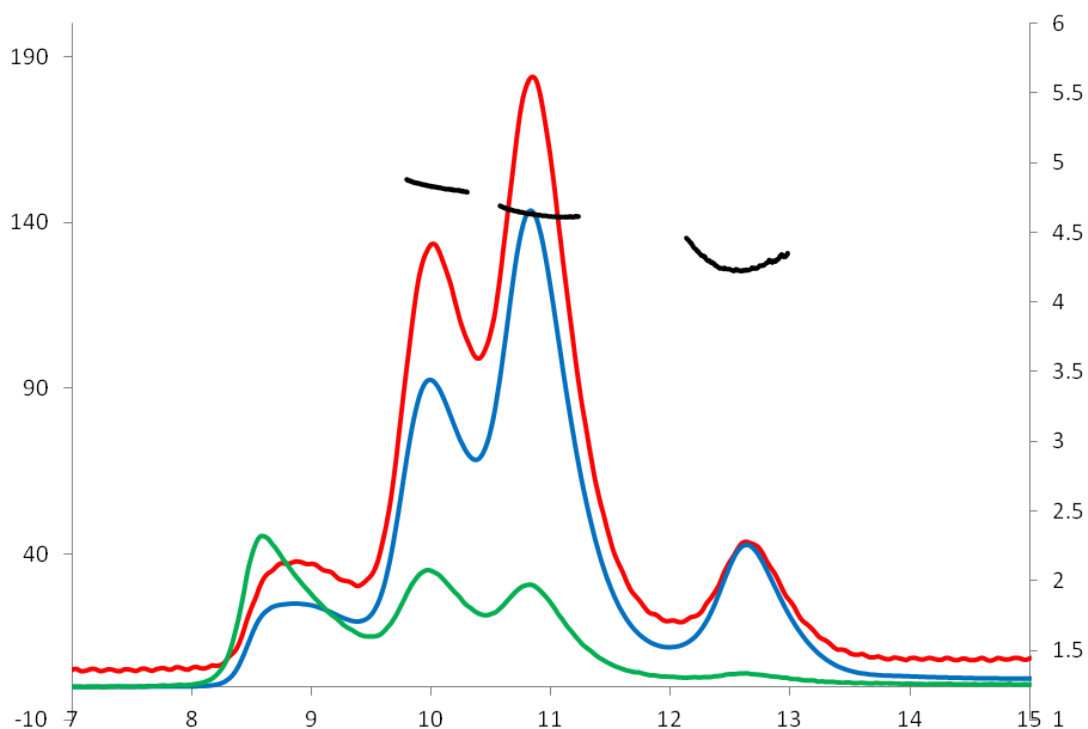
| | CD27L (wild-type) | CD27L (C238R) |
|---|---|---|
| **Data collection parameters** | | |
| Instrument | EMBL X33 beam line (DORIS-III, DESY, Hamburg) | EMBL P12 beam line (PETRA-III, DESY, Hamburg) |
| Beam geometry | 2.0 x 0.6 mm$^2$ | 0.2 x 0.12 mm$^2$ |
| Wavelength (Å) | 1.54 | 1.24 |
| $s$ range (Å$^{-1}$)[a] | 0.01-0.6 | 0.01-0.46 |
| Exposure time (s) | 8 x 15 | 1 (20 x 0.05 s) |
| Concentration range (mg/mL) | 0.9-4.0 | 1.0-8.5 |
| Temperature (K) | 283 | 283 |
| **Structural parameters[b]** | | |
| $I(0)$ (relative) [from $p(r)$] | 44 ± 2 | 3653 ± 14 |
| $R_g$ (Å) [from $p(r)$] | 33 ± 1 | 43 ± 2 |
| $I(0)$ (cm$^{-1}$) (from Guinier) | 45.6 ± 0.5 | 3664 ± 14 |
| $R_g$ (Å) (from Guinier) | 33 ± 1 | 42 ± 1 |
| $D_{max}$ (Å) | 106 | 147 |
| Porod volume estimate (Å$^3$) | 72151 ± 10000 | 91690 ± 10000 |
| Excluded volume estimate (Å$^3$) | 94000 ± 10000 | 123000 ± 10000 |
| Dry volume calculated from sequence (Å$^3$) | 39121/78219 (mon/dim) | |
| **Molecular-mass determination** | | |
| $I(0)$ (cm$^{-1}$) BSA (66,000 Da) | 71.4 ± 0.4 | 3791 ± 10 |
| Molecular mass $M_r$ (Da) [from $I(0)$] | 42150 ± 5000 | 63780 ± 5000 |
| Molecular mass $M_r$ (Da) [from Porod volume ($V_p/1.6$)] | 45094 ± 5000 | 57306 ± 5000 |
| Molecular mass $M_r$ (Da) [from excluded volume ($V_{ex}/2$)] | 47000 ± 5000 | 61500 ± 5000 |
| Calculated monomeric $M_r$ from sequence | ~32335 | |
| **Software employed** | | |
| Primary data reduction | RADAVER | |
| Data processing | PRIMUS/Qt | |
| *Ab initio* analysis | DAMMIF | |
| Validation and averaging | DAMAVER | |
| Rigid-body modeling | CORAL | |
| Equilibrium analysis | OLIGOMER | |
| Computation of model intensities | CRYSOL | |
| 3D graphics representations | PyMOL, UCSF Chimera | |

. [a]Momentum transfer $|s| = 4\pi\sin(\theta)/\lambda$. [b]Values reported for merged data sets (wild-type: 0.9 & 4.0 mg.mL$^{-1}$, C238R: 1 & 8.4 mg.mL$^{-1}$)

**Table 3** Equilibrium analysis of the SAXS data using the program OLIGOMER.

| Sample[a] | Volume Fractions | | | Fit, $\chi$ |
|---|---|---|---|---|
| | Extended Dimer | Compact Dimer | Degradation Products | |
| **CD27L** | 0.41 ± 0.02 | 0.24 ± 0.02 | 0.35 ± 0.02 | 1.5 |
| **C238R** | 0.98 ± 0.01 | 0.02 ± 0.01 | 0.0 | 1.1 |

[a]Values reported for merged data sets (wild-type: 0.9 & 4.0 mg.mL$^{-1}$, C238R: 1 & 8.4 mg.mL$^{-1}$)

| | CD27L wildtype MW estimates (RI/UV/RALS) kDa | | |
|---|---|---|---|
| | Dimer | Monomer + CBD | Monomer |
| Mean | 67.6 (3.7) | 43.0 (2.4) | 32.7 (6.6) |
| Min | 61.7 | 40.5 | 25.8 |
| Max | 76.4 | 49.5 | 53.3 |
| Mode (Frequency) | 64-64.5 | 41-41.5 | 27.5-28 |

**Supplementary Figure S1: UV/RI/RALS size exclusion chromatography traces and MW correlation estimates for wild-type endolysin**. Both the dimeric and monomer + CBD states of the protein have reasonably narrow molecular weight distributions through their respective elution profiles. The experimentally determined MW of these two species are close to that expected based on primary amino acid sequence. However, it must be noted that aggregates (that elute between 8-10 ml) or a species of a preceding peak that 'bleed into' a peak of interest will impact the intensities of the RALS data (green trace) and thus the minimum and maximum molecular weight estimates throughout the molecular weight correlations. The molecular weight distribution of the peak corresponding to the monomer is broader, likely caused by the resolution limitations of the SEC column. Although on average the last peak corresponds to the MW of a monomeric form, it is not possible to discount the possibility that this peak also contains CBD, and delta-CBD fragments. The numbers in brackets in the table refer to +/- 1 sd from the mean molecular weight. The mode (Frequency) refers to the most commonly occurring molecular weight range encountered throughout the molecular weight correlation.
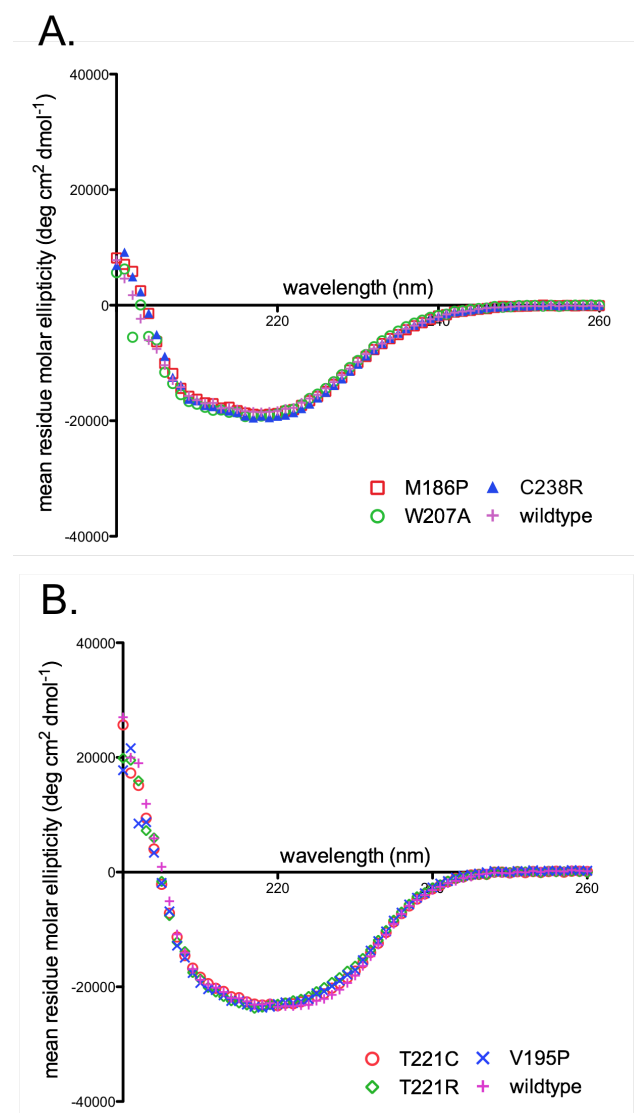
**Supplementary Figure S2: Circular dichroism measurements on CD27L/CTP1L wild-type and mutant proteins to show integrity of secondary structure:** The mean residue molar ellipticity measure by CD is shown for (A) CD27L wild-type and the mutants that prevent autoproteolytic cleavage (M186P, W207A and C238R), and (B) CTP1L wild-type and the mutants that prevent autoproteolytic cleavage (V195P, T221C and T221R).

**Supplementary Methods: CD spectropolarimetry measurements on CD27L and CTP1L endolysins and mutants:** Proteins were purified by Ni-NTA affinity purification as described above and further purified by size exclusion chromatography using an Aekta liquid chromatography system (Amersham Biosciences) and S75 10/300 GL (Tricorn) column (GE Healthcare). Size exclusion buffer for CD27L and CTP1L mutants was 20 mM HEPES, pH 7.4 with 500 mM NaCl and for CTP1L and mutants 20 mM HEPES, pH 7.4. Size excluded protein was collected and dialyzed against 20 mM HEPES pH 7.4. Protein concentration was measured by UV absorption at 280 nm. Spectra were recorded at 20°C on a Chirascan CD Spectrometer (Applied Photophysics), between 200 and 260 nm in a 0.1 cm cuvette. Machine settings were as follows: 1 nm bandwidth, 0.5-sec response, and 0.5-nm data pitch. Spectra were background-subtracted and converted into mean residue ellipticity. Each curve represents the mean of three separate measurements.