Synthesis of recombinant antibacterial proteins in the *Chlamydomonas reinhardtii* chloroplast

Maximilian Edward Alfred Blanshard

UCL

A thesis submitted for the degree of Doctor of Philosophy

September 2017

Declaration

I, Maximilian E. A. Blanshard, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

……

Acknowledgments

Firstly, I would like to thank my supervisor, Prof Saul Purton, for the opportunity to study in his lab for the last four years, and for providing help, encouragement, and ideas so freely and willingly. Huge thanks to my secondary supervisor, Dr Elaine Allan, whose support and kind words, as well as time in helping me to complete milestones, are hugely appreciated. Thank you also to my thesis committee chair, Dr Richard Hayward, who's ability to constructively criticise and encourage a methodical approach was invaluable. Dr Melinda Mayer, my collaborator, deserves huge thanks for agreeing to help out with the CD27L work and running so many turbidity reduction assays for me.

I would like to thank all members of the Purton Lab, past and present. From the beginning to the end of my PhD, Dr Laura Stoffels has taught me how to be a scientist and her amazing ability to know everything (and where everything is!) never ceases to amaze me, so I would like to thank her for all of the hours of time and frustration she has put into helping me, and resisting the urge to say "Just look it up". I would also like to thank Dr Rosie Young for teaching me so much molecular biology and for all of the encouragement when experiments failed. An enormous thank you to our lab manager, Thushi Sivagnanam, without whom I truly believe nothing would ever get done – your work is remarkable and we do not say it often enough. Thank you to Dr Henry Taunt, who established the endolysin work and has played an important part in my PhD despite supposedly leaving the lab in 2013, including buying more than his fair share of rounds and giving insight into the world of home-brewing.

To my friends and colleagues in the Purton lab with whom I have moaned, celebrated, travelled, and generally had fun – I could not have done it without you. Dr Janet Waterhouse, Dr Alice Lui, Dr Priscilla Rajakumar, and Dr Umaima Al Hoquani made me feel welcome from the day I arrived. To Juliana de Costa Ramos, Xenia Spencer-Milne, Fiona Li, Saowalak Chankgo and Marco Larrea Alvarez thank you for being great friends and best of luck with the rest of your PhDs.

Most importantly, I would like to thank my family, in particular my Mum and Dad for their exceptional generosity and unbelievable kindness throughout my life, and in their patience with all of my life decisions. I am so lucky to have the coolest parents in the world. I could not have done it without you!

Last, and most certainly not least, I would like to thank Samantha Hunt, whose love, companionship and support made even the most frustrating moments in the lab bearable, and who has filled the last seven years with adventures and fun. Thank you!

Abstract

The rise of antibiotic resistance and the decline in antibiotic discovery have been well publicised. These issues, in combination with a growing global reliance upon antibiotics for everyday modern life, urgently require the discovery of novel antibacterial drugs.

Endolysins are one potential candidate to support, or replace, conventional antibiotics. Endolysins are lytic enzymes produced by bacteriophage *in natura* to enable the release of viral progeny from inside the host bacterium. When applied exogenously, endolysins can lyse Gram-positive bacteria, and thus could be used as a novel antibacterial for this group of pathogens.

Biologically active endolysins have been successfully expressed as recombinants in the chloroplast of the green alga, *Chlamydomonas reinhardtii. C. reinhardtii*, and in particular the chloroplast, has several features as a cell factory which make it an attractive alternative to the traditional recombinant protein production platforms. *C. reinhardtii* is free of endotoxins, can be cultivated at low cost in photobioreactors, has GRAS status, and is genetically tractable.

This study initially focuses upon improving the accumulation and activity of one endolysin, Cpl-1, targeting *Streptococcus pneumoniae*. Recombinant Cpl-1 has been shown previously in the Purton lab to accumulate to moderate levels in the *C. reinhardtii* chloroplast. Here we present two transgenic lines of *C. reinhardtii* that appear to accumulate recombinant Cpl-1 to higher levels – one through the incorporation of multiple expression cassettes, and one through codon pair optimization. To improve the activity of Cpl-1 as an enzyme, Cpl-1 binding site mutagenesis, Cpl-1 dimerization, and the production of a potentially synergistic holin protein were all attempted.

Finally, an endolysin against *Clostridium difficile*, CD27L, was successfully produced in the C. *reinhardtii* chloroplast and shown to be active *in vitro*. Another endolysin, this time targeting *Propionibacterium acnes,* failed to express in *C. reinhardtii*, but was expressed in *E. coli,* albeit without obvious lytic activity.

Abbreviations

aadA	aminoglycoside-3''-adenyl-transferase gene
Amp	ampicillin
AMPS	ammonium persulphate
BHI	brain heart infusion
BME	β-mercaptoethanol
bp	base pairs
BSA	bovine serum albumin
CBD	cell binding domain
CD	catalytic domain
CES	control by epistasis of synthesis
CFU	colony forming unit
CPK	Corey, Pauling, Koltun
CTB	cholera toxin B
DNA	deoxyribonucleic acid
dNTP	2'deoxynucleoside 5'-triphosphate
EDTA	ethylenediaminetetraacetic acid (disodium salt)
e. g.	*exempli gratia* = for example
ELISA	enzyme-linked immunosorbent assay
g	gram
GC	guanine and cytosine
GFP	green fluorescent protein
GOI	gene of interest
HA	hemagglutinin

His	histidine
HSM	high salt minimal medium
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kb	kilo base pairs
kDa	kilo Dalton
LB	luria-bertani medium
mg	milligram
min	minute
ml	millilitre
mM	milimolar
mRNA	messenger ribonucleic acid
mt	mating-type
NCBI	National Centre for Biotechnology Information
OD	optical density
ORF	open reading frame
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
PG	peptidoglycan
PMF	proton motive force
PSI	photosystem I
PSII	photosystem II
psi	pounds per square inch
RCSB	Research Collaboratory for Structural Bioinformatics

RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
SD	Shine-Dalgarno consensus sequence
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
STGG	skimmed milk tryptone glycerol glucose medium
Strep	streptavidin
TAP	tris acetate phosphate medium
TBS	tris buffered saline
TBS-T	tris buffered saline – tween 20
TCP	total cell protein
TEMED	N, N, N', N'-tetramethylethylenediamine
tris	tris (hydroxymethyl) aminomethane
TRA	turbidity reduction assay
tRNA	transfer ribonucleic acid
TSP	total soluble protein
TSY	trypticase soy yeast extract medium
UTR	untranslated region
v/v	volume for volume
w/v	weight for volume
WGS	whole genome sequencing
WHO	World Health Organisation
WT	wild type

Chapter 1 Introduction

1.1 A new generation of antibiotics

1.1.1 Why are antibiotics so important?

1.1.1.1 History

Sir Alexander Fleming (1881-1955) is best known for his discovery of benzylpenicillin (Penicillin G) in 1928, an observation that shaped the modern world and opened up huge vistas of modern medical and surgical techniques that have become routine across the globe. After the initial discovery of penicillin by Fleming at St Mary's Hospital in London, it took a further 16 years of refinement, optimization and collaboration between industry and academia before penicillin could be produced in sufficient volumes to treat humans (Zaffiri et al. 2012). In contrast to the popular story of Fleming stumbling across penicillin on his return from holiday, the scaling up of production was no mean feat. After 12 years of promoting penicillin and searching for solutions to the purification and stability issues penicillin faced, Fleming had all but given up on his discovery (Aminov 2010). However, due to the pressures of the second world war, the work was continued, though still it did not come easy: as John L. Smith of Pfizer said at the time, "The mould is as temperamental as an opera singer, the yields are low, the isolation is difficult, the extraction is murder, the purification invites disaster, and the assay is unsatisfactory." (Werth 2014). By 1945 however, mass production and distribution of penicillin had been achieved – the world's first commercially available antibiotic. The chemical structure of Penicillin G and its mode of action as displayed in Figure 1.1.

Figure Removed

Figure 1.1 The small molecule benzylpenicillin (Penicillin G), known as a β-lactam due to the presence of a β-lactam ring in its molecular structure (Drugbank 2017). As with most β-lactam antibiotics, benzylpenicillin functions through disrupting cell wall biosynthesis in bacteria. The image on the right shows benzylpenicilin inhibiting formation of crosslinks in the bacterial peptidoglycan cell wall. Normally transpeptidases (also called penicilin-binding proteins) bind to D-alanyl-D-alanine pentapeptide of the peptidoglycan monomers and catalyse cross-linking via a peptide bridge. However, penicillin resembles the D-alanyl-D-alanine pentapeptide and blocks the active site of the transpeptidases, preventing monomer polymerisation. The lack of cross-linking weakens the cell wall, causing osmotic lysis. Image reproduced from Kaiser (2011).

Penicillin G was the first of more than 20 distinct antibiotic classes to be discovered from the 1940s to the 1960s, in what is now recognised as a golden era of antibiotic discovery (Arias and Murray 2015). These small molecules were all isolated from bacteria and fungi for their ability to block key enzymatic steps in bacteria, and presumably serve to give microorganisms a competitive advantage in environmental niches (Lewis 2013). However, there are also theories that these small molecules act as cell-cell communication systems at low concentrations, given their vast diversity and their ability to modulate cell transcripts (Davies 2006). Regardless of their natural role, these identified antibiotics targeted several different aspects of bacterial machinery essential for cell maintenance and reproduction, an illustration of which is presented in Figure 1.2. The incredible efficacy of antibiotics, their rapid rate of discovery and their ability to cure previously deadly infections in a matter of days led to them being hailed as a “magic bullet” (Sengupta et al. 2013). They were not only valuable as cures to bacterial infections, but in prevention too, thus opening up vast avenues for medicine and industry which will be further explored in the next section.

Figure Removed

Figure 1.2 The various modes of action of different antibiotic classes to kill bacteria. It should be noted that many classes share the same targets, for example cell-wall synthesis is inhibited by five separate classes of antibiotics. Reproduced from Lewis (2013).

1.1.1.2 Applications

The most obvious application for antibiotics is in treating bacterial infections. When penicillin was first mass produced, World War II was still raging and stocks were earmarked for soldiers on the front line, suffering from staphylococcal, streptococcal, and pneumococcal wound infections (Mailer and Mason 2001). While infections of minor cuts and injuries can still warrant antibiotics on occasion, better knowledge of hygiene, vaccinations, and after-care have reduced this market.

Prophylactic antibiotics on the other hand are still of immense importance and enable a whole suite of surgeries and chemotherapies which many take for granted.

All invasive surgeries, from organ transplants to dental work, are accompanied with a prophylactic course of antibiotics. Immunosuppressing chemotherapies for cancer treatments rely on antibiotics to stop bacterial infections while the body's immune system is out of action (O'Neill et al. 2015). In short, access to antibiotics underpins vast numbers of surgeries and medical techniques used in modern medicine.

One of the most controversial uses of antibiotics is as growth promoters in livestock. This stems from the discovery that sub-therapeutic levels of antibiotic cause increased growth. The mechanism for this remains unclear, but the most likely theory seems to be that growth is more efficient without gut bacteria competing for nutrients (Gaskins et al. 2002). Not only do low levels of antibiotics improve growth rates, but also conception rates and feed conversion ratios, altogether having a pronounced impact upon global meat production (Hao et al. 2014).

Veterinary medicine is another large consumer of antibiotics, enabling farmers to rear animals in closer proximity to one another such as in barns where previously a build-up of potentially infectious bacteria would prevent high animal densities and continuous farming. Economically, this has been advantageous – increasing farming efficiencies and lowering food prices. The same is true of fish farming: the growth of aquaculture to industrial scales has been almost wholly reliant upon prophylactic antibiotic use, enabling the high density farming of fish to meet the growing global demand (Cabello 2006).

The remarkable reliance that the modern world has built upon antibiotic consumption leads us to worry about what would happen if they ceased to be effective. We have mentioned briefly just a few of the huge industries that use, and over-use, antibiotics. As we will see in the following section, antibiotics are a finite resource due to the emergence of resistant bacteria and if these industries are to continue to grow, then new antibiotics must be sought.

1.1.2 Why do we need new antibiotics?

1.1.2.1 Resistance

The World Health Organisation (WHO) defines antibiotic resistance as a change which occurs to bacteria when exposed to antibiotics, resulting in the antibiotic becoming ineffective and the persistence of the infection (WHO 2016). This definition is valuable at a therapeutic level, but this section will give an overview of how antibiotic resistance functions a molecular level.

The five mechanisms of antibiotic action illustrated in Figure 1.2 represent the entire variability within our antibiotic arsenal. Given how important antibiotics are to modern society, this is an alarmingly small number of bacterial targets to rely upon, given that there are known to be approximately 200 conserved essential bacterial proteins (Lewis 2013). Resistance to each of these targets has arisen quickly as each new drug has been used on a large scale, with resistance to more recent antibiotics such as Levofloxacin (a second generation fluoroquinolone) being reported in 1996, the same year as it was introduced (Ventola 2015). It is therefore important to understand how resistance is acquired both at the molecular level, so that new therapies and drug combinations can be produced, and at the population level, so that preventative measures and procedures can be implemented.

Figure Removed

Figure 1.3 The six major antibiotic resistance mechanisms. Resistance mechanisms are not necessarily specific to a certain class of antibiotics, with many antibiotics being subject to multiple resistance mechanisms. Some therapies target the resistance mechanism in order to restore antibiotic efficacy, for example the application of β-lactamase inhibitors alongside β-lactam antibiotics. Reproduced from Lewis (2013).

Resistance arises due to the high selective pressure that antibiotics put upon bacterial populations. Each mode of resistance to each of the five sites of antibacterial action is illustrated in Figure 1.3. For example, penicillin is a β-lactam, a class of antibiotics that inhibit cell wall synthesis in bacteria. β-lactams bind to penicillin-binding proteins (PBPs), which are responsible for bacterial cell wall cross-linking, and inhibit their action. The resulting inability of the bacteria to correctly form a cell wall leads to autolysis (Fernandes et al. 2013). Several methods of resistance have developed towards β-lactams. Firstly, target modification: modification of PBPs to exhibit a much lower affinity to penicillin. Second is the production of β-lactamases, enzymes which degrade or modify β-lactams, and thirdly are changes to permeability in the

cell wall to β-lactams, which can either block entry or actively remove the β-lactams via efflux proteins (Fernandes et al. 2013). In this single example three different resistance mechanisms exist for a single antibiotic, highlighting the remarkable ability of bacteria to evolve under antibiotic induced selective pressure and underlining the real potential for some antibiotics to become obsolete.

It is reasonable to ask how antibiotic resistance evolves and spreads so rapidly. Two categories for resistance can be formed: endogenous resistance and exogenous resistance. Endogenous resistance refers to resistance arising within the bacteria itself i.e. by mutation and selection. Exogenous resistance involves the passing of resistance genes between different bacterial species via horizontal gene transfer (Silver 2011), each of which is explored below.

Endogenous resistance is often genetically recessive as it tends to result in a loss of function, for example reduced permeability to antibiotics, which have other (negative) consequences for normal cell operation. If antibiotic selection is continued for a long time, then secondary mutations may eventually evolve to compensate for loss of function. However, generally these mutations will eventually be lost after the removal of selective pressure.

Exogenously acquired resistance on the other hand generally involves a gain of function and is therefore more stable, for example the ability to synthesize antibiotic-degrading enzymes. Apart from a slight metabolic burden, these gains of function are not generally detrimental to the bacterium. Exogenous resistance is therefore of most concern clinically, as resistance persists after antibiotic treatment has finished (Silver and Bostian 1993).

The many uses of antibiotics described in 1.1.1.2 provide perfect opportunities for resistance to be selected. For example, the lack of distinction between “human antibiotics” and “veterinary antibiotics” has dangerous implications for emerging resistance via horizontal gene transfer. Furthermore, the manner in which antibiotics are generally delivered to livestock, such as in their food or water, can lead to sub-inhibitory concentrations of antibiotic which provides an ideal environment for resistance selection (Meek et al. 2015).

1.1.2.2 Antibiotic development

During the golden era of antibiotic discovery, a range of broad-spectrum antibiotics were quickly identified, patented and scaled to production, but today these same chemicals still form the basis of our antibiotic arsenal. From the mid-1960s onwards, only six new, approved, antibiotic classes have been discovered, and none since the discovery of daptomycin in 1986.

Figure Removed

Figure 1.4 The "Golden Era" of antibiotic discovery produced the basis for all of the antibiotics upon which we still rely today. Modifications and synthetic variants of these classes have extended the life of many of these drugs, but a lack of truly novel antibiotics persists. Very recently new bacterial culturing techniques have raised hopes of new antibiotic classes being discovered and approved (e.g. teixobactin), and are discussed later in this section (Ling et al. 2015). Image reproduced from Silver (2011).

This alarming drop-off in antibiotic discovery is due to several factors. In the early days of antibiotic research, discovery was largely through empirical screening – this amounted to testing of various microorganism extracts for bacterial growth inhibition, requiring little understanding of their action. This method worked

extremely well in the 1940s, 50s and 60s, but soon these “low hanging fruits” had all been discovered. These methods soon led to replication of discovery; common antibiotics were being repeatedly discovered, thus reducing the efficiency of the discovery system. The method deployed shifted towards target-directed screens and eventually big pharma companies reallocated resources to more lucrative therapeutics (Livermore 2011; Silver 2011). Economically, the antibiotic market has been somewhat a victim of its own success: with many competing drugs on the market, the price pharmaceutical companies can charge per prescription of antibiotics is much lower than that of many other drugs (Bax 1997).

However, the news is not all bad. Although no (or few) new classes have been discovered since the 1980s, there have been improvements within the existing classes. While such modifications do increase the efficacy of these antibiotics for a limited period of time, the fundamental resistance mechanisms still exist and resistance often quickly re-emerges (Payne et al. 2007). Other tactics have involved blocking the resistance mechanisms, such as through use of β-lactamase inhibitors (Walsh and Wencewicz 2014), and delivering these in combination with existing β-lactam antibiotics. But none of these techniques broaden the limited number of targets that our antibiotic arsenal currently possesses and have been incremental in nature.

New strategies to specifically target virulence factors, thus applying less selective pressure and with resistance being at the cost of virulence, present neat alternative targets for antibiotics, but remain a long way from the clinic (Clatworthy et al. 2007). Combinations of antibiotics are now the norm, and can achieve synergistic effects. In some cases, resistance to one antibiotic may increase susceptibility to another – for example colistin acts as a highly synergistic combination with several other antibiotics, by improving the permeability of the bacterial outer membrane in Gram negative bacteria (Tängdén et al. 2014).

One of the greatest restraints to the “Golden Era” of antibiotic discovery was the limited number of microbes that could be grown in the laboratory, and then tested for antimicrobial properties. In fact, it is estimated that 99% of bacteria are

"uncultured" and thus untapped. The development of new techniques for culturing soil microbes *in situ* enables the products of these microorganisms to be studied for the first time. This technology has already produced a new antibiotic candidate teixobactin which binds lipid II and III, the respective precursors of peptidoglycan and cell wall teichoic acid, and thus inhibits cell wall synthesis (Ling et al. 2015).

It is clear that, despite recent innovations, the dearth of recent antibiotic discovery combined with the rise of antibiotic resistance is a worrying state of affairs for our modern way of life. Better control of current antibiotic usage may slow the rate of resistance occurrence, but new antibiotics are desperately needed to compete in our evolutionary arms race against bacteria.

1.1.3 What alternatives to antibiotics currently exist?

The arms race against bacteria will never end, and even when new antibiotics with new targets are discovered, resistance will eventually arise. It is therefore vital that a robust pipeline for antibiotic production is in place, and given the lack in discovery of conventional antibiotics over the past 50 years, it makes sense to look to plug the gaps with alternative antimicrobials. With the advancement of knowledge and technology in these novel antibiotic fields, it is likely that eventually one, or many, of these "alternative antibiotics" will become the new "conventional antibiotics".

1.1.3.1 Non-lethal treatments

While not strictly "antibiotics", non-bactericidal therapies deserve a brief mention in this literature review as they provide an alternative treatment to traditional antibiotics. They also open up a different way of thinking about bacterial infections, namely that damage to host organisms when suffering from bacterial infection can be caused as much by host response as by microbial factors (Pirofski and Casadevall 2008). Efforts to prevent the host response, through tactics such as anti-inflammatory monoclonal antibodies can potentially minimise host damage while the body's immune system fights the infection.

Other methods involve hindering the activity of the bacteria through competition for resources, either by the introduction of other bacteria (probiotics) or the

sequestration of host nutrients (Spellberg 2014). Furthermore, bacterial toxin inhibitors and biofilm inhibitors which do not directly harm the bacteria, but reduce their pathogenicity offer a promising alternative path. A major advantage of these, along with host-response modulation, is that they do not apply the same level of selective pressure on the bacteria in question to evolve resistance as traditional bactericidal measures (Cheng et al. 2014).

1.1.3.2 Predatory bacteria

Some bacteria, such as *Bdellovibrio, Peredibacter* and *Vampirococcus,* prey on other bacteria and therefore may offer a mechanism for active control of bacterial infections. These bacteria go one step further than commensal 'probiotics', which generally compete for nutrients and resources thus reducing the scope for opportunistic pathogenic bacterial colonisation, by entering host bacteria and causing cell lysis (Dwidar et al. 2012). Predatory bacteria, widely known as *Bdellovibrio-and-like-organisms* (BALOs), have demonstrated killing of pathogenic bacteria *in vitro* (Van Essche et al. 2011), and recent *in vivo* studies show a reduction in bacterial burden of *Klebsiella pneumoniae* in rat lungs (Shatzkes et al. 2016).

An interesting strength of BALOs is their ability to prey upon Gram-negative bacteria, which are a major source of global disease. The outer-membrane of Gram-negative bacteria presents a barrier which several major antibiotics (e.g. vancomycin) and antibiotic-alternatives (e.g. endolysins) struggle to penetrate (Miller 2016). Another advantage is that prey resistance is extremely rare, with only one case so far reported (Varon 1979), suggesting a derived therapy could be effective for a substantial period of time.

However, safety concerns around inoculating humans with bacteria must be addressed. A potentially toxic response to the lipopolysaccharides in the Gram-negative outer-membrane of BALOs is yet to be solved. Furthermore, predatory bacteria are unable to eradicate all their prey, even when applied in very high ratios (Sockett and Lambert 2004), and are also strictly aerobic thus limiting their potential for internal human application. They may however find a place in therapies when

used together with, for example, β-lactams to which they are tolerant (Dwidar et al. 2012).

1.1.3.3 Antimicrobial peptides and bacteriocins

Antimicrobial peptides (AMPs) are small peptides (10-50 amino acids) that form part of the innate immune system, and are found in all classes of life. AMPs have been shown to act independently as antimicrobials, but also to have wide ranging immunomodulatory functions in mammals, including altering gene expression, promoting wound healing and modulating dendritic cell response (Bowdish et al. 2005; Fernebro 2011). Furthermore, AMPs can display activity against viruses and fungi too, thus earning the use of the word “antimicrobial”, rather than its subset “antibiotic”. AMPs also have activity against biofilms, which can harbour bacteria in chronic diseases and are therefore an interesting route of research. The mode of action of AMPs is still unclear, although it is believed that they cause cell lysis via permeation of the cytoplasmic membrane (Chung and Khanum 2017).

However, the high toxicity of AMPs to mammalian cells and their poor stability have led to poor results in clinical trials. This has guided research into the modification of AMPs, to produce Structurally Nanoengineered Antimicrobial Peptide Polymers (SNAPPs), which are more stable and display greater levels of antimicrobial activity (Lam et al. 2016). While as yet unproven clinically, SNAPPs have been shown to be particularly active against Gram-negative bacteria and show promising synergy with conventional antibiotics.

Antimicrobial peptides from prokaryotes are known as “bacteriocins” and tend to have a narrower activity spectrum. Bacteriocins have been used for some time in food preservatives – for example nisin is used to treat cheese, meat and vegetables to improve shelf life (Yang et al. 2014). While effective as inhibitors of microbial colonisation, the instability and potency of bacteriocins has to date limited their potential as therapeutics.

1.1.3.4 Bacteriophage

Bacteriophage, or “phage”, are the subset of viruses than infect bacteria. They are believed to be the most numerous biological entities on the planet, outnumbering bacteria 10-to-1 and displaying huge variability (Oliveira et al. 2013). Phage-therapy, which involves using phage to combat bacterial infections, has been investigated as a potential antibiotic since the early 1900s but interest in the field declined across much of the globe after the discovery of penicillin. However, phage therapy continued to be practiced in the USSR due to their limited access to conventional antibiotics, and even today the Eliava Institute in Georgia still treats patients with phage therapy (Fernebro 2011).

Phage therapy makes use of the lytic cycle of live bacteriophage to lyse infecting bacteria. The lytic cycle, illustrated in Figure 1.5, involves specific attachment of the bacteriophage to the host bacterial cell wall, injection of viral DNA into the host cell, and recruitment of the host protein synthesis machinery to produce bacteriophage progeny. After sufficient bacteriophage progeny synthesis, host cell lysis is initiated by the combined effects of holin and endolysin proteins, thus killing the host cell and releasing the phage progeny into the environment (Reindel and Fiore 2017).

Figure Removed

Figure 1.5 The bacteriophage lytic cycle. Phage specifically bind to their target bacterial species and inject viral DNA. The host cell protein synthesis machinery produces phage progeny, then holins to form a pore in the cell membrane and finally endolysins to hydrolyse the peptidoglycan layer. Reproduced from Feiner et al. (2015).

Since the rise of antibiotic resistance and the decline in new antibiotic discoveries however, interest in phage therapy has been renewed. In 2006 phage treatment of cooked meats was approved by the FDA to prevent post-processing contamination with *Listeria monocytogenes* (Fernebro 2011). However, no FDA approval has yet been made for human therapies, although a recent "emergency investigational new drug application" was approved for the treatment of a patient in San Diego, USA suffering from a multi-drug resistance bacterial infection, which resulted in the successful treatment of that patient (LaFee and Buschman 2017).

The high specificity of bacteriophage results in a narrow-spectrum of activity. This is excellent news for the protection of commensal flora in the gut for example, but relies on rapid and effective diagnosis techniques which are not readily available. Furthermore, occurrence of resistance to phage is high, and poor data exist on the efficacy of treatment and their safety profile (Fernebro 2011). However, if these obstacles can be overcome, then the sheer number, diversity and availability of bacteriophage make them a very attractive source of novel antibiotics.

1.1.3.5 Bacteriophage endolysins

Many of the hurdles faced by bacteriophage therapy can be overcome by the use of their endolysins. These are the lytic enzymes which degrade the host bacterial peptidoglycan cell wall during cell lysis. The use of these proteins as novel antibiotics holds great promise and is the focus of this thesis. Endolysins are introduced in greater detail in section 1.2.

1.2 An overview of bacteriophage endolysins

1.2.1 The biological role of bacteriophage endolysins

1.2.1.1 Background

Peptidoglycan hydrolase is a generic name given to an enzyme that catalyses the hydrolysis of the bacterial peptidoglycan cell wall. Peptidoglycan hydrolases come from many different sources, with different functions: autolysins are produced by bacteria to remodel their cell wall or cause apoptosis; exolysins are a weapon used

by bacteria to attack competing bacteria of another species; lysozymes can be produced by eukaryotes as part of their innate immune response; tail-associated lysins enable bacteriophage to inject their DNA into a new host bacterial cell; endolysins are peptidoglycan hydrolases, also of viral origin, that lyse the bacterial host cell wall from within at the end of the lytic cycle.

Upon injection of bacteriophage DNA into the host bacterium, bacteriophage viral progeny are synthesized and assembled by the host machinery. Two other proteins are also produced, which act in synchrony to optimize the timing of host bacterium lysis. One is the holin, which oligomerises in the inner membrane to form a pore. The other is an endolysin which is able to access the host peptidoglycan cell wall, via the pore formed by the holins, degrading the cell wall and causing cell lysis.

1.2.1.2 Structure

Endolysins are generally encoded by a single gene and two distinct categories exist: those from Gram-positive infecting bacteriophage, and those from Gram-negative infecting bacteriophage. Endolysins targeting Gram-positive bacteria have a modular structure and are highly species specific. Their modular structure consists of a N-terminal catalytic domain and a C-terminal cell wall binding domain (CBD), generally connected by a flexible acidic linker (Figure 1.6). The majority of endolysin specificity is conferred by the cell wall binding domain, although in some cases the catalytic domain has been shown to bestow specificity via recognition of the enzymatic substrate (Mayer et al. 2011). Removal of the CBD in some instances can improve the activity and spectrum of an endolysin, but in other cases the CBD appears to be a requirement for activity, perhaps by achieving the correct orientation of the catalytic domain to the substrate.

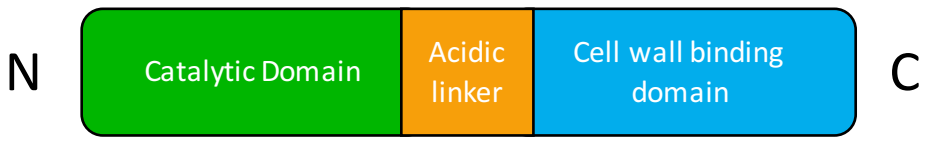


Figure 1.6 The canonical modular structure of a Gram-positive endolysin. The cell wall binding domain (CBD) is responsible for conferring specificity to the enzyme and modification or removal of the CBD has sometimes resulted in improved activity of the truncated endolysin (Cheng and Fischetti 2007).

Gram-negative endolysins on the other hand tend to be smaller with a globular structure, and lack the high binding affinity and specificity for the cell wall. The reason for this is believed to be that, while it may be advantageous for Gram-positive targeting endolysins to bind tightly to the peptidoglycan cell wall in order to prevent them from lysing surrounding bacteria which might be future hosts for the viral progeny, the surrounding Gram-negative bacteria are protected by their outer membrane so no such restraint is required (Loessner et al. 2002).

These descriptions are of course simplifications, and there are many exceptions that prove the rule. For example several staphylococcal endolysins have two N-terminal catalytic domains, the streptococcal λSA2 endolysin has a central CBD with two flanking catalytic domains at each terminal, and PlyC, another streptococcal lysin and one of the most active endolysins tested to date, is encoded by two separate genes forming a multimeric lysin (Schmelcher et al. 2012).

A few three-dimensional crystal structures of endolysins have been solved to date, but formation of crystals has proved difficult partially due to the flexible acidic linker between the domains. However, T4 lysozyme (Sagermann and Matthews 2002), T7 amidase (Cheng et al. 1994), PSA endolysin (Korndörfer et al. 2006), Cpl-1 (Hermoso et al. 2003), CTP1L (Dunne et al. 2016) and parts of CD27L (Dunne et al. 2014) have all been elucidated. The crystal structure of Cpl-1 will be discussed in greater detail in 4.3.1.

1.2.1.3 Mechanism of action

Endolysins display huge variation in their mechanism of action. While they all attack the peptidoglycan, there have been endolysins identified which target almost every bond within the peptidoglycan network. These can be broadly split into glycosidases which hydrolyse the aminosugar linkages, and amidases/peptidases which attack the cross-linking peptide stems and interpeptide bridges (Loessner 2005).

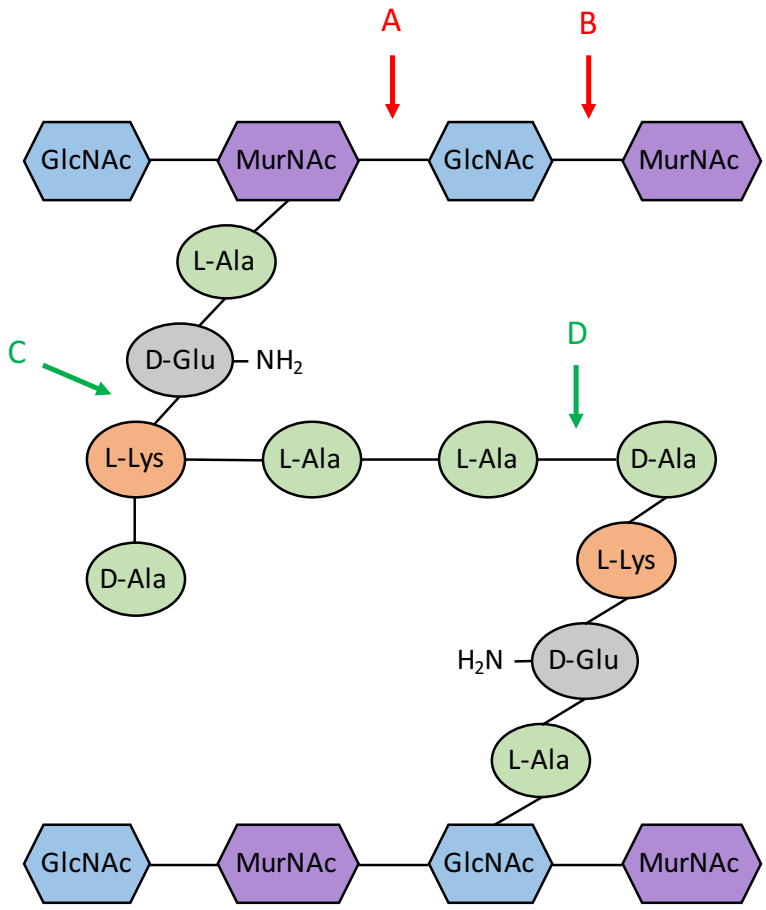


Figure 1.7 The peptidoglycan from Streptococcal cell wall (A3α). Hexagons represent sugars and ovals are amino acids. Red arrows indicate sites of action for glycosidases (**A**: N-acetyl-β-D-muramidase; **B**: N-acetyl-β-D-glucosaminidase). Green arrows indicate endopeptidase targets (**C**: D-glutamyl-L-lysine endopeptidase; **D**: D-alanoyl-L-alanine endopeptidase). Modified from Schmelcher et al. (2012).

CBD domains are present in Gram-positive targeting endolysins, and recruit the catalytic domain to the peptidoglycan cell wall. The CBD is largely responsible for the specificity observed in endolysins and is able to recognize and bind (non-covalently) a range of ligands in the cell wall. These ligands can either be a direct part of the peptidoglycan, or they can be other components of the cell wall. An example of the first is the LysM domain, one of the most common conserved binding modules, which binds the GlcNAc residues in the peptidoglycan. Meanwhile, the SH3b domain present in the Cpl-1 endolysin binds choline moieties in cell-wall associated teichoic acids of *Streptococcus pneumoniae*. The specificity conferred to endolysins by their CBD can range from an entire bacterial genus, down to serovar or strain and often the binding is exceptionally tight – in the pico- to nano-molar range (Loessner et al. 2002; Briers et al. 2009).

1.2.2 The therapeutic potential of endolysins and "exolysis"

1.2.2.1 Theory and history

Gram-positive bacteria, which have no outer membrane, are also susceptible to endolysins applied exogenously as the endolysin can directly access the peptidoglycan cell wall. The osmotic pressure inside the bacterial cell (20-50 atmospheres) then causes cell lysis. This sensitivity enables endolysins to be exploited as novel antibacterial agents, with a range of advantages over conventional antibiotics.

Gram-negative bacteria on the other hand have a protective outer membrane (Figure 1.8) which prevents the endolysins from accessing the peptidoglycan layer when applied exogenously. However, recent advancements have opened up the field for endolysin therapy against Gram-negative bacterial infections. The addition of a polycationic peptide, able to cross the lipopolysaccharide outer-membrane, to a globular Gram-negative endolysin has yielded exciting results, and the chimeric proteins have been named Artilysins® (Briers et al. 2014). However, Gram-positive targeting endolysins are the focus of this thesis.

Figure Removed

Figure 1.8 The cell wall structural differences between Gram-positive and Gram-negative bacteria. The noticeable differences are the thicker peptidoglycan cell wall in the Gram-postive bacteria, and the presence of an outer membrane in Gram-negative bacteria. In both instances the cytoplasm is at the bottom of the diagram. Reproduced from (Brown et al. 2015).

The first *in vivo* demonstration of endolysin activity was performed by Vincent Fischetti's group at Rockefeller University in 2001. The group used an endolysin prophylactically to prevent a Streptococcal infection in the mouse pharynx, showing that after a single treatment no streptococcal bacteria remained in the oral pharynx while leaving the indigenous oral flora unaffected (Nelson et al. 2001). Since then the field has grown significantly, with one commercially available endolysin therapeutic and several more in the pipeline as will be discussed in the following section.

1.2.2.2 Current therapeutic uses

The Dutch company Micreos produced the first endolysin-based therapy to receive FDA approval. The product, named Staphefekt and marketed under the brand name GladSkin, has been approved for topical treatment of *Staphylococcus aureus* on unbroken skin, treating diseases such as acne, rosacea, eczema and furunculosis (Micreos).

Another notable company in the therapeutic endolysin field is ContraFect. ContraFect are also developing an endolysin therapy against *Staphylococcus aureus* named CF-301, which is in phase II clinical trials and has been granted Fast Track Designation by the FDA having previously been put on clinical hold due to concerns over trial design (Parmley 2014). Notably, ContraFect are focusing upon endolysin purity, with the intention of treating bacteraemia systemically. ContraFect also have a Gram-negative endolysin at discovery stage (ContraFect).

To the author's knowledge, these are the only two endolysin-based therapeutics in clinical trials. However, there are other companies such as Lysando, who have licenced Artilysins®, with new products in sight.

Both Micreos and ContraFect use *E. coli* as their endolysin production platform, and ContraFect in particular have focussed on engineering their endolysin to be highly expressed, soluble and readily purified when produced in *E. coli* (Parmley 2014). While this appears to have been a suitable approach for producing endolysins targeting Gram-positive bacteria, it is unclear whether these companies will seek to use an alternative expression platform for producing endolysins targeting Gram-

negative bacteria. Some Gram-negative targeting endolysins have been shown to be highly toxic to *E. coli*, such as SPN9CC (Lim et al. 2014), and therefore potentially render *E. coli* an inefficient production platform.

1.2.2.3 Advantages of endolysins over conventional antibiotics

Endolysins hold great promise as novel antibiotics and they have numerous advantages over conventional antibiotics.

Firstly, endolysins kill bacteria at a significantly faster rate than that of conventional antibiotics. Indeed, purified lysins in nanogram quantities have been shown to reduce 10^7 *Streptococcus pyogenes* by more than 6 logs in only a matter of seconds (Fischetti 2008). *In vivo* models have shown that the endolysin ClyS achieves a 3 log reduction of bacteria in 30 minutes, while daptomycin takes 6 hours and vancomycin >24 hours to achieve similar levels of killing (Pastagia et al. 2013). Generally, endolysins act much faster than conventional antibiotics, which for blood infections in particular such as sepsis and bacteraemia, could be lifesaving.

The specificity of endolysins presents an enormous advantage over conventional antibiotics if the infecting bacteria has been diagnosed. The advantages are two-fold: firstly, a specific antibiotic will not disrupt the body's indigenous "good" bacteria, while killing the pathogen. This minimises the opportunity for secondary infections such as *Clostridium difficile* infection and yeast infections which are common complications after a course of broad-spectrum antibiotics (U.S. Department of Health and Human Services 2013). Secondly, as the vast majority of bacteria in the body or environment are unaffected by the endolysin, no selective pressure is applied to them. This is in contrast to broad-spectrum antibiotics, which put a selective pressure for resistance upon all bacterial species in the environment, thus increasing the chances of resistance emerging. High specificity is also a limitation of endolysins however, as will be discussed in section 1.2.2.4.

One of the greatest problems that conventional antibiotics face is their limited ability to kill persister cells in biofilms. Biofilms can form in various locations – from the gut lining, to water treatment plant equipment. Bacterial cells in biofilms are in a

dormant state, designed to withstand harsh environments and therefore do not undergo replication. Due to the mechanism of action of conventional antibiotics, such as disrupting cell wall formation, these biofilm-dwelling bacteria are unaffected by them. After treatment with antibiotics, the unaffected biofilm cells are then able to recolonise the environment. Endolysins on the other hand actively lyse bacterial cell walls, regardless of the state of the cell and have been shown to be extremely effective against biofilms (Gutiérrez et al. 2014).

One of the greatest and most often quoted advantages of endolysins is that, so far, bacterial resistance to them has been unreported, despite efforts to produce it (Loeffler et al. 2001). It is believed that several factors contribute to endolysin-resistance being a rare event. Firstly, the natural evolution of endolysins has led to the adoption of highly conserved targets in the peptidoglycan cell wall, which are essential for bacterial viability (Fischetti 2008). Secondly, the exogenous application and action of endolysins avoids many of the resistance mechanisms observed against conventional antibiotics (efflux pumps and deactivating enzymes). Finally, with regards to real life applications, the specificity of endolysins means that no selective pressure is being applied to non-target bacteria.

It is inevitable that one day, if endolysins become widely used, then some resistance will eventually be observed. In the case of conventional antibiotics, as we have discussed, discovering new antibiotics requires serendipity and a huge amount of time, and has recently proven very difficult. In contrast, given the diversity and size of the bacteriophage population, there is a huge store of endolysins to tap when the need arises.

Water pollution by conventional antibiotics is a growing concern. When antibiotics are not fully metabolised by the body, the high stability of some antibiotics can lead them to remain active throughout the sewerage system and, eventually, watercourses too. These low levels of antibiotics in the environment create ideal conditions for resistance emergence, and indeed antibiotic resistant bacteria have been detected in British watercourses for some time (Hughes and Meynell 1974). The proteinaceous endolysins on the other hand are relatively unstable, with a half-life

of about 20 minutes in the body, so are unlikely to survive into other environments, and given their specificity and lack of resistance, are a considerably lower risk.

Finally, though not a true advantage over conventional antibiotics, is the ability of endolysins to act synergistically with antibiotics to reduce antibiotic usage and improve therapeutic efficacy. For example, the endolysin ClyS has been shown to act synergistically with oxacillin to kill MRSA *in vivo.* As oxacillin is a cell wall synthesis inhibitor, and causes cell lysis by the resulting upregulation of autolysins, it is believed that ClyS further exacerbates the degradation of the cell wall, thus promoting even more bacterial autolysin activity (Daniel et al. 2010). Such synergistic effects may enable discontinued antibiotics to be reinstated when used with endolysins, and reduce overall antibiotic intake or course length.

1.2.2.4 Limitations

The bacteria posing the greatest threat to mankind have been given the acronym ESKAPE by the Infections Diseases Society of America, standing for *Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumanni, Pseudomonas aeruginosa,* and *Enterobacter* species (Rice 2008). Of these, the final four are all Gram-negative. Conventional antibiotics currently in clinical trials are skewed towards Gram-positive activity. One limitation of endolysins is their inability to cause Gram-negative lysis due to their inability to penetrate the outer membrane. However, as mentioned earlier, Artilysins® have made some progress towards being able to tackle Gram-negative infections, and for topical applications EDTA can be used to permeablise the Gram-negative outer membrane (Briers and Lavigne 2015).

There are two factors to be considered in terms of immune response to application of endolysins. The first is to the endolysin itself: endolysins are a protein and thus may trigger an immune response including antibody generation, potentially limiting future use of that endolysin. An attempt to reduce the postulated immune response to Cpl-1 was made by PEGylation of the protein, a well-known method of minimising immune response and increasing half-life, but this completely inhibited the lytic activity of the endolysin (Resch et al. 2011b). However, it has been observed *in vivo* that although antibodies are indeed raised against the Cpl-1 endolysin in mice, on

reapplication no adverse effects were observed and activity was only slightly slowed (Jado et al. 2003). It is postulated that the endolysin activity is much faster than the host's antibody mediated secondary immune response, and is thus largely unaffected.

The second immune response of concern is to the large volume of bacterial cell debris rapidly released by endolysin activity. Two studies have been undertaken to investigate this issue, with varying results: the first study found that the Cpl-1 endolysin produced an accentuated cytokine response when compared to the same treatment using vancomycin (Entenza et al. 2005); the latter study observed no major difference in cytokine production between Cpl-1 and amoxicillin (Witzenrath et al. 2009). This discrepancy is thought to be an issue of endolysin dosing: when a high dose is given, then the bacterial cell wall is fragmented thus producing a greater response, compared to a low dose which merely punctures the bacterial cells with minimal fragmentation (Fischetti 2008).

Bioavailability of endolysins is a concern given their large size compared to conventional antibiotics. In particular, intracellular bacterial infections such as chlamydia and tuberculosis are particularly hard to access, and present a challenge even for conventional antibiotics. However, a recent demonstration that PlyC is able to enter human throat cells and kill internalized *S. pyogenes* demonstrates the diverse abilities of the endolysins (Shen et al. 2016). In this instance it appears that a specific component of the throat cell membrane, a phosphatidylserine, grants access to the PlyC endolysin – however it is also noted in the study that other endolysins could not be expected to function in this way. This finding does however raise the prospect that even intracellular bacteria may eventually be treated with endolysins. The only endolysin currently on the market, Staphefekt, is designed for topical use and is delivered either as a cream or a gel (Micreos). It is unclear how ContraFect intend to deliver their endolysin drug, however their studies and discussions have mostly focussed around intranasal and aerosol delivery methods (Doehn et al. 2013; Loeffler et al. 2003). Oral delivery of endolysins will be harder, as the proteinaceous nature of endolysins means that they are susceptible to digestion and denaturing in

the stomach, and would need to be encapsulated to ensure release into the intestine for example.

The final limitation of endolysins, which was contrastingly presented earlier as an advantage, is their specificity. The narrow-spectrum of target bacteria killed by an endolysin presents a challenge for clinicians in that the infection must be accurately diagnosed before an endolysin treatment can be prescribed. This diagnosis is not yet easily possible, but the potential to create endolysin "cocktails", which target all the most common bacteria associated with a certain disease, may work around this.

To conclude, although various challenges have been laid in the path of endolysin research, none has yet presented a serious flaw to the technology as a therapeutic, and techniques are emerging to solve the majority of issues.

1.2.3 Other applications

The remarkable ability of endolysins to identify and strongly bind highly specific bacterial targets has produced interest in fields other than direct therapeutics.

For example, work by the Loessner group at ETH, Zurich has used endolysins to create rapid diagnostic tools for food safety. Their technology uses electrochemical impedance spectroscopy (EIS) to rapidly detect *Listeria* cells binding to endolysin-derived CBDs immobilized on a gold screen printed electrode. This enables extremely rapid, accurate, and sensitive detection of contaminating *Listeria* cells in food products (Tolba et al. 2012).

Nelson's group at the University of Maryland have designed two novel endolysin based platforms, designed to aid the body's natural immune system. The first is the generation of Infection Site Targeted Antitoxin antibodies (ISTAbs), which are endolysin CBDs fused to specific antitoxin neutralizing monoclonal antibodies. The CBD targets the antibody to the pathogen cell wall, where it sequesters the released toxins, thus reducing the toxin load and furthermore, functionally opsonises the bacterium for opsonophagocytic killing by the host immune system. The second approach, named Infection Site Targeted universal Bridging Antigen (ISTuBA), fuses

CBDs to antigens for which most people have existing immunity (e.g. childhood vaccine targets). This anchors the antigen to the pathogen cell wall, and directs the existing immunity towards the new pathogen (Nelson 2016).

Other applications include endolysin based disinfectants, which have been shown to be 1,000x more active than virkon-S against *Streptococcus equi* over 30 minutes (Hoopes et al. 2009). Food biopreservation and biofilm elimination have also been shown to be successful applications of endolysins (Ajuebor et al. 2016).

1.3 Producing recombinant therapeutic proteins

The successful administration of therapeutic insulin to a severely diabetic 14-year-old boy in 1922 marked the first use of a protein as a pharmaceutical (Joshi et al. 2007). Since then, the ability to produce proteins recombinantly in a number of biological platforms has led to lower costs of production and the rise of protein therapies. In 1977 Somatostatin, a very small inhibitor peptide, was the first human recombinant protein to be produced in *Escherichia coli,* by Genentech Inc (Brooker 2010). Other platforms have since been developed to meet the various protein requirements, such as producing larger proteins, glycosylation patterns, and disulphide bonds. Here we provide a very brief overview of some of the most common expression platforms used today.

1.3.1 Common expression platforms

E. coli, as already mentioned, was the first recombinant protein expression platform and remains widely used. Due to many years of research and development resulting in an excellent understanding of *E. coli* genetics, genetic engineering of *E. coli* is cheap, easy and extremely fast. High yields of recombinant proteins can be readily achieved and many genetic tools, such as inducible expression systems and protein secretion, enable its precise manipulation. However, the system is not well suited to producing large proteins (over 50 kDa). Furthermore, the proteins can form inactive inclusion bodies, *E. coli* does not naturally glycosylate proteins, and endotoxin contamination requires expensive down-stream purification procedures (Gupta and Shukla 2015).

Chinese Hamster Ovary (CHO) cells are probably the most commonly used expression platform for current therapeutic proteins, in particular for the production of antibodies (Dumont et al. 2015). As a mammalian system, CHO cells are able to produce large proteins (over 100 kDa) and glycosylate and fold them correctly - although they do produce some post-translational modifications that are not expressed in humans, and subsequently against which humans have antibodies (Dumont et al. 2015). On the whole, CHO cells are used to produce larger and more complex biologics than other systems. Their drawbacks include potential DNA and viral contamination, very high costs of production and lower yields than some other systems (Demain and Vaishnav 2011).

Yeast, as a single cellular eukaryote, shares some of the advantages of a microbial platform, such as cheap growth media and high yields, as well as advantages of mammalians cells, such as protein glycosylation and disulphide bond formation. Yeast therefore presents a very attractive platform for recombinant protein expression, although a few issues prevent it from being using for all proteins. For example, yeast's high-mannose type N-glycosylation pattern can lead to a shorter protein half-life and reduced therapeutic efficacy, and its lack of necessary chaperones can often result in poor protein folding (Nielsen 2013).

Plant-based platforms have lagged behind others over the past decades, despite the numerous advantages they hold as a recombinant protein production platform. Plant cells are able to fold complex proteins and perform post-translational modifications including glycosylation. Either whole plants or plant cell cultures can be used as expression platforms. Both are devoid of the endotoxins and potential viral contamination of other platforms, but otherwise have different advantages and pitfalls. Whole plants, such as tobacco (*Nicotiana tabacum*) are highly scalable, and cost effective at scale. Transient gene expression in leaf tissue of whole plants can be achieved very rapidly and has been used to successfully produce ZMapp, the only successful drug produced to treat Ebola viral infection in humans, as well as being regularly used for producing influenza vaccines (Yao et al. 2015). However, several issues have limited whole plant use to date, including the relatively long time to produce stable transgenic plants desired for long-term production, variable

expression levels, and issues of biocontainment at large scale (Park and Wi 2016). Plant cell cultures offer solutions to many of these issues, such precise monitoring and control of growth conditions, consistent cell yields and biocontainment when grown in bioreactors. Several plant cell lines are used routinely, such as those derived from *N. tabacum*, as well as those from edible crops including rice (*Oriza sativa*), carrot (*Daucus carota*) and tomato (*Lycopersicon esculentum*) (Xu and Zhang 2014), and in 2012 taliglucerase alfa, the first plant-made pharmaceutical approved for human use, was made in a carrot cell line to treat Gaucher's disease (Grabowski et al. 2014).

Like plant cell cultures, microalgae offer many similar advantages over other platforms and higher plants, but have additional advantages too. Microalgae are photosynthetic (unlike carrot cell cultures for example), enabling the potential for using cheaper, less complex media that is less susceptible to contamination. Microalgae also generally grow faster than plant cell cultures and are more robust in terms of resisting stresses from pH changes and shear stress (Sun and Linden 1999). Finally, the generation of new transplastomic lines is faster in microalgae (2-3 weeks) than plant cell cultures (4-5 weeks) (Govea-Alonso et al. 2017).

1.3.2 Microalgae as expression platforms

1.3.2.1 What are microalgae?

The term microalgae can generally refer to any unicellular organism capable of oxygenic photosynthesis, both eukaryotic (e.g. green algae) and prokaryotic (e.g. cyanobacteria) (Walker et al. 2005). Microalgae are therefore an extremely diverse group of organisms and can be found in both fresh and salt water environments, as well as terrestrially. Figure 1.9 illustrates the level of diversity observed within the microalgae.

Figure Removed

Figure 1.9 A selection of images displaying the diversity of microalgae. **(A)** *Meridion circulare* **(B)** *Merismopedia elegans* **(C)** *Acutodesmus acuminatus* **(D)** *Microcystis novacekii* **(E)** *Arthrospira jenneri* **(F)** *Dissodinium pseudolunula* **(G)** *Zygnema circumcarinatum* **(H)** *Actinocyclus normanii* **(I)** *Pediastrum duplex* **(J)** *Klebsormidium dissectum* **(K)** *Diploneis* **(L)** *Chlamydomonas reinhardtii*. **A-F** from http://nordicmicroalgae.org/. **G-K** from http://www.algaterra.org/. **L** from http://protist.i.hosei.ac.jp/. Images not to scale.

Microalgae such as *Arthrospira* (*Spirulina*) and *Aphanizomenon* have been used as a source of food for thousands of years (Spolaore et al. 2006), but only in the last 50 years have microalgae been subject to commercial cultivation. The first company to produce *Chlorella sorokiniana* on a commercial scale was Taiwan Chlorella Manufacturing Ltd, which was founded in 1964 and still produces *Chlorella* today, to be sold as dietary supplement (Spolaore et al. 2006). While *Spirulina* and *Chlorella* remain popular health-food products, the ability of microalgae to act as "sunlight-driven cell factories" (Chisti 2008) for producing high value products has spurred research in the field towards harnessing the full potential of these diverse organisms.

1.3.2.2 The history of microalgae as an expression platform

Microalgae, and in particular the green alga *Chlamydomonas reinhardtii,* have been used as a model organism for many years for the investigation of the genetic mechanisms behind cellular processes (Harris 2009). Microalgae gained prominence as a recombinant protein production platform with their potential to produce biofuels without affecting food production - as a photosynthetic organism it was postulated that algae could be grown in areas unsuitable for crop production, thus producing biomass from cheap substrates on unused land. However, despite the successful production of biodiesel in microalgae, the economics of this strategy currently fail to compete with that of fossil-fuels (Scranton et al. 2015).

Despite the current shortcomings of microalgae for the production of biodiesel, the platform still holds promise for high-value proteins and has several advantages over higher-plants. Microalgae have the ability, like higher plants, to synthesize and correctly fold human antibodies, and many are Generally Regarded As Safe (GRAS). But in addition to these attributes, microalgae can be grown in closed containment photobioreactors, removing the risk of gene flow into the environment via pollen; genetically engineered strains can be produced in a matter of weeks rather than sometimes years for higher-plants; and scale up from bench to production scale is straightforward (Franklin and Mayfield 2004). Microalgae therefore present an extremely attractive platform for recombinant protein expression.

Several species of microalgae have to date been genetically transformed, including *C. reinhardtii*, *Haematococcus pluvialis*, *Chlorella* and diatoms (Spicer and Purton 2016). *H. pluvialis* is a natural producer of astaxanthin, a carotenoid with high economic and health-related value, and it has been successfully genetically engineered to produce higher levels of astaxanthin (Steinbrenner and Sandmann 2006). The green alga *Chlorella vulgaris* is a source of several products believed to be beneficial as health supplements, including high-quality proteins, fatty acids, and essential amino acids, and a method of agrobacterium-mediated transformation has been established (Cha et al. 2012).

However, while various species of microalgae have been genetically modified, the species with the most advanced genetic toolkit is *C. reinhardtii*. *C. reinhardtii* has been shown to express multiple recombinant proteins using a variety of transformation methods, and in all three genomes (nuclear, mitochondrial and chloroplast) (Scaife et al. 2015). The potential of *C. reinhardtii* will be explored further in the following section.

1.3.3 *Chlamydomonas reinhardtii* as an expression platform

1.3.3.1 An introduction to *Chlamydomonas reinhardtii*

Chlamydomonas reinhardtii is a unicellular green alga found widely in soil and freshwater and belongs to the Chlorophyta division of green algae. Figure 1.10 shows a schematic of a *C. reinhardtii* cell as well as a scanning electron microscopy image, highlighting the main morphological features of the cell. It is approximately 10 µm in diameter and has two flagella enabling phototaxis, and a large chloroplast filling approximately 70% of the cell volume (Harris 2009). *C. reinhardtii* can grow phototrophically or, in the presence of acetate, mixotrophically or heterotrophically. All three genomes have been sequenced (Vahrenholz et al. 1993; Maul et al. 2002; Merchant et al. 2010), and a simple (and controllable) sexual cycle exists in which nuclear genes are inherited according to Mendelian rules, chloroplast genes inherited uniparentally from the mating-type (mt) plus parent, and mitochondrial genes from the mt minus parent. Furthermore, the haploid nature of the nuclear genome and the fact that many nuclear genes are single copy (rather than members of gene families) makes the recovery and classical genetic analysis of mutants very straightforward. This has led to *C. reinhardtii* being refered to as the "photosynthetic yeast" (Rochaix 1995). Given these attributes and the fact that it is relatively fast growing (5 - 8 hour doubling time, depending on nutrient, CO_2 and light availability (Hosler et al. 1989)), it has been an attractive model for studying photosynthesis, circadian rhythms, mating and other processes for many years (Harris 2009).

Figure Removed

Figure 1.10 (**A**) The *Chlamydomonas reinhardtii* cell represented as a schematic diagram with labelled physiological features (Shi 2013). (**B**) A Scanning Electron Microscopy image of *C. reinhardtii* clearly showing the main morphologies of the cell including the two flagella (Smith et al. 1996).

1.3.3.2 Genetic engineering of *Chlamydomonas reinhardtii*

C. reinhardtii is the most genetically well studied of the microalgae, and methods to transform the nuclear, chloroplast and mitochondrial genomes have all been

developed. The sexual cycle can be precisely controlled, enabling genetic crosses to be carried out and introduced traits to be combined (Grossman et al. 2003).

The nuclear genome can be transformed using a number of methods including electroporation and particle bombardment to randomly incorporate DNA into the genome (Jinkerson and Jonikas 2015). Antibiotic markers and rescuing of nutritional auxotrophs have been successful methods for nuclear transformation selection, but positional effects lead to a spectrum of gene expression levels (Rochaix 1995). The nuclear genome is very GC rich (64%), and codon optimization has been shown to play an important role in gene expression levels, improving GFP expression levels approximately 5-fold (Fuhrmann et al. 1999). However, gene silencing and low protein yields have been problematic for *C. reinhardtii* nuclear transformants, although recent work, such as that by Barahimipour et al. (2016) demonstrating high yields of HIV antigen P24, show that these issues are being gradually overcome. Furthermore, Lauersen et al. (2015) have created the pOptimized vector system, which allows the rapid swapping of reporter genes, selectable markers and regulatory elements to speed up the optimisation of nuclear protein expression, while Scranton et al. (2016) have demonstrated the successful identification and use of synthetic nuclear promoters to improve heterologous nuclear gene expression.

The chloroplast on the other hand offers a more attractive target for genetic modification. Very simple transformation processes exist, for example transient pores can be formed by vortexing cell wall-less mutants with glass beads, which enables DNA entry to the chloroplast (Kindle et al. 1991; Maliga 2014). Homologous recombination then allows precise targeting of expression cassettes into the plastome, and generally results in greater protein accumulation than nuclear transformations (Franklin et al. 2002). Furthermore, by transforming photosynthesis-deficient mutants, antibiotic-free selection is possible by using the rescue of photosynthesis as a selection mechanism (Wannathong et al. 2016). The precise method of transformation is described in more detail in 2.3.13. Gene silencing, thought to be a defence mechanism against viral DNA insertion into the genome, has not been observed in the *C. reinhardtii* chloroplast. However, the chloroplast does not glycosylate proteins.

A summary of the strengths and limitations of the *C. reinhardtii* chloroplast compared to the nucleus can be seen in Figure 1.11.

Figure Removed

Figure 1.11 The nucleus and chloroplast genomes represent the major locations for transgene insertion and recombinant protein synthesis in *C. reinhardtii*. While the mitochondrial genome has been transformed it is not commonly used for recombinant protein production. Reproduced from Rasala & Mayfield (2015).

1.3.3.3 The suitability of the *C. reinhardtii* chloroplast for endolysin synthesis

Over one billion years ago an endosymbiotic event, resulting in the encapsulation of a photosynthetic prokaryote inside a nonphotosynthetic eukaryotic host, is believed to have been the genesis of the chloroplast organelle (Purton 2007). Since then the prokaryotic genome has shed much of its unnecessary genetic material, resulting in a simplified circular plastome. The *C. reinhardtii* chloroplast, sequenced in 2002 (Maul et al. 2002), contains 99 genes and has a low GC content (34%), in contrast to the nuclear genome.

The prokaryotic beginnings of the chloroplast result in an environment well suited for the production of endolysins, which are naturally produced in bacteriophage-infected prokaryotes. For example, the chloroplast appears to contain proteases that

are direct homologues of prokaryotic proteases (Adam et al. 2006). This is expected to lead to a respectable half-life of the endolysins inside the chloroplast, as they are adapted to be relatively resistant to prokaryotic proteases. Furthermore, the chloroplast is non-glycosylating, which may often be seen as a disadvantage, but for the production of endolysins, and other prokaryotic proteins, is beneficial (Almaraz-Delgado et al. 2014). The expression of endolysins to high levels in *E. coli* has been reported to have a detrimental effect upon bacterial cell growth, presumably because of low-level activity against its peptidoglycan cell wall despite the specificity of endolysins and lack of holin. However, plant and green algal chloroplasts have lost the cell wall during evolution, and extremely high levels of endolysin accumulation (~70 % TSP) have been reported in the chloroplast of *Nicotiana* (Oey et al. 2009). Therefore, the prokaryotic environment, lack of glycosylation, and lack of peptidoglycan structures makes the chloroplast a well suited platform for endolysin production (Oey et al. 2009). The *C. reinhardtii* chloroplast has successfully produced therapeutic proteins in the past, including endolysins, a summary of which can be seen in Table 1.1.

C. reinhardtii has been designated GRAS status and there is speculation in the field that this may lead to reduced downstream processing costs, or even the ability to use crude extracts of recombinant protein therapeutically (Almaraz-Delgado et al. 2014). In fact, the two companies closest to therapeutic endolysin production have both been hampered by issues of product purity and solubility with their bacterial production systems. ContraFect, attempting to produce an endolysin for systemic use, have so far failed to receive FDA approval, while Micreos have instead opted to produce an endolysin for topical use, with less regulation around product purity (Parmley 2014).

To date, the major limitation to microalgae as a recombinant protein production platform has been their low production yields. CHO cells are capable of producing 4-6 g/L of recombinant protein (Wurm 2004), yeasts 9-12 g/L, *E. coli* 15-17 g/L, transgenic animals (e.g. goat milk) up to 23 g/L, and fungal expression systems have been reported to achieve 35 g/L (Demain and Vaishnav 2011). These are clearly optimized yields, and differ depending on the protein in question. However, the

highest level of recombinant protein production in *C. reinhardtii* is 5% TSP under laboratory conditions (Manuell et al. 2007), which would equate to approximately 8 mg/L culture volume when scaled to pilot scale (Gimpel et al. 2015). As it is, pilot scale production of recombinant bovine milk amyloid A (MAA) in *C. reinhardtii* achieved 3.28 mg/L (Gimpel et al. 2015). It is clear from these data that the *C. reinhardtii* platform has a long way to go before being able to compete with the platforms named above. However, rate of development can be fast: between 1986 and 2004 the recombinant protein yield of CHO cells rose from 50 mg/L to 4.7 g/L (Wurm 2004). This vast improvement was achieved through a variety of improved techniques such as better selection strategies for high-productivity clones and identification and subsequent targeting to transgenes to transcriptional "hotspots" (Kim et al. 2012). Furthermore, different platforms have different protein recovery rates, for example, due to inclusion bodies in *E. coli*, production rates may be high, but recovery rates and renaturation rates can be extremely low, leading to the possibility that another platform with a lower headline production rate is more suitable (Datar et al. 1993).

While the GRAS status of many microalgae gives hope that subsequent purification of protein products will be cheaper and less necessary in comparison to, for example, *E. coli*, there are other limitations to their use as a protein production platform, such as the presence of chlorophyll. Chlorophyll is a problematic pigment present in algal extracts that requires further purification, although this can be performed relatively cheaply by use of column chromatography with activated carbon (Santillan-Jimenez et al. 2016).

RECOMBINANT Protein	Function	Use	Reference
HSV8-lsc	Human antibody against glycoprotein D from Herpes simplex virus	Therapeutic	(Mayfield et al. 2003)
CTB-VP1	Protein VP1 from foot-and-mouth disease virus fused to cholera toxin B (CTB)	Vaccine	(Sun et al. 2003)
hTRAIL	Tumour necrosis factor-related apoptosis-inducing ligand	Therapeutic	(Yang et al. 2006)
M-SAA	Bovine mammary-associated serum amyloid	Therapeutic	(Manuell et al. 2007)
CSFV-E2	Swine fever virus structural protein E2	Vaccine	(He et al. 2007)
hGAD65	Human glutamic acid decarboxylase	Diagnostics	(Wang et al. 2008)
VP28	White spot syndrome virus protein 28	Vaccine	(Surzycki et al. 2009)
CTB-D2	D2 fibronectin-binding domain of *Staphylococcus aureus* fused to CTB	Vaccine (oral)	(Dreesen et al. 2010)
VEGF	Human vascular endothelial growth factor	Therapeutic	(Rasala et al. 2009)
Pfs25 and Pfs28	*Plasmodium falciparum* surface proteins	Vaccine (oral)	(Gregory et al. 2012)
αCD22PE40	Antibody against DC22 surface protein from B-cells fused to two domains of an exotoxin from *Pseudomonas aeruginosa*	Immunotoxin	(Tran et al. 2013)
CtxB-Pfs25	*Plasmodium falciparum* surface protein 25 fused to CTB	Vaccine (oral)	(Gregory et al. 2013)
TPS4	Bifunctional diterpene synthase, for therapeutic terpenoid synthesis	Bioprocessing	(Zedler et al. 2015)
V_HH	Camelid heavy chain-only antibodies	Therapeutic	(Markey et al. 2015)
Ara h 1	Major peanut allergen	Therapeutic	(Gregory et al. 2016)
Cpl-1 and Pal	*Streptococcus pneumoniae* targeting endolysins	Therapeutic	(Stoffels et al. 2017)
$CD27L_{(1-179)}$	*Clostridium difficile* targeting endolysin	Therapeutic	This study

Table 1.1 A selection of therapeutic recombinant proteins successfully produced in the *C. reinhardtii* chloroplast. Modified and updated from (Stoffels 2015), originally adapted from (Specht et al. 2010) and (Almaraz-Delgado et al. 2014).

1.4 Targets/applications

1.4.1 Streptococcus pneumoniae

Although *S. pneumoniae* is a common commensal bacterium found in the nasopharynx, and is usually benign, colonisation and infection can cause a host of diseases, including pneumonia, meningitis, febrile bacteraemia and otitis media (Bogaert et al. 2004). Pneumococcal infection accounts for the death of approximately 1.2 million people every year, most of them children under the age of five (Obaro and Adegbola 2002).

Figure Removed

Figure 1.12 Scanning electron microscope image of an encapsulated strain (R6) of *S. pneumoniae*, in its streptococcal form with 2 % choline (left) and diplococcal form without choline (right). Scale bars are 2 μm. Image reproduced from (Paik et al. 1999).

S. pneumoniae is a mesophilic, non-motile, Gram-positive bacterium and was named after its role in pneumonia. It usually takes a diplococcal form but can form streptococci in liquid medium. Depending upon the strain, a polysaccharide capsule sometimes encases the cell, below which is the peptidoglycan cell wall. Presence of this capsule can increase virility of the strain and interferes with leukocyte phagocytosis (Hyams et al. 2010). However, it has been shown that encapsulated

strains are no more or less susceptible to lysis by endolysins than un-encapsulated strains (Loeffler et al. 2001).

The peptidoglycan cell wall is approximately six layers thick and teichoic acid is attached to approximately every third N-acetylmuramic acid. Specifically to *S. pneumoniae*, both the teichoic acid and lipoteichoic acid (attached to the cell membrane via a lipid moiety) are decorated with phosphoryl-choline residues. Choline is an essential requirement for pneumococcal activity, being the ligand by which it attaches to human cells via the platelet-activating-factor receptor and choline-binding protein, CbpA (Dworkin and Falkow 2006; Gehre et al. 2008). Choline is also the ligand by which the *S. pneumoniae* autolysin, LytA, and the endolysin, Cpl-1, bind to the cell wall.

Figure Removed

Figure 1.13 A schematic of a generic peptidoglycan cell wall. Some strains of *S. pneumoniae* also have an outer capsule, and both teichoic and lipoteichoic acids are decorated with choline. Reproduced from Madigan & Martinko (2006).

While highly effective vaccines have been produced to protect humans from *S. pneumoniae* infections, their success has been limited due to poor coverage, and the threat of non-vaccine serotypes requires an active solution (Bogaert et al. 2004). Studies agree that prevention of nasopharyngeal colonization is the most successful method of averting infection, however the rise of antibiotic resistant strains of *S.*

pneumoniae, in particular penicillin-resistant *S. pneumoniae* (PRSP), chloramphenicol resistance, macrolide resistance, and multiresistant strains present a major global health threat (Appelbaum 1992; Adam 2002; Bogaert et al. 2004; Stanek et al. 2011).

The presence of a virulence-factor target (choline) in *S. pneumoniae*, and the rise of multiresistant strains makes it an excellent candidate for endolysin treatment. Furthermore, for non-systemic infections, for example topical applications or inhalers to reduce nasopharyngeal colonization, regulatory requirements are less stringent thus being an attractive target for endolysin proof-of-concept.

1.4.2 Clostridium difficile

Clostridium difficile, which inhabits the gastrointestinal tract, was first isolated in 1935 (Hall and O'Toole 1935) and has had its whole genome sequenced (Sebaihia et al. 2006). It is a Gram-positive anaerobe, has the ability to form spores, and has shown itself to be particularly adept at resisting antibiotics. *C. difficile* infections (CDI) generally occur nosocomially, and are a common problem after antibiotic treatment, causing antibiotic-associated diarrhoea and pseudomembranous colitis.

11 % of the *C. difficile* genome has been shown to consist of mobile genetic elements, including conjugative transposons and phage, conferring high mobility to the genome. This highly mobile genome has enabled *C. difficile* to survive in the highly variable gastrointestinal tract environment, and has benefitted from a niche where opportunities to exchange DNA with other organisms are frequent. It is believed that many of the antibiotic resistance genes present in *C. difficile* are encoded by these mobile genetic elements (Sebaihia et al. 2006). Resistance and multiresistance has been reported to: metronidazole; vancomycin; clindamycin; erythromycin; tetracycline; moxifloxacin; rifampicin; and fusidic acid, via a number of mechanisms including ribosomal modification, antibiotic efflux and drug inactivation (Huang et al. 2009). Furthermore, the ability of *C. difficile* to form spores when it comes under environmental stress causes huge healthcare issues. The resultant spores are long-lived and resistant to many disinfectants and antibiotics, making them a source of infection in many hospital environments (Gerding et al. 2008).

Spore-formation, alongside huge potential for antibiotic resistance, frequently results in a residual population of *C. difficile* surviving in the human gut after a course of antibiotics. This population is then able to rapidly colonise the gut due to lack of competition from natural gut flora, and the result is CDI. Further treatment using antibiotics is then required to treat the recurrent infection which can increase the rate of future recurrence. Faecal transplant therapy, a method of repopulating the gut with the natural microbiota to compete with *C. difficile* and prevent CDI, as well as probiotics have been used relatively successfully to combat this issue (Allen et al. 2014). Endolysin therapy as a treatment for CDI is particularly attractive as it would not disturb the natural gut flora against which *C. difficile* is usually unable to compete, while conventional antibiotics often exacerbate the situation.

1.4.3 Propionibacterium acnes

Propionibacterium acnes is a Gram-positive bacilliform and forms a significant part of the human commensal microbiome. It can be found on the skin, particularly in fatty acid rich areas such as the face and scalp, as well as on internal membranes such as the large intestine and urinary tract (Dessinioti and Katsambas 2017). *P. acnes* is best known for its role in the common skin affliction acne vulgaris, but is also associated with a wide range of inflammatory diseases and implant-associated infections (Portillo et al. 2013). Acne vulgaris is one of the most common global skin diseases, affecting 45 million people in the USA and accounting for approximately 20 % of all dermatologist appointments (Bhatia et al. 2004). Currently treatments include topical and systemic antibiotics, as well as chemical treatments such as benzyl peroxide. However, growing antibiotic resistance and the sometimes severe side effects of chemical therapies call for a new solution.

Due to the prevalence and unsightly nature of acne vulgaris, *P. acnes* has been widely treated with antibiotics for many years, and cross-resistance to clindamycin and erythromycin was first reported in 1979. Since then, resistance to oxytetracycline, doxycycline, and minocycline have also been reported world-wide (Dessinioti and Katsambas 2017). While rarely life-threatening in the form of acne vulgaris, other less-common *P. acnes* infections benefit from resistance commonly acquired when

treating acne, for example, implant-associated infections are a serious threat (Portillo et al. 2013). It is therefore highly desirable to reduce the selective pressure upon *P. acnes* during the treatment of acne.

As a topical infection with a high prevalence of antibiotic resistance and large afflicted population, acne vulgaris, and more precisely *P. acnes,* makes a suitable target for endolysin therapy. Furthermore, the targeted killing of *P. acnes* in the affected region is preferable to elimination of the bacterium across the body, as *P. acnes* has been demonstrated to be a nonspecific immune stimulant, so may be beneficial at some level, although precisely how this mechanism works is not yet understood (Bhatia et al. 2004).

1.5 Summary, aims and objectives

The importance of antibiotics to modern society cannot be overstated. They have wide ranging uses, from enabling invasive surgeries, to large scale food production. For the past 70 years we have become increasingly dependent upon them, but the efficacy of the old drugs is dwindling and new drugs are becoming harder to find.

A number of potential alternative antibiotics exists, each with strengths and weaknesses. From reviewing the literature, it has become evident that our future antibiotic use will require a different strategy to that which we have become accustomed. We will not be able to rely on a handful of broad-spectrum antibiotics, but are likely to require a variety of antimicrobial products to best suit the problem in hand.

One of the possible tools against bacteria that we will be able to call upon is the bacteriophage endolysins. The question remains of how to produce them in a cost-effective and safe manner. Recombinant expression in the *C. reinhardtii* chloroplast provides a potential solution to this problem: featuring GRAS status; lack of glycosylation; quick and precise genetic transformation; and ability to scale to production rapidly. However, low recombinant protein yields have so far held back this platform. Three pathogens, *S. pneumoniae*, *C. difficile,* and *P. acnes* have been selected as suitable targets for endolysin therapy.

This thesis sets out to address the issues discussed above, with the aims of advancing both the use of the *C. reinhardtii* chloroplast as a recombinant protein platform, and the potential for endolysin therapy, using the following objectives:

1. To improve the level of recombinant endolysin accumulation in the *C. reinhardtii* chloroplast.
2. To improve the activity of endolysins produced in the *C. reinhardtii* chloroplast against *S. pneumoniae*.
3. To demonstrate the synthesis and activity of two endolysins never previously expressed in the *C. reinhardtii* chloroplast.

Chapter 2 Materials and Methods

2.1 Strains, media, culture conditions and cell density quantification

2.1.1 Chlamydomonas reinhardtii

2.1.1.1 Strains

Strains of *C. reinhardtii* in this thesis are named in a similar manner to previous work (Taunt 2013; Stoffels 2015), and the convention is illustrated in Figure 2.1.

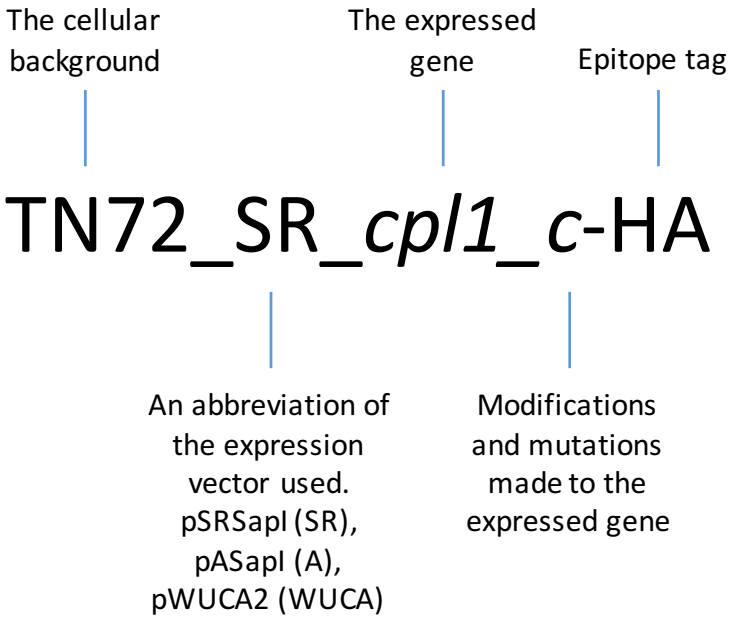


Figure 2.1 A naming convention modified from Taunt (2013) and in line with previous work by Stoffels (2015). This convention is used throughout this thesis, and any minor modifications are noted in the following sections.

STRAIN	DESCRIPTION	REFERENCE
TN72	A recipient strain based upon the cell wall-less strain cw15 (mt+). Partial deletion of *psbH* gene via *aadA* cassette insertion. PSII-deficient mutant. Spectinomycin resistant.	Ninlayarn (2012)
TN72_SR_Control	TN72 transformed with pSRSapI expression vector, containing no GOI, *psbH* rescued.	Young & Purton (2014)

TN72_ SR_*cpl1*-HA	TN72 transformed with pSRSapI containing HA-tagged *cpl1*, *psbH* rescued.	Taunt (2013)
TN72_ SR_*cpl1*_CPO1-HA	TN72 transformed with pSRSapI containing HA-tagged *cpl1* with a single codon change, *psbH* rescued.	This study
TN72_ SR_*cpl1*_CPO2-HA	TN72 transformed with pSRSapI containing HA-tagged *cpl1* with a single codon change, *psbH* rescued.	This study
TN72_ SR_*cpl1*_W202A-HA	TN72 transformed with pSRSapI containing HA-tagged *cpl1* with tryptophan-202 mutated to an alanine, *psbH* rescued.	This study
TN72_ SR_*cpl1*_W209A-HA	TN72 transformed with pSRSapI containing HA-tagged *cpl1* with tryptophan-209 mutated to an alanine, *psbH* rescued.	This study
TN72_ SR_*cpl1*_W230A-HA	TN72 transformed with pSRSapI containing HA-tagged *cpl1* with tryptophan-230 mutated to an alanine, *psbH* rescued.	This study
TN72_ SR_*cpl1*_W202A_W209A-HA	TN72 transformed with pSRSapI containing HA-tagged *cpl1* with tryptophan-202 and -209 mutated to alanine residues, *psbH* rescued.	This study
TN72_ SR_*cpl1*_W223A_W230A-HA	TN72 transformed with pSRSapI containing HA-tagged *cpl1* with tryptophan-223 and -230 mutated to alanine residues, *psbH* rescued.	This study
TN72_ SR_*cpl1*_W202A_W209A_ W223A_W230A-HA	TN72 transformed with pSRSapI containing HA-tagged *cpl1* with tryptophan-202, -209, -223 and -230 mutated to alanine residues, *psbH* rescued.	This study
TN72_ SR_*cpl1*_dimer-HA	TN72 transformed with pSRSapI containing HA-tagged *cpl1* with mutations to putatively enable dimerization, *psbH* rescued.	This study

TN72_ A_*cpl1_c*-HA	TN72 transformed with pASapI containing HA-tagged *cpl1* encoded by a different DNA nucleotide sequence (*cpl1_c*), *psbH* rescued.	This study
TN72_SR_*cpl1-HA*_A_*cpl1_c*-HA	TN72 transformed with a plasmid containing both the HA-tagged *cpl1* under the *psaA* exon 1 promoter, and the HA-tagged *cpl1_c* under the *atpA* promoter, *psbH* rescued	This study
TN72_ SR_*cpl1*-StrepII	TN72 transformed with pSRSapI containing StrepII-tagged *cpl1*, *psbH* rescued	This study
TN72_SR_*cpl1*-StrepII_A_*cpl1*-HA	TN72 transformed with a plasmid containing both the StrepII-tagged *cpl1* under the *psaA* exon 1 promoter, and the HA-tagged *cpl1-c* under the *atpA* promoter, *psbH* rescued	This study
TN72_SR_*cd27l*-HA	TN72 transformed with pSRSapI containing HA-tagged *cd27l*, *psbH* rescued	This study
TN72_ SR_*cd27l*-StrepII	TN72 transformed with pSRSapI containing StrepII-tagged *cd27l*, *psbH* rescued	This study
TN72_ WUCA2_Control	TN72 transformed with pWUCA2 expression vector, containing no GOI, *psbH* rescued.	Young & Purton (2016)
TN72_ WUCA2_*cph1*-HA	TN72 transformed with pWUCA2 containing HA-tagged *cph1*, *psbH* rescued	This study
TN72_A_*gp20*-HA	TN72 transformed with pASapI containing HA-tagged *gp20*, *psbH* rescued	Taunt (2013)
TN72_ WUCA2_*gp20N*-HA	TN72 transformed with pWUCA2 containing N-terminal HA-tagged *gp20*, *psbH* rescued	This study

2.1.1.2 Media

2.1.1.2.1 Tris-acetate phosphate (TAP) medium

NH_4Cl	0.4	g/L
$MgSO_4•7H_2O$	0.1	g/L
$CaCl_2•2H_2O$	0.05	g/L
K_2HPO_4	0.108	g/L
KH_2PO_4	0.056	g/L
Tris	2.42	g/L
Trace element stock solution	1	ml/L
Glacial acetic acid	-	to pH 7.0
Bacto agar (Optional)	2	g/L

2.1.1.2.2 High salt medium (HSM)

NH_4Cl	0.4	g/L
$MgSO_4 • 7H_2O$	0.1	g/L
$CaCl_2 • 2H_2O$	0.05	g/L
K_2HPO_4	0.717	g/L
KH_2PO_4	0.363	g/L
Trace element stock solution	1	ml/L
HCl	-	to pH 6.9
Bacto agar (optional)	2	g/L

2.1.1.2.3 Trace element stock solution (Kropat et al. 2011)

A detailed recipe can be found in Kropat et al. (2011). In brief seven 1000x stock solutions were prepared and then combined into a single solution. 1 ml of each individual stock (7 ml of mixture) was added to each litre of medium. This trace recipe was devised to support a cell density of 2 x10^7 cells/ml (a typical stationary phase cell density), with Cu, Fe, Mn and Zn ions in 3-fold excess and molybdate in 2-fold excess.

STOCK SOLUTION	CONCENTRATION IN STOCK	COMPOSITION
EDTA-Na_2	25 mM	Titrated to pH 8.0 with trace element grade KOH
$(NH_4)_6Mo_7O_{24}$	28.5 µM	
Na_2SeO_3	0.1 mM	
Zn·EDTA	2.5 mM	2.5 mM $ZnSO_4$ in 2.75 mM EDTA
Mn·EDTA	6 mM	6 mM $MnCl_2$ in 6 mM EDTA
Fe·EDTA	20 mM	20 mM $FeCl_3$ and 22 mM Na_2CO_3 in 22 mM EDTA
Cu·EDTA	2 mM	2 mM $CuCl_2$ in 2 mM EDTA

2.1.1.3 Culture conditions

Chlamydomonas reinhardtii was cultured by streaking the cells under aseptic conditions using a sterile toothpick onto plates containing TAP supplemented with 2% bacto agar. The plates were then incubated at 20 °C under a light intensity of 50 µmol/m^2/s.

Liquid starter cultures of *C. reinhardtii* were started in 50 ml flasks, with 25 ml liquid TAP medium. *C. reinhardtii* cells were scrapped from an actively growing colony from a TAP agar plate and smeared onto the inside of the flask, just above the water-line. The cells were then resuspended by swirling the flask to avoid clumping. The flasks were incubated at 25°C, 120 rpm agitation, 24-hour illumination at 50 µmol/m^2/s

light intensity, to mid-log phase. A 100-fold dilution was then performed for larger sub-cultures.

2.1.1.4 Cell density quantification

Cell counting was performed using a haemocytometer. 1 ml sample was extracted from a flask and 10 µl tincture of iodine (19.7 mM iodine in 95 % (v/v) ethanol) was added to immobilise the cells. The number of cells on two grids was then counted by eye at 400x magnification, the mean taken, and multiplied by 10,000 to provide an estimation of the number of cells/ml of culture volume.

Alternatively, the optical density of the culture was measured using a Unicam UV/Vis Spectrometer at 750 nm, with a path-length of 1 cm.

2.1.1.5 Storage

Inoculated TAP plates were stored at 20°C, under dim light conditions (5 µmol/m^2/s) for up to 5 weeks before being re-streaked on fresh plates.

Cryopreservation of *C. reinhardtii* strains was based upon a protocol from Crutchfield et al. (1999). *C. reinhardtii* strains were re-streaked on TAP agar plates every 3 days for a week to ensure active growth states. The cells were then resuspended in 10 ml TAP-Sorbitol (1 %), with the sorbitol acting as an osmoprotector for the cell-wall deficient strains, to a concentration of approximately 10^7 cells/ml. The suspension was gently shaken at 120 rpm to ensure cell dissociation for 30 minutes, and then diluted to 6 x10^6 cells/ml. The cell suspension was then gently mixed in a 1:1 ratio with TAP containing 10 % methanol, under low light conditions. 2 ml aliquots of the final cell mixture were then cooled at an approximate rate of -1 °C/min to -80 °C for 90 minutes, before being stored permanently in a Dewar containing liquid nitrogen.

2.1.2 Escherichia coli

2.1.2.1 Strains

STRAIN	DESCRIPTION	REFERENCE
DH5α	*fhuA2* Δ(*argF-lacZ*)U169 *phoA gln*V44 Φ80 *Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	Hanahan (1983)
DH5α_ProEX_*cpl1_dimer*-HA-StrepII	Carrying pProEX HTb containing HA-tagged *cpl1* with mutations to induce dimerization	This study
DH5α_ProEX_*gp20*-HA	Carrying pProEX HTb containing *gp20* tagged on the C-terminus with an HA tag	This study

2.1.2.2 Media

2.1.2.2.1 Lysogenic broth (LB) (Bertani 1951)

Bacto-tryptone	10	g/L
Yeast Extract	5	g/L
Sodium chloride	10	g/L
Bacto agar (optional)	15	g/L

Where necessary, ampicillin was added to a final concentration of 100 µg/ml. Ampicillin stocks were created at 100 mg/ml in H_2O and stored in aliquots at -20 °C.

2.1.2.3 Culture conditions

E. coli was cultured either on solid LB agar plates at 37 °C, or in liquid LB at 37 °C with constant shaking.

2.1.2.4 Storage

25 % (v/v) glycerol stocks were prepared for long term storage of *E. coli* by mixing 500 µl of stationary phase bacterial culture with 500 µl 50 % (v/v) sterile glycerol. The aliquots were then snap frozen in liquid nitrogen and stored at –80 ℃.

2.1.3 Streptococcus pneumoniae

2.1.3.1 Strains

STRAIN	DESCRIPTION	REFERENCE
D39	Serotype 2, Mouse-passaged clinical isolate	McDaniel et al. (1987)
AL2	D39 LytA^{-}. Defined autolysin-deficient lytA mutant of D39; Em^{r}	Berry et al. (1989)

2.1.3.2 Media

2.1.3.2.1 Trypticase Soy yeast (TSY) extract medium

Trypticase soy broth	30	g/L
Yeast Extract	3	g/L

2.1.3.2.2 Columbia blood agar plates (SIGMA)

Agar	15	g/L
Special nutrient substrate	23	g/L
Sodium chloride	5	g/L
Starch	1	g/L

2.1.3.3 Culture conditions

S. pneumoniae were cultured in liquid culture in TSY medium at 35 ℃ in an anaerobic CO_2 rich environment, using an anaerobic jar from Oxoid and Oxoid AGC CO_2 Gen Compact gas packs. *S. pneumoniae* cultures were grown for 16 h in Trypticase Soy yeast extract medium, sub-cultured the next morning and grown to mid-log phase. Inoculated Columbia blood agar plates were sealed and incubated at 35°C for 18 h.

2.1.3.4 Cell density quantification

Optical density was measured in 96-well plates using a FLUOstar Omega plate reader (BMG Labtech) at 595 nm.

Serial dilution followed by a colony forming unit assay were also used to estimate cell density.

2.1.3.5 Storage

Aliquots of bacterial culture were snap frozen in liquid nitrogen and stored at –80 °C.

2.1.4 Clostridium difficile

2.1.4.1 Strains

STRAIN	DESCRIPTION	REFERENCE
NCTC 11204	Isolated from human neonatal meconium	(Hafiz and Oakley 1976)

2.1.4.2 Media

2.1.4.2.1 Robertsons Cooked Meat Medium (SGL, Corby)

Heart muscle	454.0	g/L
Peptone	10.0	g/L
`Lab-Lemco' powder	10.0	g/L
Sodium chloride	5.0	g/L
Glucose	2.0	g/L

pH 7.2 ± 0.2 @ 25°C

2.1.4.2.2 Liquid Brain Heart Infusion + complements (Mayer et al. 2008)

Brain infusion solids	12.5	g/L
Beef heart infusion solids	5.0	g/L
Proteose peptone	10.0	g/L
Glucose	2.0	g/L
Sodium chloride	5.0	g/L
Disodium phosphate	2.5	g/L
pH 7.4 ± 0.2 @ 25°C		
Complements:		
Vitamin K	50	mg/L
Hemin	5	mg/L
Resaurin	1	mg/L
L-cysteine	0.5	g/L

2.1.4.3 Culture conditions

The *C. difficile* NCTC 11204 spore stock was stored in Robertson's cooked meat medium (SGL, Corby). The stock was then grown in 25 ml of liquid Brain Heart Infusion (BHI) + complements (Mayer et al. 2008). After 16 h of growth at 37 °C, 200-500 µl of *C. difficile* culture was used to inoculate fresh broth and grown to an OD_{595nm} of 1-1.4 (late log phase).

2.1.5 Propionibacterium acnes

2.1.5.1 Strains

STRAIN	DESCRIPTION	REFERENCE
ATCC 6919	Type strain. Strain Designations: NCTC 737 [VPI 0389]	From Laura Stoffels

2.1.5.2 Media

2.1.5.2.1 Liquid Brain Heart Infusion

See 2.1.4.2.2, without complements.

2.1.5.2.2 Columbia blood agar plates (SIGMA)

See 2.1.3.2.2.

2.1.5.3 Culture conditions

P. acnes was grown at 37 °C for 4-5 days in liquid heart brain infusion or on blood agar plates in an anaerobic CO_2 rich environment, using an anaerobic jar from Oxoid and Oxoid AGC CO_2 Gen Compact gas packs.

2.1.5.4 Storage

Aliquots of bacterial suspension were mixed in equal volumes with 50 % (v/v) sterile glycerol and snap frozen in liquid nitrogen before being stored at –80 °C.

2.2 Plasmids

PLASMID	DESCRIPTION	REFERENCE
pSRSapI	Based on the pUC8 backbone, containing the endogenous *psaA* promoter and 5' UTR, with the *rbcL* terminator and 3' UTR. Includes flanking regions around the expression cassette for homologous recombination with the chloroplast genome, and a functional psbH gene	Young and Purton (2014)
pSRSapI_Control	pSRSapI expression vector, containing no gene of interest	Young & Purton (2015)
pSRSapI _*cpl1*-HA	pSRSapI containing HA-tagged *cpl1*	Taunt (2013)
pSRSapI _*cpl1*_CPO1-HA	pSRSapI containing HA-tagged *cpl1* with a single codon change	This study
pSRSapI _*cpl1*_CPO2-HA	pSRSapI containing HA-tagged *cpl1* with a single codon change	This study
pSRSapI _*cpl1*_W202A-HA	pSRSapI containing HA-tagged *cpl1* with tryptophan-202 mutated to an alanine	This study

pSRSapI _*cpl1*_W209A-HA	pSRSapI containing HA-tagged *cpl1* with tryptophan-209 mutated to an alanine	This study
pSRSapI _*cpl1*_W230A-HA	pSRSapI containing HA-tagged *cpl1* with tryptophan-230 mutated to an alanine	This study
pSRSapI _*cpl1*_W202A_W209A-HA	pSRSapI containing HA-tagged *cpl1* with tryptophan-202 and -209 mutated to alanines	This study
pSRSapI _*cpl1*_W223A_W230A-HA	pSRSapI containing HA-tagged *cpl1* with tryptophan-223 and -230 mutated to alanines	This study
pSRSapI _*cpl1*_W202A_W209A_ W223A_W230A-HA	pSRSapI containing HA-tagged *cpl1* with tryptophan-202, -209, -223 and -230 mutated to alanines	This study
pSRSapI _*cpl1*_dimer-HA	pSRSapI containing HA-tagged *cpl1* with mutations to induce dimerization	This study
pASapI	Based on the pUC8 backbone, containing the endogenous *atpA* promoter and 5' UTR, with the *rbcL* terminator and 3' UTR. Includes flanking regions around the expression cassette for homologous recombination with the chloroplast genome, and a functional psbH gene	(Wannathong et al. 2016)
pASapI_*cpl1_c*-HA	pASapI containing HA-tagged *cpl1* with a different DNA nucleotide sequence (*cpl1_c*	This study
pSRSapI _*cpl1-HA*_A_*cpl1_c*-HA	pSRSapI containing both the HA-tagged *cpl1* under the *psaA* exon 1 promoter, and the HA-tagged *cpl1-c* under the *atpA* promoter	This study
pSRSapI _*cpl1*-StrepII	pSRSapI containing StrepII-tagged *cpl1*	This study
pSRSapI _*cpl1*-StrepII_A_*cpl1*-HA	pSRSapI containing both the StrepII-tagged *cpl1* under the *psaA* exon 1 promoter, and the HA-tagged *cpl1-c* under the *atpA* promoter	This study

pSRSapI _*cd27l*-HA	pSRSapI containing HA-tagged *cd27l*	This study
pSRSapI _*cd27l*-StrepII	pSRSapI containing StrepII-tagged *cd27l*	This study
pWUCA2	Based upon pSRSapI, with the introduction of a sythetic $trnW_{UCA}$ gene	Young & Purton (2015)
pWUCA2_Control	pWUCA2 expression vector, containing no GOI	Young & Purton (2015)
pWUCA2_*cph1*-HA	pWUCA2 containing HA-tagged *cph1*	This study
pWUCAC2_*gp20N*-HA	pWUCA2 containing N-terminal HA-tagged *gp20*	This study
pProEX HTb	An expression vector confering Ampicilin resistance and under trc promoter	Gift from J. Santini, Birkbeck College
pProEX_*cpl1_dimer*-HA-StrepII	pProEX HTb containing *cpl1* with mutations to induce dimerization. Tagged on the C-terminus with an HA-tag, followed by a StrepII tag	This study
pProEX_*gp20*-HA	pProEX HTb containing *gp20* tagged on the C-terminus with an HA tag	This study

2.3 Molecular biology

2.3.1 Buffers

2.3.1.1 Sodium phosphate (Na-Pi) buffer

NaH_2PO_4 59.99 g/L

Adjust pH to 7.4 with NaOH

2.3.1.2 Tris-HCl

Tris 121.14 g/L

Adjust pH to 8.0 (unless otherwise stated) with HCl

2.3.2 Gene design and synthesis

Gene sequences were acquired from Genbank (www.ncbi.nlm.nih.gov/genbank) and verified with Blastn. The sequence for an epitope tag was added before the stop codon, unless otherwise indicated.

EPITOPE TAG	AMINO ACID SEQUENCE	NUCLEOTIDE SEQUENCE
HA tag	YPYDVPDYA	5' TAC CCA TAC GAT GTT CCA GAT TAC GCT 3'
StrepII tag	WSHPQFEK	5′ TGG TCT CAC CCA CAA TTC GAA AAA 3′
His tag	HHHHHH	5′ CAT CAC CAT CAC CAT CAC 3′

A SapI and a SphI restriction site were included at the 5' and 3' ends of the gene, respectively, for subsequent cloning unless otherwise indicated. The sequences were then codon optimised using *Codon Usage Optimizer Beta 0.92* (Kong 2013) software developed specifically for *C. reinhardtii* chloroplast genes in the Purton lab by Khai Kong. Genes were then synthesised by Integrated DNA Technologies.

2.3.3 Polymerase chain reaction (PCR)

DNA amplification, when carried out *in vitro,* was performed using the Polymerase Chain Reaction (PCR). The reaction was generally carried out in 50 µl volumes, and using the Fermentas Phusion polymerase. Mix:

5x HF Buffer	10 µl
Template DNA	50-100 ng
Primers	10 pmol each
20mM mixed dNTPs	1 µl
Phusion polymerase	0.5 µl
ddH_2O	35.5 µl

Polymerase Chain Reaction cycling parameters were varied slightly depending upon the precise nature of the reaction being performed: the annealing temperature was calculated depending upon the melting temperature (Tm) of the primers being used, and the extension time was altered depending upon the length of the expected amplicon, according to the Fermentas Phusion polymerase instructions. As an example, for amplifying the entire Cpl-1 gene using a Techne TCF3000X thermocycler the following conditions were used:

98°C	2 min	
98°C	10 s	X 30 cycles
60°C	30 s	
72°C	30 s	
72°C	5 min	
10°C	Hold	

2.3.4 PCR DNA Purification

Following PCR, if the product was not to be loaded onto an agarose gel, it was "cleaned up" by the Qiagen PCR clean-up kit, as per the manufacturer's instruction. The clean-up removed unwanted primers, buffer and polymerase from the reaction mixture. It was also used following restriction digests to remove small (<50 bp) fragments of DNA.

2.3.5 Agarose gel Electrophoresis

1 % agarose gels were used for gel electrophoresis. 10 µg of DNA sample and 1x loading dye were run at 10 volts per centimetre until the ladder is well separated. Either Thermo Scientific Generuler 1kb ladder or Generuler Mix ladder were used to confirm DNA length, for expected large or small bands respectively.

2.3.6 DNA gel extraction

Following agarose gel electrophoresis, DNA bands of interest were excised from the gel using a sterile scalpel and placed into a microcentrifuge tube of predetermined

mass. The excised slice of agarose was the weighed and extracted using the QIAquick Gel Extraction Kit from Qiagen, following the manufacturer's instructions.

2.3.7 Site directed mutagenesis by PCR

Primers were synthesised by Eurofins and used in conjunction with Polymerase Chain Reaction to change a required DNA nucleotide.

Figure 2.2 describes the process.

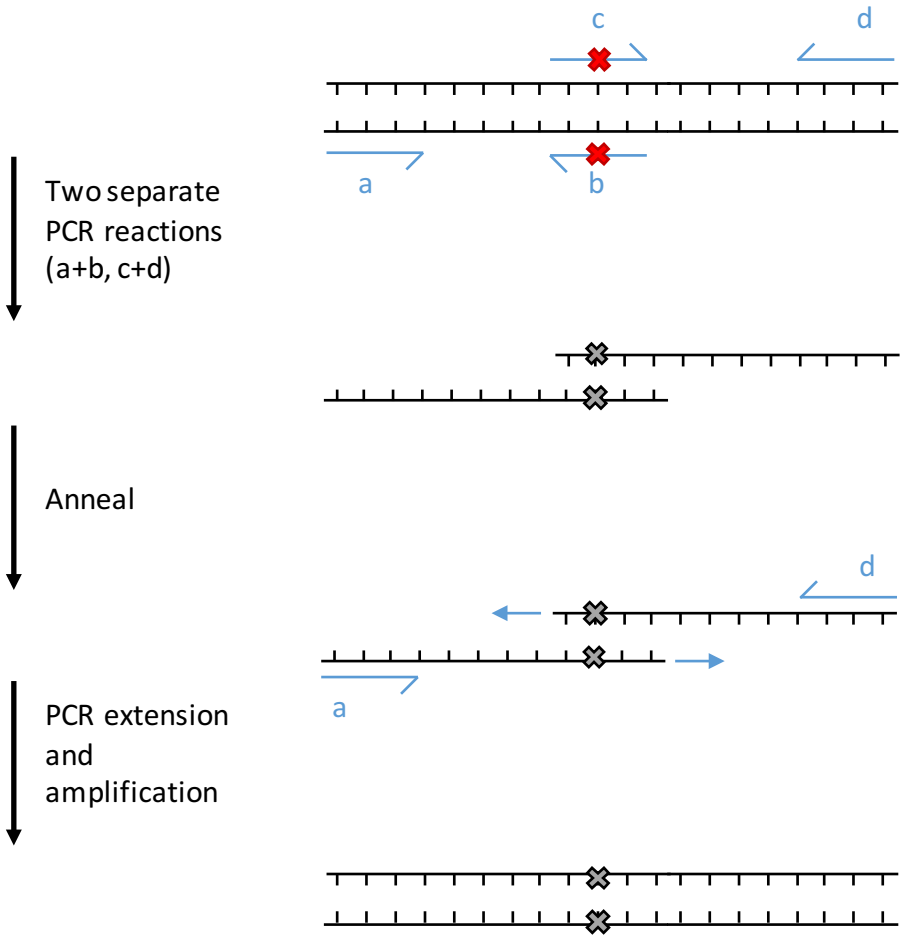


Figure 2.2 Process for site directed mutagenesis. Template DNA is in black and primers are in blue. The 'X' indicates a point mutation. Primers a + b and c + d were initially use in separate reactions. Then the two reactions were combined and 10 cycles of annealing was performed. Finally, a further 25 cycles of extension with primers a and d were performed.

2.3.8 Dephosphorylation using Antarctic Phosphatase

In order to prevent the ligation of linearized plasmid DNA after restriction digestion without the required DNA insertion, the 5' phosphate was removed. For 1 pmol of DNA ends, 5 units of Antarctic phosphatase were added in 1X Antarctic phosphatase reaction buffer. After incubation at 37 °C for 30 minutes, the reaction was stopped by heat inactivation at 80 °C for 2 minutes.

2.3.9 Blunt-end DNA cloning

Blunt-end DNA cloning into the positive selection cloning vector pJET1.2/blunt was carried out using a Thermo Scientific CloneJET PCR Cloning Kit according to the manufacturer's instructions.

2.3.10 DNA Restriction endonuclease digestion

Linearization of the pSRSapI expression vector, and excision of the Gene of Interest (GoI) from pJet was carried out as shown in the table below, followed by 1 hour incubation at 37 °C.

Tango Buffer	4 µl
LguI (SapI)	1.2 µl
PaeI (Sph1)	1 µl
DNA	2 µl
ddH_2O	31.8 µl

2.3.11 DNA Ligation

DNA insert and vector concentrations were adjusted such that there is a 3:1 molar ratio and approximately 20 ng of vector. Then 2 µl of 10x T4 DNA ligase buffer, 1 µl of T4 ligase (New England Biolabs) and ddH_2O up to 20 µl total volume was added. This mixture was then incubated at room temperature for 1 hour.

2.3.12 Competent *E. coli* transformation – Heat shock

Competent *E. coli* (DH5α) stocks, stored at -80°C, were thawed on ice. For transformation using ligation mixes, the whole 10 µl ligation reaction mix was added

to 100 µl of *E. coli* stock in a 1.5 ml microcentrifuge tube. For intact plasmid transformations, approximately 10 ng of DNA was sufficient. An additional tube of competent *E. coli* was also included as a negative control. The tubes were then incubated on ice for 25 minutes, followed by a heat shock for 70 s in a 42 °C water bath, followed by an additional 2 min on ice. 0.5 ml of liquid LB was added to each tube before incubation at 37 °C for 1 hour with shaking. 100 µl from each tube was then spread onto LB amp^{100} or LB kan^{100} plates (depending on antibiotic resistance gene in the plasmid) and incubated at 37 °C overnight.

2.3.13 *C. reinhardtii* transformation – glass bead

TN72 is a photosynthesis-deficient, cell wall-deficient, *C. reinhardtii* recipient strain that can be grown heterotrophically on media containing a carbon source e.g. TAP medium. It is also photosensitive so cultures grow better when kept in the dark. TN72 colonies were regularly restreaked on fresh TAP agar plates over three weeks until growing healthily before starting a 400 ml liquid TAP culture.

The vector pSRSapI (Young and Purton 2014) contains an expression cassette into which a gene of interest can be inserted between SapI and SphI restriction sites. pSRSapI can then rescue photosynthetic activity in TN72 after transformation as it carries the photosystem II gene *psbH,* which is non-functional in TN72. Successful transformants can then be selected for on minimal media under bright light, by their ability to grow phototrophically.

The protocol followed for glass bead transformation was based upon Kindle et al. (1991). The TN72 cells were harvested by centrifugation at 3,800 x g for 15 minutes, once the culture density reached 1-2 $x10^6$ cells/ml. The pellet was then resuspended in liquid HSM to give a final concentration of 2 $x10^8$ cells/ml. 300 µl of resuspended cells were transferred to an autoclaved glass test tube containing approximately 300 mg of 0.4 mm diameter pre-washed glass beads. 10 µg of DNA were added per tube to 11 tubes, and one was used as a negative control - with no DNA added. The tubes were vortexed for 15 s, and then 3.5 ml of warm soft HSM agar were added, quickly swirled and poured onto HSM plates. After 20 minutes the plates were inverted and

sealed with Parafilm, making two small piercings in the film with a sterile pipette tip. The plates were then placed under a bright light for 21 days. All colonies present after this time were twice restreaked onto fresh HSM plates.

2.3.14 *E. coli* colony PCR

E. coli colonies were picked using a sterile pipette tip and mixed well in a tube containing Taq Polymerase PCR mix containing relevant primers. PCR was carried out as described previously, but with an extra preliminary 2 minutes of cell lysis at 98°C. Confirmed transformants were picked again from the plate and cultured for further work.

2.3.15 *E. coli* plasmid isolation – Crude and Kit

For plasmid preparations of less than 12 µg a Qiagen Miniprep kit was used as per the manufacturer's instruction, while for larger preparations (<200 µg) a Qiagen Midiprep kit was used.

2.3.16 *E. coli* test digest

Tango Buffer	1 µl
LguI (SapI)	0.2 µl
PaeI (SphI)	0.2 µl
DNA sample	2 µl
ddH_2O	6.6 µl

A digestion was performed on plasmid mini/midi preparations as above and incubated at 37 °C for 1 hour. Gel electrophoresis was used to ensure the correct insertion had been excised by digestion.

2.3.17 *C. reinhardtii* genomic DNA isolation

This method allows checking for homoplasmicity, that is, to ensure that all ~80 copies of the chloroplast plastome have been successfully transformed. *C. reinhardtii* cells were picked using a sterile pipette tip and resuspended in 20 µl of H_2O. Then 20 µl of

100 % ethanol and 200 µl of 5 % CHELEX were added and mixed. After incubation for 10 minutes in a 98°C heatblock and then centrifugation at 13,000 g for 10 minutes, the supernatant was used in a PCR using the following primers (unless stated otherwise):

Flank1:	GTCATTGCGAAAATACTGGTGC	
rbcL.Fn:	CGGATGTAACTCAATCGGTAG	
RY-psaR:	CATGGATTTCTCCTTATAATAAC	(For pSRSapI transformants)
atpA.R:	ACGTCCACAGGCGTCGT	(For pASapI transformants)

Successful recombination: primers Flank1 and RY-psaR gave a 1,135 bp product, and Flank1 and atpa.R gave a 1,204 bp product. Parental sequence (e.g. TN72): primers Flank1 and rbcL.Fn gave an 850 bp product.

2.3.18 DNA Sequencing

Source Biosciences carried out DNA sequencing upon providing DNA samples and primers according to their instructions.

2.4 Protein analysis

2.4.1 Preparation of total protein extracts

The OD_{740} of *C. reinhardtii* culture was measured. The culture was then centrifuged (3,795 g, 20 minutes) and the pellet resuspended in a volume corresponding to the obtained OD. For example, a 10 ml *C. reinhardtii* culture, with an OD of 0.569, would be resuspended in 0.569 ml of buffer (20 mM NaPi buffer + Protease inhibitor unless otherwise stated). Then 10 µl of 4x protein loading dye (200 mM Tris-HCl pH 6.8, 8 % (w/v) SDS, 40 % (v/v) glycerol, 4 % (v/v) β-mercaptoethanol, 50 mM EDTA, 0.08 % (w/v) bromophenol blue) was added to 30 µl of the sample to a final 1x concentration. The mixture was then boiled for 3 min at 98 °C and centrifuged at 11,000 g for 2 min. The resulting supernatant was used as the prepared sample.

2.4.2 Sodium dodecyl sulphate polyacrylamide gels (SDS-PAGE)

The Bio-Rad mini-PROTEAN Tetra System was used to cast SDS-PAGE gels. The gels consisted of a 15 % polyacrylamide resolving layer, topped with a 3.75 % polyacrylamide stacking layer, as detailed below.

RESOLVING GEL (FOR 2 GELS)

Acrylamide/bisacrylamide (Sigma) 40% stock, acrylamide:bisacrylamide = 37:1	3.75	ml
Resolving buffer (8x stock) 0.25M Tris, 1.92M glycine, 1% SDS, pH8.3	1.25	ml
10 % SDS	0.1	ml
ddH_2O	4	ml
10 % ammonium persulphate	0.375	ml
Tetramethylethlenediamine	3.75	µl

STACKING GEL (FOR 2 GELS)

Acrylamide/bisacrylamide (Sigma) 40 % stock, acrylamide:bisacrylamide = 37:1	0.47	ml
Stacking buffer (4x stock) 0.5M Tris-HCl (pH 6.8)	1.25	ml
10 % SDS	0.05	ml
ddH_2O	3	ml
10 % ammonium persulphate	0.25	ml
Tetramethylethlenediamine	3.75	µl

RESERVOIR BUFFER (10X STOCK)

Tris	0.25	M
Glycine	1.92	M
SDS	1	%
HCl	To pH 8.3	

The prepared sample (20 µl for 10-well gel or 12 µl for 15-well gel) and 5 µl of protein ladder (Color Prestained Protein Standard, Broad Range (11-245 kDa), (New England Biolabs)) were loaded onto the gel and run in reservoir buffer for 120 minutes at 120 V.

2.4.3 Protein preparation for dot blot

The *C. reinhardtii* culture was resuspended as for the SDS-PAGE preparation. The freeze-thaw method was then used to mechanically break the cells (3 x freezing in liquid nitrogen, thawing at 35 °C). Then the samples were centrifuged (16,200 g, 2 min) and 2 µl of supernatant was pipetted directly onto the Hybond-ECL nitrocellulose membranes (GE Healthcare). The membrane was allowed to dry for 10 min and then immuno-detection was performed as below.

2.4.4 Coomassie Brilliant Blue R staining

The SDS gel was incubated in Coomassie Brilliant Blue R solution for 1 hour, with gentle agitation, followed by washing in destaining solution overnight. Gels were photographed using the LiCor Odyssey® CLx imaging system.

Coomassie Blue Solution:

Coomassie Brilllant Blue R250	2.5 g/L
Acetic acid	100 ml/L
Methanol	450 ml/L
ddH_2O	450 ml/L

Destaining Solution:

Acetic acid	100 ml/L
Methanol	450 ml/L
ddH_2O	450 ml/L

2.4.5 Western blot analysis

Semi-dry electrophoresis was used to transfer separated proteins from SDS-PAGE onto Hybond-ECL nitrocellulose membranes (GE Healthcare). Initially the gel, 3MM Whatman paper and nitrocellulose were soaked in Towbin buffer (25 mM Tris, 192 mM glycine and 20 % (v/v) methanol) at room temperature for 10 minutes. A transfer stack was built, consisting (in order from top to bottom) of 4 sheets of Whatman paper, SDS-PAGE gel, nitrocellulose membrane, 4 more sheets of Whatman paper. This was then placed into a Bio-Rad Trans-Blot SD semi-dry electrophoretic transfer system, and all air bubbles were rolled out and excess buffer removed. Transfer was carried out according to the manufacturer's instructions, at a constant voltage of 20 V (Fisons FEC 570 power pack) for 1 hour.

2.4.6 Immuno-detection

2.4.6.1.1 Antibodies

ANTIBODY	SOURCE	DILUTION
Primary antibodies		
Anti-HA (polyclonal, rabbit)	Sigma Aldrich, #H6908	1:2,000
Anti-rbcL (polyconal, rabbit)	Gift from J. Gray, University of Cambridge	1:20,000
Anti-StrepII (polyclonal, rabbit)	Abcam, #ab76949	1:2,000
Secondary antibodies		
Goat anti-Rabbit IgG (H+L) Secondary Antibody, DyLight 488	ThermoFisher Scientific, #35552	1:20,000

2.4.6.1.2 Protocol

After transfer, the nitrocellulose membrane was blocked in 3% (w/v) skimmed milk powder in TBS-T (20 mM Tris base, pH adjusted to 7.4 with 5 M HCl, 137 mM NaCl, 0.1 % (v/v) Tween-20) for 1 hour at room temperature, or overnight at 4°C. This was followed by primary antibody treatment in 1.5 % skimmed milk powder in TBS-T for 1 hour at room temperature while shaking at 60 rpm. The membrane was then rinsed twice in TBS-T, followed by a 10 min wash in TBS-T at 60 rpm. The secondary antibody treatment was then applied 1.5 % milk solution, but diluted according to the manufacturer's instructions. Detection was performed using a LiCor Odyssey® CLx System and analysed using Odyssey® Image Studio Lite Version 5.5.

2.4.7 Protein quantification assay

In order to determine the total soluble protein concentration of samples, a Bradford assay was performed in a 96 well plate. This was performed according to the manufacturer's instructions (Sigma Aldrich). Bovine serum albumin (BSA) dilutions were used to generate a standard curve. Then various dilutions of the unknown sample were prepared to find one with a concentration in the range of 0.1-1.4 mg/ml. 250 µl of Bradford reagent was added to each well containing the BSA or unknown sample and the plate was incubated at room temperature for 10 minutes. A plate reader was used to measure absorbance at 595 nm. The net absorbance vs. the protein concentration of each standard were plotted, and then by comparing the net A_{595} values of the sample against the standard curve it is possible to determine the concentration of protein in the unknown sample.

2.5 *C. reinhardtii* protein purification

2.5.1 Crude extract preparation

C. reinhardtii cell cultures were grown to an OD_{740nm} of 2 – 3 and harvested by centrifugation at 5,000 *g* for 20 minutes. The pelleted cells were then resuspended in 20 mM Na Pi-buffer (NaH_2PO_4, the pH was adjusted to 6.9 with NaOH), or Binding buffer (20 mM NaH_2PO_4, 280 mM NaCl, 6 mM potassium chloride, pH 7.4) if being

used for later column purification, to 50x concentration. Roche cOmplete, EDTA-free protease inhibitor was also added. The cells were either broken through free-thawing (3 cycles of freezing in liquid nitrogen and thawing in a 37 °C water bath), or via glass bead abrasion (1 g of 0.4 mm diameter glass beads was added per 10 ml of concentrated cells, followed by vigorous shaking at 4°C for 10 minutes). Subsequently the broken cell suspension was centrifuged at 21,000 *g* for 10 minutes and the supernatant collected. The supernatant is referred to as crude protein extract.

To further remove cell debris and insoluble proteins from the supernatant, ultracentrifugation at 100,000 g for 1 hour was performed, producing a supernatant referred to as ultracentrifuged crude protein extract.

2.5.2 Ammonium sulphate precipitation

Ultracentrifuged crude protein extracts were stirred on ice and powdered ammonium sulphate was added until the desired percentage of saturation was reached. After 30 minutes of stirring the solution was centrifuged for 30 minutes at 3,000 g. The pellet was collected for further analysis and the process was repeated to the supernatant to collect all the desired protein fractions.

2.5.3 Anti-StrepII purification column

Ultracentrifuged crude protein extracts, in binding buffer, were passed through a 1ml StrepTrap HP HiTrap™ column (GE Healthcare) at 1 ml per minute using a peristaltic pump and eluted using 2.5 mM desthiobiotin in binding buffer, as per the manufacturer's instructions.

2.6 *E. coli* protein preparation

Three different methods of protein preparation were used for *E. coli*-produced recombinant proteins:

2.6.1 IPTG induction

For *E. coli* transformed with the pProEX HTb plasmid, transgene expression must be induced using Isopropyl β-D-1-thiogalactopyranoside (IPTG). An overnight culture of

E. coli was diluted 1:100 and grown in 2 ml LB + amp^{100} for 4 hours at 37 °C. After 4 hours, 1 ml was snap frozen in liquid nitrogen to provide the "uninduced" control. 1 ml of LB + amp^{100} + 1 mM IPTG was then added to the culture and incubated at 20 °C overnight, to provide an induced culture which was then analysed by western blotting or used for further investigation.

2.6.2 BugBuster™ (Novagen)

BugBuster Protein Extraction Reagent was used as according to the manufacturer's instructions. In brief, *E. coli* cells were harvested by centrifugation (16,000 g for 10 minutes) and resuspended in BugBuster reagent (2.5 ml of BugBuster for every 50 ml of bacterial culture). The suspension was incubated for 10 minutes at room temperature with gentle shaking. Centrifugation at 16,000 g for 20 minutes at 4 °C removed the insoluble fraction and the supernatant was used for SDS-PAGE analysis or purification.

2.6.3 High Pressure Homogenizer

E. coli were harvested by centrifugation (16,000 g for 10 minutes) and resuspended in $1/5^{th}$ culture volume of NaPi buffer. A pre-equilibrated high pressure homogenizer (Stansted Fluid Power) was used to lyse the cells at 14 psi and washed with buffer between runs. The resulting cell extract was centrifuged (16,000 g for 30 minutes at 4 °C) to remove cell debris and the resulting supernatant was filter sterilised (0.22 µm) before being used for colony forming unit assays or SDS-PAGE analysis.

2.6.4 Sonication

Overnight cultures of *E. coli* were harvested by centrifugation (16,000 g for 10 minutes). The resulting pellet was resuspended in $1/5^{th}$ culture volume of 10mM Tris-HCl buffer (pH 8.0) and placed in a polystyrene microcentrifuge tube. Sonication was performed using a Q700 cup-horn Sonicator (QSonica). The bath was pre-cooled to 4 °C and sonication performed at 100 % amplitude for a total of four minutes (30 seconds on, 30 seconds off). The lysed cells were centrifuged (16,000 g for 30 minutes

at 4 °C) and the pellet and supernatant analysed by SDS-PAGE or used for further analysis.

2.6.5 Disulphide bond formation *in vitro*

While disulphide bond formation *in vitro* was unsuccessful in this study, the following method was used, adapted from Hyvonen (2002). Sonication was used to lyse induced *E. coli* cells resuspended in 10mM Tris-HCl buffer (pH 8.0) as described in 2.6.4, and the resulting supernatant was used for this analysis. To create an oxidising environment conducive to disulphide bond formation, glutathione was used as a redox reagent, both in its oxidised and reduced forms (GSSG and GSH respectively). Solutions of 10mM GSSG and GSH were made fresh from powdered stocks (SIGMA). 25 µl of *E. coli* cell extract was combined with GSSG and GSH in order to achieve a range of ratios between the two. In general, the final concentration of GSH was 2-5 mM, with 10-fold less of the oxidised form. The mixtures were incubated overnight at 20 °C before being separated by SDS-PAGE and analysed by western blotting. No β-mercaptoethanol was included in the loading dye and the mixture was not boiled before running on the gel.

2.7 Antibacterial activity assays

2.7.1 Colony forming unit assay

2.7.1.1 *Streptococcus pneumoniae*

S. pneumoniae cultures were grown as described above. 100 µl of culture was mixed with endolysin samples and incubated at 37 °C for 1 hour. The mixture was subsequently diluted in ten 10-fold dilution steps. 20µl of dilution numbers 6-10 were spotted, in triplicate, onto Columbia blood agar plates and incubated at 37 °C overnight. The number of colonies in each spot were counted by eye and the number of colony forming units per ml was calculated.

2.7.1.2 *Propionibacterium acnes*

Cultures of *P. acnes* were grown anaerobically as described above. The cultures were centrifuged at 5,000 g for 20 minutes and resuspended in 20 mM NaPi buffer. 50 µl

of bacterial suspension and 50 µl of filter sterilized cell lysate or buffer were incubated at 37 °C for 0, 30, 60, 90, 120 or 150 minutes. 10 µl samples were taken after the incubation time from each combination and spotted in triplicate onto Columbia blood agar plates. After anaerobic incubation at 37 °C for one week a comparison was made subjectively between growth as individual colony growth was not visible.

2.7.2 Turbidity reduction assay

2.7.2.1 *Streptococcus pneumoniae*

S. pneumoniae cultures were harvested by centrifugation (15 min, 5,500 g) and the cells were resuspended in 20 mM NaPi-buffer to an OD_{600nm} of ~0.8.

The turbidity reduction assays were performed in 96-well microtitre plates with a final assay volume of 300 µl. An ELx 808 microplate reader (BIO-TEK INSTRUMENTS INC.) measured the OD_{595nm} every minute for 200 minutes, including 2 s of shaking before each measurement, at 37 °C.

Unless otherwise stated, 20 µl of antimicrobial sample, with varying total soluble protein quantities were added to 280 µl of suspension of the target bacterium. As a positive lysis control for *S. pneumoniae* turbidity reduction assays, 20 µl of deoxycholate was added. As a negative control, samples were prepared from TN72_SR_Control.

2.7.2.2 *Escherichia coli*

E. coli cultures were grown in LB to stationary phase and pelleted by centrifugation (3,800 g, 5 min). The following protocol to disrupt *E. coli* membranes is based on Nakimbugwe et al. (2006):

Chloroform containing Tris-HCl buffer was prepared by adding 30 % (v/v) chloroform to Tri-HCl (50 mM, pH 7.0) and shaking for 30 min, 200 rpm, at 25 °C. The top phase of this mixture was then used to resuspend the *E. coli* pellet in its original volume. This was then shaken for 45 min at 25 °C and then centrifuged again (8000 g, 20 min, 4 °C). The pellet was then washed with Tris-HCl and resuspended in Tris-HCl buffer to

an OD_{600nm} between 0.7-0.8. 100 U of egg white lysozyme was used per well as a positive lysis control.

The assay was then performed as described for *S. pneumoniae* above.

2.7.2.3 *Clostridium difficile*

*C. difficile c*ells were harvested by centrifugation (2 min at 5,000 g) and resuspended in a volume of 20 mM Na Pi buffer equivalent to their OD (e.g. 1.2 ml if an OD_{595nm} of 1.2 was measured). 10-50 µg of *C. reinhardtii* extract in 30 µl final volume was added to 270 µl of prepared *C. difficile* cells in buffer. Positive controls were 100 U/well of mutanolysin and 30 µl of CD27L NiNTA-purified protein (unknown concentration) from the Mayer lab. A Labsystems Bioscreen C (Thermo Fisher Scientific) measured the OD every 2 minutes for 18 h, at 37 °C, in a Thermo Fisher 100-well honeycomb plate.

Chapter 3 Improving accumulation levels of Cpl-1 in the chloroplast of *Chlamydomonas reinhardtii*

3.1 Introduction

3.1.1 Introduction to bacteriophage lysin Cpl-1

The bacteriophage Cp1 was first isolated in 1981, from throat samples of healthy children (Ronda et al. 1981). Cp1 infects the bacterium *Streptococcus pneumoniae*, which causes a range of globally significant diseases. *S. pneumoniae* is a causative agent of pneumonia, meningitis and febrile bacteraemia, causing a total annual death toll of 1.6 million people, mainly young children and the elderly in developing countries (WHO 2007).

In 1987 the 39 kDa protein of Cp1 was characterised as a muramidase and named Cpl-1 (García et al. 1987). The same study noted that choline was required in the pneumococcal cell wall for Cpl-1 to be active, and in 2003 the crystal structure of Cpl-1 was determined, showing two active choline binding sites (Hermoso et al. 2003). Cpl-1, displayed in Figure 3.1, is a 339-residue polypeptide chain consisting of an N-terminal catalytic domain and a C-terminal cell wall binding domain (CBD), joined by a 10 residue acid linker.

Cpl-1 belongs to the glycosyl hydrolase (GH) 25 family, based upon the amino acid sequence similarity of its catalytic domain, and is built of modules homologous to LytC, a major *S. pneumoniae* autolysin (Monterroso et al. 2008). The catalytic domain displays a characteristic peptidoglycan binding site and a deep hole of highly negative electrostatic potential located at the C-terminal end of the catalytic barrel that constitutes the active site.

The CBD is known to bind the choline moiety in the teichoic and lipoteichoic acid of the *S. pneumoniae* cell wall (see Figure 1.13), with two identified accessible choline binding domains on the protein. The addition of free choline, or the substitution of choline for ethanolamine in the *S. pneumoniae* cell wall, inhibits Cpl-1 activity. The CBD is composed of two well-defined regions, named CI and CII, containing a total of

six homologous repeats of about 20 amino acids. CI consists of four repeats in a left-handed superhelical arrangement and includes both active choline binding sites, while CII has two repeats arranged as an antiparallel six-stranded β-sheet.

Figure 3.1 Cpl-1 crystal structure. N-terminal catalytic domain (green) is linked via a flexible acidic linker (orange) to the CI domain (blue) and the CII domain (pink). The bound choline molecules are displayed in situ, in ball-and-stick representation. Image produced using Jmol (PDB 1oba).

Production of Cpl-1 is attractive due to its antimicrobial potential. Vincent Fischetti was the first to propose using Cpl-1 as a novel antibiotic and his group showed its efficacy in bacteremic mice (Jado et al. 2003). However, as previously discussed in 1.1.2, the antibiotic market is one of notoriously low margins, and producing biologics is currently an expensive undertaking. This is due in no small part to the expensive purification process involved in removing antigens and endotoxins associated with traditional protein production platforms, such as *E. coli*. Given the GRAS status of *C.*

reinhardtii, we speculate that purifying recombinant Cpl-1 from the *C. reinhardtii* chloroplast may not be required to be as rigorous, and therefore be cheaper.

3.1.2 Factors to consider in protein accumulation

The relatively low levels of recombinant protein accumulation achieved in the *C. reinhardtii* chloroplast, when compared to existing platforms such as *E. coli,* offer an area for improvement. This chapter discusses two techniques investigated to improve accumulation of the Cpl-1 protein in the *C. reinhardtii* chloroplast.

When protein accumulation is considered, one must break the issue down into two key factors, protein synthesis and protein degradation, and investigate how the balance of the two is leading to the observed accumulation.

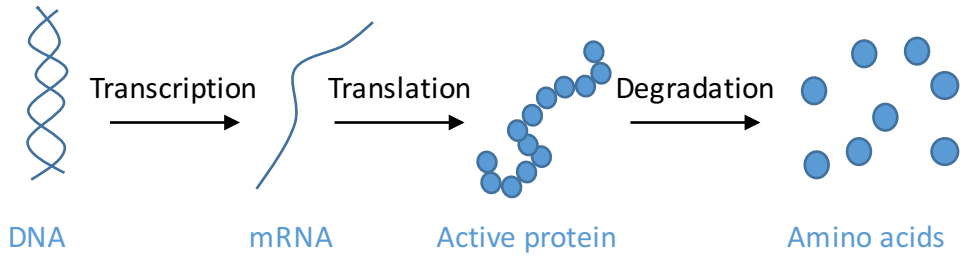


Figure 3.2 Protein turnover is composed of protein synthesis (transcription and translation) and protein degradation.

The literature has described that an increased quantity of mRNA, i.e. by increasing levels of transcription, does not appear to improve the final levels of protein accumulation (Coragliotti et al. 2011). Given the polyploid nature of the *C. reinhardtii* chloroplast and the strength of promoters from chloroplast genes encoding photosynthetic components, one can imagine that the rate of synthesis of transgene transcripts is likely to be high (Hosler et al. 1989; Blowers et al. 1990). The same study points to translation, and in particular ribosome binding and translation elongation, as likely limiting factors. Work by Taunt (2013) indicates that Cpl-1 degradation is minimal in the *C. reinhardtii* chloroplast.

This chapter, therefore, investigates two potential molecular methods for improving protein accumulation. Firstly, constructs containing two similar genes under different

promoters were inserted into the chloroplast in order to potentially overcome the process of Control by Epistasy of Synthesis (Choquet et al. 1998). Secondly, continuing upon previous work by Taunt (2013), codon-pair optimization is investigated in order to improve translation.

Experiments and calculations by Taunt (2013) estimated that a mid-log culture of *C. reinhardtii* can accumulate Cpl-1 to approximately 8 mg L^{-1}, or 17 mg g^{-1} dry weight. A comparable endolysin, Pal, produced by Stoffels (2015) achieved slightly lower levels of accumulation: 5 mg L^{-1} or 13 mg g^{-1} dry weight. These offer a bench mark upon which we should aim to improve. However, in this thesis dry cell weight was not considered, as generally we were interested in relative improvements, rather than actual yields of protein. However, moving forward a clear understanding of dry cell weights achieved and their reliable quantification will be important in understanding the commercial viability of the platform.

3.1.3 Epistasy of synthesis in the *C. reinhardtii* chloroplast

It has been reported that in the *C. reinhardtii* chloroplast the 5' UTR of *petA*, which encodes the apoprotein of cytochrome *f*, mediates its autoregulation of translation via interaction with the C-terminal domain of the unassembled protein (Choquet et al. 1998). This process is called Control by Epistasy of Synthesis (CES) and effectively means that proteins have the ability to control the rate of translation of their own mRNA. Therefore, when using an endogenous 5' UTR for transgene expression, the native associated protein levels may control the level of the transgene product.

We hypothesize in this chapter that the CES process is limiting the level of recombinant protein synthesis, due to the presence of associated endogenous proteins. By introducing two copies of *cpl1* and expressing them under the control of different 5' promoter/UTRs in the *C. reinhardtii* chloroplast we aim to increase overall levels of Cpl-1 accumulation.

3.1.4 Codon usage in the *C. reinhardtii* chloroplast genome

Redundancy in the genetic code results in multiple codons coding for the same amino acid. It is well established that there is a difference in preference amongst organisms for which particular codon codes for which amino acid, and this is known as codon bias. By optimizing nucleotide sequences such that preferred codons are used, it is possible to improve gene expression compared to un-optimized versions (Franklin et al. 2002).

Protein coding sequence analysis in *Escherichia coli* by Gutman & Hatfield (1989) revealed that there is also a strong bias in codon pairs, in that they do not appear to be randomly distributed. This is believed to be due to the varying compatibility of adjacent aminoacyl-tRNA acceptors at the A and P-sites of a ribosome. This, in turn, led to the suggestion that codon pairs can be optimized in a similar way to codons, and thus further improve gene expression levels. Indeed, Irwin et al. (1995) produced evidence to suggest that the influences of codon pair choice are more important than the choice of individual codons in *E. coli*, exemplified by the fact that ACC and CUG are frequently used codons, but ACC:CUG and CUG:ACC are translated at very different rates.

Previous work by Taunt (2013) recognised a similar, non-random distribution of codon pairs in the *C. reinhardtii* chloroplast plastome. One pair in particular stood out as appearing extremely non-random, the phenylalanine:arginine pair, UUU:CGU. Under a random codon distribution, one would expect to observe this pair approximately 28 times in the chloroplast plastome, but in fact it never occurs. Interestingly, the UUU:CGU pair is present in the well expressed *cpl1* gene. This project investigates the impact of this codon pair upon Cpl-1 expression.

3.2 Aims and objectives

1. To identify and optimize a method for comparing protein accumulation levels between different transgenic *C. reinhardtii* lines.
2. To produce transgenic lines of *C. reinhardtii* in which two copies of *cpl1* are introduced under the control of different 5’ UTR sequences.

3. To show improved levels of protein accumulation in the new transgenic lines compared with existing lines.
4. To produce two transgenic lines of *C. reinhardtii* with different codon pair optimization strategies.
5. To compare protein accumulation levels of the two codon pair optimized versions.

3.3 Results and discussion

3.3.1 Investigation into measuring protein accumulation levels

The first hurdle to overcome when attempting to improve protein accumulation levels in the *C. reinhardtii* chloroplast was to identify and optimize an accurate method for comparing the protein accumulation levels between two cell lines. Given the almost impossible task of starting two or more cultures with an identical number of cells, in the same growth phase, and growing them under identical physiological conditions, it was apparent that a suitable normalisation technique was required. Of particular concern is the ability to estimate the error introduced at each stage of the protein preparation and quantification process.

3.3.1.1 Quantifying cells per millilitre of culture

When comparing two or more algal cultures, one must be sure to inoculate the culture with the same number of cells. This immediately introduces an opportunity for error, which here we aim to quantify. There are essentially two practised methods to estimate the cell density of an algal culture – either direct counting using a haemocytometer and microscope, or by using optical density (OD) as a proxy for cell number. Both of these methods are investigated and this section examines the error associated with each method.

The following set of experiments were carried out as follows: three flasks containing 400ml TAP medium were inoculated with equal starter volumes from a single *C. reinhardtii* starter culture. These were incubated and measured in the Algem photobioreactor for 96 hours. After this, the three flasks (labelled: a, b and c) were

used to provide samples to test all of the subsequent techniques. Therefore, the following set of experiments were performed using precisely the same algal cultures, enabling them to be compared.

Two errors will be measured with each technique. Firstly, the error between identical samples – e.g. measuring the same cuvette three times. This gives an indication of the machine/technique error. Secondly, the error between flasks, which will provide an understanding of the pipetting error, a combination of measuring out of the medium and in adding the starter culture.

3.3.1.1.1 Algem Photobioreactor

The Algem Photobioreactor automatically measures Optical Density at 740 nm, while maintaining near identical physiological conditions in three separate flasks. By starting three *C. reinhardtii* cultures, and comparing their OD_{740} over 96 hours, it is possible to see the error introduced.

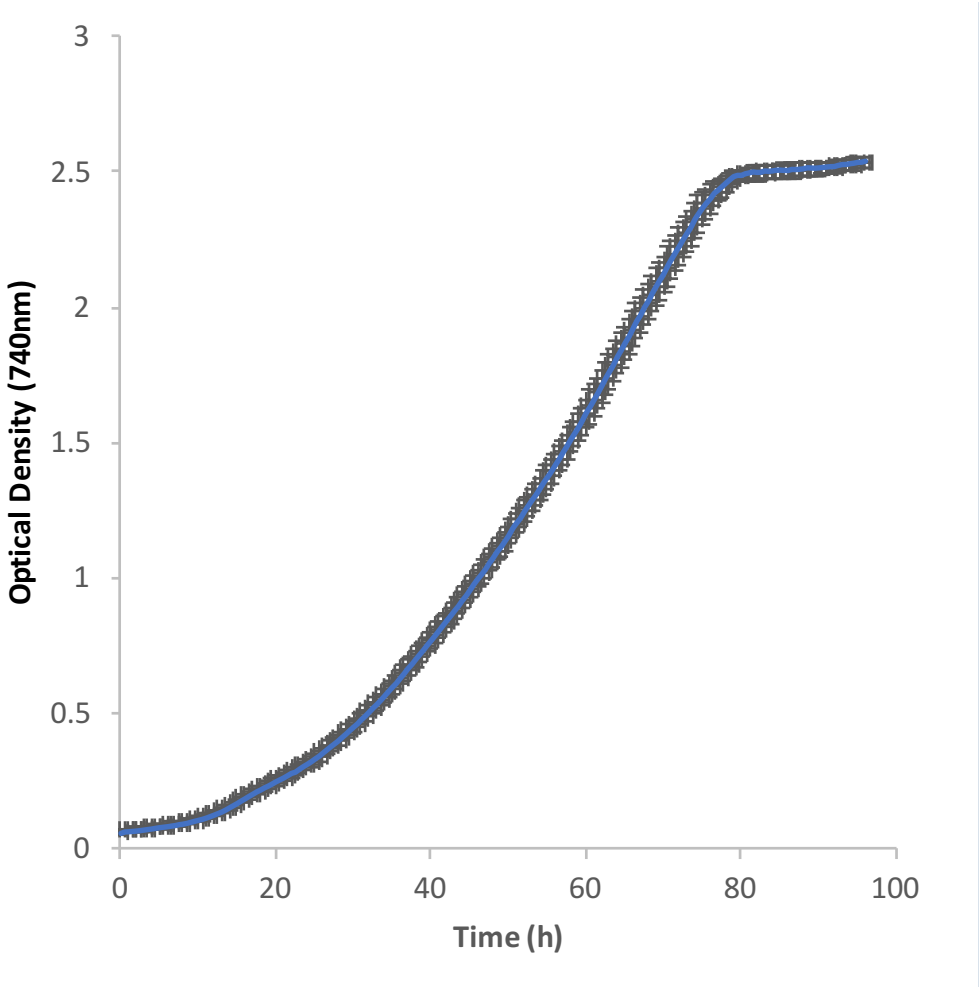


Figure 3.3 Three flasks containing the same volume of TAP medium and the same volume of starter culture (400 ml TAP medium with 2 ml of starter culture) were grown in the Algem Photobioreactor for 96 hours. The average OD740 of the three flasks was calculated and the error bars indicate the standard deviation between the flasks.

In order to estimate the error introduced by the Algem photobioreactor, the standard deviation between every three measurements in flask A was calculated. As each measurement was taken at 30-minute intervals, then for the purposes of estimation, three consecutive measurements will be considered to be from the same time point. The average percentage error calculated in this manner was 1.99 %.

Meanwhile, the average error between the three separate cultures, shown by the error bars in Figure 3.3, was 4.17 %.

3.3.1.1.2 Optical density error

Three biological replicates of *C. reinhardtii* cultures, a, b and c, were sampled and measured three times each using a Unicam UV/Vis Spectrometer (Thermo Electron Corporation, USA) at 750 nm (OD_{740}):

		OD750			Calculations	
		1	2	3	St. dev.	% error
Culture	a	0.541	0.532	0.540	0.0040	0.25%
	b	0.479	0.499	0.480	0.0092	0.63%
	c	0.499	0.503	0.505	0.0025	0.17%
				Av.	0.0052	0.35%

Three different samples were then taken from the same flasks and their OD_{740} measured in the same manner:

		OD750			Calculations	
		1	2	3	St. dev.	% error
Culture	a	0.532	0.569	0.538	0.0162	0.99%
	b	0.509	0.499	0.522	0.0094	0.62%
	c	0.503	0.521	0.499	0.0095	0.63%
				Av.	0.0117	0.74%

These results indicate that a Unicam UV/Vis Spectrometer (Thermo Electron Corporation, USA) has an average error margin of 0.35 % when measuring the exact same sample three times. Taking different samples from the same flask, at the same time-point, raises the error to 0.74 % – potentially due to the different settling of cells and non-uniform distribution of cells through the culture. This is lower than the Algem photobioreactor, but here the time points are identical, while to average the Algem they were 90 minutes apart. It is expected that the true error of the Algem photobioreactor is similar to this, as it uses a similar technology.

The average error measured between the cultures is 4.15 %, virtually identical to that observed in the Algem photobioreactor.

3.3.1.1.3 Haemocytometer error

A haemocytometer can be used to count the number of *C. reinhardtii* cells in a certain volume and then that number can be used to estimate the overall concentration of cells in the culture, as explained in 2.1.1.4.

In order to investigate the error introduced when counting *C. reinhardtii* cells, the same samples as above were diluted in a ratio of 1:4 with sterile TAP media, and then iodine tincture was added to immobilize the cells.

		Count			Calculations	
		1	**2**	**3**	**St. Dev.**	**% error**
Culture	**a**	252	322	267	36.86	4.38%
	b	223	281	220	34.39	4.75%
	c	248	275	220	27.50	3.70%
				Av.	32.91	4.28%

To calculate the total cells/ml of culture, the counted number was multiplied by the dilution factor and then by 10,000 as according to the manufacturer's instructions – this does not affect the percentage error.

The average standard deviation within the three counts of the same samples is 1.65 $x10^6$ cells/ml, representing ~4.3 % error.

Meanwhile, the average error calculated between the cultures is 8.16 %, almost twice that of the Algem photobioreactor and Unicam Spectrometer.

3.3.1.1.4 Cell density quantification conclusions

These investigations indicate that the haemocytometer introduces the largest degree of error to cell density estimations, and the Unicam UV/Vis Spectrometer introduces the least. This is most likely to be due to the slightly objective nature of counting cells on a grid, to decide whether the cell is on or off the grid, and also the human error when counting large numbers. It could also be due to the uneven dispersal of cells across the haemocytometer, or a settling of cells in the flask before the sample was

taken. However, the Algem photobioreactor has a level of error at around 2 %, but as mentioned earlier, this is certainly an overestimation due to the fact that three time points across 90 minutes were used, and with a *C. reinhardtii* doubling rate of approximately 5 hours, this is a significant period of time. Additionally, the Algem photobioreactor is able to automatically measure up to four cultures simultaneously across an extended period of time.

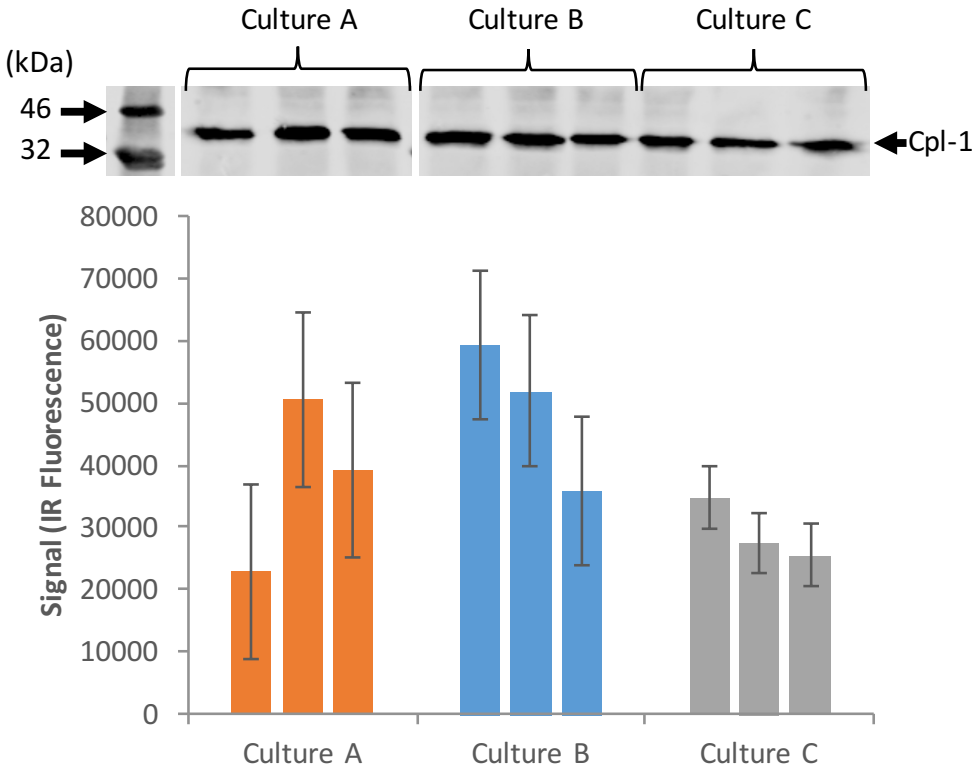
Therefore, the Unicam UV/Vis Spectrometer will be used for cell density estimations and comparisons for infrequent and one-off measurements. Otherwise, the Algem photobioreactor will be used where possible.

The manual counting method of the haemocytometer still has a value in offering a real number for cell density, but for the purposes of comparing two or more cultures, an optical density reading is superior.

3.3.1.2 Western blot quantification

3.3.1.2.1 Anti-HA primary antibody for detection and quantification

To investigate the error introduced when quantifying western blot signals, the three separate cultures of TN72_SR_*cpl1*-HA, grown in the Algem Photobioreactor, were prepared for western blotting. Equal volumes from each culture were loaded in three technical repeats onto an SDS gel and probed with anti-HA primary antibodies, followed by Goat anti-Rabbit IgG (H+L) Secondary Antibody IRDye® 800CW (referred to henceforth as IRDye®). The resulting signal was quantified using the LiCor Odyssey® CLx System.



		Culture A	Culture B	Culture C	
Replicate	1	22800	59300	34800	
	2	50700	51900	27400	
	3	39400	35800	25500	Average
% Error		37.3%	24.5%	16.8%	26.2%

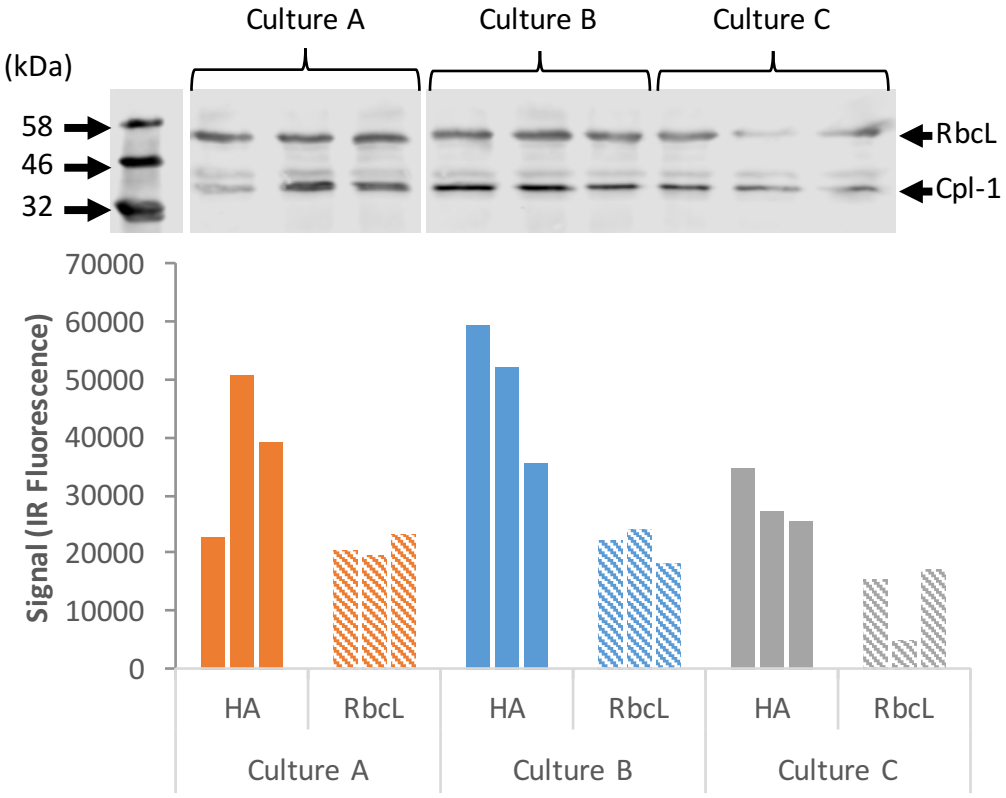
Figure 3.4 The western blot was quantified using Odyssey® Image Studio Lite v5. The bar chart shows the raw data and percentage errors. The table shows the signal from each replicate within each culture and the corresponding percentage error.

In Figure 3.4 a very high percentage error was observed between technical replicates, despite an equal volume of sample being loaded onto the SDS gel. It is possible that different quantities of protein were included in each replicate, due to pipetting and loading error.

3.3.1.2.2 Anti-RbcL primary antibody as loading control

In order to test the assertion that pipetting and loading error caused the high level of error in Figure 3.4, the endogenous protein RbcL (which is the large subunit of Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCo), the product of the *rbcL* gene), was probed with an anti-RbcL primary antibody on the same blot and

quantified in the same manner. If loading error is indeed to blame, then one would expect to see correlating variation in detection of this control.



		Culture A		Culture B		Culture C			
		HA	RbcL	HA	RbcL	HA	RbcL		
Rep.	1	22800	20600	59300	22500	34800	15600		
	2	50700	19700	51900	24000	27400	4930		
	3	39400	23000	35800	18000	25500	17100	HA Av.	RbcL Av.
% Error		37.3%	8.1%	24.5%	14.5%	16.8%	52.9%	26.2%	25.2%

Figure 3.5 The western blot was quantified using Odyssey® Image Studio Lite v5. The bar chart shows the raw data. The table shows the signal from each replicate within each culture and the corresponding percentage error.

There appears to be little correlation between the RbcL and HA detection levels, although similar levels of error are observed between replicates. This could be due to a number of factors. Possibly the blotting efficiency varied across the membrane, so proteins were unevenly blotted from the SDS gel onto the nitrocellulose membrane. It could also be due to uneven antibody decoration of the blot, suggesting a longer blotting time is necessary. Finally, the quantification technique using the

Image Studio Lite software requires the drawing of a discreet box around the band, which requires some judgement and introduces error.

3.3.1.2.3 Anti-PsbO antibody as loading control

To further confirm the result above, the same blot was additional probed with a PsbO primary antibody, which is expected to bind to another photosynthesis protein: the 33 kDa subunit of the oxygen evolving system of photosystem II. As a highly conserved protein, the aim was to use its detection as a loading control. However, as seen in Figure 3.6, no 33 kDa protein could be detected.

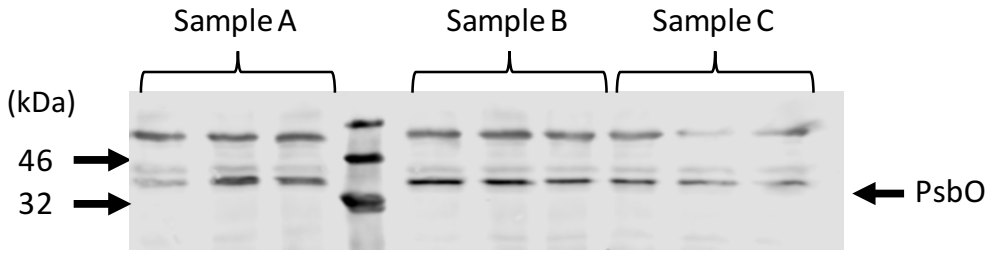


Figure 3.6 Further probing with an anti-PsbO antibody failed to provide a useful band for normalization.

Failure to detect PsbO left the question of why such high variation was observed between western blot technical repeats unanswered, but the RbcL result suggests that using a single protein as a loading control is unreliable.

3.3.1.2.4 Total Protein Stain as loading control

REVERT™ Total Protein Stain is specifically designed for western blot normalization and can be easily quantified using the LiCor Odyssey® CLx system, in a separate channel to the protein of interest.

To investigate the usability and accuracy of this system, a test was run using a sample of TN72_SR_*cpl1*-HA in various dilutions, as shown in Figure 3.7.

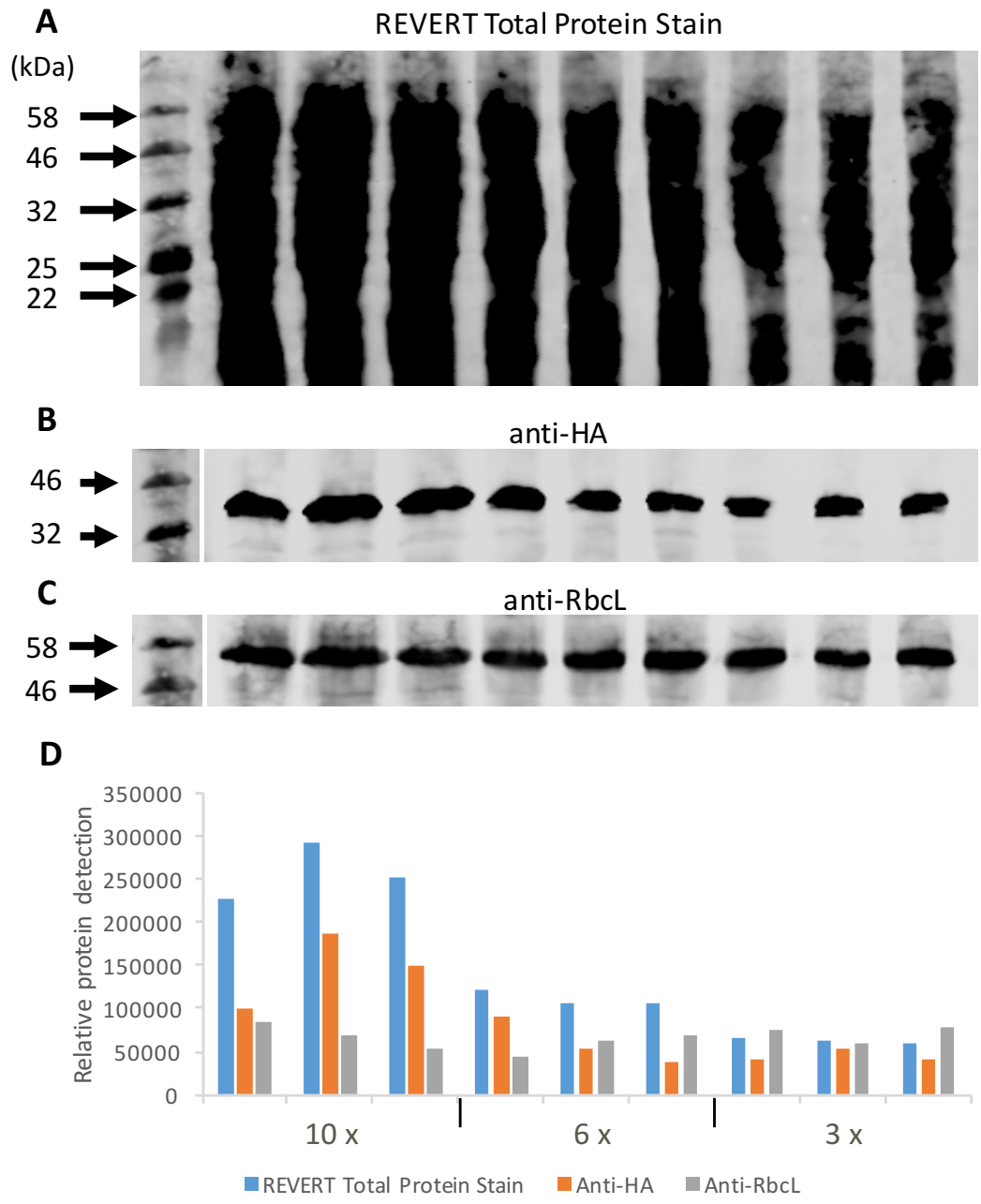


Figure 3.7 REVERT Total Protein Stain analysis. Three dilutions of TN72_SR_*cpl1*-HA extract were separated by SDS-PAGE and transferred to nitrocellulose membrane. Whole protein stained, analysed (A) washed, the probed for HA tagged proteins (B), then washed again and re-probed using anti-RbcL (C). Imaged with the LiCor Odyssey® CLx system. (D) Comparison of the signal from whole protein stain, HA detection and RbcL detection. Note: Whole protein stain is shown as 1/20th total so as to enable comparison to other bars.

Figure 3.7 shows that the REVERT® Total Protein Stain detects similar levels of total protein within the three replicates of each dilution, and appears to be well correlated with Cpl-1 detection at higher concentrations, but more variable at lower concentrations. RbcL detection on the other hand yet again displays a poor

correlation to anti-HA tagged protein detection. This is most likely due to the interaction of the RbcL protein and anti-RbcL antibody – for example, the RbcL protein may require more boiling to fully denature and enable the anti-RbcL antibody to bind, and we are therefore seeing erratic binding.

Overall, this analysis suggests that REVERT® Total Protein Stain is a good method of normalising western blot signals provided the signal is relatively strong. Therefore, a ten times (10 x) concentration of *C. reinhardtii* cultures, at a growth stage with an OD_{740} of approximately 0.750 using the Unicam UV/Vis Spectrometer, will be used for protein quantification purposes unless otherwise specified.

3.3.1.2.5 Using a protein standard for quantification

A protein standard enables observed protein accumulation levels to be quantified by comparing them to a standard of known concentration. Furthermore, by including the same standard in all western blot analyses, one can normalize between membranes, if for example they have been blotted for slightly different amounts of time. A standard encoding two or more epitopes will also enable comparison of differently tagged proteins.

Previous attempts at quantifying accumulation of HA-tagged recombinant protein in *C. reinhardtii* in the Purton lab have used a MultiTag protein (AbCam) that includes an HA epitope, or a commercially available HA-tagged protein (Calcium Regulated Heat Stable Protein 1 (CARHSP1)) as protein standards against which the HA-tagged proteins in *C. reinhardtii* cell extracts have been compared. However, the results obtained in this manner have been unreliable due to several factors including degradation of the protein standards, the unknown efficiency of blotting to nitrocellulose membrane and unknown accessibility of the HA tag to the anti-HA antibodies. Furthermore, estimated levels of a protein obtained in this way, when expressed as 'percentage of total soluble protein in cell extracts' were not supported by Coomassie Brilliant Blue stained SDS-PAGE analysis of the extracts (Stoffels 2015) and appeared to overestimate the quantity of accumulated protein in *C. reinhardtii* extracts.

3.3.1.2.6 Producing a bespoke Cpl-1 protein standard

In order to minimise the problems stated above with commercially available protein standards, a bespoke Cpl-1 protein standard was produced in *E. coli*.

A slightly modified version of Cpl-1 was used (cpl1_dimer), for compatibility with other projects, see 4.3.2, and both HA and StrepII epitope tags were included on the C-terminus. The gene was codon optimized for *E. coli* and synthesized by Integrated DNA Technologies. NcoI and NotI restriction sites were used to clone the gene in-frame into the IPTG inducible expression vector pProEX-HTb. The resulting plasmid was sequenced to confirm the correct insertion and used to transform DH5α competent *E. coli* cells.

The resulting *E. coli* strain was named DH5α_ProEX_*cpl1_dimer*-HA-StrepII, and its recombinant protein product, purified using StrepTactin affinity to the StrepII epitope tag and quantified using a Bradford assay, would, therefore, produce a more reliable protein standard for quantifying Cpl-1 accumulation in *C. reinhardtii* than the Multi-tag or CARHSP1 standards.

IPTG induction was performed (2.6.1) and *cpl1_dimer*-HA-StrepII accumulation was confirmed by SDS-PAGE and Coomassie Brilliant Blue R staining (Figure 3.8).

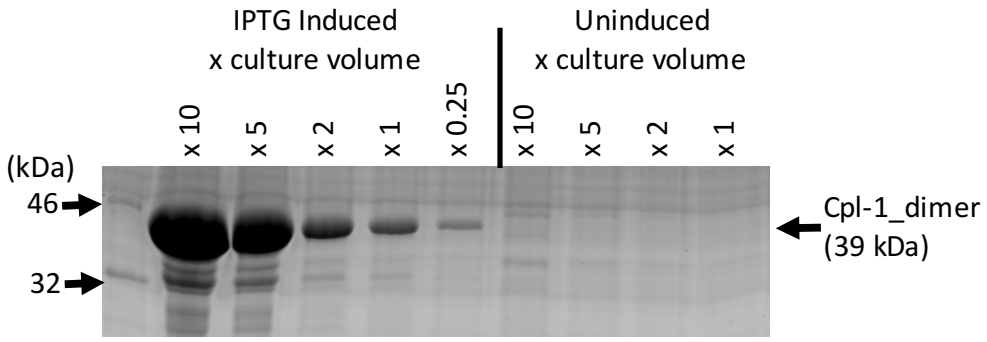


Figure 3.8 Whole cell extracts were separated by SDS-PAGE and total protein staining was performed using a Coomassie Brilliant Blue R stain. Note: despite the name Cpl1_dimer, this is a standard 39 kDa Cpl-1 monomer (the name "dimer" refers to its ability to dimerize in non-denaturing conditions, as discussed in 4.3.2).

The IPTG-induced DH5α_ProEX_*cpl1_dimer*-HA-StrepII cells were lysed using BugBuster™ (2.6.2). The StrepII epitope tag on the Cpl1_dimer-HA-StrepII protein

was then used for protein enrichment via a commercially available StrepTactin column (Thermo Fisher Scientific). The purification fractions are shown in Figure 3.9.

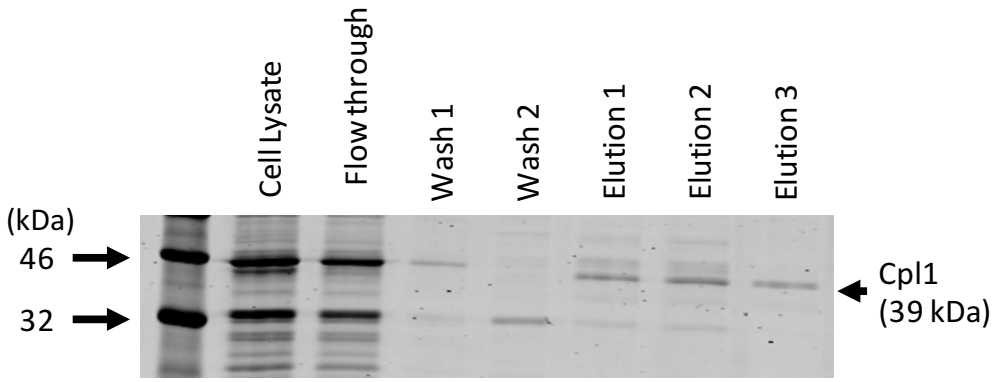


Figure 3.9 The various fractions involved using the StrepTactin purification column were separated by SDS-PAGE and stained with Coomassie Brilliant Blue R stain.

Coomassie Brilliant Blue staining in Figure 3.9 indicates that other proteins aside from Cpl-1 are still present in the eluted fractions, so the Cpl1-enriched elution was then further purified and concentrated by ammonium sulphate precipitation – the fractions of which are shown in Figure 3.10.

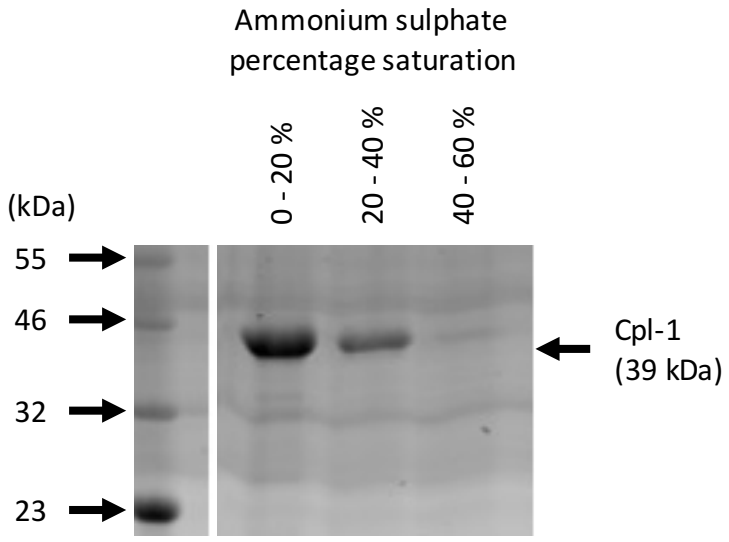


Figure 3.10 Coomassie Brilliant Blue R stained SDS-PAGE. The 20 – 40 % ammonium sulphate fraction was chosen as the fraction with the most Cpl-1 compared to other visible proteins.

Three dilutions of the 20 - 40 % ammonium sulphate fraction were then quantified using a Bradford protein assay. Absorbance at 595 nm (A_{595}) was measured using a Nanodrop 2000, enabling very small volumes to be used. These measurements were compared to a standard curve measured in the same way, shown in Figure 3.11. The calculations using the standard curve to measure the protein concentration are shown in Figure 3.12.

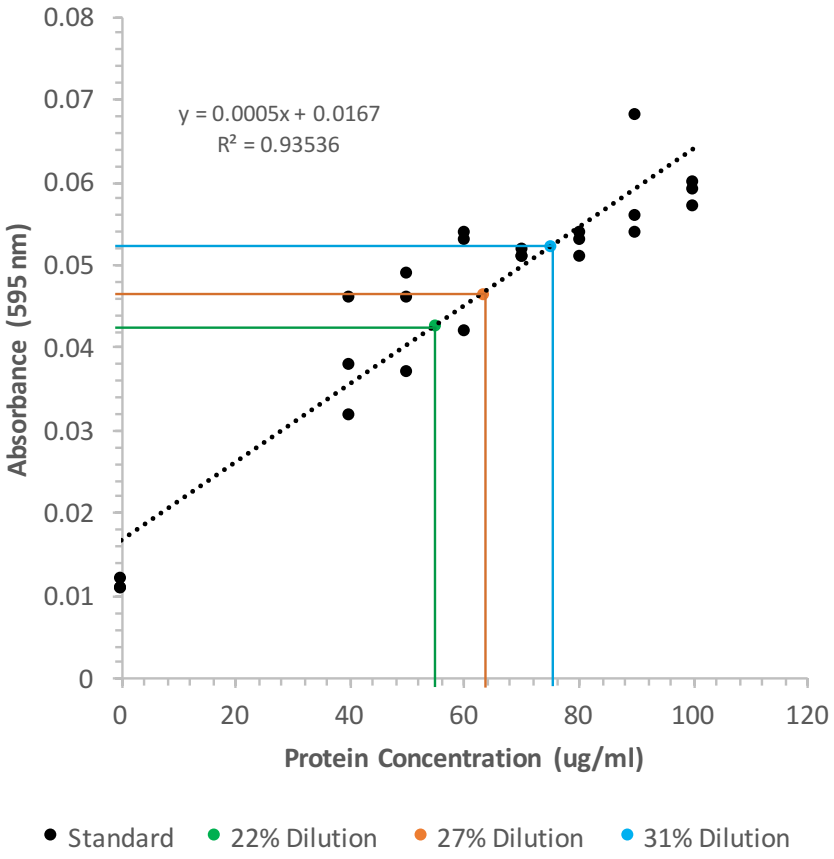


Figure 3.11 A Bradford assay was performed, and the three dilutions compared to the protein standards. Each coloured point shown on this graph is the average of 4 measurements.

	A595	Average A595	St. Dev. A595	(y = 0.0005x + 0.0167) Average Protein	St Dev. Protein	Stock Conc. Average	Stock Conc. St. Dev.
Dilution 1 (22.3 %)	0.050						
	0.038						
	0.043						
	0.039	0.0425	0.128	51.6	6.61	231.9	29.7
Dilution 2 (26.7 %)	0.051						
	0.053						
	0.037						
	0.045	0.0465	0.155	59.6	9.21	223.2	34.5
Dilution 3 (31.2%)	0.057						
	0.053						
	0.046						
	0.053	0.0523	0.088	71.1	6.22	228.3	20.0
					Average:	**227.8**	**28.1**

Figure 3.12 Bradford Assay calculations. The equation was derived from the line of best fit from the graph in Figure 3.11.

It was inferred from Figure 3.12 that the stock solution of purified cpl1_dimer-HA-StrepII had a concentration of 230 ± 30 µg/ml.

The synthesis of this protein standard will enable comparison of different western blot analyses as well as compare the accumulation levels of proteins tagged with different epitopes, in this case HA and StrepII.

3.3.2 Investigating control by epistasy of synthesis in protein accumulation

The low yields of recombinant protein accumulation in the *C. reinhardtii* chloroplast are a major factor impeding it as a production platform. Here we attempt to improve protein accumulation levels of Cpl-1 by expressing the *cpl1* gene under the control of two different promoter/5'UTRs.

3.3.2.1 Existing *C. reinhardtii* lines

This section uses two *C. reinhardtii* lines created by Taunt (2013) as controls. One is a negative control that will be used throughout this thesis, named TN72_SR_Control, as shown in Figure 3.13 this is the recipient strain of *C. reinhardtii* transformed with an empty pSRSapI plasmid. The second *C. reinhardtii* line is TN72_SR_*cpl1*-HA which will be used as the "benchmark" for this chapter - i.e. a line displaying a moderate level of Cpl-1 accumulation upon which we aim to improve.

3.3.2.1.1 *C. reinhardtii* line TN72_ SR_Control

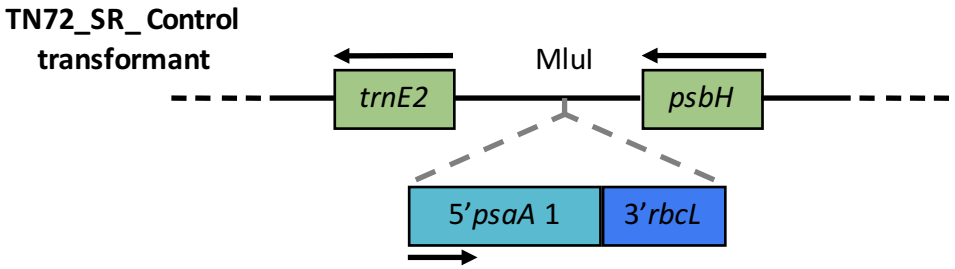


Figure 3.13 TN72 transformed with an empty pSRSapI plasmid. Unlike the original TN72 recipient, the control is wildtype for photosynthesis, like the *cpl1* transformants, as the *psbH* deletion has been rescued.

3.3.2.1.2 *C. reinhardtii* line TN72_ SR_*cpl1*-HA

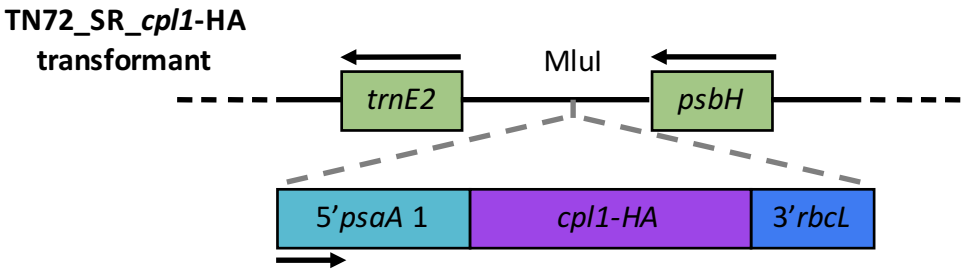


Figure 3.14 TN72 transformed with the pSRSapI expression plasmid containing HA-tagged *cpl1*.

3.3.2.1.3 Western blot confirmation

To show that Cpl-1 accumulation occurs in TN72_SR_*cpl1*-HA and that no protein of a similar size is detected in TN72_SR_Control, whole cell extracts of each were prepared and separated by SDS-PAGE. A blot of the gel was then probed with an anti-HA primary antibody and an anti-RbcL primary antibody, followed by IRDye® secondary antibodies. The LiCor Odyssey® CLx System was then used for detection. The 55 kDa RbcL band acts as an approximate loading control in this instance, to show that cell extract was present in the negative control lane. The 39 kDa Cpl-1 protein can only be seen in the TN72_SR_*cpl1*-HA transformed line.

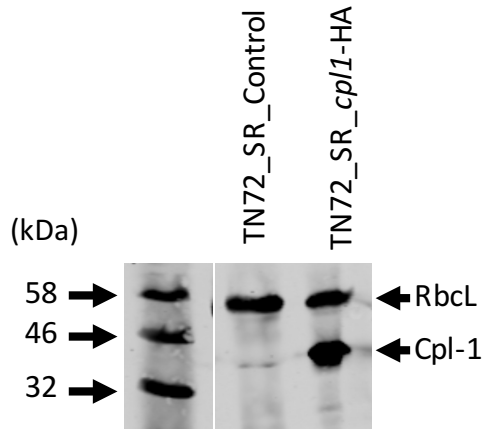


Figure 3.15 Western blot analysis of whole cell extracts. A Color Prestained Protein Standard, Broad Range (New England Biolabs, Inc.) was used to estimate protein size.

3.3.2.2 The production of *C. reinhardtii* lines with improved Cpl-1 accumulation

We have hypothesised that CES limits the level of recombinant protein accumulation, due to the use of endogenous 5'UTRs for recombinant protein expression in the *C. reinhardtii* chloroplast. By introducing the gene of interest into the chloroplast inside two different expression cassettes, we speculate that the level of protein accumulation per cell can be improved. In this case, we will use the endogenous *psaA* exon 1 5'UTR and the *atpA* 5'UTR, which have both been shown to achieve moderate levels of recombinant protein accumulation in the *C. reinhardtii* chloroplast (Taunt 2013).

However, given that we wish to transform *C. reinhardtii* using the convenient TN72 system for targeted insertion into the chloroplast, we have only one plastome insertion site; namely, the neutral site between *psbH* and *trnE2*. The two expression cassettes, containing the same gene of interest, will, therefore, need to be adjacent to one another in the transformation plasmid, and be integrated by a single transformation event into the plastome by homologous recombination.

The gene of interest in this instance is *cpl1* which is ~1 kb in length. Homologous recombination between DNA direct repeats in the *C. reinhardtii* chloroplast is efficient – indeed, it is this propensity to recombine which we exploit for the targeted insertion of transgenes when performing *C. reinhardtii* chloroplast transformations.

However, two identical copies of *cpl1* in close proximity to one another are considered to be at very high risk of undergoing recombination between the two copies, resulting in a plastome rearrangement and loss of one copy of the Gene of Interest (GOI) (Figure 3.16). Mudd et al. (2014) suggest that direct repeats of ~600 bp or more result in very high frequencies of homologous recombination. Additionally, such a recombination event would still produce recombinant protein, so may go un-noticed if it occurs in a transformant line subsequent to molecular analysis confirming the two-GOI insertion.

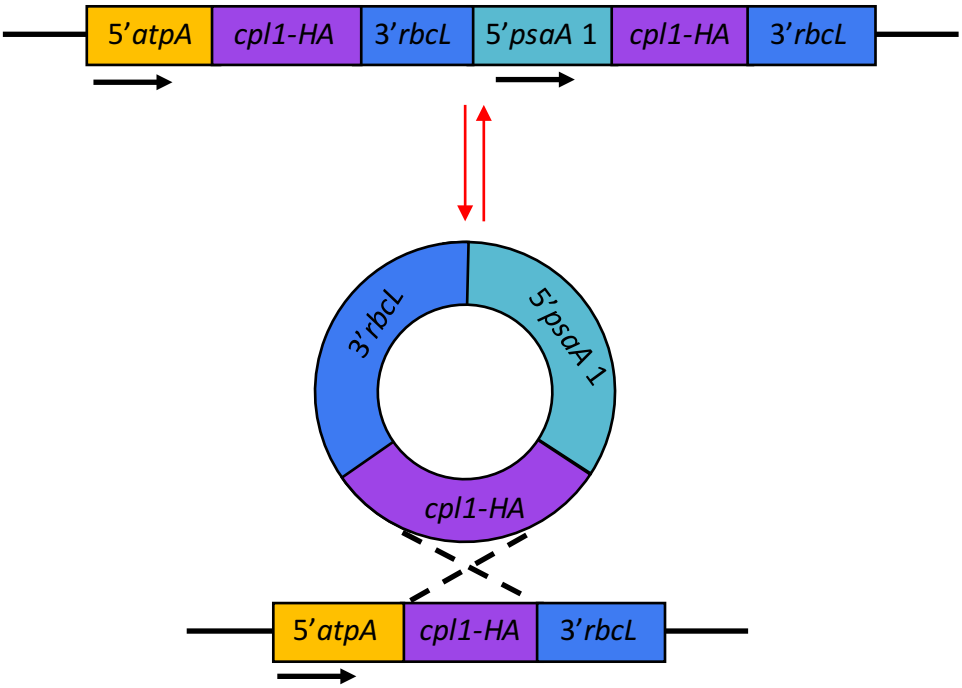


Figure 3.16 Two expression cassettes in parallel orientation might homologously recombine to leave a viable expression cassette in the chloroplast plastome, thus still produce Cpl-1 protein.

However, as illustrated in Figure 3.17, by arranging the expression cassettes in antiparallel directions, recombination events will merely result in a switching of the promoter regions and no loss of DNA will occur.

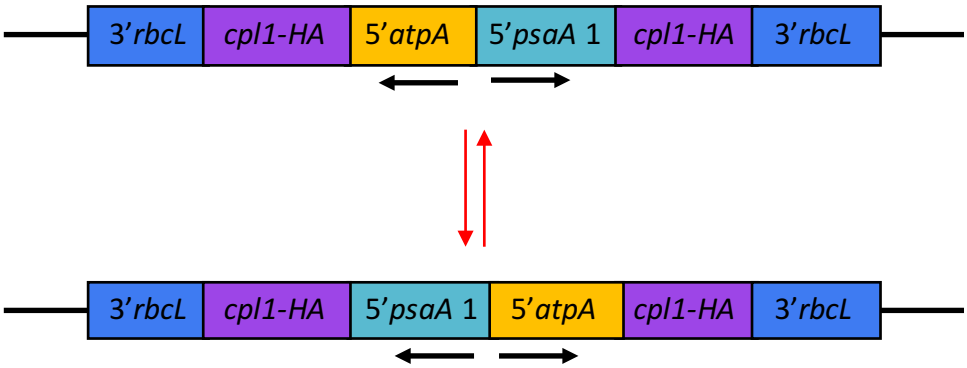


Figure 3.17 When arranged in antiparallel directions, the recombination event would not result in any loss of DNA from the transformed plastome.

Therefore, the antiparallel configuration was used when transforming *C. reinhardtii* with two expression cassettes in a single plasmid, in particular when the two genes are highly similar. Furthermore, Coragliotti et al. (2011) report no detectable difference in expression as a result of transgene orientation. In fact, the *C. reinhardtii* chloroplast contains two large inverted repeat regions which allow similar "flip-flop" recombination of the entire plastome, and it is believed that the two possible orientations exist in approximately equal proportion, with no apparent adverse effect (Harris 2009).

3.3.2.2.1 *C. reinhardtii* line TN72_ A_*cpl1_c*-HA

However, to reduce the chance of DNA recombination events between the directly repeated *cpl1* genes, one gene was differently codon optimized to reduce the sequence similarity. The different version of *cpl1* was named *cpl1_c*, with the "c" standing for codon changed.

3.3.2.2.1.1 *Cpl-1_c* gene design

The original genetic sequence for *cpl1* as used by Taunt (2013), and present in TN72_SR_*cpl1*-HA, was uploaded to *Codon Usage Optimizer Beta 0.92* (Kong 2013) software and codons were manually edited to ones of similar "scores" for the *C. reinhardtii* chloroplast. In total 148 nucleotides were changed out of 1050 (14 %) without affecting the overall score for codon usage. SapI and SphI restriction sites were included for cloning.

3.3.2.2.1.2 Expression cassette production and *E. coli* transformation

The full-length *cpl1_c* gene was synthesized by Integrated DNA Technologies. SapI and SphI restriction enzymes were used to clone the gene into the pASapI expression vector in-frame, producing the plasmid pASapI_*cpl1_c*-HA. DH5α competent *E. coli* cells were transformed with the resulting plasmid and a plasmid midi-prep was performed to produce suitable quantities of plasmid for glass bead transformation.

3.3.2.2.1.3 *C. reinhardtii* transformation

The pASapI_*cpl1_c*-HA plasmid was used to transform recipient *C. reinhardtii* strain TN72 via the glass bead method as described in 2.3.13. Figure 3.18 illustrates the resulting *C. reinhardtii* cell line, TN72_A_*cpl1_c*-HA

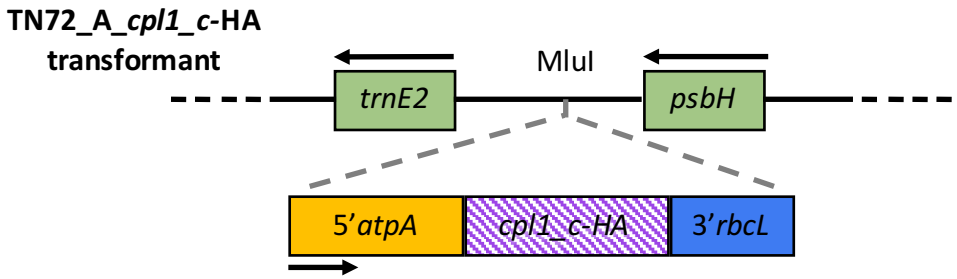


Figure 3.18 TN72 recipient strain successfully transformed to carry *cpl1_c* under the *atpA* 5' UTR in the chloroplast, and with *psbH* functionality restored.

The plastomes of the resulting transformants were confirmed to be homoplasmic as described in 2.3.17. In this instance, the three primers used were Flank1, atpa.R and psbH.R. Figure 3.19 shows the resulting PCR products separated by gel electrophoresis. TN72 was used as a negative control. As a final step, the sequence integrity of the cassette was confirmed by DNA sequencing of the PCR-amplified plastome region, as in 2.3.18.

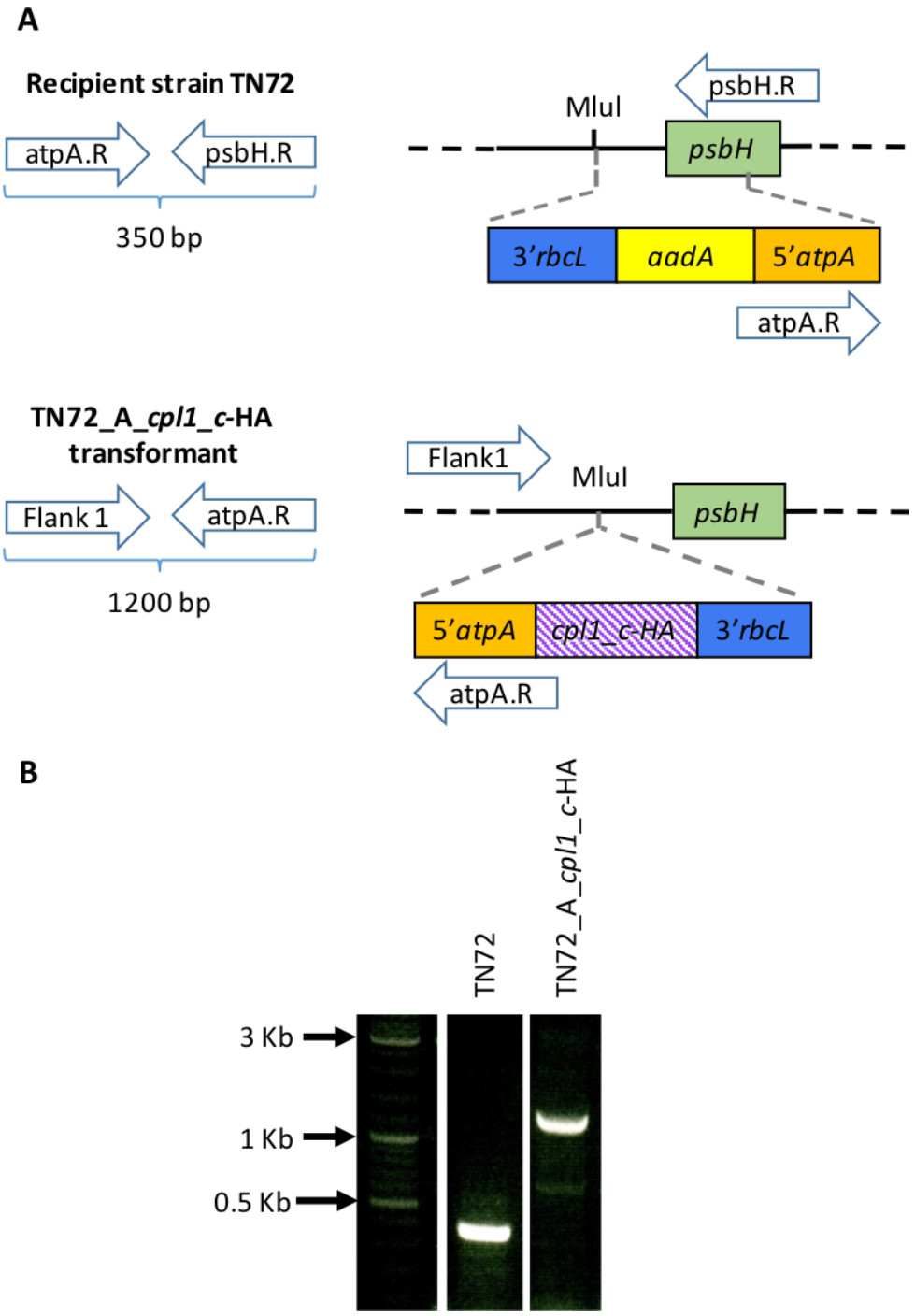


Figure 3.19 PCR screening for successful transformation of TN72 and reaching a homoplasmic state. **(A)** Schematic showing the product expected with each primer combination. **(B)** Gel electrophoresis of the PCR products confirms that the transformant is homoplasmic as it lacks the 350 bp band of the recipient strain.

3.3.2.2.1.4 Western blot quantification

A whole cell extract of TN72_A_*cpl1*_*c*-HA was prepared and separated by SDS-PAGE. After blotting to nitrocellulose membrane as described in 2.4.5, the membrane was

probed with anti-HA and anti-RbcL primary antibodies, followed by IRDye® secondary antibodies and fluorescence was observed using the LiCor Odyssey® CLx system.

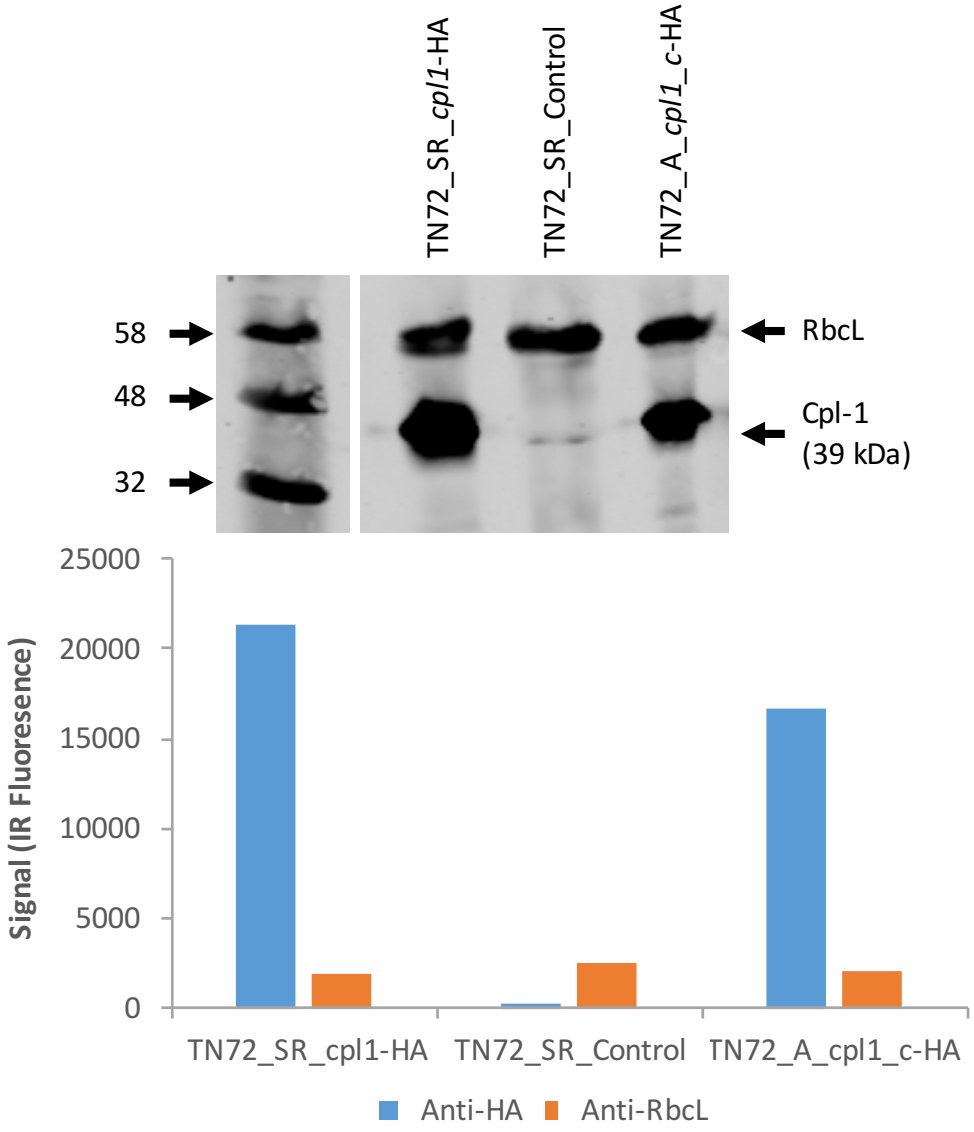


Figure 3.20 Western blot analysis shows that similar levels of Cpl-1 protein accumulation are detected in the two *C. reinhardtii* lines. RbcL is an approximate loading control to show that similar quantities of cell lysate were loaded onto the gel and a negative control, TN72_SR_Control is also included.

The results show similar levels of Cpl-1 protein accumulation, and the RbcL loading control confirms that similar quantities of cell lysate were used.

3.3.2.2.2 *C. reinhardtii* line TN72_SR_*cpl1*-HA_A_*cpl1_c*-HA

To see if higher levels of protein accumulation could indeed be achieved with two different *cpl-1* expression cassettes inserted into the *C. reinhardtii* chloroplast, the SR_*cpl1*-HA and the A_*cpl1_c*-HA cassettes were combined into a single plasmid and used to transform *C. reinhardtii* recipient strain TN72.

3.3.2.2.2.1 Expression plasmid production

The final assembly of pSRSapI_*cpl1*-HA_A_*cpl1_c*-HA was achieved as described in Figure 3.21. The existing MluI site in pSRSapI_*cpl1*-HA was used to open the vector, and the insert was created by site-directed mutagenesis PCR to introduce a BssHII restriction site at the 3' end of the A_*cpl1-c*-HA cassette. As MluI and BssHII form compatible restriction sites, the *cpl1-c* containing expression cassette was ligated into the open vector, producing one conserved MluI site. The resulting plasmid was called pSRSapI_*cpl1*-HA_A_*cpl1_c*-HA. It is believed that the alternative arrangement, i.e. opening the pASapI_*cpl1-c*-HA plasmid and inserting the SR_*cpl1*-HA cassette, would have also been possible, but when both were attempted the cloning succeeded with the former first and therefore this route was continued.

The primers used for the PCR reaction are indicated below. The BssHII site is underlined, and the mutated nucleotide is shown in lowercase.

Cpl1.F: CAGGCAATTTGCTTACACCTTT

Cpl1_BssHII.R: TTAAGATAAgCGCGCTACATCC

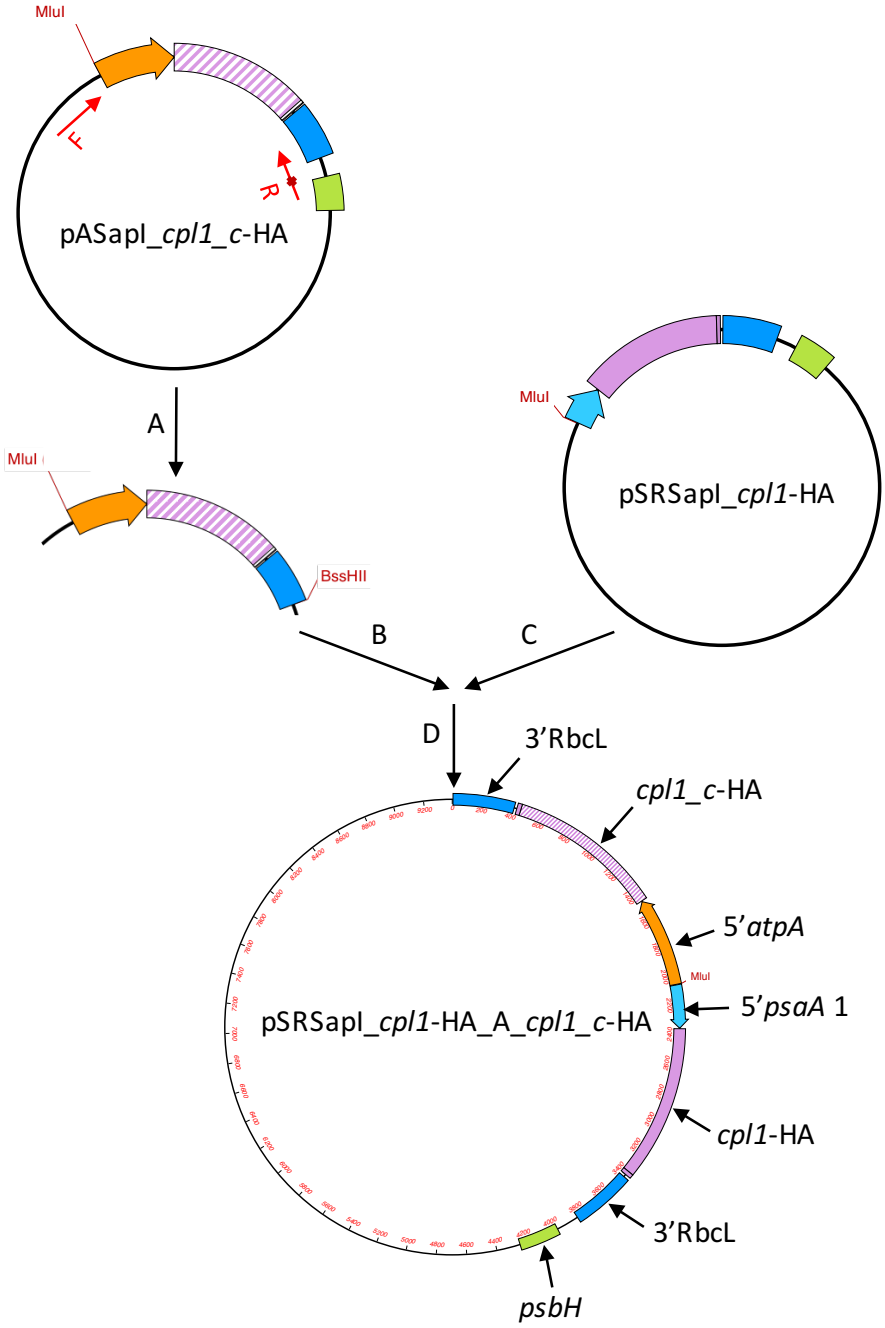


Figure 3.21 Schematic showing the stages in the construction of pSRSapI_*cpl1*-HA_*cpl1_c*-HA. **(A)** Site-directed mutagenesis PCR was carried out upon pASapI_cpl1_c-HA to amplify the expression cassette and to introduce a point mutation to create a BssHII restriction site. The red arrows represent the primers used for this reaction. **(B)** The PCR product resulting from (A) was digested with MluI and BssHII before being cleaned up using a PCR purification kit. **(C)** pSRSapI_*cpl1*-HA was digested using MluI before being treated with Antarctic Phosphatase to prevent re-circularization of the plasmid. **(D)** The resulting insert (A_*cpl1_c*-HA) and vector (pSRSapI_*cpl1*-HA) were then ligated using T4 DNA ligase.

DH5α *E. coli* were transformed with the resulting plasmid. Sequencing confirmed the successful insertion and orientation of the insert.

3.3.2.2.2.2 *C. reinhardtii* transformation

The pSRSapI_*cpl1*-HA_A_*cpl1*_*c*-HA plasmid was used to transform the recipient *C. reinhardtii* strain TN72 via the glass bead method as described 2.3.13. Figure 3.22 illustrates the resulting *C. reinhardtii* cell line, TN72_SR_*cpl1*-HA_A_*cpl1*_*c*-HA.

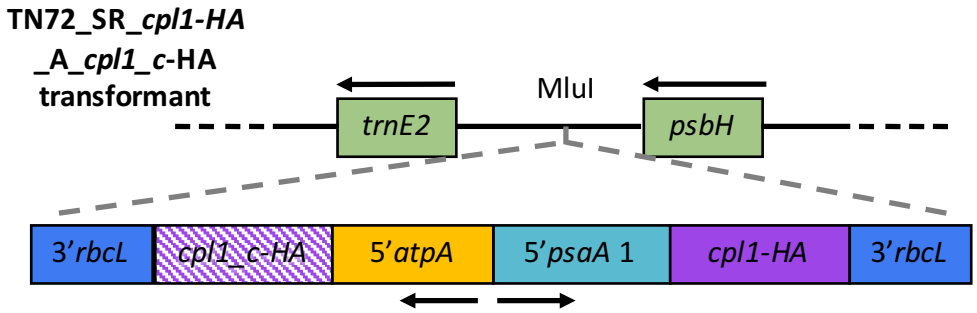


Figure 3.22 TN72 recipient strain transformed with the pSRSapI_*cpl1*-HA_A_*cpl1*_*c*-HA plasmid, thus rescuing *psbH* activity. Carrying *cpl1*-HA under the *psaA* 5' UTR, and *cpl1*_*c*-HA under the *atpA* 5' UTR in the chloroplast.

The resulting transformants were confirmed to be homoplasmic as described in 2.3.17 and represented in Figure 3.23. Finally, correct insertion into the plastome was confirmed by PCR amplification from transformant genomic DNA and DNA sequencing, as in 2.3.18.

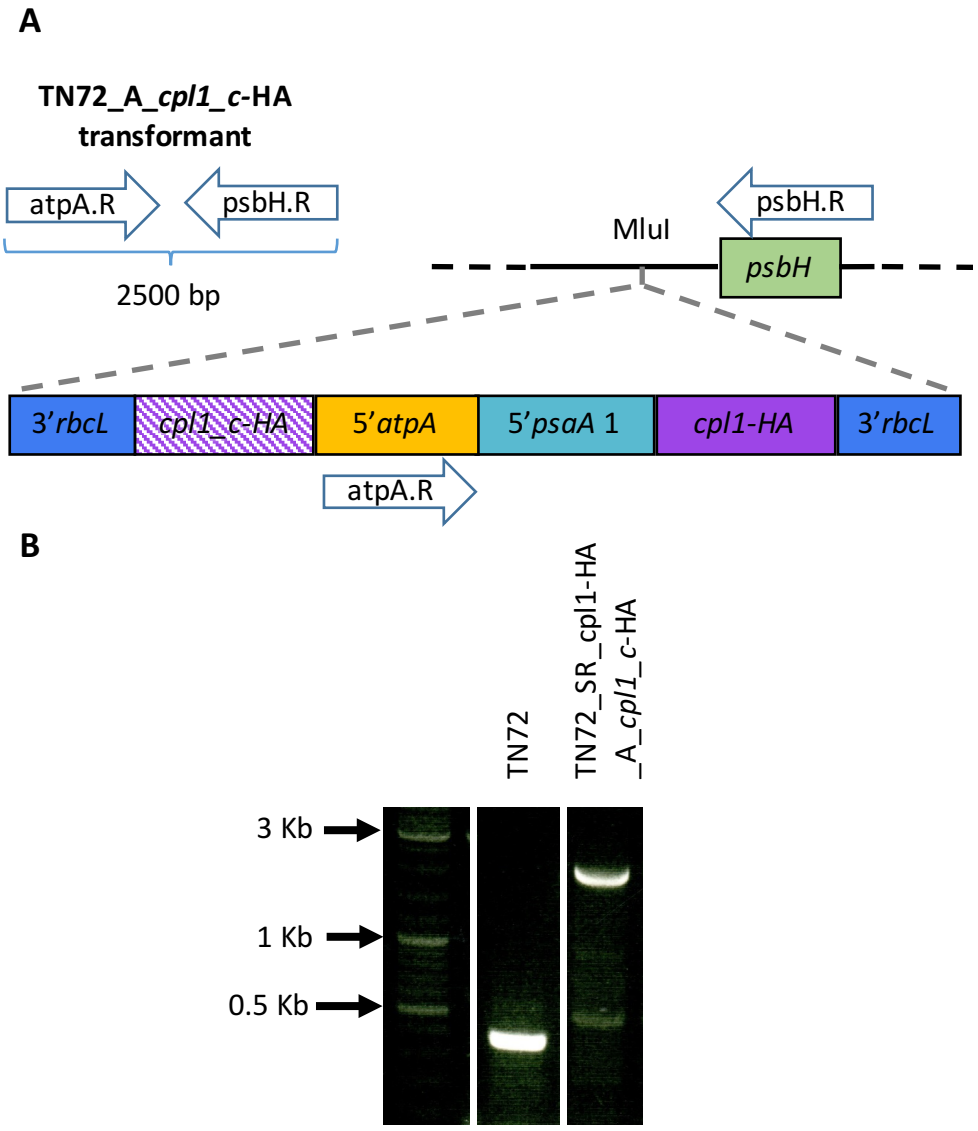


Figure 3.23 **(A)** Schematic describing the expected band when testing homoplasmic transformants. The combination which produces the negative band can be seen in Figure 3.19. It is the absence of the negative, 350 bp, band that indicates homoplasmicity i.e. no copies of the untransformed plastid remain **(B)** PCR products separated by gel electrophoresis show the correct expected lengths and thus show that the new transformant is now homoplasmic.

3.3.2.2.2.3 Western blot quantification

In order to compare the relative Cpl-1 accumulation levels between TN72_SR_*cpl1*-HA and TN72_SR_*cpl1*-HA_A_*cpl1*_*c*-HA, and investigate whether two cassettes with the same gene does indeed improve overall protein accumulation levels in the *C. reinhardtii* chloroplast, western blot analysis was performed. Whole cell extracts of TN72_SR_*cpl1*-HA and TN72_SR_*cpl1*-HA_A_*cpl1*_*c*-HA were prepared, and five technical replicates of each were separated by SDS-PAGE. After blotting to

nitrocellulose membrane as described in 2.4.5, the membrane was stained with REVERT™ Whole Protein Stain which confirmed similar quantities of cell preparation were loaded onto the gel (Appendix A) and was used to normalize small loading differences. After removal of the REVERT stain and blocking, the membrane was probed with the anti-HA primary antibody, followed by IRDye® secondary antibody and fluorescence was observed using the LiCor Odyssey® CLx system. The western blot image and analysis are shown in Figure 3.24.

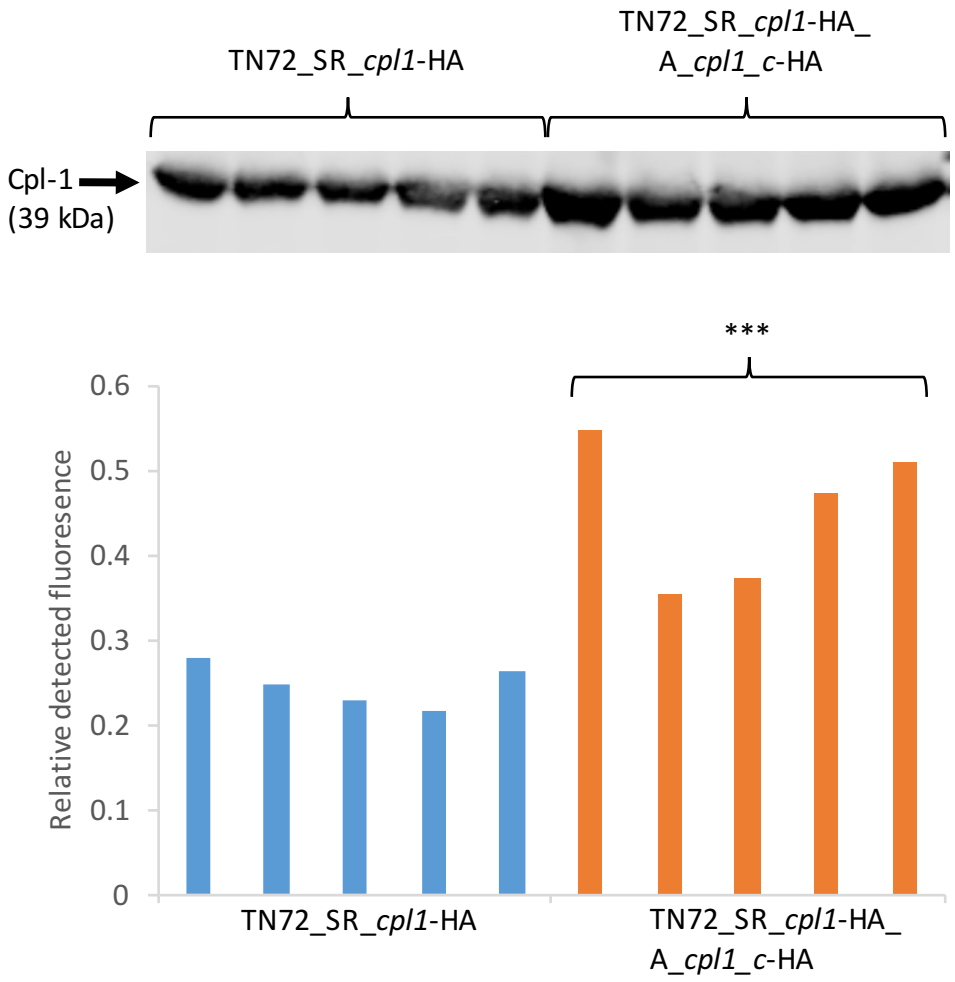


Figure 3.24 Western blot analysis enabled the quantification of Cpl-1 accumulation in the two *C. reinhardtii* lines. Cpl-1 detection was normalised using the signal from REVERT Total Protein Stain.

A two sample t-test was carried out to compare the levels of protein accumulation in TN72_SR_*cpl1*-HA and TN72_SR_*cpl1*-HA_A_*cpl1*_*c*-HA, in order to identify any statistical significance. This is a comparison of the mean protein accumulation levels,

and it is a two-tailed test. The null hypotheses (H_0) is that the means are the same. The alternative hypothesis (H_1) is that the means are different.

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

The full results can be seen in Figure 3.25, but in brief, there was a significant difference in accumulation for TN72_SR_*cpl1*-HA (M = 0.247, Var = 0.001) and TN72_SR_*cpl1*-HA_A_*cpl1_c*-HA (M= 0.451, Var = 0.007); t(8) = -5.14, p < 0.001.

t-Test: Two-Sample Assuming Equal Variances		
	TN72_A_*cpl1*-HA	**TN72_SR_*cpl1*-HA_A_*cpl1_c*-HA**
Mean	0.2472	0.4512
Variance	0.0006	0.0073
Observations	5	5
Pooled Variance	0.0039	
Hypothesized Mean Difference	0	
df	8	
t Stat	-5.1423	
P(T<=t) one-tail	0.0004	
t Critical one-tail	1.8595	
P(T<=t) two-tail	0.0009	
t Critical two-tail	2.3060	

Figure 3.25 Statistical report generated using Microsoft Excel to compare the levels of protein accumulation in the two lines of *C. reinhardtii*.

This result supports our original hypothesis - that protein accumulation can be improved in the *C. reinhardtii* chloroplast by introducing several copies of the GOI as different expression cassettes. It suggests that CES is limiting the level of translation for each endogenous 5' UTR and that by having the same gene under the control of two different 5' UTRs, this can be improved.

3.3.2.2.3 *C. reinhardtii* line TN72_ SR_*cpl1*-StrepII

In the previous experiment, it is not possible to determine how the protein synthesis load is being shared by the two expression cassettes, as the proteins are both tagged with an HA epitope and are therefore indistinguishable on a western blot.

We therefore produced a similar *C. reinhardtii* cell line but with one *cpl1* encoding a StrepII epitope tag instead of an HA tag. This enables the Cpl-1 protein from the two different promoters/5'UTRs to be distinguished on a western blot and the contribution of the two cassettes compared.

The first stage in this process was to produce the *C. reinhardtii* cell line TN72_SR_*cpl1*-StrepII, to be used as a control but also to ensure that this tag can be detected on western blots and that the protein accumulates to detectable levels.

3.3.2.2.3.1 Expression cassette production

The full-length *cpl1*-StrepII gene was synthesized by Integrated DNA Technologies. SapI and SphI restriction enzymes were used to clone the gene in-frame into the pSRSapI expression vector, producing the plasmid pSRSapI_*cpl1*-StrepII. DH5α competent *E. coli* cells were transformed with the resulting plasmid.

3.3.2.2.3.2 *C. reinhardtii* transformation

The pSRSapI_*cpl1*-StrepII plasmid was used to transform the recipient *C. reinhardtii* strain TN72 via the glass bead method as described in 2.3.13. Figure 3.26 illustrates the resulting *C. reinhardtii* cell line, TN72_SR_*cpl1*-StrepII.

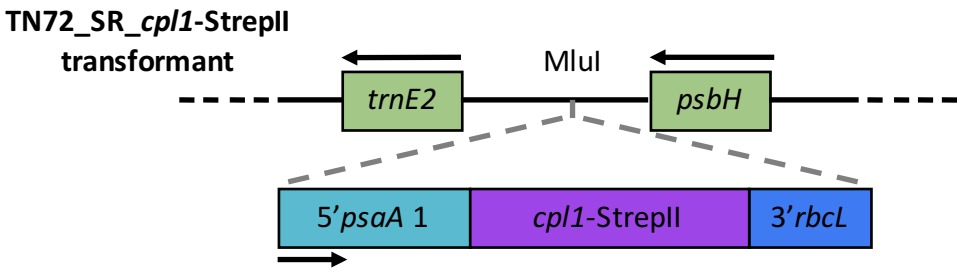


Figure 3.26 TN72 recipient strain successfully transformed to carry *cpl1*-StrepII under the *psaA* 5'UTR in the chloroplast, and with *psbH* functionality restored.

The resulting transformants were confirmed for homoplasmicity as described in 2.3.17 (Appendix B). Finally, the sequence integrity of the inserted cassette was confirmed by DNA sequencing as in 2.3.18.

3.3.2.2.3.3 Western blot detection

Whole cell preparations of TN72_SR_*cpl1*-HA (as a control) and TN72_SR_*cpl1*-StrepII were prepared and separated, in duplicate, by SDS-PAGE. After blotting to nitrocellulose membrane and blocking as described in 2.4.5, one membrane was probed with an anti-HA primary antibody, while the second membrane was probed with anti-StrepII primary antibody. Both membranes were then incubated by IRDye® secondary antibody and fluorescence was observed using the LiCor Odyssey® CLx system. The resulting images are shown in Figure 3.27.

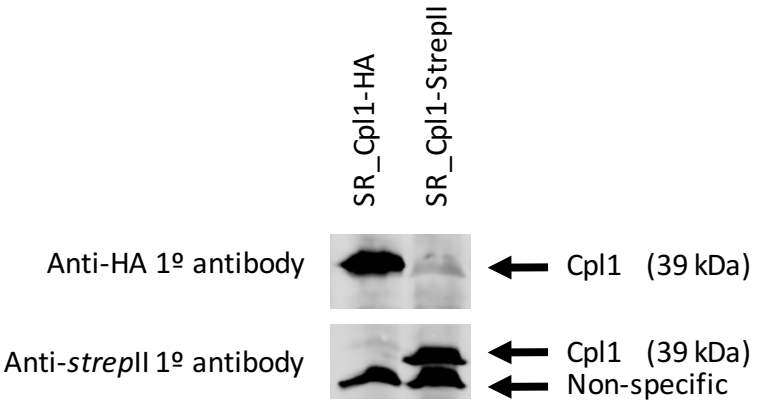


Figure 3.27 Two duplicate membranes were probed with anti-HA or anti-StrepII primary antibodies and Cpl-1 accumulation was detected in the expected cell lines i.e. Cpl-1-HA was detected with the anti-HA antibody and not with the anti-StrepII antibody and *vice versa*. The non-specific band seen in the lower image confirms that a similar quantity of extract was loaded on to the gel for both strains.

3.3.2.2.4 *C. reinhardtii* line TN72_SR_*cpl1*-StrepII_A_*cpl1*_*c*-HA

3.3.2.2.4.1 Transformation plasmid production

The plasmid was produced in the same manner as in 3.3.2.2.2.1, but using pSRSapI_*cpl1*-StrepII instead of pSRSapI_*cpl1*-HA. The resulting plasmid was named pSRSapI_SR_*cpl1*-StrepII_A_*cpl1*_*c*-HA.

3.3.2.2.4.2 *C. reinhardtii* transformation

The pSRSapI_SR_*cpl1*-StrepII_A_*cpl1*_*c*-HA plasmid was used to transform the recipient *C. reinhardtii* strain TN72 via the glass bead method as described in 2.3.13. Figure 3.28 illustrates the resulting *C. reinhardtii* cell line, TN72_SR_*cpl1*-StrepII.

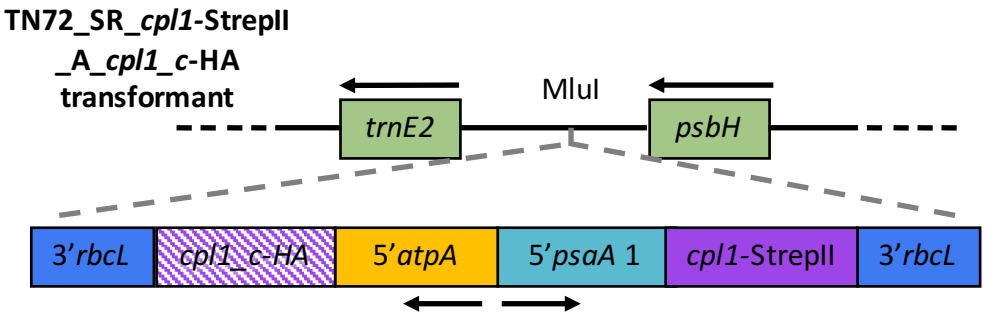


Figure 3.28 TN72 recipient strain successfully transformed to carry *cpl1*-StrepII under the *psaA* 5' UTR, and *cpl1*_*c*-HA under the *atpA* 5' UTR in the chloroplast, and with *psbH* functionality restored.

Correct insertion of the two GOIs into the plastome and homoplasmy of the transformant lines were confirmed as described above.

3.3.2.2.4.3 Western blot quantification

In order to compare accumulation of proteins tagged with different epitope tags, and therefore detected via different antibodies, a protein standard tagged with both HA and StrepII epitope tags was used to normalize the signal. This standard was produced specifically for this purpose, as described in 3.3.1.2.6.

C. reinhardtii whole cell extracts from the various cell lines constructed in this chapter were separated by SDS-PAGE in duplicate. The proteins were blotted to nitrocellulose membranes as described in 2.4.5 and REVERT™ Total Protein stain was used to normalise loading error (Appendix C). One membrane was probed with anti-HA primary antibody, while the second membrane was probed with anti-StrepII primary antibody. Both membranes were then incubated with IRDye® secondary antibody and fluorescence was observed using the LiCor Odyssey® CLx system. The resulting images are shown in Figure 3.29.

The protein standard, which is tagged with both HA and StrepII epitope tags was then used to normalize the two membranes, as was the REVERT™ total protein stain. The relative protein accumulation levels of each *C. reinhardtii* strain were compared in Figure 3.29 (B). TN72_SR_*cpl1*-HA, which has been used throughout this chapter as a benchmark for Cpl-1 protein accumulation level, upon which to improve, was set to 100% and other levels are represented as percentages of this.

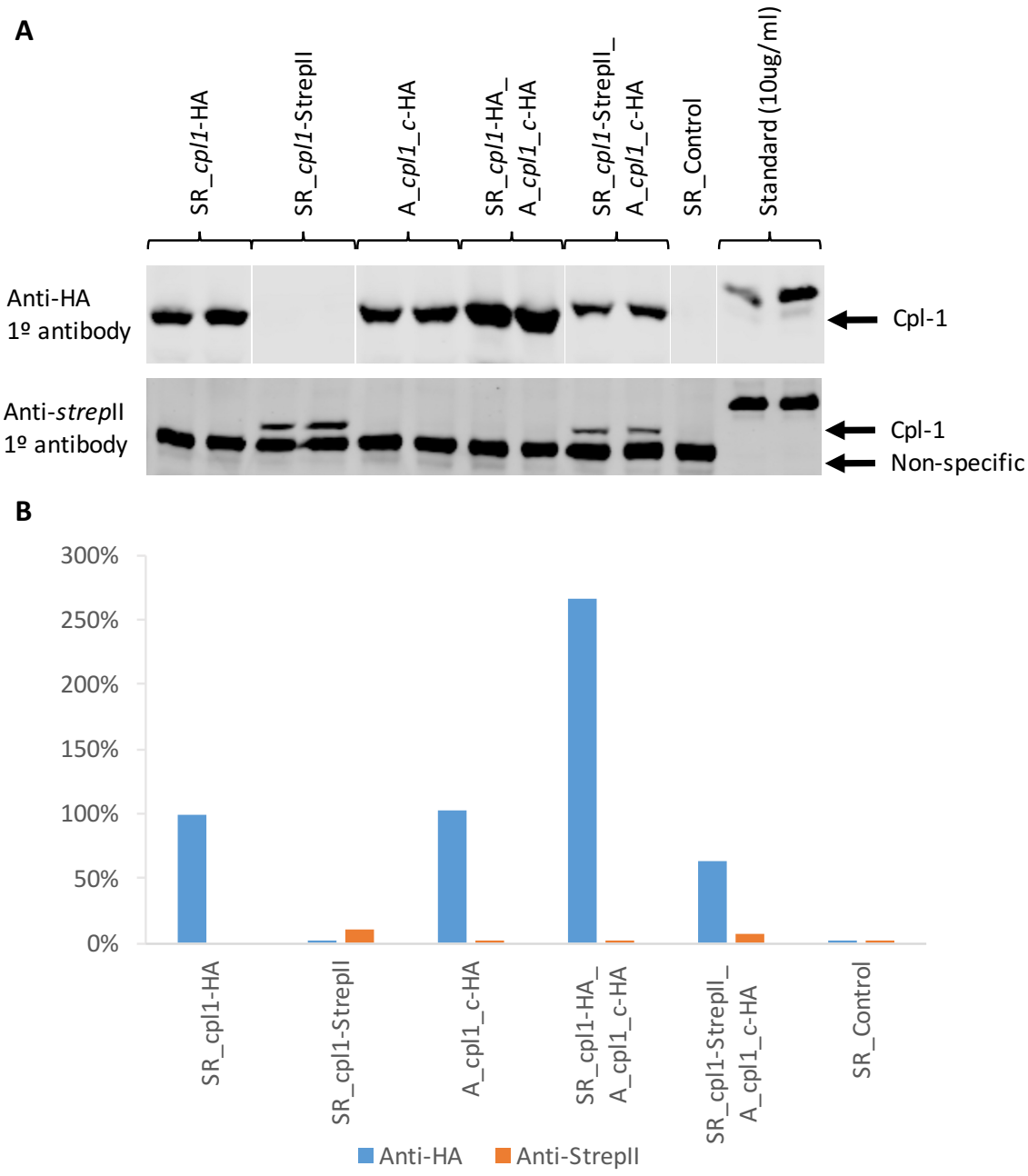


Figure 3.29 **(A)** Western blot analysis of *C. reinhardtii* whole cell extracts probed with anti-HA or anti-StrepII primary antibodies. IR detection was performed by the LiCor Odyssey® CLx. **(B)** Protein accumulation levels were quantified using the Odyssey® Image Studio v 5.0 and the signal was normalised using the protein standard and REVERT™ total protein staining. TN72_SR_*cpl1*-HA is the benchmark strain in this investigation and was therefore set to 100% accumulation. Other strains are represented relative to this. Protein accumulation of the StrepII tagged Cpl-1 appears to be considerably lower than that of the HA tagged protein.

As observed in Figure 3.24, the level of Cpl-1 protein accumulation in the TN72_SR_*cpl1*-HA_A_*cpl1*_c-HA is approximately equal to the sum of TN72_SR_*cpl1*-HA and TN72_A_*cpl1*_c-HA. However, it is interesting to note that accumulation of StrepII-tagged Cpl-1 is considerably lower than its HA-tagged equivalent. This could

be due to either the StrepII tag destabilizing the protein, or it could possibly be an artefact of the normalization method.

The addition of any peptide to a protein has the potential to destabilise it by affecting the ability or rate of folding of the protein into a stable form, and this may be the case with the StrepII tagged Cpl-1.

Alternatively, it is possible that the StrepII tag on the protein standard is more easily accessible than that on Cpl-1, or conversely that the HA tag is less accessible on the standard than on Cpl-1. This scenario would lead to an underestimation of StrepII tagged proteins, or an overestimation of HA tagged proteins, respectively. It seems plausible that a little of both scenarios may be occurring, given that the protein standard is tagged with an HA tag at the C-terminal end, followed by the StrepII tag beyond that. Anti-HA antibodies could therefore be slightly hindered by the presence of the StrepII tag, while anti-StrepII antibodies may find the epitope to be more accessible due to its slightly increased protuberance from the Cpl-1 protein.

To investigate which of these theories is correct, two separate protein standards could be created for each tagged Cpl-1. If much lower levels of StrepII tagged Cpl-1 are still observed, then it could be concluded that the StrepII tag is destabilizing Cpl-1, but if accumulation levels appear to be similar then it was merely an artefact of the normalization process.

In either case, it appears that the insertion of multiple expression cassettes into the *C. reinhardtii* chloroplast results in a simple addition of the protein levels observed when cassettes are inserted singly.

3.3.3 Codon pair optimization for improvement of protein accumulation levels

The second aspect of this chapter looks at using codon pair optimization as a technique for improving recombinant protein accumulation levels in the *C. reinhardtii* chloroplast. Codon optimization is a well-known approach to improve recombinant protein yields, and has been shown to be particularly important for transgene expression in the *C. reinhardtii* chloroplast – Franklin et al. (2002b) increased *gfp*

expression by 80-fold in a codon optimized version compared to a unoptimized version. The codon frequency chart for the *C. reinhardtii* chloroplast is shown in Appendix D. Taunt (2013) investigated the role that codon pairs play and whether these too could be optimized to improve expression. The study identified several codon pairs which had high predicted occurrence but were rarely or never observed in the *C. reinhardtii* chloroplast plastome. These pairs were terms Unexpectedly Unseen Zero Scoring Codon Pairs (uuZSCP). Two such uuZSCP stood out as being particularly significant, a Phenylalanine:Arginine pair, and an Aspartic Acid:Leucine pair, shown in Figure 3.30. The Phe:Arg pair was chosen for investigation in this project as it happened to be included in the initial codon optimization for *cpl1,* and is therefore present in TN72_SR_*cpl1*-HA.

	First Codon		Second Codon		Codon Pair		
Amino Acid Pair	**Code**	**Relative Freq.**	**Code**	**Relative Freq.**	**Combined Relative Freq.**	**Predicted Occurrence**	**Actual Occurrence**
Phe:Arg	UUU	0.67	CGU	0.73	0.49	28.29	0
Asp:Leu	GAU	0.76	CUU	0.14	0.1	9.1	0

Figure 3.30 Two uuZSCP are particularly rare in the *C. reinhardtii* chloroplast, despite their constitutive codons being relatively common. The relative frequency refers to the frequency with which this codon would be chosen to represent this amino acid in the *C. reinhardtii* chloroplast. The Predicted Occurrence is how often, given the frequency of amino acid pairs, one would expect to see this codon pair in the total 2579 codon pairs in the *C. reinhardtii* chloroplast.

MVKKNDLFVDVSSHNGYDITGILEQMGTTNTIIKISEST

TYLNPCLSAQVEQSNPIGFYHFARFGGDVAEAEREAQFF

LDNVPMQVKYLVLDYEDDPSGDAQANTNACLRFMQMIAD

AGYKPIYYSYKPFTHDNVDYQQILAQFPNSLWIAGYGLN

DGTANFEYFPSMDGIRWWQYSSNPFDKNIVLLDDEEDDK

PKTAGTWKQDSKGWWFRRNNGSFPYNKWEKIGGVWYYFD

SKGYCLTSEWLKDNEKWYYLKDNGAMATGWVLVGSEWYY

MDDSGAMVTGWVKYKNNWYYMTNERGNMVSNEFIKSGKG

WYFMNTNGELADNPSFTKEPDGLITVAYPYDVPDYA

Figure 3.31 The amino acid sequence of Cpl-1. The Phe:Arg pair is highlighted in yellow, while the purple box indicates the HA tag.

According to Maul et al (2002), the *C. reinhardtii* chloroplast plastome contains 30 identified tRNA genes, providing a small yet sufficient set for protein synthesis. After examination of the data and checking the chloroplast genome (GenBank: BK000554.2) for tRNA genes using tRNAscan-SE (http://lowelab.ucsc.edu/cgi-bin/tRNAscan-SE), a couple of inaccuracies in the tRNA data were noticed. Firstly, the *trnE* gene is counted three times, when in fact there appear to be only two copies identified on the plastome. There are therefore only 29 tRNA genes in the *C. reinhardtii* chloroplast. Secondly, the *trnR2* is labelled with the incorrect anti-codon – it should read *trnR2*(UCU), not *trnR2*(AGA) – likely just a confusion between codon and anticodon. Thirdly, while not a mistake per se, the nomenclature is unclear for tRNA genes and seems to be inconsistent. For example, there are *trnR* and *trnR2* genes, which identify tRNAs with different anticodons, but no *trnR1* exists. A table of all the tRNA genes in the *C. reinhardtii* chloroplast has been compiled in Figure 3.33. The *C. reinhardtii* chloroplast chromosome map by Maul et al. (2002) is shown in

Figure 3.34. It shows the location of all the tRNA genes on the plastome and explains that the two copies of *trnA* and *trnI* are due to their location on the inverted repeat sequences.

In order to test whether the presence of the uuZSCP Phe:Arg codon pair is affecting the level of protein accumulation in TN72_SR_*cpl1*-HA, two *cpl1* gene mutants were created by site-directed mutagenesis. The first, *CPOCv1* (Codon Pair Optimized Cpl-1 version 1), was a mutation of the phenylalanine codon from UUU to UUC – there is only one *trnF* gene, so this codon would be served by the same tRNA "wobbling" as *cpl1.* The second mutant, *CPOCv2*, includes a mutation of the arginine codon from CGU to AGG – this was designed to be served by the tRNA encoded by *trnR2*, rather than that encoded by *trnR1* which serves *cpl1*.

N.B. AGA is a slightly more frequently seen Arginine codon than AGG, as shown in Figure 7.4, and would not rely on tRNA wobble. However, UUU:AGA is also a codon pair that is never observed in the chloroplast, so for this reason, UUU:AGG which is observed, was chosen.

Phe	Arg	CPAI	Present in…
UUU	CGU	0	*cpl1*
UUU	CGC	0	
UUU	CGA	0	
UUU	CGG	0	
UUU	AGA	0	
UUU	AGG	0.0625	*cpov2*
UUC	CGU	1	*cpov1*
UUC	CGC	0	
UUC	CGA	0.0625	
UUC	CGG	0	
UUC	AGA	0.0625	
UUC	AGG	0	

Figure 3.32 All possible codon combinations for a Phe:Arg pair. The codon pair adaptation index (CPAI) is the ratio of codon pair frequency to the frequency of the most observed codon pair. The data above are from a "hand-pick" database in the *Codon Usage Optimizer Beta 0.92* (Kong 2013) software, whereby the codon pair usage in highly expressed genes is collected.

Given the theory for codon pair bias laid out in 3.1.4, it was hypothesised that *CPOCv1* would be unaffected by the codon change, while *CPOCv2* may display a different, possibly higher, level of Cpl-1 protein accumulation.

No.	Name	Anticodon	Codon	Amino Acid	
1	trnA	UGC	GCA	Ala	
	trnA	UGC			
2	trnC	GCA	UGC	Cys	
3	trnD	GUC	GAC	Asp	
4	trnE	UUC	GAA	Glu	
	trnE	UUC			
5	trnF	GAA	UUC	Phe	
6	trnG1	UCC	GGA	Gly	Phenylalanine
7	trnG2	GCC	GGC	Gly	
8	trnH	GUG	CAC	His	
9	trnI	GAU	AUC	Ile	
	trnI	GAU			
10	trnK	UUU	AAA	Lys	
11	trnL1	UAG	CUA	Leu	
12	trnL2	UAA	UUA	Leu	
13	trnM1	CAU	AUG	fMet	
14	trnM2	CAU	AUG	Met	
15	trnM3	LAU	AUA	Ile	
16	trnN	GUU	AAC	Asn	
17	trnP	UGG	CCA	Pro	
18	trnQ	UUG	CAA	Gln	
19	trnR1	ACG	CGU	Arg	Arginine
20	trnR2	UCU	AGA	Arg	
21	trnS1	UGA	UCA	Ser	
22	trnS2	GCU	AGC	Ser	
23	trnT	UGU	ACA	Thr	
24	trnV	UAC	GUA	Val	
25	trnW	CCA	UGG	Trp	
26	trnY	GUA	UAC	Tyr	

Figure 3.33 A list of tRNA genes identified by Maul et al. (2002). The yellow rows indicate duplicate tRNA genes, i.e. they have the same anticodon. Green rows indicate tRNA genes with different anti-codons, but coding for the same amino acid.

Several naming changes have been made in Figure 3.33 to the original names published in Maul et al. (2002). The new names fall in line with the annotations made to the chloroplast genome available at NCBI (https://www.ncbi.nlm.nih.gov/nuccore/BK000554). The names mean that identical, duplicate genes (*trnA*, *trnI* and *trnE*) have the same anticodon, while genes which encode similarly charged tRNAs, but with differing anticodons, are numbered e.g. *trnL*1 and *trnL*2. The only change made from the NCBI naming is to name both *trnE* genes the same, as they are identical duplicates. It should also be noted that some discrepancy exists between the defined gene length on NCBI and the tRNA genes detected using the tRNAscan software, most likely due to incorrect annotation of NCBI.

Figure Removed

Figure 3.34 The *C. reinhardtii* chloroplast chromosome. tRNA genes are identified on the outer ring in black and labelled with a single letter. e.g. V, which indicates *trnV*. Two tRNA genes (*trnA* and *trnI*) are located in the Inverted Repeat (IR) regions of the plastome, indicated in thick black, thus explaining their duplication. The second copy of *trnL* (2 o'clock) should be black to indicate tRNA, as should *trnI* and *trnA* on the right-hand IR. Reproduced from Maul et al. (2002). Note: As an aside, the neutral site used for chloroplast transformation throughout this thesis is located between *trnE* and *psbH*, 5 o'clock.

3.3.3.1 Codon Pair Optimized Cpl-1 version 1 and 2 (CPOCv1 and CPOCv2)

CPOCv1 was designed such that a different codon coded for phenylalanine, but that codon would still be served with the same tRNA as was used in the original, through the phenomenon known as tRNA wobble (Alkatib et al. 2012). CPOCv1 would therefore act as a control for this investigation, as the theory being tested is that tRNAs interact in the A and the P site of the ribosome in a manner that can influence translation efficiency. Here, the same tRNAs will be used in this mutant as in the original version and thus we expect to see no change in translation efficiency.

CPOCv2 was designed such that a different codon coded for arginine, and that that codon would be served with a different tRNA to the original, as there are two *trnR* genes in the *C. reinhardtii* chloroplast with different anti-codons. CPOCv2 was produced in an attempt to influence accumulation levels of Cpl-1, as different tRNAs will be interacting in the A and the P site of the ribosome compared to those in the original version, and therefore translation may progress at a faster rate.

3.3.3.1.1 Codon change by site-directed mutagenesis

To create CPOCv1, primers were designed to mutate U → C in the third position of the phenylalanine codon through PCR, and for CPOv2 to mutate C → A in the first position and T → G in the third position of the arginine codon. The steps taken to perform this mutagenesis are shown in Figure 3.35 and Figure 3.36 respectively. The primers used are indicated below. The lower case letters show the sites for mutation.

cpl1.F:	GCTCTTCAATGGTTAAAAAAAATGATTTATTC
cpocv1_SDM.F:	AGGTGGTGGTTcCGTCGTAA
cpocv1_SDM.R:	TTACGACGgAACCACCAACCT
cpocv2_SDM.F:	GTTGGTGGTTTaGgCGTAATAATG
cpocv2_SDM.R:	CATTATTACGcCtAAACCACCAAC
cpl1.R:	GCATGCTTATTAAGCATAATCTGGAAC

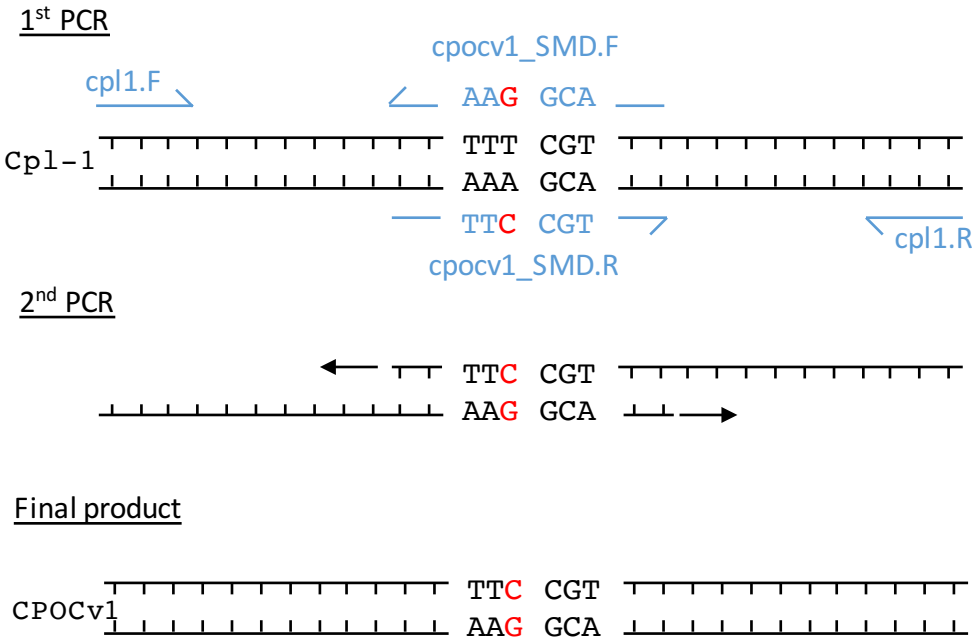


Figure 3.35 Site-directed mutagenesis was performed as above to modify the phenylalanine codon to make CPOCv1. Primers are shown in blue and PCR conditions are described in 2.3.7.

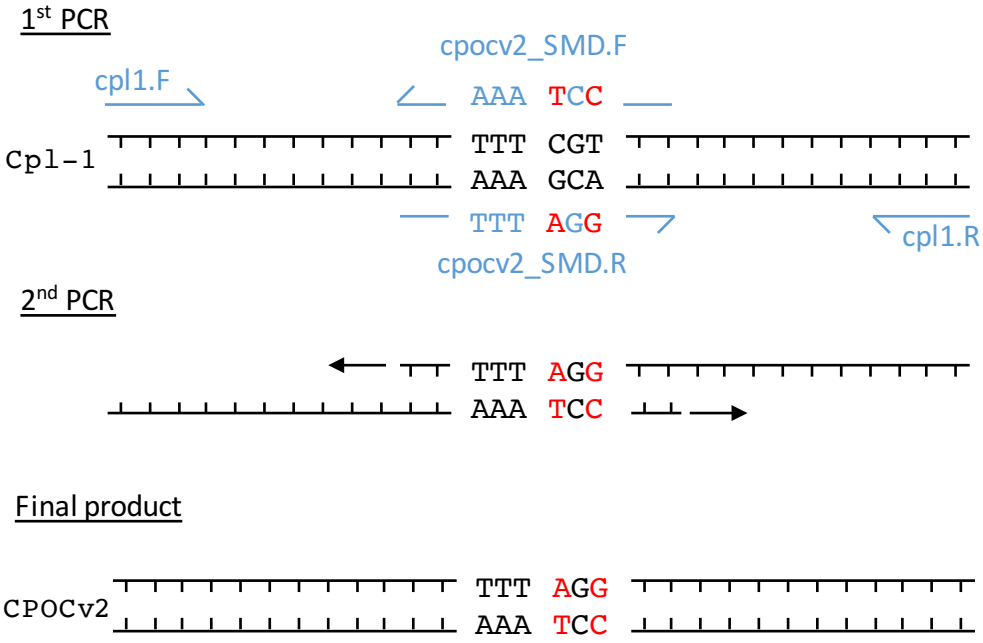


Figure 3.36 Site-directed mutagenesis was performed as above to modify the arginine codon to make CPOCv2. Primers are shown in blue and PCR conditions are described in 2.3.7.

3.3.3.1.2 Expression cassette production

The expression cassette was assembled as with previous examples. The gene, including the HA tag at the C terminus, was cloned into the pSRSapI vector in frame using SapI and SphI restriction sites. The insertion was confirmed by DNA sequencing

of the plasmid as described in 2.3.18. The resulting plasmids were called pSRapI_*CPOCv1*-HA and pSRapI_*CPOCv2*-HA.

3.3.3.1.3 *C. reinhardtii* transformation

Two TN72 transformations were performed, using pSRSapI_*CPOCv1*-HA and pSRSapI_*CPOCv2*-HA plasmids. Transformed *C. reinhardtii* lines were checked for homoplasmicity by PCR and finally confirmed by sequencing.

3.3.3.2 Western blot quantification of TN72_SR_*CPOCv1*-HA and TN72_SR_*CPOCv2*-HA

It is hypothesised that the codon changes made above will result in different levels of protein accumulation between the two transformant lines. To test this, four *C. reinhardtii* cultures, TN72_SR_*cpl1*-HA, TN72_SR_*CPOCv1*-HA, TN72_SR_*CPOCv2*-HA and TN72_SR_Control, were grown in an Algem Photobioreactor simultaneously, with the same starting cell density and with the same physiological conditions. OD_{740} was measured automatically every 30 minutes and samples were taken at intervals for analysis by western blotting. 1 ml samples were collected from the culture volume and immediately snap-frozen in liquid nitrogen and stored at -80 °C. The culture was grown for 145 hours in total. At the end of this time the frozen samples were analysed together by western blotting. Whole-cell extracts were separated by SDS-PAGE, with no compensation made for cell density. After blotting to nitrocellulose membrane and blocking, the membranes were probed with anti-HA primary antibody, followed by IRDye® secondary antibody and fluorescence was observed using the LiCor Odyssey® CLx system.

Figure 3.37 shows the western blot analysis – a rise in protein accumulation over time as cell density increases is to be expected.

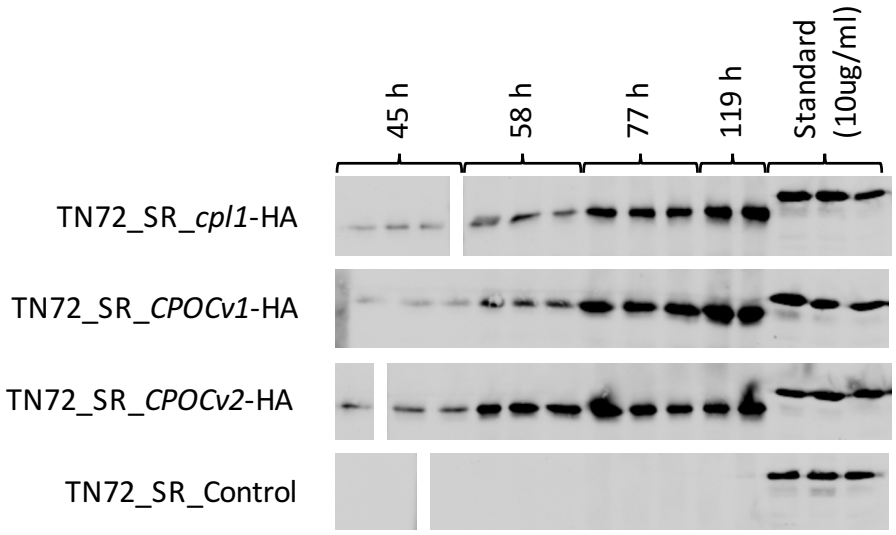


Figure 3.37 Western blot analysis of four *C. reinhardtii* strains at four time points. The known standard was used to normalise any difference between blotting on separate membranes.

The quantified and normalised data is displayed in Figure 3.38, alongside optical density data collected by the Algem photobioreactor. The whole experiment was repeated a second time, and the final results can be seen in Figure 3.39, with western blot analysis in Appendix E.

3.3.3.2.1 Statistical analysis of western blot data

A one-way ANOVA test was carried out on the recombinant protein yields of TN72_SR_*cpl1*-HA, TN72_SR_*CPOCv1*-HA and TN72_SR_*CPOCv2*-HA in order to identify any statistical difference between the levels of protein accumulation at different time points, detailed in Appendix F. In the first run of the experiment, it was found that the protein accumulation at 45 hours and 58 hours was significantly higher in TN72_SR_*CPOCv2*-HA than in the other cell lines, but that there was no significant difference beyond this point. On the other hand, the second performance of the experiment was run for longer and likewise, suggests that protein accumulation of TN72_SR_CPOCv2 is significantly higher the 45 and 56-hour time points. However, CPOCv1 and CPOCv2 then continue to rise, while the protein accumulation in TN72_SR_*cpl1*-HA rises at a much lower rate, being significantly lower than the codon pair optimised versions at the 112, 139 and 194-hour time points. The statistical method and results are detailed in Appendix G.

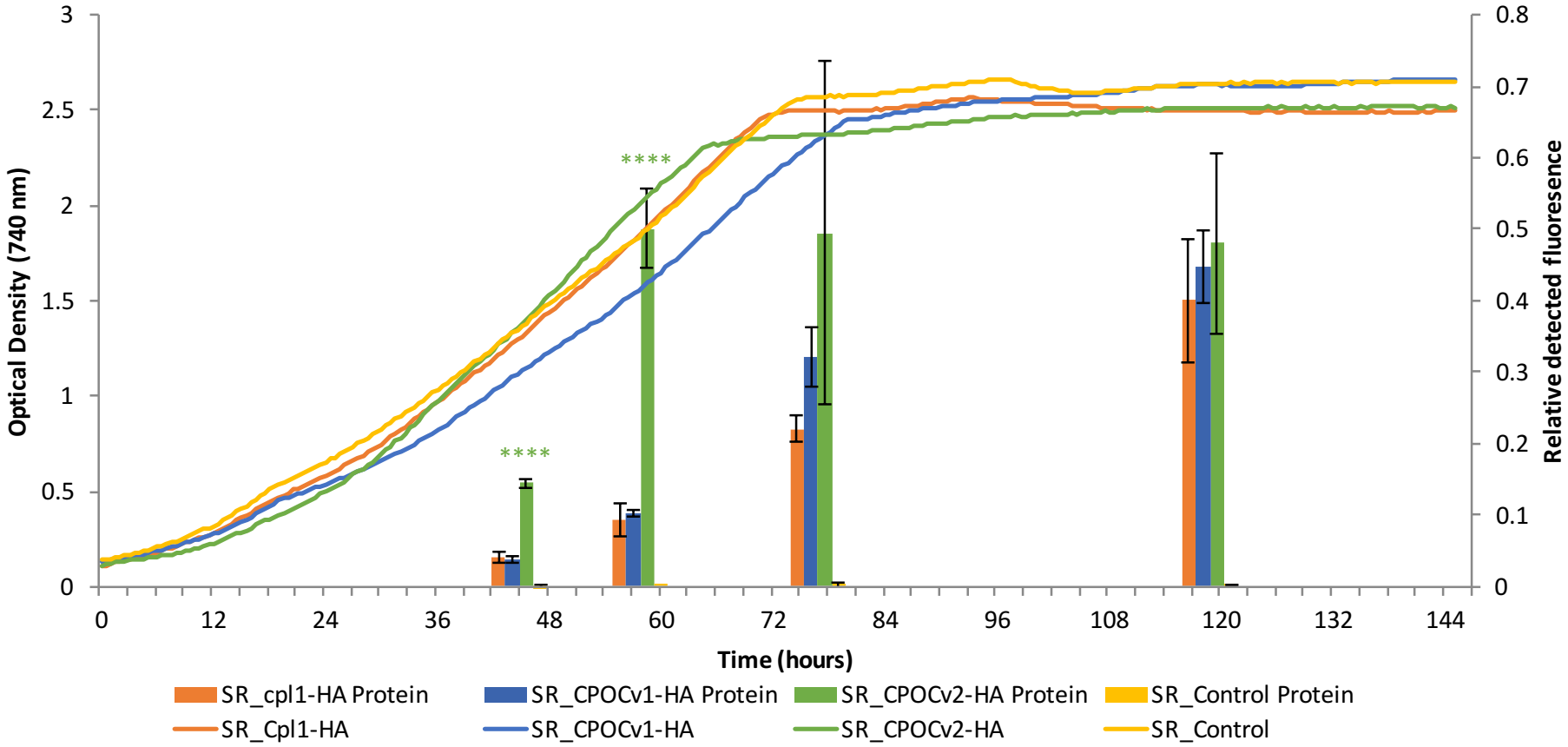


Figure 3.38 Algem growth data (line chart, left y-axis) and western blot quantification (bar chart, right y-axis). The Algem photobioreactor recorded optical density at 740 nm every 90 seconds. At OD_{740} = ~2.5, the photobioreactor is unable to record any higher OD due to the long path-length across a 400 ml culture. However, as indicated by the protein accumulation levels, the cultures are reaching their stationary phase around this point anyway. The relative detected fluorescence for HA-tagged protein accumulation is taken from Figure 3.37, and normalised using the protein standard and the OD_{740}.

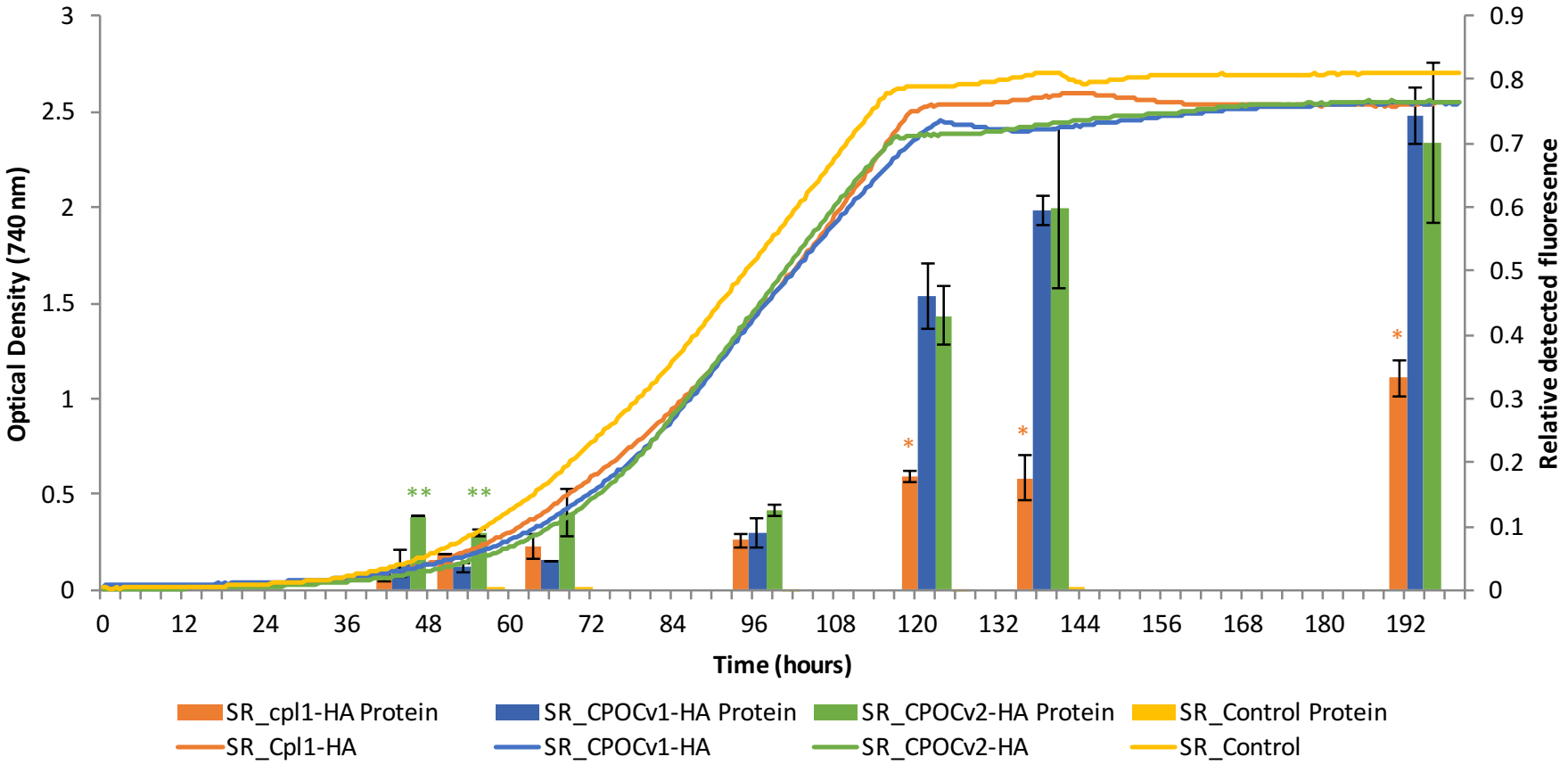


Figure 3.39 Algem growth data (line chart, left y-axis) and western blot quantification (bar chart, right y-axis). The starting OD was lower in this experiment, so the culture takes longer to reach the logarithmic growth stage. The original western blot analysis is shown in Appendix E, in this instance the 25 kDa band of the protein ladder was used to normalise the membranes, this enabled more samples to be fitted onto a single gel and thus measure accumulation over a longer time period.

Figure 3.38 indicates that CPOCv1 is growing slightly slower than the two controls, and CPOCv2 is growing slightly faster. This is most likely due to a slightly different number of cells used to inoculate the cultures, despite efforts to ensure they were the same. However, to compensate for this when comparing protein yields from western blot analysis, the detected signal was divided by the OD, thus normalising the data for cell density. Each of the four membranes used for western blot analysis also showed slightly different signals for the protein standard, probably due to non-uniform transfer from the gel to the membrane, slightly different antibody incubation times and slightly different dilutions of milk and antibody. Assuming that these variables are the same within membranes, but can vary from membrane to membrane, the data were normalised by division of the average level of protein standard detected. The same is true for Figure 3.39, but this time using the 25 kDa band of the protein ladder to normalise for inter-membrane differences, as there was no room on the membrane to include the protein standard too.

While the two graphs cannot be compared directly, due to the longer lag time in the latter and different inter-membrane normalisation techniques, they do appear to display two trends. Firstly, that CPOCv2 seems to accumulate faster than the other Cpl-1 variants throughout the logarithmic growth stage, and secondly that the un-optimised version of Cpl-1 consistently accumulates to a lower level than either of the CPO versions.

This is an interesting observation, as according to the theory of codon pair optimisation, CPOCv1 accumulation should be no different to the unchanged Cpl-1, as the tRNAs serving the mRNA are identical. However, particularly in the later growth stages in Figure 3.39, there is a consistently and significantly higher accumulation of CPOCv1 and CPOCv2.

Given the evidence from section 3.3.1, in which data was collected on the error introduced at various points of measuring protein accumulation, these differences seem too large to be artefacts of variability in western blotting and cell density measuring techniques. Therefore, alternative explanations for this result must be sought.

The first thing to consider is the fact that in CPOv2, another codon pair is being unintentionally altered. The amino acid immediately following the codon optimized Arginine is another Arginine, as can be seen in Figure 3.31. This Arg:Arg codon pair is being unintentionally altered when the preceding Phe:Arg codon pair is optimized in CPOv2. This presents another variable that has not been accounted for, and as shown in Figure 3.40, the new pair is never observed in the "hand-picked" database, consisting of highly expressed proteins in the *C. reinhardtii* chloroplast, generated by Taunt (2013). This presents an issue, as although one uuZSCP (UUU:CGU) has been replaced, another unseen pair is introduced (AGG:CGU). It is, therefore, possible that the beneficial effects of the former are negated by the detrimental effects of the latter. There is no similar effect made by CPOv1, as all Phenylalanine codons are served by a single tRNA.

Arg	Arg	CAI	Present in…
CGU	CGU	1	*cpl1 & cpov1*
CGC	CGU	0	
CGA	CGU	0	
CGG	CGU	0	
AGG	CGU	0	
AGA	CGU	0	*cpov2*

Figure 3.40 Data taken from the "hand-pick" database of *Codon Usage Optimizer Beta 0.92* (Kong 2013). This is a database of codon usage in highly expressed proteins, a common method of selecting the best codon usage.

Another possibility is that while gene expression appears to be primarily controlled at the level of translation in the *C. reinhardtii* chloroplast, the transcriptional level still may play a part and as such the levels of mRNA produced for CPOCv1 and CPOCv2 should be investigated. If these are found to remain similar then it can be discounted, but should be investigated nevertheless.

Additionally, the unexpectedly unseen nature of UUU:CGU could be just a coincidence. Although this seemed unlikely, given the convincing statistical analysis performed by Taunt (2013), the assertion that this pair is purposefully avoided is slightly undermined by the frequent presence of UUC:CGU. UUC:CGU accounts for 33 of the 59 observed instances of Phe:Arg in the *C. reinhardtii* chloroplast, but given

the fact that UUU and UUC are served by the same tRNA, the different codon should confer no difference in translational efficiency or codon selection. It could be therefore that some neutral mechanism has resulted in this bias, not selection, for example, preferential mutation bias during DNA replication or repair (Lucks et al. 2008). Another alternative is that the single base change in the *cpl1* transgene has created an unexpected local area of secondary structure within the mRNA, thus altering the translational efficiency (Gaspar et al. 2013). This kind of interaction is very difficult to predict, as the mRNA folding events may involve two distant sequences, and changing one base might strengthen/weaken these interactions.

The final and possibly most likely explanation is one which is little understood. This is the frequently observed and reported phenomenon of variable transgene expression between "identical" *C. reinhardtii* chloroplast transformants (Mayfield and Schultz 2004; Surzycki et al. 2009; Michelet et al. 2011). The differences can be quite remarkable: Surzycki et al. (2009) reported that recombinant protein expression levels of VP28, a White Spot Syndrome virus protein, varied between transformants from 0.88 % to 20.9 % TCP. When only a low number of transformants are generated, as is frequently the case with the glass bead method regularly used in the Purton group, a significant difference in transgene expression could be incorrectly apportioned to a factor which really has little effect. The reason put forward by Surzycki et al. (2009) for this observation is transformation-associated genotypic modifications, termed the "transformosome". The paper postulates that during the process of genetic transformation, fragments of the transformation plasmid are unknowingly inserted into the plastome or nuclear genome, thus altering nuclear genes that regulate chloroplast gene expression, protease expression, or sites of regulatory gene action in the chloroplast. Evidence is presented that the level of steady state transgene mRNA is unaffected, giving further weight to the belief that gene regulation occurs primarily at the translation level in *C. reinhardtii*.

C. reinhardtii is extremely good at resource management, and as such is able to prioritise the production of key proteins over others, for example when Copper is limiting the cell will preferentially produce cytochrome C oxidase over plastocyanin for use in the electron transfer chain of PS1 (Merchant and Bogorad 1986). However,

as Cpl-1 is a recombinant protein, it is not involved in this hierarchy and is therefore likely to be the first protein to be sacrificed in the case of a metabolic limitation. As the value for protein accumulation is relative (i.e. it has been normalised for the optical density of the culture), Figures 3.38 and 3.39 both show increasing levels of protein accumulation as the culture progresses into stationary phase and then it plateaus. Therefore, metabolic limitation does not appear to be an issue in this instance.

To conclude, it appears that a difference in Cpl-1 accumulation has been observed, but it is unclear as to whether this difference is due to the codon pair optimisation, as the results do not support the theory that codon pairs can be optimised due to the different interactions of tRNAs at the A and P-sites of the ribosome, and the observed differences are small enough to be accounted for by other factors.

3.4 Conclusion and future work

This chapter has investigated two putative methods of improving recombinant protein expression in the *C. reinhardtii* chloroplast: the introduction of multiple expression cassettes; and the application of codon pair optimisation. The former appears to be a method to robustly improve protein accumulation, while the latter method remains unclear as the complex and incremental nature of the change appears to be masked by other factors in play. Interestingly the concern addressed at the start of the chapter, namely how to accurately compare protein accumulation between transformant lines, re-emerged as an issue towards the end of the investigation.

3.4.1 Multiple expression cassette strategy

The overall conclusion from this study was that *C. reinhardtii* lines transformed with multiple expression cassettes produce a recombinant protein yield approximately equal to the sum of the yields observed in lines transformed with each cassette individually. This finding could potentially be used to double existing yields in the *C. reinhardtii* chloroplast. However, some further work into what the limitations of this method are will enable us to better understand its potential.

The first and most obvious test is to investigate whether this result can be repeated with other transgenes. This would be a relatively straightforward test to perform, but it would be interesting to test both another highly expressed gene, as well as a gene which is currently only weakly expressed in the *C. reinhardtii* chloroplast. The *cpl1* gene is well expressed, and therefore it is possible that CES mechanisms are limiting its accumulation under a single 5' UTR, thus making it a good candidate for this technique. On the other hand, proteins displaying poor accumulation might be limited by protease activity for example, and will therefore remain at low levels of accumulation despite different mRNA transcripts.

The second question that one would like to ask is to what point can this theory be extended. Will three or four cassettes triple or quadruple the protein accumulation, for instance? Going beyond two copies of the gene raises the issue discussed in section 3.3.2.2 – when two duplicate genes are in parallel directions to each other, there is a potential for DNA excision to occur between the repeats. Avoiding this by increasing the distance between the copies would most likely require another neutral site to be used and targeted for transformation, requiring the use of a second selectable marker, enabling sequential transformations. Two options exist in this case, either the addition of a positive selectable marker, such as *aadA* and maintaining the algae on spectinomycin containing media, or the addition of a negative selectable marker, such as *codA*, to the recipient line which is then replaced in the second transformation event (Young and Purton 2014). The latter would result in the formation of a marker-less mutant which is desirable for further transformations but requires more time to make the constructs. As an initial test of how far the multiple-cassette theory can be taken, it is suggested that using an antibiotic resistance marker is more straightforward. Furthermore, the practice of using codon manipulation to minimise the potential for DNA excision, as performed in 3.3.2.2.1.1, may enable more cassettes to be inserted into the same site. However, given that codon optimisation has been shown to be very influential in protein accumulation the *C. reinhardtii* chloroplast (Surzycki et al. 2009), this can only be taken so far without a considerable reduction in translation.

As discussed in 3.3.2.2.4.3, an investigation into why the StrepII epitope appears to be affecting protein accumulation so profoundly should be performed. The simplest explanation is that normalising to the protein standard is skewing the results by having a different epitope accessibility than that of the singly tagged proteins. This was an issue encountered previously in the Purton group (Taunt 2013; Stoffels 2015), whereby the commercial multi-tagged protein standard suggested levels of protein accumulation that could not be confirmed by SDS-PAGE and Coomassie Blue staining. Furthermore, a similar result in 5.3.2.4.3 suggests that this is the case, where another endolysin, CD27L, also appears to accumulate to a much lower level when StrepII tagged. However, it may be that the endolysin is extremely sensitive to C-terminal epitope tags in terms of stability, in which case an investigation into the accumulation using different epitope tags could be carried out and one should also consider tagging the N-terminus instead.

Finally, a calculation of recombinant protein yield as a percentage of Total Soluble Protein (TSP) or Total Cell Protein (TCP) would be possible once the issue surrounding the protein standard has been resolved. This would allow comparison to other yields achieved in the literature. Alternatively, an ELISA assay could be performed but antibodies would have to be raised against Cpl-1, which is expensive. A robust method of estimating protein yield using an epitope tag and western blot analysis would be cheaper and more flexible, enabling quantification of other proteins in the future.

3.4.2 Identical transformant transgene variability

The issue of large variability in transgene expression between supposedly identical transformants, also known as “clonal variation”, has the potential to mask or exaggerate investigations into other methods of improving protein accumulation. Until this issue has been addressed, “fine-tuning” techniques will remain difficult to assay and are unlikely to be as effective as simply screening many transformants and selecting the best, as many protein production companies working with *Pichia pastoris* currently do (Aw et al. 2017).

The most useful tool to aid investigation into this phenomenon would be the development of a sensitive and robust reporter gene for *C. reinhardtii*. Reporter genes such as GFP have been used for this purpose in many other organisms, as fluorescence is graded and easily measured *in vivo*, enabling rapid quantitative screening of many transformants. However, while some studies have produced reasonable levels of GFP in the *C. reinhardtii* chloroplast (Franklin et al. 2002), due to the high levels of auto-fluorescence in *C. reinhardtii*, GFP is unsuitable as it must be produced at high levels to be detectable (Mayfield and Schultz 2004). Luciferase has been developed as a viable alternative, with fluorescence shown to be proportional to protein accumulation, and no interference by cell auto-fluorescence (Mayfield and Schultz 2004). Surzycki et al. (2009) demonstrated that co-introduced selectable marker genes are expressed in a manner proportional to the GOI. Therefore, the co-introduction of luciferase with a GOI should enable its fluorescence to be used as a reliable proxy for recombinant protein accumulation. This in turn will enable the rapid screening of many different transformants, and the production of a bell-curve for transgene expression. Once this method can be used to create a reproducible bell-curve of luciferase detection, then statistical analysis can reliably deduce how smaller changes are influencing protein accumulation levels.

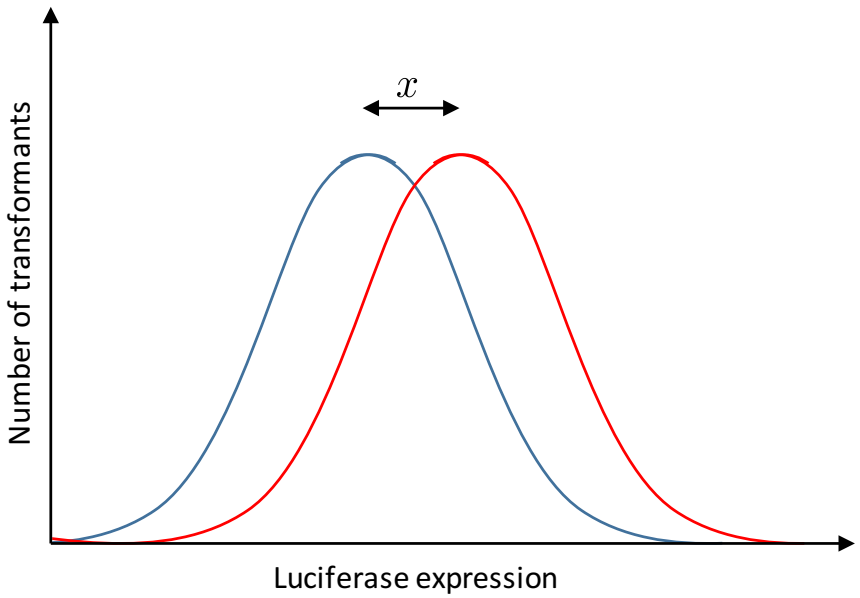


Figure 3.41 A comparison of means. By producing two bell curves with the luciferase expression level measured from many transformants, it will be possible to identify subtler improvements in protein accumulation. Here x is the difference between two transformant

lines, with the blue line representing all the transformants of one treatment, and the red line all the transformants of another.

Development of the above will enable more precise testing of other strategies including codon and codon pair optimization. The above also depends upon the development of a more efficient system of transforming *C. reinhardtii* chloroplast, as many transformants need to be rapidly produced for this method to be effective. Current work in the Purton lab by Saowalak Changko aims to improve current transformation methods. Another future advantage of such a system is that it would allow the identification of outliers, cells which produce very high or very low levels of recombinant protein. With the cost of whole genome sequencing (WGS) falling all the time, these outliers could be subjected to WGS in an attempt to understand what factors are influencing protein expression so heavily, and then attempt to specifically introduce those changes. WGS has already been used to successfully identify point mutations in *C. reinhardtii* (Dutcher et al. 2012). This will improve our understanding of the subtleties of current transformation methods and the complex role of nuclear factors in chloroplast gene expression.

3.4.3 Codon pair optimisation

As discussed in the previous section, attempting to manipulate and detect differences in codon pair optimisation strategies may be premature at this stage of developing the *C. reinhardtii* chloroplast as a protein production platform. However, given the importance of codon optimization often reported in the literature (Franklin et al. 2002; Surzycki et al. 2009; Coragliotti et al. 2011), and the clear existence of a codon pair bias in the *C. reinhardtii* chloroplast (Taunt 2013), it is possible that codon pair optimisation will become an important tool in the future.

Codon and codon pair optimisation are currently just a ‘best guess’ at how to express a recombinant protein to a high level. It appears that codon optimisation by the host in its own endogenous genes enables controlled, delicate adjustments of expression. The discovery of codon pair optimisation further muddies the waters around deriving which codons to use in a transgene. It seems intuitive that just using the most common codons with the largest pool of available tRNAs will result in the fastest

translation and therefore the greatest level of protein production. However, there is evidence to suggest that more subtle mechanisms are in play, such as the use of rare codons to slow down translation in sections of the protein that require complex co-translational folding for example (Sherman and Qian 2013).

When one is trying to express proteins for commercial uses, generally the highest possible level of protein accumulation is desired, something that the host cell rarely attempts to do *in natura*. On the other hand, bacteriophage, which infect host cells and whose genomes follow similar codon usage patterns to their host, often do attempt to produce as much protein as possible, to make as many phage progenies as possible (Lucks et al. 2008). In the case of endolysins, as they are designed to be highly expressed *in natura,* it can be assumed that they are already coded in a near-optimal manner for the host bacterium. This raises the possibility of generating an inter-species codon map, by which one can map a rare codon in *Streptococcus pneumoniae* to a rare codon in *C. reinhardtii* for example, thus conserving the pattern of “fast” and “slow” codons that has been optimised over millennia.

Overall, this study demonstrates that there is huge potential to improve protein accumulation yields in the *C. reinhardtii* chloroplast, and the technique of using multiple expression cassette insertions appears to be a robust method of improving protein yields. However, other methods of improving protein accumulation such as codon pair optimisation firstly require improvements in understanding of clonal variation and transformosome effects before they can be properly assayed and implemented.

Chapter 4 Improving the activity of Cpl-1 against *Streptococcus pneumoniae*

4.1 Introduction

This chapter aims to improve the activity of the *Streptococcus pneumoniae* targeting endolysin, Cpl-1, in its ability to lyse *S. pneumoniae* cells. Several strategies were tested with this aim in mind, namely: manipulating the cell wall binding site of Cpl-1; attempting to form pre-dimerised Cpl-1 endolysins in the *C. reinhardtii* chloroplast; and investigating holin production in the *C. reinhardtii* chloroplast. While Cpl-1 is already a highly active endolysin, with proven efficacy *in vivo* using mouse and rat models (Loeffler et al. 2003; Entenza et al. 2005; McCullers et al. 2007; Grandgirard et al. 2008), an increase in activity would enable a reduction in dose, thus minimising some of the potential limitations discussed in 1.2.2.4 involving immunogenic reactions.

4.1.1 Manipulating Cpl-1 binding activity

The function of an endolysin, when produced by a bacteriophage, is to lyse the host bacterium, thus allowing the release of phage progeny into the environment in search of another bacterial host. The endolysin, therefore, should lyse only the current host cell, and not neighbouring bacterial cells, which could be potential future hosts for the progeny. One characteristic of endolysins is that they tend to bind strongly, often in the pico- to nanomolar range (Loessner et al. 2002), to their substrate - it has been postulated that this is a tethering mechanism to reduce unintentional damage to nearby bacteria (Chapot-Chartier 2014).

This feature, while beneficial to bacteriophage, often reduces the activity level of the endolysin as an enzyme as it restricts its lytic activity to a localised area of the peptidoglycan cell wall (Nelson et al. 2001). For some endolysins, such as Cpl-1, removal of the C-terminal cell wall binding domain (CBD) results in an almost complete loss of function (Sanz et al. 1992), but for others such as CD27L (a clostridial endolysin) removing the CBD can improve lytic activity and broaden the spectrum of

bacteria against which it is active (in this case, to include two more species, in addition to the six species susceptible to the full-length CD27L protein) (Mayer et al. 2011).

While it is clear from the literature (Sanz et al. 1992; García et al. 1999; Monterroso et al. 2008) that a complete removal of the Cpl-1 CBD results in a reduction in activity, we propose that if binding could be weakened, activity may be improved. With weaker binding, the endolysin may be able to release after lysis and repeat its lytic actions elsewhere, thus making the endolysin more suited to a therapeutic role. This project involved mutating the cell wall binding site of Cpl-1 via site-directed mutagenesis, and comparing the activity of the resulting proteins with that of the original.

It is worth noting that in this case of manipulating the cell wall binding site, we are using a slightly different definition of "activity". The SI unit for enzyme activity is the katal, 1 katal = 1 mol s^{-1}, i.e. the number of moles of substrate converted per second. However, in this instance, we are more interested in the number of moles converted per endolysin molecule, which is what is termed "activity" in this chapter.

From a protein engineering perspective, Gram-positive-targeting endolysins are particularly attractive due to their modular structure, and the decoupling of the binding and catalytic domains. In many globular enzymes, the binding and catalytic sites are the same, such that changes to binding sites often reduce the ability of the enzyme to catalyse the reaction. While the endolysin catalytic domain must still bind its substrate, the majority of binding specificity and affinity is conferred by the binding domain (Mayer et al. 2011).

4.1.2 Dimerisation as a method to improve protein stability

Cpl-1 exhibits extensive similarities with both major pneumococcal autolysins, LytA and LytC. In the presence of choline, LytA dimerises – becoming stable and active, while LytC is active as a monomer but is stabilized in the presence of choline. Cpl-1 is believed to dimerise through its C-terminal (cell wall binding) domain when choline is present. The additional activity of the dimers is likely to be due to the doubling of

active sites spatially orientated next to the pneumococcal cell wall (Monterroso et al. 2008).

Figure Removed

Figure 4.1 The stable and active LytA dimer. Cell binding domains are shown in orange, the catalytic domain in blue and the 10 choline ligands shown in CPK colour scheme. Resch et al. (2011) postulate that Cpl-1 could dimerise in a similar manner given the homology of the final 13 amino acids at the C-terminus. Image reproduced from Maestro & Sanz (2016).

Cpl-1 is known to be cleared from the blood plasma of mice in approximately 20 minutes (Witzenrath et al. 2009), presenting a possible issue for systemic use. Resch et al. (2011) hypothesised that a "pre-dimerised" Cpl-1 homodimer would be a more stable and active therapeutic than the native monomer. LytA dimerisation occurs through the 13 amino acids at the C-terminus (Fernández-Tornero et al. 2002), and 10 of these 13 amino acids were found to be identical to the C-terminus of Cpl-1. By the mutation of each of the 13 amino acids in turn, Resch et al. (2011) produced a series of mutated Cpl-1 proteins, in an attempt to find one which forms a disulphide bridge between two monomers to produce a homodimer. The successful mutated Cpl-1 dimer was shown to have a 10-fold reduction in plasma clearance and a two-fold increase in activity, compared to the monomeric version (Resch et al. 2011a).

Cpl-1 naturally dimerises on binding choline, the crystal structure of which has been reported by Buey et al. (2007) and is displayed in Figure 4.2. This dimerisation occurs not at the C-terminus, as the pre-dimerising mutant created by Resch et al. (2011)

does, but between the final choline binding repeats at the N-terminus of the cell wall binding domain – i.e. the choline binding repeat closest to the acidic linker and catalytic domain.

Figure Removed

Figure 4.2 The C-terminus of the CBD is circled in red in each monomer. The box shows the side-chain of the aromatic residue involved in dimerisation. Reproduced from Buey et al. (2007).

Figure 4.2 does not show the acidic linker joining the two domains of each monomer as this could not be accurately modelled. This observed mode of dimerisation is similar to that of the dimer-dimer interactions of LytA to form a tetramer (Buey et al. 2007).

The quaternary structure of LytA and LytC is crucial to their hydrolytic activity, and given the homology each displays to Cpl-1, it is reasonable to believe that the quaternary structure of Cpl-1 is also important. However, given the stability of Cpl-1

both as a monomer and without the presence of choline, it may not be as important as with the pneumococcal autolysins.

4.1.3 The role of holins in the bacteriophage lytic cycle

In the canonical two-gene model of bacteriophage-mediated cell lysis, the holin plays an essential role in enabling the endolysin to access the peptidoglycan cell wall, by creating a hole in the cytoplasmic membrane of the bacterium (Young 1992). The timing of this event is critically important to the success of the bacteriophage by controlling the length of the cycle, and thus the holin gene is under enormous selective pressure (Wang et al. 2000). It is perhaps due to this selective pressure that holins are such a diverse functional group, comprising seven superfamilies encompassing 52 recognised families with no clear conserved sequence motif (Reddy and Saier 2013). However, all holins have certain characteristics including the possession of at least one transmembrane α-helical sequence and a highly charged hydrophilic C-terminal domain (Gu et al. 2014).

The release of endolysins beyond the cytoplasmic membrane is a carefully orchestrated event, and one which is not wholly understood. Several theories exist on how this timing is controlled, but the most well-established theory is that there is a late expression of holins, which accumulate in the cytoplasmic membrane. When the concentration of these holins reaches a critical threshold, a raft is formed and then a conformational change causes the formation of a hole (White et al. 2011). This sequence of events is illustrated in Figure 4.3.

Figure Removed

Figure 4.3 An illustration of the holin triggering mechanism, and the consequential depolarisation of the membrane. pmf = proton motive force.

A further method of holin control exists in some bacteriophage systems, involving an antiholin. An antiholin is almost identical to the holin, but with one or two additional N-terminal amino acids, one of which is a positively charged residue (usually lysine) and is translated from a dual start motif encoding: Met-Lys-(x)-Met-…. The antiholin then binds to the holin resulting in heterodimers unable to form the raft displayed in Figure 4.3. However, this binding is PMF (proton motive force)-dependent: as the antiholin is produced at a slightly lower rate than the holin, a hole will eventually form, resulting in a lowering of the PMF and the unbinding of the anti-holin and a positive-feedback resulting in the collapse of the PMF and the 'all or nothing' release of endolysins beyond the cell (Pang et al. 2013; Lella et al. 2016).

The gene *cph1*, encoding a holin, is located immediately upstream of *cpl1* in the Cp-1 bacteriophage genome. The two genes are transcribed from two tandem promoters upstream of *cph1* and are expressed from late phage transcripts (Martín et al. 1998). The *cph1* gene does not display the classical dual-start motif, although it does encode another methionine residue at position six, but no lysine residue lies between them. Cph-1 belongs to the Holin IV superfamily and has three predicted transmembrane α-helical segments (Reddy and Saier 2013).

Recent literature has suggested that holins can be applied exogenously to *Streptococcus suis* (Shi et al. 2012) and *Staphylococcus aureus* (Song et al. 2016) in

conjunction with endolysins to produce an enhanced antibacterial effect, improving the killing efficacy and also broadening the spectrum of susceptible pathogens (Song et al. 2016). The studies by Shi et al. (2012) demonstrate that the transmembrane domains of the holin are essential for its antimicrobial activity, and it is assumed that the formation of large holes in the membrane causes cell death. However, it is not understood how the holin gains access to the cell membrane given the presence of the peptidoglycan layer, nor how capsules and slime-layers affect holin access. Further investigation of holin structure and its potential as a novel antibacterial has been hindered by the difficulties involved in successfully expressing holins in prokaryotic systems, given the cytotoxicity of holins (Garrett et al. 1990). Therefore we aim to produce holins in the *C. reinhardtii* chloroplast to enable further research.

4.2 Aims and objectives

1. To identify and optimize a method for comparing activity levels between different mutant Cpl-1 endolysins against *S. pneumoniae*.
2. To produce seven transgenic *C. reinhardtii* lines expressing *cpl-1* genes with different mutations in their choline binding sites (CBS).
3. To confirm accumulation of the CBS mutant Cpl-1 proteins in the *C. reinhardtii* chloroplast by western blotting.
4. To compare the activity of the CBS mutant Cpl-1 proteins against *S. pneumoniae*.
5. To produce a transgenic *C. reinhardtii* line expressing a variant *cpl-1* gene encoding a dimerising mutant.
6. To confirm accumulation of a Cpl-1 dimer in the *C. reinhardtii* chloroplast by western blotting.
7. To produce the holin Cph-1 in the *C. reinhardtii* chloroplast and confirm accumulation by western blotting.

4.3 Results and discussion

4.3.1 Manipulation of the Cpl-1 choline binding sites to improve activity

The two choline binding sites on the C-terminus of the Cpl-1 protein recruit the endolysin to the cell wall by binding to the integral (lipo)teichoic acids. Here the key residues involved in Cpl-1 choline binding are identified and mutated by site directed mutagenesis. The resulting Cpl-1 mutants are assayed for lytic activity.

4.3.1.1.1 Identification of potentially important residues in choline binding

The binding affinity of the Cpl-1 cell binding domain to choline is relatively weak for an endolysin, $K_d \approx 3.6$ mM (Monterroso et al. 2008), however the requirement of bacteriophages to tether their endolysins to only the host's cell wall suggests that it is unlikely to completely dissociate, and Cpl-1 possibly transits across the glycopeptide network to reach new cleavable bonds (Jervis et al. 1997; Monterroso et al. 2008). In order to test whether there is an optimal binding affinity that improves activity, and lies between the two extremes of complete CBD removal and the native tethering CBD, a graded weakening of the CBD was tested.

The crystal structure of Cpl-1 binding choline, shown in Figure 3.1, was analysed using Jmol[1]. Interactions of less than 5 angstroms from the choline moiety were considered to be of importance for hydrogen bonding during endolysin binding and two tryptophan residues and one tyrosine were identified in each binding site to fit these criteria. This confirmed the original analysis by Hermoso et al. (2003) which also identified these three residues as important, forming a hydrophobic cavity into which the choline methyl group sits. Hydrogen bonds form between the three residues and the choline and these interactions are shown in Figure 4.4. For the purposes of this investigation, the two tryptophan residues in each binding site were selected for mutation.

[1] Jmol is an open-source Java viewer for chemical structures in 3D. http://www.jmol.org/

The cell wall binding domain of Cpl-1 has six choline binding repeats (p1-6, see Figure 4.5), at the interface of which a putative choline binding site exists. However, the crystal structure of Cpl-1 suggests that only the first two choline binding sites are in fact active, namely: p1/p2 and p2/p3 (Hermoso et al. 2003). At the p1/p2 site, there are two key interactions: the nitrogen atoms on Trp202 and Trp209 interacting with the choline moiety. At the second choline binding site, the p2/p3 site, it is Trp223 and Trp230 that interact with the choline. When the fact that an O-H bond and N-H bond are approximately one Angstrom in length is accounted for, the hydrogen bonds formed in these interactions cover approximately 3-4 Angstroms. According to Jeffrey (1997), hydrogen bonds over this distance range from "moderate, mostly electrostatic" to "weak, electrostatic".

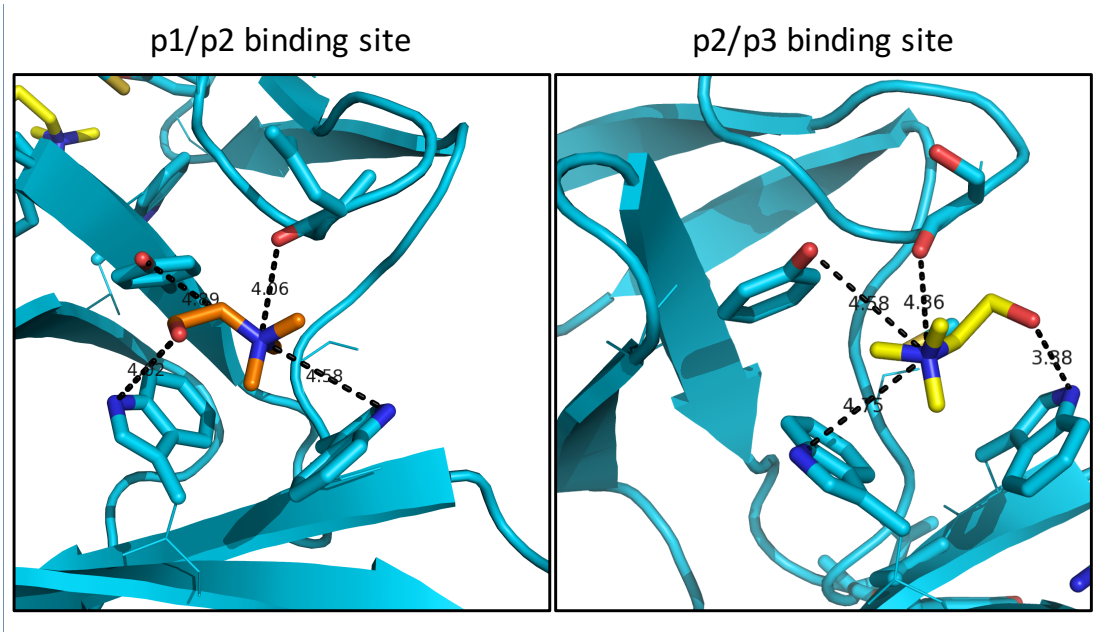


Figure 4.4 The two active choline binding sites in the CBD of Cpl-1. Distances are in Angstroms. It should be noted that the distances are between the nitrogen (blue) and oxygen (red) atoms, so are not a direct representation of a hydrogen bond length as the hydrogen must be accounted for too in each case. However, these remain "moderate/weak and electrostatic" hydrogen bonds.

Given that tryptophan is a polar amino acid, and it is this polarity which creates hydrogen bonds, it was decided to mutate these to the amino acid alanine, which is non-polar. This method of systematically substituting residues for alanine is known as 'alanine scanning' and is a widely used mutagenesis approach (Kristensen et al.

1997). The disruption of these hydrogen bonds was designed to weaken the affinity of Cpl-1 to choline, and potentially improve the lytic activity of the endolysin.

4.3.1.2 Production of *C. reinhardtii* lines expressing *cpl1* CBS mutants

4.3.1.2.1 Site-directed mutagenesis of *cpl1*

Of the four identified tryptophan residues, shown as part of the whole Cpl-1 amino acid sequence in Figure 4.6, none offered a more or less obvious target for mutation and therefore seven *C. reinhardtii* cell lines expressing *cpl1* with either individual tryptophan mutations or combinations of tryptophan mutations were generated.

Figure Removed

Figure 4.5 The six putative choline binding repeats of Cpl-1. Asterisks indicate strictly conserved residues, while a colon indicates conservative substitutions. Red circles highlight the mutated residues. Reproduced and annotated from Monterroso et al. (2008).

Figure 4.6 identifies the four mutated tryptophan residues in the *cpl1* gene. The seven gene mutants which were intended to be created are displayed in Figure 4.7, with each tryptophan codon (TGG) being mutated by site directed mutagenesis to an alanine codon (GCT).

MVKKNDLFVDVSSHNGYDITGILEQMGTTNTIIKISEST

TYLNPCLSAQVEQSNPIGFYHFARFGGDVAEAEREAQFF

LDNVPMQVKYLVLDYEDDPSGDAQANTNACLRFMQMIAD

AGYKPIYYSYKPFTHDNVDYQQILAQFPNSLWIAGYGLN

DGTANFEYFPSMDGIRWWQYSSNPFDKNIVLLDDEEDDK

202 209 223 230

PKTAGTWKQDSKGWWFRRNNGSFPYNKWEKIGGVWYYFD

SKGYCLTSEWLKDNEKWYYLKDNGAMATGWVLVGSEWYY

MDDSGAMVTGWVKYKNNWYYMTNERGNMVSNEFIKSGKG

WYFMNTNGELADNPSFTKEPDGLITVAYPYDVPDYA

Figure 4.6 The four tryptophan residues identified to be of importance in choline binding are highlighted in yellow in the context of the entire protein sequence. W202 and W209 form part of the p1/p2 choline binding site, and W223 and W230 form part of the p2/p3 site. The HA tag used for western blot detection is highlighted in purple.

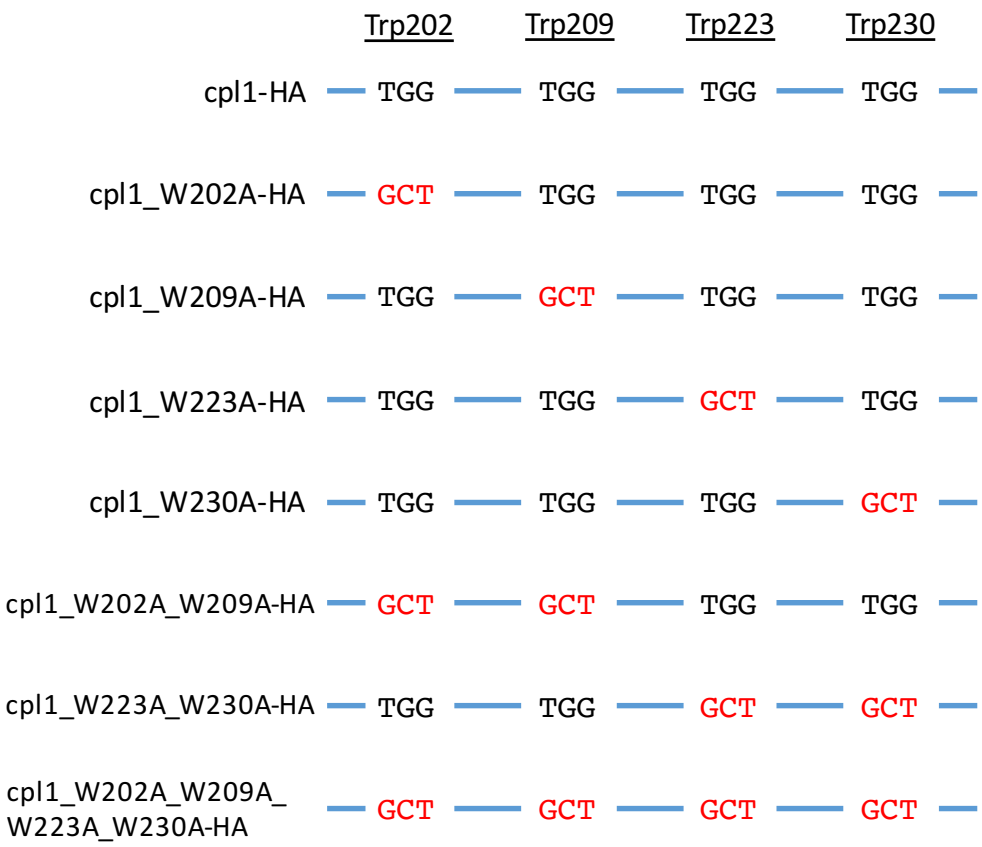


Figure 4.7 The gene *cpl1* was subjected to site directed mutagenesis to create seven unique mutants. The naming of each construct indicates the mutation location and nature, e.g. *cpl1*_W202A-HA is the *cpl1* gene with tryptophan-202 mutated to an alanine residue, and tagged with an HA tag. Multiple mutations are separated by an underscore.

A pair of primers was designed for each mutation, as displayed below, with the mutation site in lowercase, and start and stop codons underlined. cpl1.F and cpl1.R are the primers at the start and end of the gene.

cpl1.F:	GCTCTTCAATGGTTAAAAAAAATGATTTATTC
cpl1.R:	GCATGCTTATTAAGCATAATCTGGAAC
W202A_SDM.F:	GTACAgctAAACAAGATTCAAAAG
W202A_SDM.R:	CTTTTGAATCTTGTTTagcTGTAC
W209A_SDM.F:	CAAAAGGTgctTGGTTTC
W209A_SDM.R:	GAAACCAagcACCTTTTG
W223A_SDM.F:	CCATATAATAAAgctGAAAAAATTG

W223A_SDM.R: CAATTTTTTCagcTTTATTATATGG

W230A_SDM.F: GTGGTGTTgctTATTATTTCG

W230A_SDM.R: CGAAATAATAagcAACACCAC

The first four mutant genes, which have only one W→A change, were produced using site directed mutagenesis (SDM) as shown in Figure 4.8. These mutated DNA sequences were then used as template DNA in subsequent rounds of SMD PCR to produce the final three genes, which have more than one mutated codon.

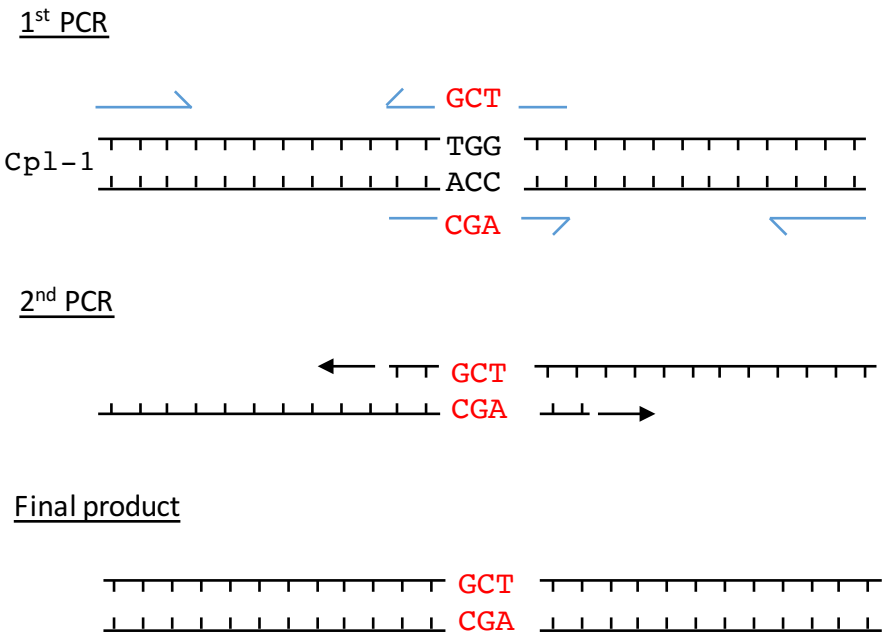


Figure 4.8 Site directed mutagenesis was performed as above to mutate the tryptophan encoding codon, TGG, to the alanine encoding one, GCT. Primers are shown in blue and PCR conditions are described in 2.3.7.

Of the seven mutants, six were successfully created. The transformation of *E. coli* with *cpl1*_W223A-HA in an intermediate vector, pJet, was repeatedly unsuccessful. It is not clear why this particular mutant was so problematic, and it is discussed further in 4.4.1, but due to time constraints it was not possible to investigate this further.

4.3.1.2.2 Expression cassette production

The resulting PCR products were digested with SapI and SphI and ligated in-frame into the pSRSapI vector. The insertion was confirmed by DNA sequencing.

4.3.1.2.3 *C. reinhardtii* transformation

The *C. reinhardtii* recipient line TN72 was transformed by the glass bead method, using the pSRSapI_*cpl1* plasmids (2.3.13). The resulting transformed lines are represented in Figure 4.9. Photosystem II functionality is restored by *psbH* and the resulting transformants were capable of growing phototrophically on minimal medium. Each of the six transformed lines was shown to reach a homoplasmic state by PCR as described in 2.3.17. As a final step, the integrity of the cassette was confirmed by DNA sequencing of the PCR-amplified plastome region, as in 2.3.18.

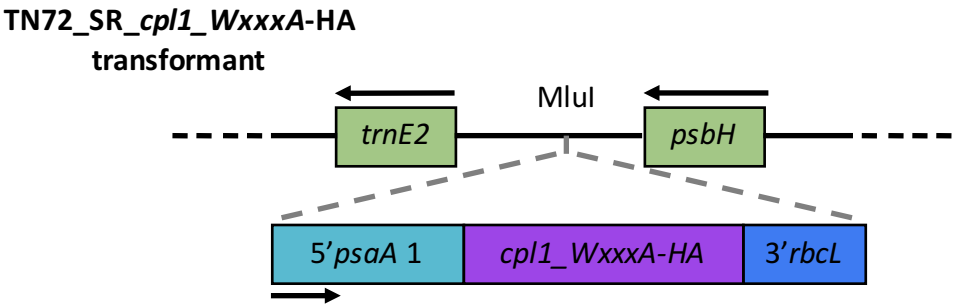


Figure 4.9 The resulting *C. reinhardtii* lines express the mutated *cpl1* gene under the endogenous *psaA* exon 1 5'UTR/promoter. The successfully transformed colonies are selected by their ability to grow phototrophically on minimal medium.

4.3.1.2.4 Western blot confirmation

Whole cell extract from each of the six *C. reinhardtii* cell lines was separated by SDS-PAGE and blotted to nitrocellulose membrane for western blot analysis. The membrane was probed with anti-HA primary antibodies followed by IRDye® secondary antibody to enable detection using the LiCor Odyssey® CLx system. An image of the western blot analysis is shown in Figure 4.10 and relative quantification of Cpl-1 accumulation is shown in Figure 4.11. The error bars in Figure 4.11 represent the standard deviation in signal detection when this blot was repeated.

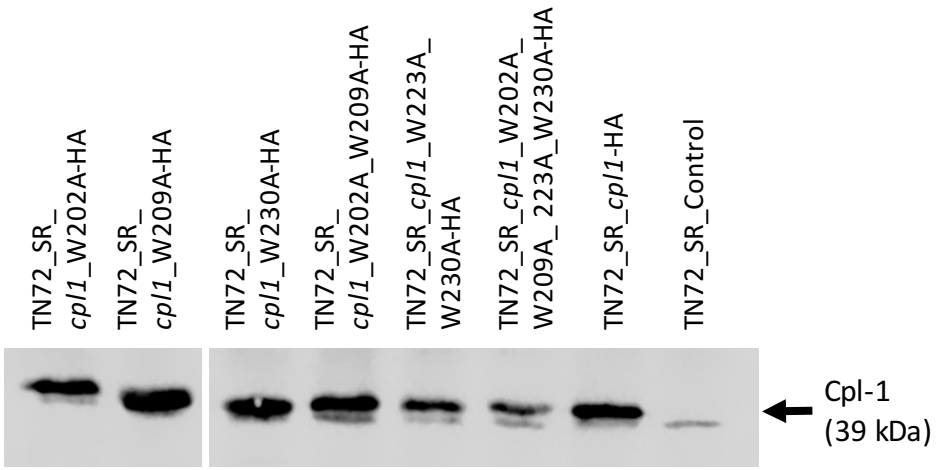


Figure 4.10 Western blot analysis of each of the six *C. reinhardtii* lines expressing mutated *cpl1*. Visible levels of accumulation are seen in all cases. TN72_SR_*cpl1*-HA and TN72_SR_Control are included as positive and negative controls, respectively.

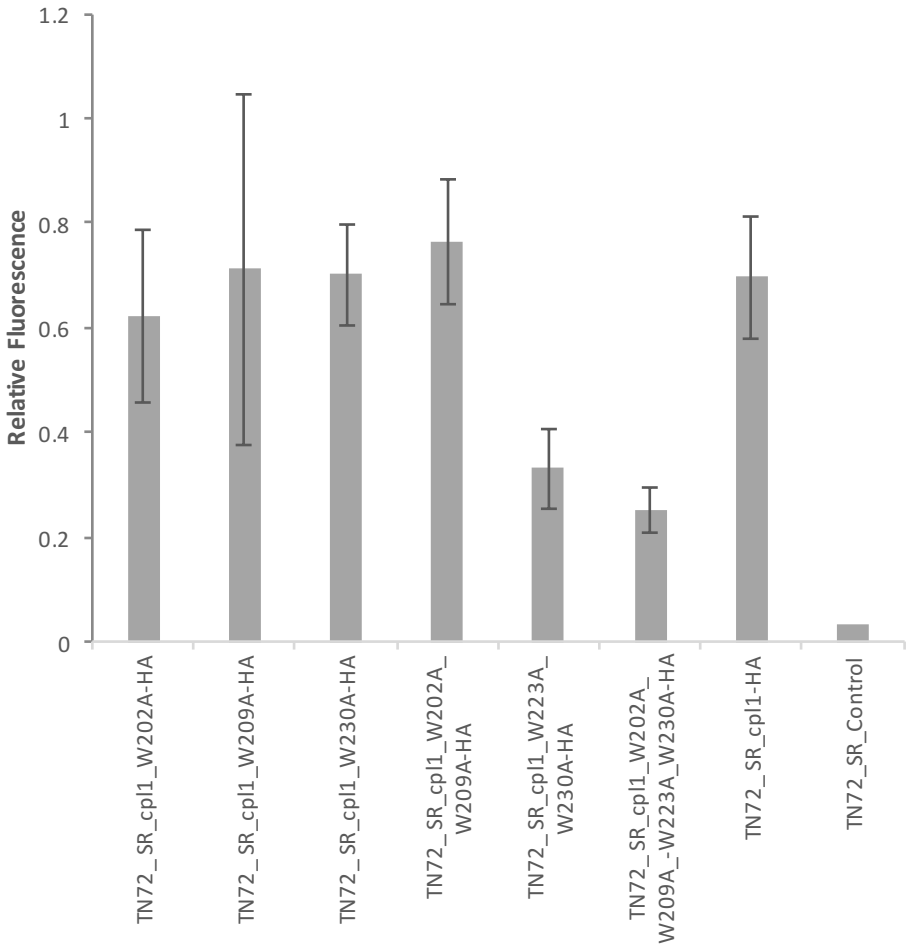


Figure 4.11 Quantification of two western blot analyses. In general, the first four cell lines produce levels of protein accumulation comparable to the unmodified version, while TN72_SR_*cpl1*_W223A_W230A-HA and TN72_SR_*cpl1*_W202A_W209A_W223A_W230A-HA accumulate to lower levels, approximately 30% of that observed for the others.

As discussed at length in 3.3.2.2.4.3, the reason for the difference in protein accumulation levels is disputable. However, for the purposes of this investigation, only the relative accumulation levels are necessary to ensure that equal quantities of protein were used in future assays.

4.3.1.3 Activity of Cpl-1 CBS mutants against *S. pneumoniae*

The six *C. reinhardtii* lines described above were created with the intention of altering Cpl-1 activity. To measure this activity several approaches were taken, namely: turbidity reduction assays and colony forming unit assays, using both crude cell extracts, and extracts enriched and concentrated by ammonium sulphate precipitation or epitope tag enrichment.

4.3.1.3.1 Activity of crude extracts – turbidity reduction assay

One of the proposed advantages of using *C. reinhardtii* as a recombinant endolysin production platform is that, due to its GRAS status, simple crude extracts could be used therapeutically, particularly for topical applications, thus removing the expensive requirement to purify the final protein.

It was therefore investigated whether crude extracts of the six Cpl-1 CBS mutants described in this section display different levels of lytic activity towards *S. pneumoniae*, when compared to the activity of the un-mutated Cpl-1 protein.

Crude extracts were prepared as described in 2.4.1. In brief, the six *C. reinhardtii* cell lines were grown simultaneously and their OD_{740} measured after 72 hours. The cultures were centrifuged and the pellet resuspended in 20 mM NaPi buffer to give suspensions of equivalent cell density measured at OD_{740}. The cells were lysed by freeze-thaw and ultracentrifuged to remove the cell debris. The supernatant was stored at 4 °C and this was considered the "crude extract". Fresh *S. pneumoniae* cultures were prepared as described in 2.7.1.1 and resuspended in 20 mM NaPi buffer to an OD_{595} = ~0.8.

In a 96-well plate, *S. pneumoniae* cells and *C. reinhardtii* crude extract were combined and the change in OD_{595} measured with a plate reader over a period of 150 minutes at 37 °C. 200 µl of resuspended *S. pneumoniae* cells were incubated with ~50 µl of

crude extract. The precise volume was adjusted according to data shown in Figure 4.11, to ensure an equal quantity of Cpl-1 protein was included in each well. Crude TN72_SR_Control was used as a negative control.

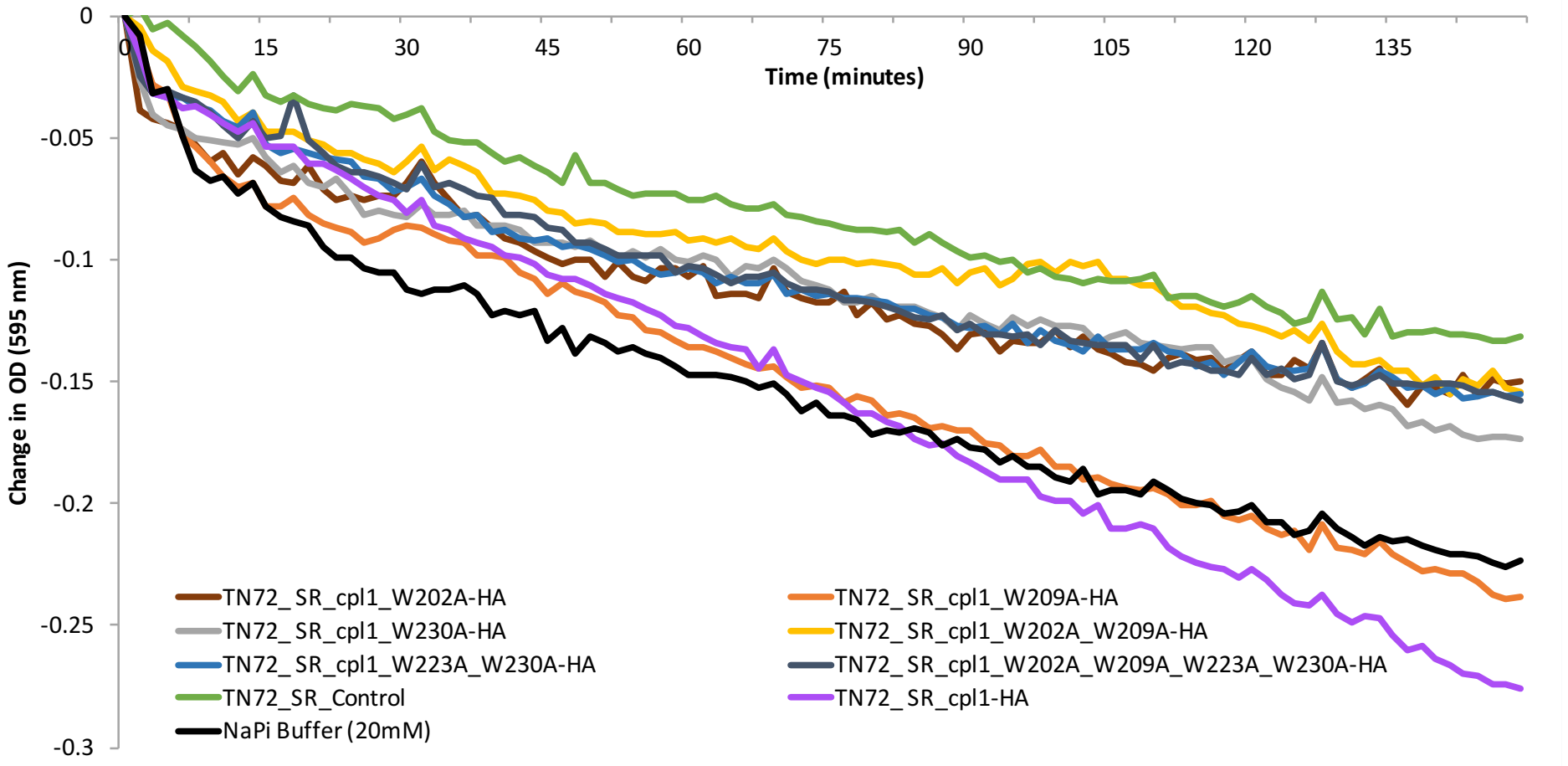


Figure 4.12 Turbidity reduction assay of *S. pneumoniae* when challenged with crude *C. reinhardtii* extracts containing different Cpl-1 CBS mutants. Change in OD_{595} is shown rather than measured OD_{595} due to the different volumes of crude extract in each well. Each combination was performed in triplicate and the average is shown.

From Figure 4.12, it appears that the un-modified TN72_SR_*cpl1*-HA is the most active endolysin, while all the mutated endolysins showed activity intermediate between this and the negative control, TN72_SR_Control. The high level of lysis observed by the buffer alone is surprising. However, it has been noted in previous endolysin activity assays using crude extracts that an increase in osmotic pressure due to the presence of large quantities of protein in the suspension can lead to an inhibition of cell lysis (Mayer et al. 2010) - this is exemplified here by the high level of autolysis observed when NaPi buffer (20mM) is added, compared to that seen in the negative control.

First inspection of the data shown in Figure 4.12 suggests that altering the CBD of Cpl-1 has a detrimental effect upon enzymatic activity. While this result is not entirely unexpected, given the issue of osmotic pressure affecting bacterial cell lysis, combined with the different levels of protein accumulation observed in Figure 4.11, it was decided that turbidity reduction assays with crude extracts are insufficiently precise to enable a comparison of the various Cpl-1 CBS mutant activities.

4.3.1.3.2 Enrichment of CBS mutant Cpl-1 by ammonium sulphate precipitation

In order to address the above concerns, ammonium sulphate precipitation was used to partially purify and concentrate the Cpl-1 CBS mutants, providing an enriched crude extract. The enriched extracts were separated by SDS-PAGE, blotted to nitrocellulose membrane and probed with antibodies for western analysis. The resulting analysis and relative quantification are displayed in Figure 4.13.

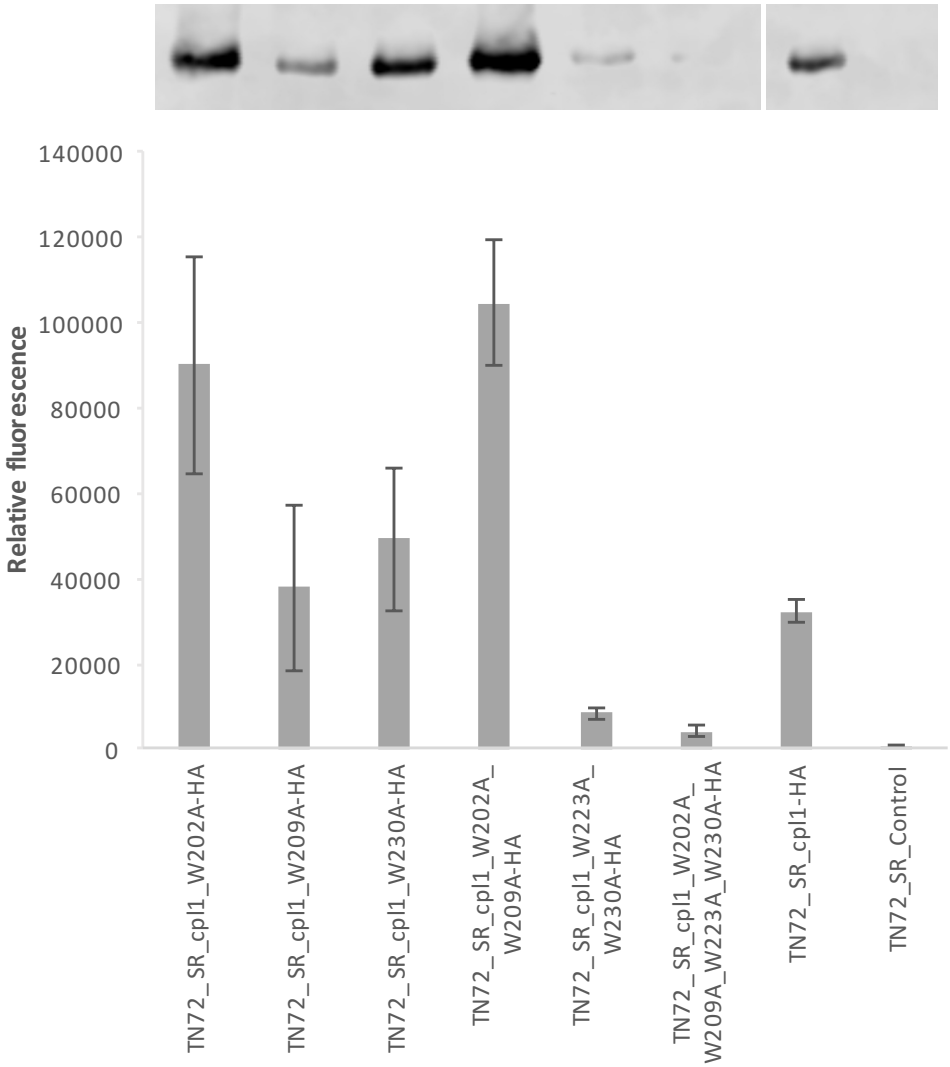


Figure 4.13 The western blot analysis of six Cpl-1 CBS mutants concentrated and enriched by ammonium sulphate precipitation, alongside positive and negative controls (TN72_SR_*cpl1*-HA and TN72_SR_Control, respectively). Image Studio Lite 5.0 software and LiCor Odyssey® CLx system enabled the quantification of each band, shown below each lane as a bar chart.

4.3.1.3.3 Activity of extracts enriched by ammonium sulphate precipitation - Colony forming unit assay

The turbidity reduction assay shown in Figure 4.12 required fresh *S. pneumoniae* cultures to be grown on the morning of each assay, which often leads to difficulties in the reproducibility of the assay. On the other hand, colony forming unit assays, as described in 2.7.1.1, were performed using snap-frozen aliquots of *S. pneumoniae* suspensions, thus enabling a much more reproducible and robust assay. The number of colony forming units was counted by eye and when an uncountable lawn of indistinguishable colonies was present, this was counted as "100". Serial dilutions of

bacterial suspension containing endolysin were performed after one hour of incubation at 37 °C, and three 10 µl drops of each dilution were spotted onto blood agar plates. The mean count across the three spots was calculated. The data from Figure 4.13 enabled similar quantities of HA-tagged protein to be added to the bacterial suspension.

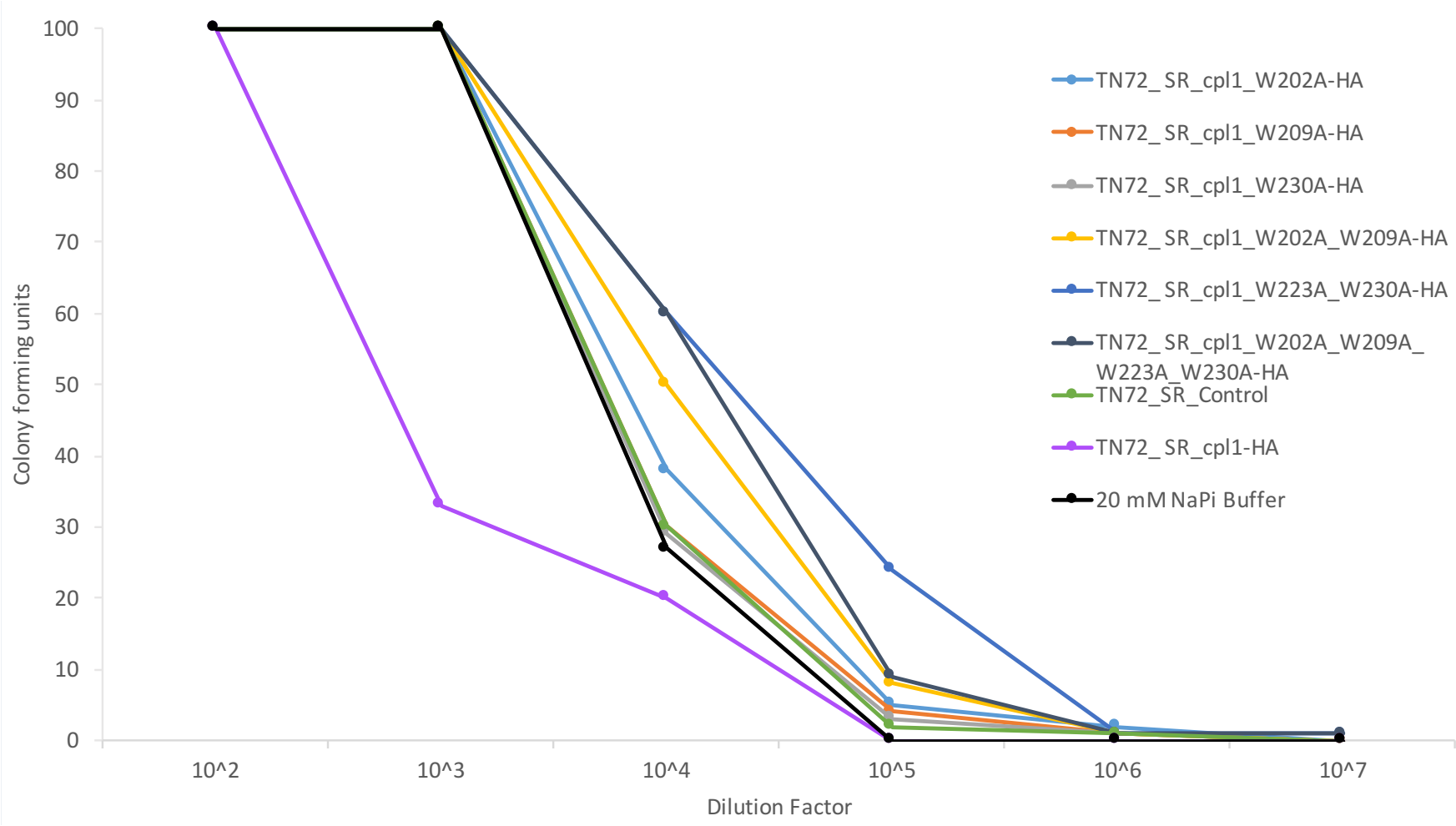


Figure 4.14 Colony forming unit assay of *S. pneumoniae* cells treated with Cpl-1 CBS mutant ammonium sulphate-enriched crude extracts.

The colony forming unit assay in Figure 4.14 confirms that TN72_SR_*cpl1*-HA is more active than any of the Cpl-1 CBS mutants, however there still exists the problem of varying osmotic pressure due to the varying concentrations of protein present. While equal quantities of endolysin have been achieved, this has probably resulted in a different level of total protein, because ammonium sulphate precipitation enables the enrichment of all proteins within a range of solubility, which undoubtedly includes many other proteins. In order to include similar quantities of the HA-tagged Cpl-1 CBS mutant protein, differing volumes of the ammonium sulphate enriched extract were used, thus altering the osmotic pressure on the bacterial cells in the suspension. In this case too, it is worth noting that the 20 mM NaPi buffer is causing lysis to a greater extent than all of the extracts except for the unmodified TN72_SR_*cpl1*-HA.

4.3.1.3.4 Protein enrichment via epitope tags – HA tags

All Cpl-1 CBS mutants were tagged with an HA epitope tag for detection in western blot analysis, but it is also possible to use this epitope to purify the protein. Each of the six Cpl-1 CBS mutants was purified using an anti-HA resin, and the fractions were spotted onto a nitrocellulose membrane for a dot blot, as described in 2.4.3. After probing with anti-HA antibodies and IR secondary antibody, the membrane was analysed using the LiCor Odyssey® CLx system, shown in Figure 4.15

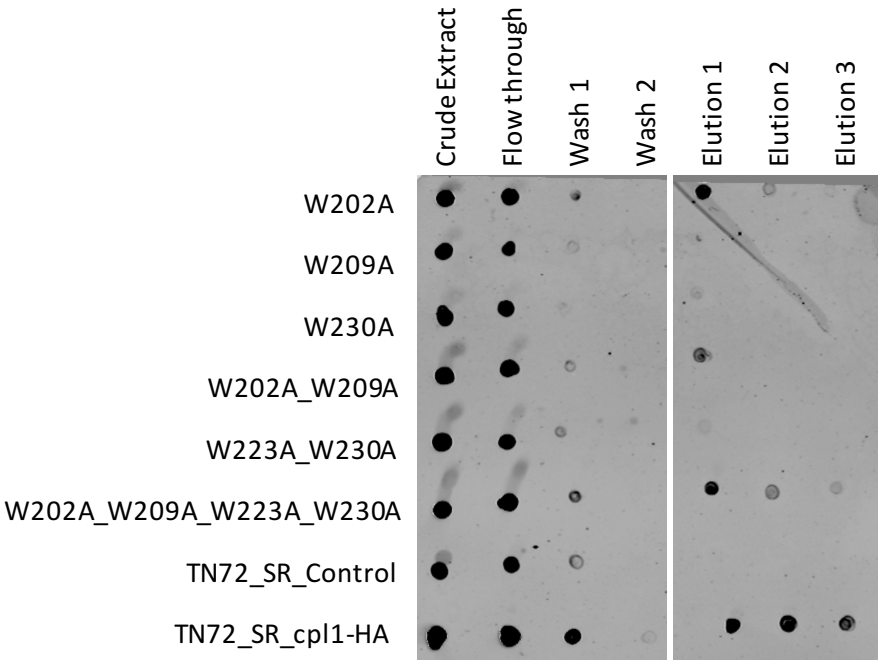


Figure 4.15 Dot blot analysis of HA-resin purification fractions. The *C. reinhardtii* lines were abbreviated to just the mutation for clarity e.g. 'TN72_SR_*cpl1*_W202A-HA' is just labelled 'W202A'. Mixed results were achieved, with W202A, W202A_W209A, W202A_W209A_W223A_W230A, and the un-mutated Cpl-1 being the only strains where the elution had visible quantities of HA-tagged protein.

It is noteworthy that TN72_SR_Control gives a strong signal for the crude extract and flow-through, which suggests that there are background proteins fluorescing and disguising the HA-tagged proteins in these fractions. However, these background proteins are not present in the elution fractions. To confirm that the fluorescing protein is indeed the endolysin, the most promising examples from the dot blot were separated by SDS-PAGE. Western blot analysis allows relative quantification of the yield achieved with this purification method, shown in Figure 4.16.

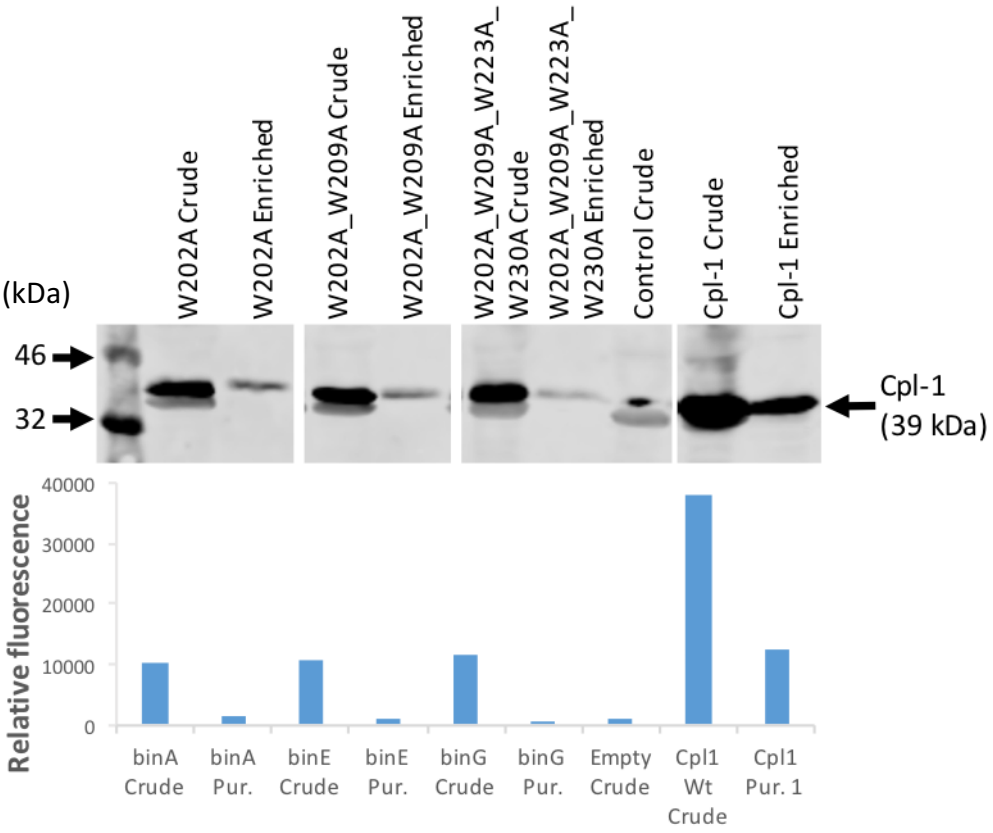


Figure 4.16 Western blot analysis of both crude *C. reinhardtii* extracts and then the same extract enriched using the HA tag. Only mutants which have visible enriched bands are shown. The bands were quantified used the LiCor Odyssey® CLx system and the results are shown in the bar chart. The percentage of recovered protein ranged from 33% for the unmodified Cpl-1 to only 6% recovery for the Cpl1-W202A_W209A_W223A_W230A-HA protein.

Only three of the Cpl-1 CBS mutants achieved sufficient levels of protein accumulation to enable a colony forming unit assay to be performed.

The colony forming unit assay was performed using the same batch of frozen *S. pneumoniae* cells as in Figure 4.14. The *C. reinhardtii* crude extracts and the HA-purified elutions were diluted such that they contained equal concentrations of HA-tagged protein before being used. The method can be found in 2.7.1.1. The number of colony forming units was counted by eye and when an uncountable lawn of indistinguishable colonies was present, this was counted as 100. Serial dilutions of bacterial suspension and endolysin were performed after one hour of incubation at 37 °C and three 10 µl drops of each dilution were spotted onto blood agar plates. The count was then averaged across the three spots.

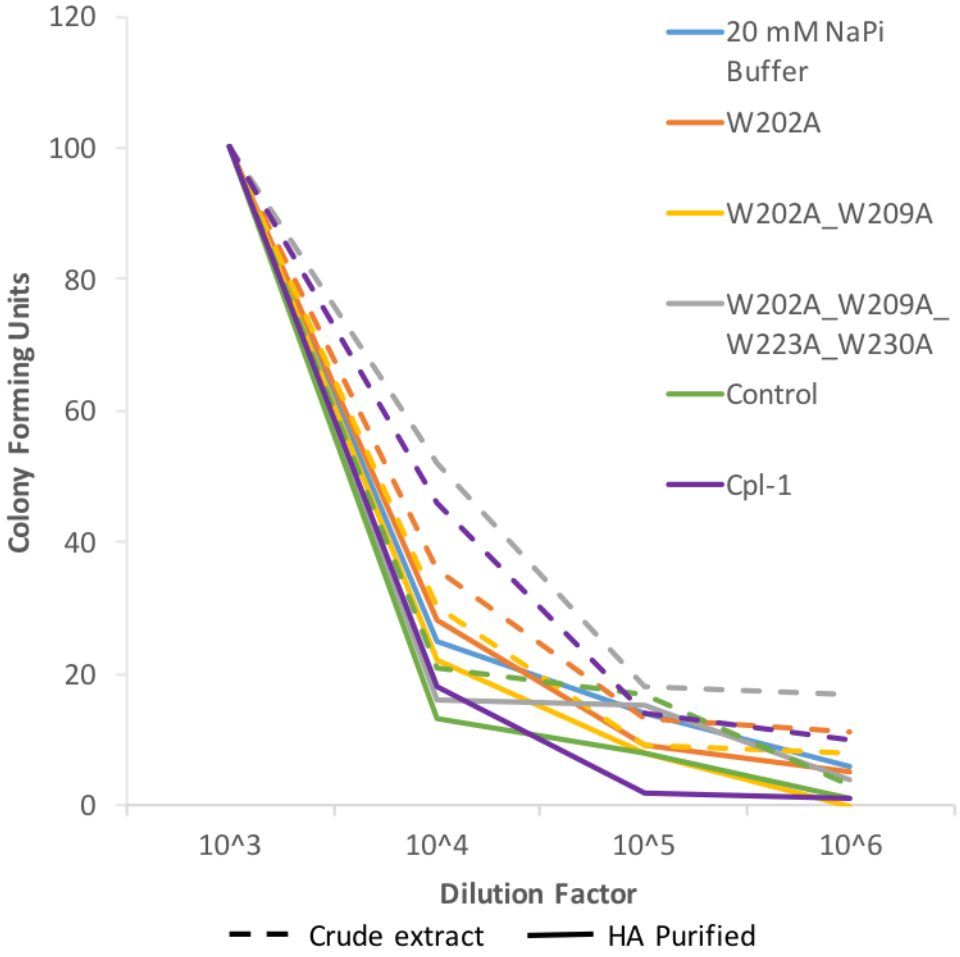


Figure 4.17 Colony forming unit assay of *S. pneumoniae* cells treated with Cpl-1 CBS mutant crude extracts (dotted line) and the HA-resin purified elution (solid line).

Generally, the crude extracts are less active than the purified extracts – although the purified TN72_SR_Control appears to achieve the second highest level of antibacterial activity. The buffer control sits nearly in the middle of the crude extract and purified extract activities, which lends weight to the argument that osmotic pressure is reducing cell lysis in the crude extracts. The un-mutated Cpl-1 protein has the greatest antibacterial effect, while the Cpl-1 CBS mutants are no more effective than the purified TN72_SR_Control, which expresses no recombinant protein, suggesting a complete loss of activity in the mutant proteins. Given the poor results achieved using anti-HA purification resin, enabling us to only test three of the six CBS mutants, a different system is required.

4.3.1.3.5 Protein enrichment via epitope tags – StrepII tags

The HA epitope has been widely used in the Purton group, as well as other algal research groups. The original expression of *cpl1* in *C. reinhardtii* encoded an HA epitope tag on its C-terminus (Taunt 2013), and this was continued into this investigation. However, while convenient for western blot detection, using the HA-tag for protein purification is expensive compared to some other epitopes, as it involves affinity purification using an anti-HA monoclonal antibody. The StrepII tag was chosen as an alternative as it has been used in the algal field to purify native chloroplast proteins from *C. reinhardtii* (Derrien et al. 2012; Derrien 2013). Furthermore, as the system is based on the affinity of biotin and streptavidin, purification columns are relatively cheap. 3.3.2.2.3 describes the production of the *C. reinhardtii* line TN72_SR_*cpl1*-StrepII, which is used again here to investigate the StrepII-tag for protein purification and enrichment. Polyhistidine tags (His-tags) are the most common epitope tag used for this purpose but a StrepII tag was chosen in this instance due to its previous success in the *C. reinhardtii* chloroplast, as well as the fact that it can be readily used for detection in western blotting, while antibodies to His tags are more problematic.

4.3.1.3.6 Enrichment of Cpl1-StrepII using HiTrap StrepII Column

C. reinhardtii cell extracts were prepared as described in 2.5.1, and ultracentrifuged at 100,000 g for 1 hour to remove cell debris. This ultracentrifuged (UC) extract was then enriched for StrepII-tagged proteins using the commercially available HiTrap StrepII Column (GE Lifesciences) and eluted with desthiobiotin, as described in 2.5.3.

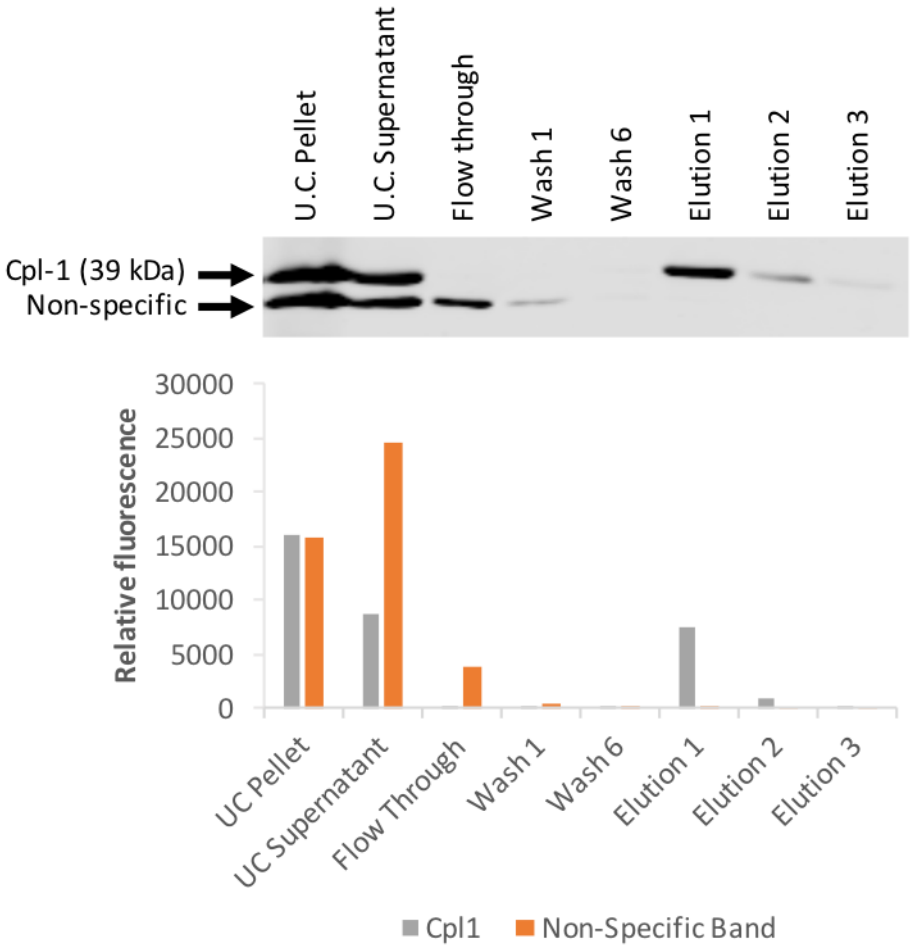


Figure 4.18 Samples from extract preparation and the HiTrap StrepII column were collected and separated by SDS-PAGE. Western blot analysis and relative quantification using the LiCor Odyssey® CLx system enabled the StrepII-tagged protein recovery to be estimated as a percentage of the ultracentrifuged (UC) extract passed through the column.

Approximately one-third of the total protein remains in the supernatant on ultracentrifugation, but virtually 100% of this was recovered using the HiTrap StrepII column. Further optimisation of ultracentrifugation speeds and durations could increase the proportion remaining in the supernatant further, but for this investigation maximising total protein yield is not the aim.

As a negative control, HiTrap StrepII purification was also performed with an extract of TN72_SR_Control. This controls for any *C. reinhardtii* cell components that may have non-specifically bound to the column and be present in the eluted fraction.

4.3.1.3.7 Activity of purified Cpl1-StrepII

A *S. pneumoniae* colony forming unit assay was performed as described previously (4.3.1.3.4), using the eluted fraction, described here as an "enriched extract", from both TN72_SR_*cpl1*-StrepII and TN72_SR_Control. Bacterial suspension and extract were combined in a 1:1 ratio and serial dilutions of bacterial suspension and endolysin were performed after one hour of incubation at 37 °C, whereupon three 10 µl drops of each dilution were spotted onto blood agar plates.

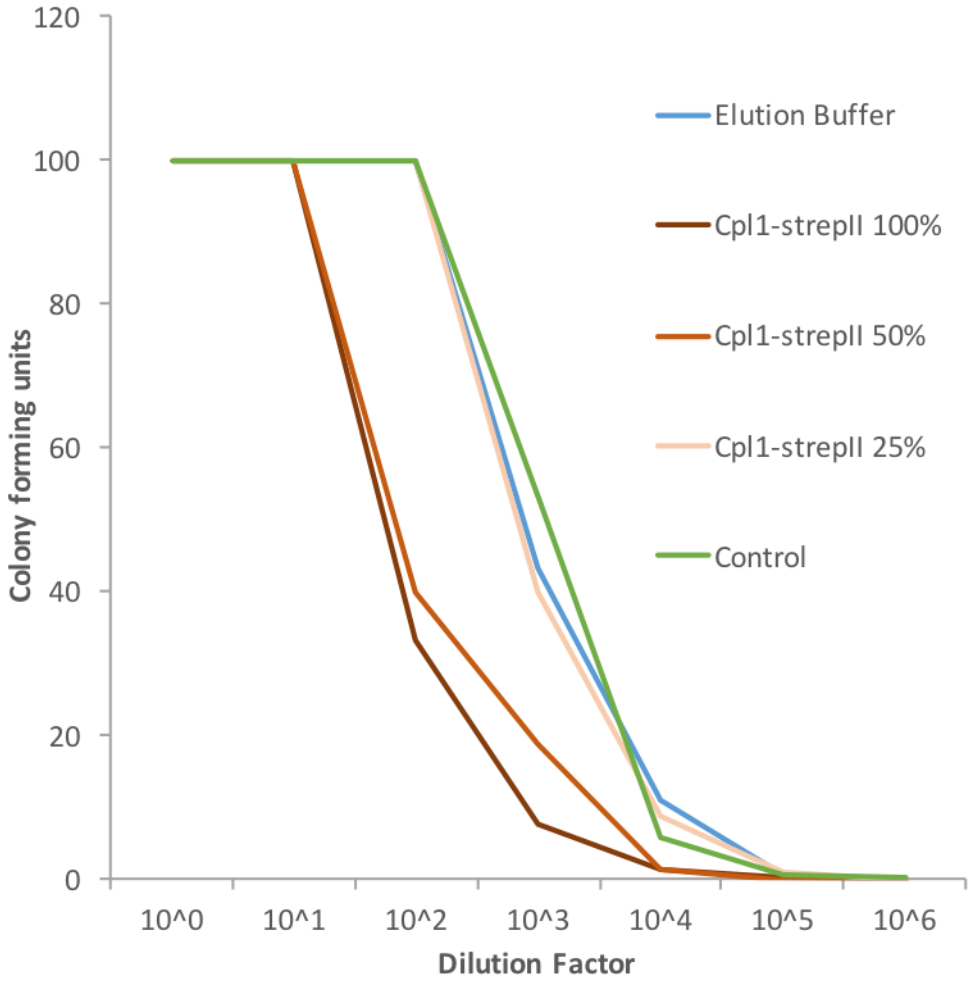


Figure 4.19 Colony forming unit assay of *S. pneumoniae* cells treated with Cpl1-StrepII enriched extracts. "Control" refers to a purified extract from TN72_SR_Control. The number of colony forming units was counted by eye and when an uncountable lawn of indistinguishable colonies was present, this was counted as 100. The count was then averaged across the three spots. Three dilutions of the enriched extract were tested (100 %, 75 % and 25 %), and the percentage indicates the initial concentration, not the final concentration (i.e. prior to mixing with the bacterial suspension).

The colony forming unit assay in Figure 4.19 indicates a measurable difference between the three Cpl-1 dilutions, and they are clearly more active than the extract from TN72_SR_Control at the two higher concentrations. These data suggest that this method of enrichment removes both the mild antibacterial activity of the *C. reinhardtii* crude extract and the osmotic pressure effect that seemed to be masking results in previous trials.

Unfortunately, time limitations prevented the application of this more robust assay to the investigation of binding site mutation. By reproducing each of the *C. reinhardtii* lines expressing CBS mutants, this time encoding a StrepII epitope instead of an HA epitope, it would be possible to more accurately compare the relative activities of the mutants. However, given the problems encountered above and the longer transformation time of *C. reinhardtii* lines, in this particular case investigations could be more easily carried out using recombinant proteins produced in *E. coli*. *E. coli* is an excellent platform for rapidly producing recombinant proteins to a high level and in this case would allow easy comparison of different mutants. The best of these mutants, if indeed cell wall binding site mutagenesis can improve endolysin activity, could then be expressed in *C. reinhardtii* for the final therapeutic drug production. In addition, proteins toxic to *E. coli*, such as the endolysin SPN9CC, which can be expressed in *C. reinhardtii* (Young and Purton 2015) could be investigated using this method.

4.3.2 Improving Cpl-1 stability through dimerisation

The *C. reinhardtii* chloroplast is known to produce and regulate the formation of disulphide bonds (Kim 1997), and research suggests that yields of proteins containing disulphide bonds can be improved in the *C. reinhardtii* chloroplast by the addition of selenocystamine to the culture medium (Ferreira-Camargo et al. 2015). Furthermore, Resch et al. (2011) showed that Cpl-1 can be modified to "pre-dimerise", thus decreasing plasma clearance and improving its clinical potential.

Cpl-1 is stable as a monomer, but Cpl-1 dimerisation occurs naturally on choline binding (Buey et al. 2007). We expect that the addition of a disulphide bridge between two Cpl-1 monomers will reduce the rate of degradation in storage by limiting both protease access and the denaturing of the protein, as well as achieving the decreased plasma clearance reported by Resch et al. (2011). The aim is also to show that the *C. reinhardtii* chloroplast is a suitable platform for producing disulphide-bonded endolysins as therapeutics.

4.3.2.1.1 Stability of Cpl-1 under varying conditions

An initial investigation into the stability of Cpl-1 in storage was conducted. An extract from the *C. reinhardtii* line TN72_SR_*cpl1*-HA was prepared as described in 2.5.1. The cells were either resuspended and lysed in 20 mM NaPi buffer or 20 mM NaPi + Protease Inhibitor buffer, then both were ultracentrifuged for one hour at 100,000 g. Each supernatant was then split into two tubes and stored at 25 °C or 4 °C for seven days. The four tubes were sampled at 24-hour intervals and the samples were snap-frozen in liquid nitrogen and stored at -80 °C. At the end of the seven-day period, all samples were thawed and separated by SDS-PAGE before being blotted to nitrocellulose membrane and analysed by western blotting. Relative quantification using the LiCor Odyssey® CLx system produced the data displayed in Figure 4.20.

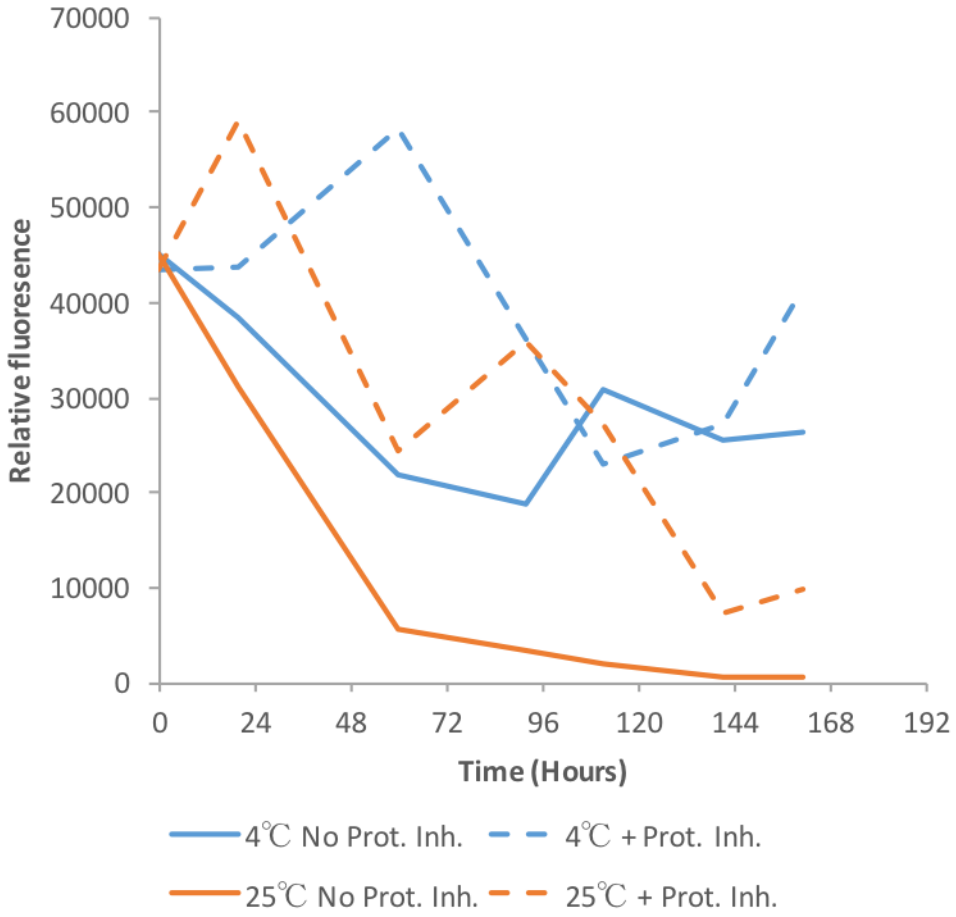


Figure 4.20 Cpl-1 protein degradation over a seven-day period when stored at 4 °C or 25 °C with and without protease inhibitor.

The data displayed in Figure 4.20 appears very erratic as this was only performed once, but in general it appears that the protein is best conserved at 4 °C and with protease inhibitor, which was to be expected. This has established a benchmark for the stability of monomeric Cpl-1 when stored under lab conditions. If Cpl-1 dimerisation can be achieved, then this stability may be improved, thus enabling the addition of less protease inhibitor which would reduce costs and simplify the therapeutic regulatory procedure.

4.3.2.2 Production of a *C. reinhardtii* line expressing *cpl-1* dimer

The pre-dimerised Cpl-1 protein produced by Resch et al. (2011) was achieved through site-directed mutagenesis, and their experiments are briefly summarised here: site-directed mutagenesis firstly changed a cysteine residue near the N-terminus (C45) to a serine (see Figure 4.21). This first step removed the possibility of

interactions with this solvent-accessible cysteine. Then an aspartic acid residue (D324), in a position 13 amino acids away from the C-terminus, was mutated to a cysteine. The 13 final amino acids are highly conserved and, in the case of the homologous lysin LytA, responsible for LytA dimerisation, so were not modified.

4.3.2.2.1 Site-directed mutagenesis of *cpl1*

Site directed mutagenesis was performed as described in 2.3.7. The primers used in this instance are below, and mutation sites are in lowercase:

cpl1.F:	GCTCTTCAATGGTTAAAAAAAATGATTTATTC
cpl1_C45S.F	ACATATTTAAATCCATcaTTATCAGCTC
cpl1_C45S.R	GAGCTGATAAtgATGGATTTAAATATGT
cpl1_D324C.F	GTGAATTAGCAtgtAATCCTTCATTT
cpl1_D324C.R	AAATGAAGGATTacaTGCTAATTCAC
cpl1.R:	GCATGCTTATTAAGCATAATCTGGAAC

MVKKNDLFVDVSSHNGYDITGILEQMGTTNTIIKISEST

45

TYLNPCLSAQVEQSNPIGFYHFARFGGDVAEAEREAQFF

LDNVPMQVKYLVLDYEDDPSGDAQANTNACLRFMQMIAD

AGYKPIYYSYKPFTHDNVDYQQILAQFPNSLWIAGYGLN

DGTANFEYFPSMDGIRWWQYSSNPFDKNIVLLDDEEDDK

PKTAGTWKQDSKGWWFRRNNGSFPYNKWEKIGGVWYYFD

SKGYCLTSEWLKDNEKWYYLKDNGAMATGWVLVGSEWYY

MDDSGAMVTGWVKYKNNWYYMTNERGNMVSNEFIKSGKG

324

WYFMNTNGELADNPSFTKEPDGLITVAYPYDVPDYA

Figure 4.21 The two amino acids selected for mutation are highlighted in yellow. C45 was mutated to a serine to prevent unwanted interactions, as this cysteine is believed to be solvent-accessible. D324 was mutated to a cysteine to form a disulphide bridge in a homodimer. The HA tag is highlighted in purple.

4.3.2.2.2 Expression cassette production

The final PCR product, named cpl1_dimer-HA, was inserted into the pSRSapI vector using SapI and SphI as previously described (4.3.1.2.2). Competent DH5α *E. coli* were then transformed with the resultant plasmid and DNA sequencing confirmed the correct insertion and that the intended mutations had been successfully introduced.

4.3.2.2.3 *C. reinhardtii* transformation

The TN72 recipient line was transformed with the pSRSapI_*cpl1_dimer*-HA transformation plasmid. Transformed *C. reinhardtii* lines, named TN72_SR_*cpl1_dimer*-HA were checked for homoplasmicity by PCR (Appendix I) and finally confirmed by sequencing.

4.3.2.2.4 Western blot confirmation

Whole cell extracts of TN72_SR_*cpl1*-HA and TN72_SR_Control (as controls) and TN72_SR_*cpl1_dimer*-HA were prepared, both with and without β-mercaptoethanol (BME), and separated by SDS-PAGE. The BME reduces disulphide bonds so was included to test if the dimerised Cpl-1 could be detected when BME was absent but not when present – this would confirm disulphide presence. After blotting to nitrocellulose membrane and blocking as described in 2.4.5, the membrane was probed with anti-HA primary antibody, then incubated with IRDye® secondary antibody. Fluorescence was observed using the LiCor Odyssey® CLx system. The resulting analysis is shown in Figure 4.22.

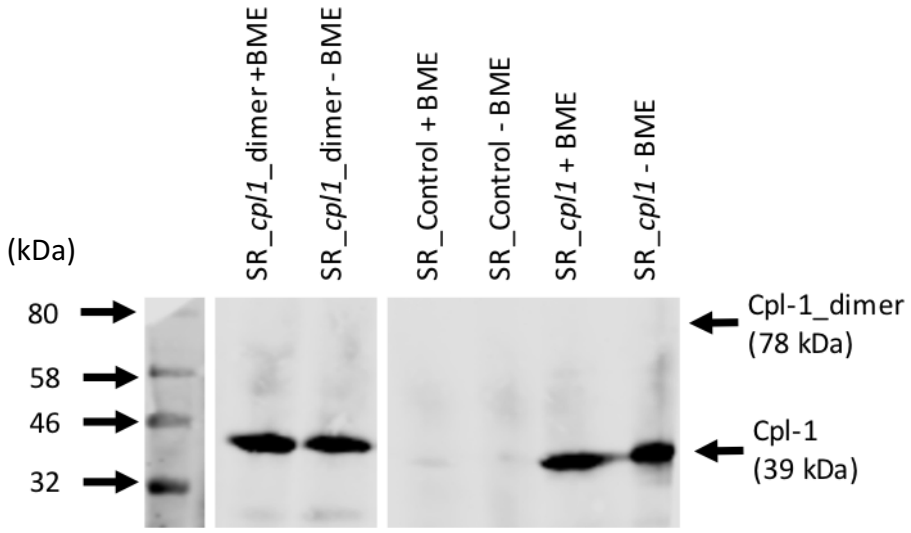


Figure 4.22 Western blot analysis of TN72_SR_*cpl1_dimer*-HA. The loading mixture either contained β-mercaptoethanol or did not, indicated by +/- BME. The predicted band location for Cpl-1_dimer is indicated but no bands were observed.

No spontaneously forming Cpl-1 dimers were observed in the whole cell extract. However, monomeric Cpl-1 still accumulates to similar levels as seen in TN72_SR_*cpl1*-HA. The presence of the monomer suggests that there is no issue with synthesis, but for some reason dimerisation is not occurring, perhaps due to steric issues involving the C-terminal HA tag, or simply that, at the concentrations observed, the equilibrium is shifted towards monomers and dimers are not formed.

4.3.2.2.5 Addition of selenocystamine

The addition of the small molecule selenocystamine to the growth medium of *C. reinhardtii* was reported to catalyse disulphide bond formation in recombinant proteins (Ferreira-Camargo et al. 2015).

Selenocystamine was added to *C. reinhardtii* cultures to a final concentration of 2 µM, the optimal concentration observed by Ferreira-Camargo et al. (2015). Whole cell extracts of TN72_SR_*cpl1*-HA and TN72_SR_Control (as controls) and TN72_SR_*cpl1_dimer*-HA, all grown both with and without 2 µM selenocystamine were prepared, without β-mercaptoethanol (BME), and separated by SDS-PAGE. The western blot analysis, carried out as previously, is displayed in Figure 4.23.

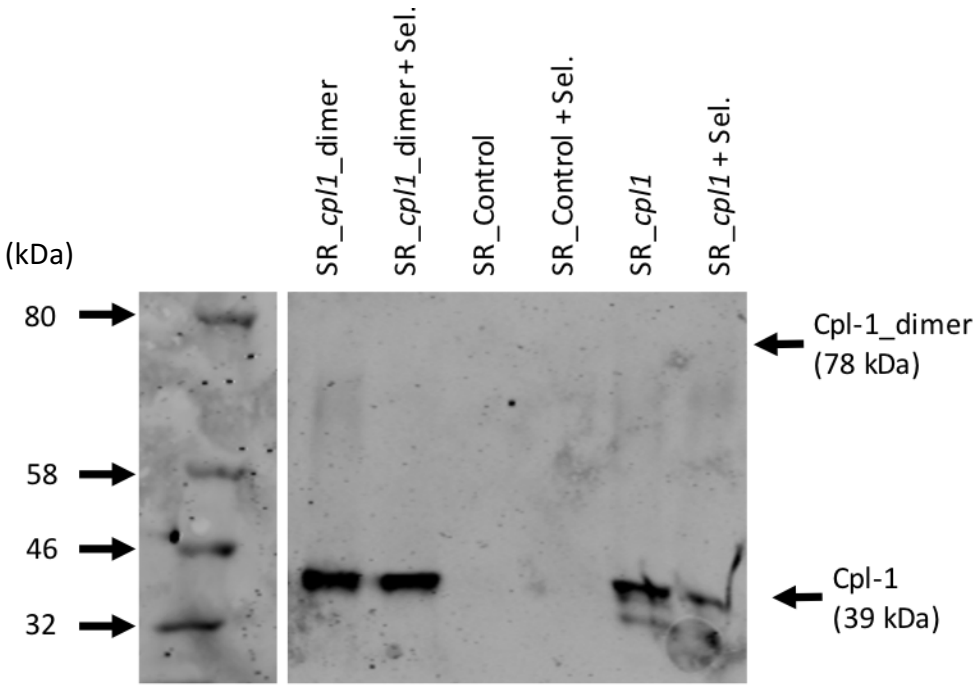


Figure 4.23 Western blot analysis of whole cell *C. reinhardtii* extracts, grown with selenocystamine (+ Sel.) and without. All were prepared in the absence of β-mercaptoethanol and the predicted Cpl-1 dimer band location indicated.

The addition of selenocystamine to the growth medium of TN72_SR_*cpl1_dimer*-HA failed to catalyse spontaneous Cpl-1 dimerisation, nor did it significantly affect levels of monomeric Cpl-1 accumulation.

4.3.2.2.6 Addition of choline

The investigation by Resch et al. (2011) used DEAE-Sepharose™ affinity chromatography to purify Cpl-1 from *E. coli* extracts. DEAE acts as a choline analogue and therefore this method exploits the choline binding ability of Cpl-1. Cpl-1 was then eluted from the column using a buffer containing 10 % choline. However, no crude lysate analysis was performed by Resch et al. (2011) and only the purified Cpl-1 was shown on SDS-PAGE. While it is most likely that in the experiments by Resch et al. (2011) Cpl-1 dimerisation occurred *in vivo* in the *E. coli* periplasm, it is possible that the dimerisation occurred *in vitro* when the addition of free choline recruited Cpl-1 monomers together, enabling spontaneous *in vitro* dimerisation.

To test this hypothesis, free choline was added to cell extract containing Cpl-1_dimer, at varying concentrations, and incubated for one hour. The samples were separated by non-reducing SDS-PAGE and analysed by western blotting, displayed in Figure 4.24.

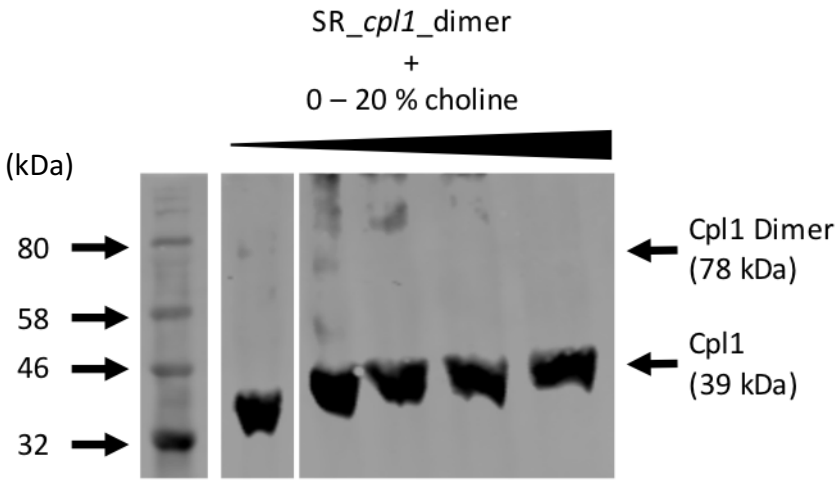


Figure 4.24 Western blot analysis of whole cell *C. reinhardtii* extracts, incubated with choline. The first band has no choline added. All were prepared in the absence of β-mercaptoethanol and the predicted Cpl-1 dimer band location indicated.

Incubation with choline failed to induce Cpl-1 dimerisation. The presence of choline was expected to recruit Cpl-1 towards the free choline and enable the formation of a disulphide bond between the two monomers. The native dimerisation of Cpl-1 upon

the addition of choline reported by Monterroso et al. (2005) was not expected to be observed, given the denaturing conditions of SDS-PAGE.

4.3.2.2.7 Production of Cpl-1_dimer in *E. coli* and enrichment

As shown in Figure 3.8, the *E. coli* line DH5α_pProEX_*cpl1_dimer*-HA-StrepII was shown to express Cpl-1 to a high level, and the application of HiTrap StrepTactin purification columns enabled the production of an enriched extract. This was performed to reproduce the work of Resch et al. (2011), as well as to act as a protein standard as described in 3.3.1.2.6. When separated by SDS-PAGE under non-reducing conditions, no dimerisation could be observed, however this was to be expected as disulphide bond formation requires an oxidising environment in order to oxidise a pair of cysteines. In Gram-negative prokaryotes, such as *E. coli*, this occurs in the periplasmic space, where enzymes catalyse the reaction. Recombinant proteins therefore require an N-terminal signal peptide to target them to the periplasm for disulphide bond formation, which was not included on our *cpl1_dimer* transgene. Resch et al. (2011) expressed *cpl1* using a plasmid based upon pIN-IIIA, which attaches an outer membrane protein A (OmpA) signal peptide to the N-terminus of cloned proteins to promote secretion of the recombinant protein into the periplasm. Given that the plasmid used here, pProEX_HTb, does not attach an N-terminal signal peptide, it was necessary to perform dimerisation *in vitro*.

4.3.2.2.8 Dimerisation of *E. coli*-produced Cpl-1 using oxidised glutathione *in vitro*

Disulphide bond formation can occur spontaneously if the environment is oxidising, and can therefore be formed *in vitro* (Berkmen 2012). Glutathione is a commonly used reagent in its reduced (GSH) and oxidised (GSSH) form to produce a redox environment beneficial to disulphide bond formation. The protocol is described fully in 2.6.5, adapted from Hidaka et al. (1998) and Hyvonen (2002).

A crude cell lysate of *E. coli* cells induced to express *cpl1_dimer* was obtained by sonication, and ultracentrifugation removed the insoluble fraction. An existing Cpl-1_StrepII enriched extract from *C. reinhardtii* was used as a control (from 4.3.1.3.6). The soluble fraction and the control were then incubated overnight at 20 °C with varying GSH:GSSG ratios. SDS-PAGE was used to separate the fractions, with no β-

mercaptoethanol and no heat denaturation prior to running the gel. The resulting western blot analysis is shown in Figure 4.25.

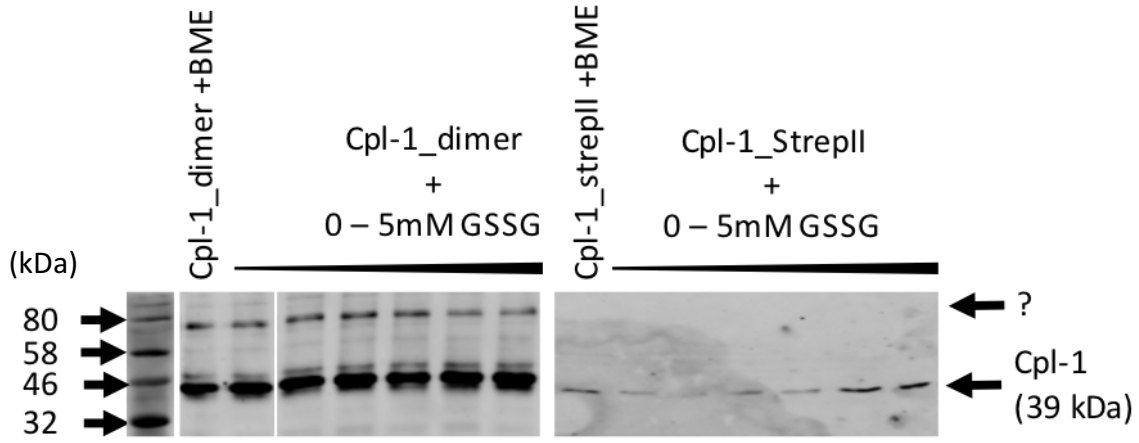


Figure 4.25 Western blot analysis attempting to detect *in vitro* dimerisation of Cpl-1_dimer, with Cpl-1_StrepII being used as a negative control. The Cpl-1_dimer runs slightly larger than the Cpl-1_StrepII due to the additional epitope tags and slightly different charge. The sample in the first lane of each blot was boiled for 3 minutes in the presence of β-mercaptoethanol, which should reduce any disulphide bonds. The subsequent lanes all contain 2.5 mM GHS and increasing quantities of GSSG, from 0 mM to 5 mM. The '?' marks the expected location of the 78 kDa Cpl-1 dimer band, but as discussed below, the observed band is not believed to be the dimer.

Despite the presence of a 78 kDa band, which is the predicted size of a Cpl-1 dimer (Resch et al. 2011a), this is not believed to be a Cpl-1 dimer. The presence of the 78 kDa band in the first lane, where the sample has been boiled in the presence of β-mercaptoethanol, suggests that this is an artefact of non-specific binding of the anti-StrepII antibody rather than dimerised Cpl-1. It is possible that this band is masking Cpl-1 dimer formation in subsequent lanes, however this seems unlikely, as one would expect to be able to see two bands even if they were located close together or the single band would become brighter. No non-specific band is seen in the negative control, but this is an extract from *C. reinhardtii* that has been purified using a StrepTrap column as previously described, so has a different background protein make-up to the *E. coli*-derived crude extract. It serves its purpose of showing that no dimerisation occurs with the un-mutated Cpl-1, while the β-mercaptoethanol lane is a negative control for the mutated Cpl-1.

There are a couple of factors that could be preventing dimerisation of Cpl-1_dimer, even *in vitro*. The first and most likely issue is the presence of C-terminal epitope tags causing steric hindrance. The HA tag in the *C. reinhardtii* version of the protein and the HA and StrepII tags in the *E. coli* version are both on the C-terminus of the protein. In hindsight, locating these on the N-terminus may have been a better approach as the Cpl-1 protein has been engineered to form a disulphide bridge close to the C-terminus. Resch et al. (2011) note that the positioning of the new cysteine residue near the C-terminus is of crucial importance to dimerisation. Given all of this, it is not surprising that the inclusion of an 8-9 amino acid epitope tag at the C-terminus would interfere with dimerisation.

The same Cpl-1_dimer protein was expressed as a recombinant in *E. coli* after the failure to achieve dimerisation in *C. reinhardtii* to test whether this failure was a property of the protein *per se* or a result of the expression platform. In essence, if dimerisation of Cpl-1_dimer could be achieved in *E. coli* but not in *C. reinhardtii* then we would conclude that the *C. reinhardtii* chloroplast is unsuitable for pre-dimerised Cpl-1 production. However, if dimerisation could not be achieved in *E. coli* either, then the conclusion would be that the protein design was in some way preventing dimerisation.

In this study, the lack of targeting to the *E. coli* periplasm likely prevented *in vivo* dimerisation from occurring. Future studies should use vector such as pIN-III, which includes an N-terminal OmpA target peptide and may rectify this, or consider using an engineered strain of competent *E. coli* such as Rosetta-gami™ (Novagen) which displays enhanced disulphide bond formation in the cytoplasm. The failure of *in vitro* dimerisation could be due to a lack of condition-optimisation of the dimerisation protocol, as only a relatively narrow range of conditions were tested. For example, different reagents could be used, such as cysteine and cystine, and in different ratios to create the redox environment, as well as different temperatures and incubation times. Optimisation of this protocol for Cpl-1_dimer may have enabled dimerisation.

4.3.3 Expression of the Cpl-1 associated holin, Cph-1

4.3.3.1 Production of *C. reinhardtii* lines expressing *cph-1*

The *C. reinhardtii* derived *psaA* 5' UTR/promoter is functional in *E. coli*, as are several chloroplast promoters probably due to the chloroplast's prokaryotic origins (Young and Purton 2015). Given the ability of holins to form holes in the cytoplasmic membrane of *E. coli*, and thus inhibit cell growth (Catalão et al. 2011), the use of the pSRSapI plasmid to clone these in *E. coli* is unlikely to be possible.

A strategy developed by Young & Purton (2015) makes use of a 'spare' opal stop codon, UGA, that is unused in the *C. reinhardtii* chloroplast. The codon is reassigned to code for a tryptophan by the co-introduction of a modified tryptophan tRNA gene (*trnW*) into the *C. reinhardtii* chloroplast, along with the gene of interest. The gene of interest is modified to contain several UGA codons in the place of UGG (Tryptophan) – these are read as "Stop" codons by the *E. coli* translational machinery, but as tryptophan codons in the chloroplast. A plasmid, named pWUCA2 was developed to perform this cloning strategy. The plasmid is very similar to pSRSapI (2.2) but with the addition of a $trnW_{UCA}$ gene upstream of the gene of interest (GOI) expression cassette, illustrated in Figure 4.26.

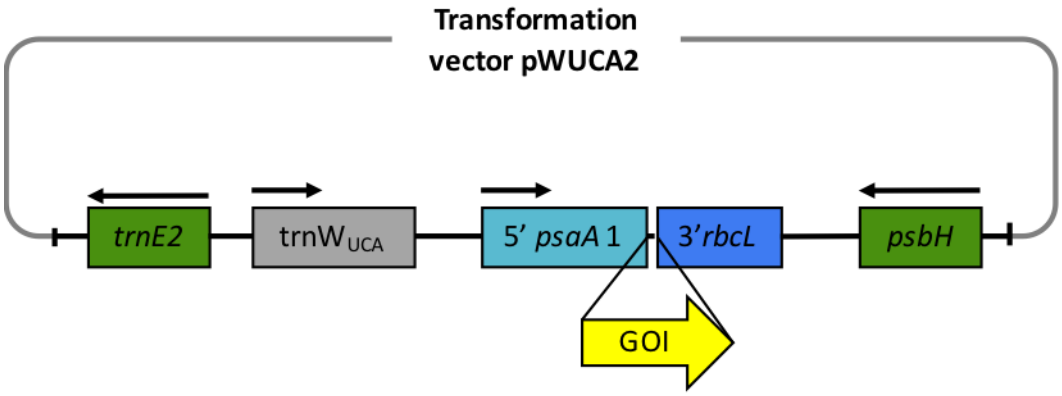


Figure 4.26 The pWUCA2 transformation vector is based upon pSRSapI and has the same cloning sites for the GOI and the same flanking regions for homologous recombination. However, an additional synthetic gene, $trnW_{UCA}$, has been inserted upstream of the GOI expression cassette.

This strategy enables the cloning in *E. coli* of genes under functional/constitutive promoters, which would otherwise produce proteins that are toxic or detrimental to

E. coli growth. In this case, the *E. coli* translational machinery will stop short of producing the whole length protein due to the presence of stop codons in the middle of the GOI.

Producing holins in the *C. reinhardtii* chloroplast may avoid the problem of *E. coli* lysis that occurs when holins are expressed as recombinant proteins in bacterial platforms (Smith et al. 1998; Shi et al. 2012; Song et al. 2016). This problem has been largely overcome by the use of induction systems, but allows only 10 – 15 minutes of accumulation before the metabolic activity fails (Smith et al. 1998). Furthermore, the nature of the protein means that it must be purified from the bacterial membrane to which it localises, which further reduces yields and increases costs. While *C. reinhardtii* achieves only low levels of recombinant protein expression compared to *E. coli* for many proteins, given these two issues it may be that it can compete as an expression platform when producing holins.

4.3.3.1.1 Gene design and codon optimization

Codon Usage Optimizer Beta 0.92 (Kong 2013) was used to optimize the *cph1* gene for *C. reinhardtii* chloroplast expression. Two tryptophan codons early in the gene were changed from UGG to UGA. *SapI* and *SphI* restriction sites were included for cloning into the pWUCA expression vector. The full-length *cph1*-HA gene was synthesized by Integrated DNA technologies (Appendix H).

4.3.3.1.2 Expression cassette production

SapI and SphI restriction enzymes were used to digest and ligate the *cph1* gene into pWUCA2 in-frame, producing the plasmid pWUCA2_*cph1*-HA. DH5α competent *E. coli* cells were transformed with the resulting plasmid and sequencing confirmed the correct plasmid assembly.

4.3.3.1.3 *C. reinhardtii* transformation

The pWUCA2_*cph1*-HA plasmid was used to transform the recipient *C. reinhardtii* strain TN72 via the glass bead method as described in 2.3.13. Figure 4.27 illustrates the resulting *C. reinhardtii* cell line, pWUCA2_*cph1*-HA.

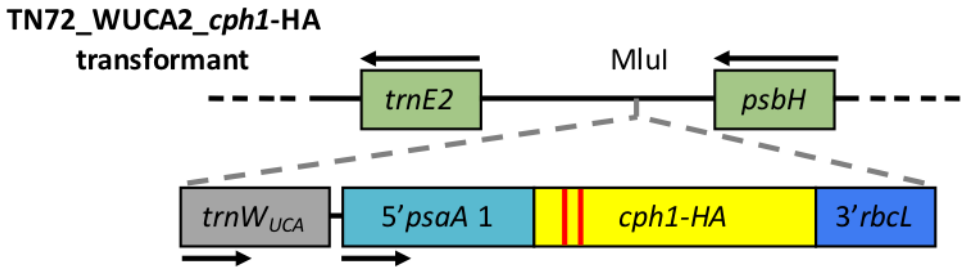


Figure 4.27 TN72 recipient strain successfully transformed to carry *cph1*-HA under the *psaA* exon 1 promoter/5'UTR in the chloroplast, and with *psbH* functionality restored. A modified *trnW* gene has been co-introduced to correctly code for the introduced UGA codons in the *cph1* gene. The approximate positions of the UGA codons are indicated in red.

The resulting transformants were confirmed for homoplasmicity as described in 2.3.17. In this instance, due to the addition of the $trnW_{UCA}$ gene, the expected band size for homoplasmicity will be 291 bp larger, giving a total expected size of 1421 bp (Primers: Flank1 & psa.R (Appendix O: List of Primers)). Gel electrophoresis of the resulting PCR product from three transformants is shown in Figure 4.28.

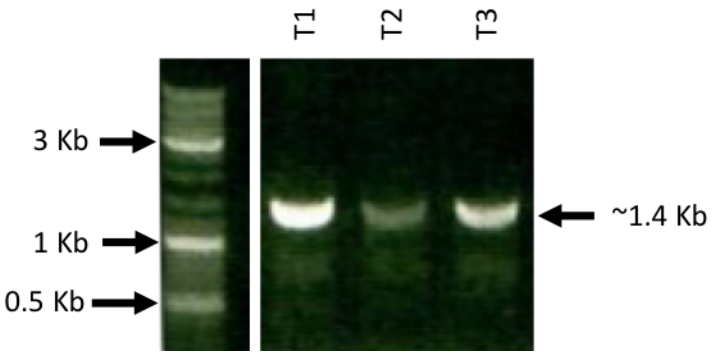


Figure 4.28 The resulting PCR products are approximately 1.4 kb in size as expected. The lack of an 850 bp product confirms that all three of these transformants are homoplasmic.

Finally, the integrity of the inserted cassette was confirmed by DNA sequencing as in 2.3.18.

4.3.3.1.4 Western blot confirmation

Whole cell extracts of TN72_WUCA2_*cph1*-HA and TN72_SR_Control (as a negative control) were prepared and separated by SDS-PAGE. After blotting to nitrocellulose membrane and blocking as described in 2.4.5, the membrane was probed with anti-HA primary antibody, then incubated with IRDye® secondary antibody. Fluorescence

was observed using the LiCor Odyssey® CLx system. The resulting image is shown in Figure 4.29.

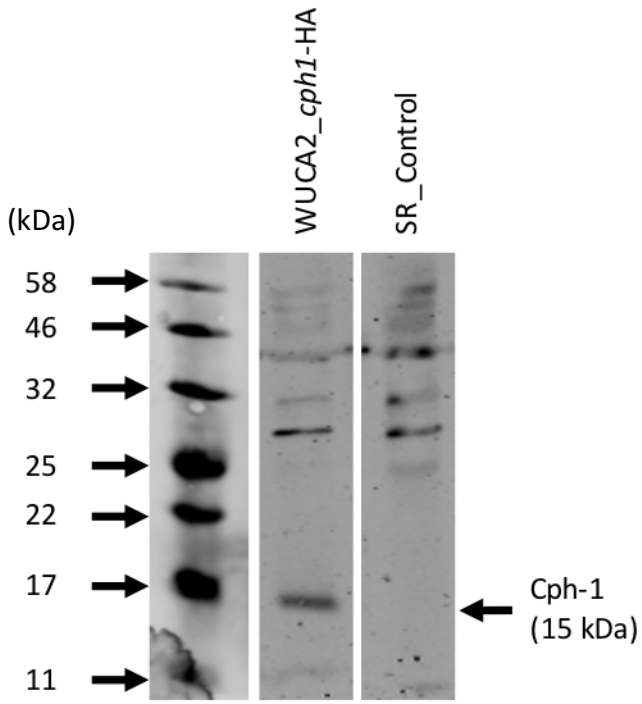


Figure 4.29 Western blot analysis indicates that a ~15 kDa protein is detected by anti-HA antibodies in the TN72_WUCA2_*cph1*-HA extract, and not in the negative control. This is likely to be Cph-1 protein.

Cph-1 appeared to be present in TN72_WUCA2_*cph1*-HA whole cell extract. However, it was noted that the transformed *C. reinhardtii* lines did not appear to be as healthy as other lines, in that they took a longer time to reach a bright green culture, taking approximately six days rather than the usual three or four. This was merely an observation, and no data was collected at the time to objectively investigate a change in growth rate.

4.3.3.1.5 Investigation into holin localisation

Given the non-specific insertion of holins into bacterial cell membranes, it is likely that the holin would insert into chloroplast membranes when produced as a recombinant. In order to investigate whether the holin was integrating into the thylakoid membrane or chloroplast envelope, TN72_WUCA2_*cph1*-HA was lysed by sustained glass bead agitation at 4 °C and subjected to ultracentrifugation at 100,000 g for 1 hour at 4 °C. This was designed to separate the membrane fraction (pellet)

and cytoplasmic fraction (supernatant). However, when the two fractions were separated by SDS-PAGE and analysed by western blotting as before, no 15 kDa protein could be detected.

In an attempt to detect Cph-1 accumulation again, very high concentrations of cell extract were separated on SDS-PAGE and analysed by western blotting, as well as using ammonium sulphate precipitation to concentrate large volumes of cell extract, but no Cph-1 was detected. Again, only observationally, it was noted that at the time of these further experiments, the TN72_WUCA2_*cph1*-HA cell line seemed to grow at a normal rate and it had recovered from its earlier unhealthy appearance. This observation and the lack of detectable Cph-1 accumulation suggested that the gene had been silenced in some way, either through point mutations or recombination events between similar sequences.

4.3.3.1.6 Investigating *cph1* silencing

Figure 4.28 demonstrates that the three TN72_WUCA2_*cph1*-HA lines were homoplasmic, meaning that all of the ~70 chloroplast genome copies contained the desired expression cassette. This was achieved by several rounds of selection upon minimal medium, thus selecting for photosynthetically-competent cells and promoting the presence of the transformed plastome copies. However, once homoplasmicity had been achieved and confirmed, cells were grown mixotrophically upon acetate and under light which promotes faster growth. Had the cell been heteroplasmic at this stage and the GOI was sufficiently toxic, it is possible that selection could be driven the other way, back to the photosynthetically deficient recipient line. However, the selection for photosynthesis is so strong that this seems quite unlikely and the loss of *psbH* functionality also confers photosensitivity, further increasing the pressure on the plastome to gain *psbH* function. However, given that the lines were homoplasmic, this cannot be the case. The sensitivity of PCR means that even if one plastome copy were without the expression cassette, this is likely to be visible in Figure 4.28.

There are several sites in the inserted expression cassette that could be subject to recombination events within the chloroplast as they are of endogenous origin,

namely the *psaA* exon 1 promoter and 5'UTR, *rbcL* 3'UTR and *trnW.* However, the native *rbcL* gene and the *psaA* exon 1 are respectively 46 and 47 kbp away from the insertion site, and the *trnW* gene is 61 kbp away, which would be expected to reduce recombination rates, although this relationship has not been confirmed (Mudd et al. 2014). Furthermore, each of these repeated sequences is relatively small: *trnW* is 73bp; *psaA* exon 1 promoter and 5'UTR is 294 bp; and *rbcL* 3'UTR is 407 bp. For context, Mudd et al. (2014) suggest that repeats of 650 bp separated by 5 kbp results in efficient recombination, but no strict rule has been suggested. Another alternative is that point mutations in either the modified *trnW* gene or *cph1* gene have resulted in failed translation or instability of the protein.

To test the above theories, the cassette region was amplified by PCR from genomic DNA and sequenced (regions b and c in Figure 4.30), as performed after the initial confirmation of homoplasmicity. Sequencing confirmed the presence of both the holin and trn_{WUCA} genes, with no mutations. The psbH.R primer binds inside the downstream flanking region of the transformation plasmid, and 123a binds upstream of the trn_{WUCA} gene, which therefore suggests that the whole cassette is present. However, the Flank1 primer lies outside the transformation plasmid, upstream of the cassette, and no product could be seen when this primer was used.

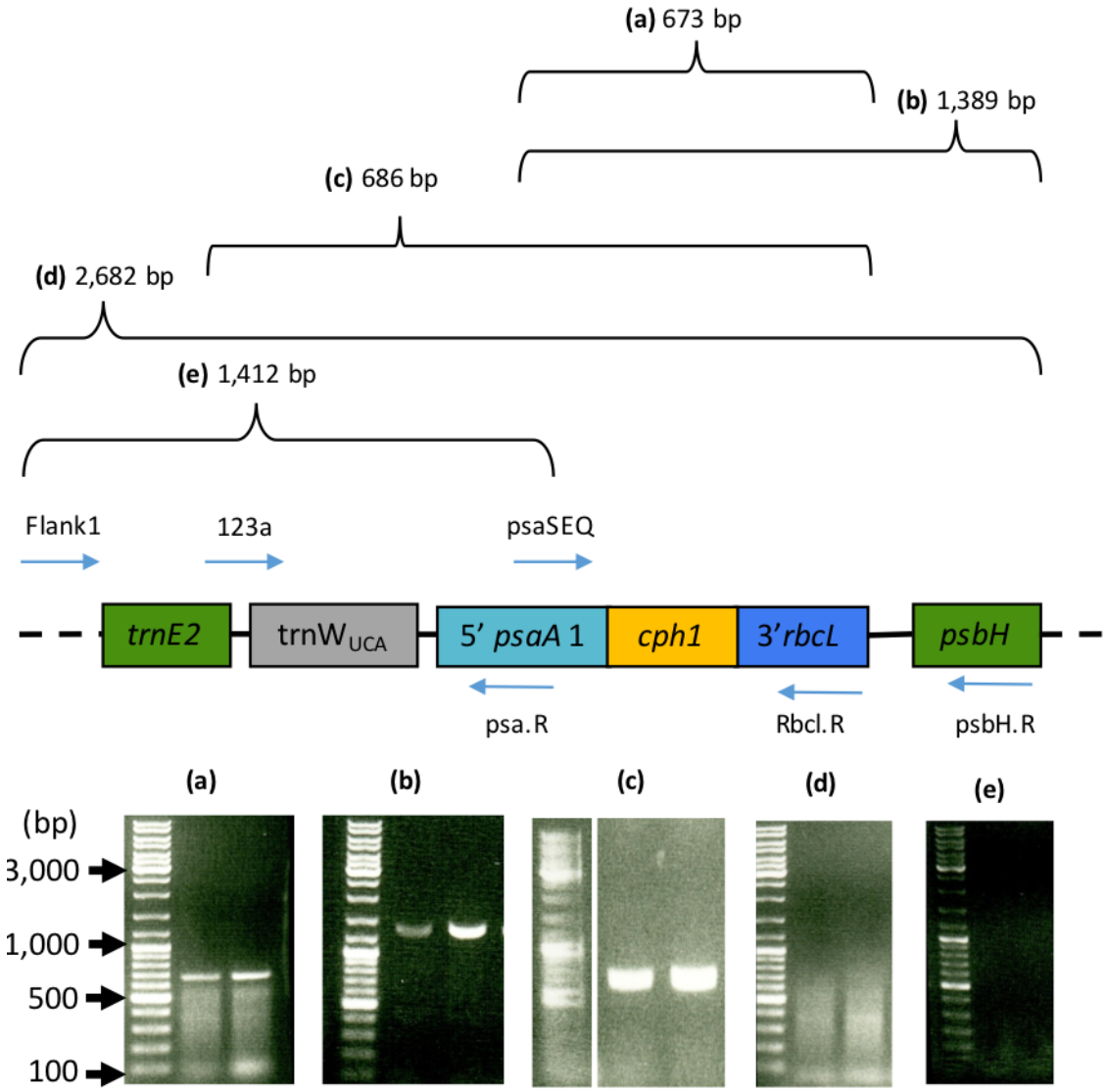


Figure 4.30 Reactions **a-c** produced the expected amplified PCR product, while **d** and **e** failed to produce any product. The successful amplification of **a-c** shows that the cassette is present somewhere in the genome. The success of amplification in Figure 4.28 (the same primers as **e** above) suggests that the cassette was originally in the correct locus. However, the failure of **d** & **e** here suggests that the cassette is no longer present in the intended neutral site between *trnE2* and *psbH* in the chloroplast. The two lanes show two different TN72_WUCA_*cph1*-HA transformants. Primer sequences can be found in Appendix O: List of Primers.

It is possible that a very small number of plastome copies remain with the correct cassette insertion, but that the majority of copies have recombined in a manner so as to silence gene expression. However, as mentioned previously, the sensitivity of PCR would normally enable even a single copy to be detected. Mudd et al. (2014) does note that recombination has been observed with relatively small repeats in *Chlamydomonas*, for example, 216 bp direct repeats in the *chlL* gene are able to

recombine. The most likely explanation is therefore that some combination of recombination events has led to the relocation of the expression cassette to a silent site in the plastome, which was then highly selected for, given the apparent toxicity of Cph-1.

4.4 Conclusion and future work

This chapter has examined three strategies to improve Cpl-1 activity against *S. pneumoniae*, the first two of which deal with direct protein engineering and the third attempts to exploit synergy with another protein, a holin.

4.4.1 Cell wall binding site mutagenesis

Six binding site mutated Cpl-1 proteins were expressed in *C. reinhardtii* to detectable, albeit varying, levels. However, difficulties arose when attempting to compare the activity of these proteins as lytic enzymes against *S. pneumoniae*. Firstly, the potential for improvements around the mutations themselves will be discussed, followed by a discussion of how the activity assays might be adapted and suggestions of possible alternative assays.

It is noted in 4.3.1.1.1 and illustrated in Figure 4.4 that there are three residues believed to be of key importance in each choline binding site, two tryptophan residues and a tyrosine residue. Hermoso et al. (2003) also suggest that a lysine is involved in capping the site and stabilising the phosphate group of phosphoryl-choline. This investigation only considered the two tryptophan residues of each site, simply due to time constraints and the fact that they appear to be closer to the choline residue. It appears that mutating the tryptophan residues did indeed negatively affect choline binding, but that this resulted in reduced lytic activity. It would, therefore, be interesting to consider the tyrosine and lysine in each binding site as they may enable a subtler mutation to be produced. Alanine scanning is a common approach to mutating residues of interest, as it eliminates the role of the polar side chain, however, in this instance, this may result in too drastic a change in polarity. Therefore, a more conservative mutation might be considered, for example, the mutation of tryptophan to tyrosine. The binding site is clearly sensitive to

mutations and highly conserved, and the alanine changes appear to have been too severe and resulted in reduced activity.

Furthermore, the Cpl1_W223A-HA construct was never successfully synthesised, for reasons unknown. It is interesting to note that the Cpl1_W223A_W230A-HA and Cpl1_W202A_W209A_W223A_W230A-HA, which both contain the W223A mutation in addition to others, produce consistently detectable levels of protein accumulation. This is somewhat unexpected, as W223 is not a strictly conserved residue in choline binding repeats (Figure 4.5) while W209 and W230 are, though no problems were observed when these were mutated. The issues arose after PCR when attempting to insert the gene into pJet, a vector from a commercially available PCR cloning kit (ThermoFischer), enabling blunt end cloning and transformation of *E. coli*. No *E. coli* colonies were observed in any of the attempts to produce *cpl1*_W223A-HA, suggesting this mutated gene is either highly toxic to *E. coli* or for some reason unable to ligate into the pJet vector. In future either a different intermediate vector, such as pGemTEasy could be used or the intermediate step could be skipped and attempt to clone directly into pSRSapI using a digested PCR product.

An interesting and related point of discussion is why the inactive choline binding repeats (p4-p6) remain so highly conserved. As these repeats appear to be sterically unable to bind choline, it would be reasonable to assume that there was little evolutionary pressure for them to remain unaltered. The fact that they remain so highly conserved indicates that they also have some other important function. This may suggest that the three-dimensional shape of the protein is highly dependent on the interaction of these repeats and that mutations to them have effects more wide-ranging that merely altering choline binding.

Given that initial attempts to disrupt the choline binding sites appeared to result in a reduction of activity, and that the p4-p6 repeats are highly conserved despite their purported inactivity, an alternative protein engineering focus could be taken. For example, the flexible acidic linker that joins the cell wall binding domain and catalytic domain would be an interesting site for improvement. If it is indeed the case that the binding domain remains bound to choline in the cell wall, then the number of

peptidoglycan cleavages that can be performed by the catalytic domain is limited by the length and nature of this linker. When producing Artilysins®, Briers et al. (2014) investigated the effect of increasing linker length between the peptide and endolysin. It was found that antibacterial activity increased in a stepwise fashion as the linker was extended, then began to plateau at around 18 amino acids in length (Figure 4.31).

Figure Removed

Figure 4.31 L0, L1, L2, L3, and L4 correspond to linker amino acid sequences of AGAS (4), AGAGAS (6), AGAGAGAGAS (10), AGAGAGAGAGAGAS (14), and AGAGAGAGAGAGAGAGAS (18). Reproduced from Briers et al. (2014).

The acidic linker for Cpl-1 comprises residues 189-199, with amino acid sequence DDEEDDKPKTA. While not directly analogous to the linker investigated by Briers et al. (2014), it would be interesting to attempt to extend the Cpl-1 linker to see if any improvement in activity can be achieved. However, it is unlikely that much improvement to activity can be gleaned by extending the acidic linker, as this would have probably been highly selected for throughout evolution. This is in contrast to choline binding which, as discussed previously, appears to act as a tethering mechanism to prevent accidental lysis of potential future host bacterial cells.

DEAE cellulose acts as a choline analogue and has therefore been exploited for ion exchange chromatography to purify Cpl-1, using a protocol first described by Loeffler et al. (2001). Cpl-1 binds to DEAE and can then be eluted using a free choline. This method was not used in this investigation, as choline binding was the object of investigation. However, the method could be adapted into an assay for choline binding ability. For example, a mutated Cpl-1 protein displaying weakened choline binding would be eluted from the DEAE column more easily than strongly binding Cpl-1. This assay could be used to confirm that any change in activity was indeed due to choline binding alteration, and not due to some other unknown factor or three-dimensional shape change.

The protein engineering steps taken in this section were highly speculative and would have benefitted from testing a larger number of mutants more quickly. For this reason, the choice of *C. reinhardtii* as a platform limited these initial investigations, as transformations take more time, recombinant protein yields are lower, and background antimicrobial effects disrupted activity assays compared to other platforms such as *E. coli*. This work would have been better suited to *E. coli,* which enables rapid prototyping and testing of recombinant proteins. *C. reinhardtii* should then be used further down the production line, once the protein sequence has been optimised. At that point, the many benefits of using *C. reinhardtii* as a platform for the production of therapeutic recombinant proteins can still be harnessed.

4.4.2 Cpl-1 dimerisation

Cpl-1 dimerisation was attempted for two reasons. Firstly, a report by Resch et al. (2011) demonstrated that pre-dimerised Cpl-1 had a 10-fold reduction in plasma clearance and a two-fold increase in activity. Secondly, the *C. reinhardtii* chloroplast is known to control redox reactions and is capable of catalysing disulphide bonds in recombinant proteins (Tran et al. 2009), which can be further increased by the addition of selenocystamine to the growth medium (Ferreira-Camargo et al. 2015). However, our attempt to produce a Cpl-1 dimer in the *C. reinhardtii* chloroplast was unsuccessful, as was *in vitro* formation of disulphide bonds using an *E. coli* produced Cpl-1.

The most likely reason for this failure is the presence of epitope tags on the protein C-terminus, as previously discussed. The first step towards testing this theory is to redesign the *cpl1*_dimer gene such that it encodes an N-terminal tag, with no additional amino acids on the C-terminus. The gene should also be transformed into *E. coli* using a plasmid such as pIN-III which will target the protein to the periplasm for disulphide bond catalysis, or into an engineered strain of competent *E. coli* such as Rosetta-gami™. If this successfully produces a Cpl-1 dimer, which it should as it is a simple replication of the work by Resch et al. (2011), then the same gene should be transformed into the *C. reinhardtii* chloroplast. At this point, it will be clear that if no dimer can be seen in the *C. reinhardtii* transformants, then it is an issue with the platform, rather than the protein.

If Cpl-1 dimers can accumulate in the *C. reinhardtii* chloroplast, then investigations into the use of selenocystamine can be untaken. It may be that this strategy can be used to increase accumulation levels of Cpl-1 in total, particularly if the dimer is more stable than the monomer. An investigation can also be carried out into whether dimerised Cpl-1 is less susceptible to degradation whilst in storage, as it would potentially be less accessible to proteases and less prone to denaturation. This would be an important property for the future commercialisation and distribution of Cpl-1 as a therapeutic.

4.4.3 Holin production

Holins enable endolysins to cross the cell membrane during the bacteriophage lytic cycle. The treatment of some bacteria with both endolysins and holins has been shown to have a synergistic effect upon cell lysis (Shi et al. 2012; Song et al. 2016). However, the production of holins in bacterial systems is problematic as they insert into cell membranes and depolarise the cell. Here we have reported the successful accumulation of the holin Cph-1 in the *C. reinhardtii* chloroplast but found that accumulation quickly ceased.

The ‘silencing’ of Cph-1 is worthy of investigation, as chloroplast transformations are not usually susceptible to gene silencing, although it is common in nuclear transformations. The most likely explanation is that in fact homoplasmicity of the

plastome was never achieved and that some non-specific PCR amplification occured, which led to Figure 4.28 appearing to confirm homoplasmicity. Sequencing of the PCR product could have confirmed this and a TN72 negative control should have been included to demonstrate that heteroplasmicity would have been detected. However, as discussed previously, even if this was the case it seems unlikely that the strain would revert to the parental sequence, as loss of *psbH* function results in photosensitivity, so growth under bright light would also encourage homoplasmicity. Furthermore, an investigation by PCR of the genomic DNA is displayed in Figure 4.30, and clearly demonstrates that the cassette is present somewhere in the *C. reinhardtii* genome. However, the same investigation failed to confirm, or refute, that the cassette was located in the targeted location, namely between the *trnE* and *psbH* genes in the chloroplast. A primer located beyond the *psbH* gene is required so that the entire region can be amplified to confirm that the cassette is not present or otherwise - this would be very straightforward.

We know that over the millennia since the first photosynthetic prokaryote was engulfed into a eukaryote to become a chloroplast, the vast majority of the genetic information from the chloroplast genome has been eliminated or migrated to the nucleus, leaving only genes essential for photosynthesis in the chloroplast (Howe and Purton 2007). It is possible that the highly undesirable *cph1*-containing cassette has been translocated to the nucleus under a similar mechanism, and once there, silencing mechanisms are prolific. The potential mechanism for such a DNA relocation is not understood, but such a mechanism is known to exist (Cullis et al. 2009). Furthermore, such transfers have been shown to occur at high frequencies in tobacco: more than 1 in 5 million cells were able to successfully transfer a marker from the chloroplast to the nucleus (Stegemann et al. 2003). Such an event would explain the presence of the cassette, yet lack of expression. A method such as inverse PCR could then be used to discover the location of the cassette.

Regardless of the mechanism used by the *C. reinhardtii* cells to prevent Cph-1 accumulation, it clearly is not tolerated by the cell. It is likely that the holin was inserting into the thylakoid membranes, given their similarity to bacterial cell membranes. As shown in Figure 4.3, a large number of holin monomers are required

to form a raft and then oligomerise into a hole before any depolarisation will occur. It may be that the holin was only weakly expressed and therefore not lethal to the cell, but caused enough selective pressure that it was highly undesirable. This would explain why transformants were obtained but quickly lost their Cph-1 expression.

It is likely that any repeated attempt to express Cph-1 in the *C. reinhardtii* chloroplast will result in the activation of a mechanism to prevent its expression. However, holins, unlike endolysins, are not species specific in nature and are generally able to insert into the membranes of a wide range of bacterial species (Savva et al. 2008). It would be interesting to express a holin and anti-holin pair, such as S105 and S107 from the lambda bacteriophage, as well as the holin alone, to see whether expressing the pair prevents silencing of the holin gene. This would enable a continuation of the investigation into whether the addition of a holin alongside Cpl-1 can improve its lytic activity.

Chapter 5 Extending the *Chlamydomonas reinhardtii* chloroplast platform to other endolysins

5.1 Introduction

5.1.1 Previous endolysin expression in the *C. reinhardtii chloroplast*

Dr Henry Taunt, working in the Purton Lab, was the first to successfully express an active endolysin in the *C. reinhardtii* chloroplast (Taunt 2013). The work showed that Cpl-1 accumulated to a relatively high level, approximately 1% of total soluble protein, in the chloroplast and that it was active against *S. pneumoniae in vitro.* However, Dr Taunt was unable to show expression of two other endolysins, Gp20 and Lys16 that target *Propionibacterium acnes* and *Staphylococcus aureus*, respectively. The work by Taunt (2013) also attempted to identify elements of endolysin genes that resulted in successful or unsuccessful expression in the *C. reinhardtii* chloroplast. However, while several fusion proteins were created to test this theory, no clear conclusions could be drawn. Additionally, work by Dr Laura Stoffels, also of the Purton group, achieved expression and activity of the *S. pneumoniae* endolysin Pal, but failed to show activity of the *Staphylococcus aureus* targeting endolysin ϕ11, despite detecting expression (Stoffels 2015). Another endolysin SPN9CC, targeting *Salmonella typhimurium,* was expressed by Dr Rosie Young as a proof of concept that recombinant proteins highly toxic to *E. coli* can be synthesised in the *C. reinhardtii* chloroplast using a synthetic tRNA system, but the activity of SPN9CC was not investigated (Young and Purton 2015). As a Gram-negative targeting endolysin, it is unlikely to be active against the target bacterium without modification owing to the presence of the outer membrane.

Current work in the Purton Group by Juliana de Costa Ramos (unpublished) is focussed upon producing endolysins targeting Gram-negative bacteria. Two such endolysins (LysAB2 and LoGT-023) have been successfully expressed in the *C. reinhardtii* chloroplast, and although the 'target bacteria' for both endolysins is *Acinetobacter baumannii,* Gram-negative endolysins tend to be much less specific than those targeting Gram-positive bacteria and therefore have a much broader

spectrum of lytic activity. For example, the endolysin LysAB2 has been reported to effectively lyse *Staphylococcus aureus, Bacillus subtilis, Streptococcus sanguis, Escherichia coli, Citrobacter freundii* and *Salmonella enterica*, all in addition to *Acinetobacter baumannii,* which the bacteriophage producing LysAB2 specifically infects (Lai et al. 2011). Table 5.1 lists all known attempts at expressing endolysins in the *C. reinhardtii* chloroplast to date.

Endolysin	Target pathogen	Successfully expressed in *C. reinhardtii chloroplast*	Active against target pathogen	Reference
Cpl-1	*Streptococcus pneumoniae*	Yes	Yes	(Taunt 2013)
Gp20	*Propionibacterium acnes*	No	N/A	(Taunt 2013)
Lys16	*Staphylococcus aureus*	No	N/A	(Taunt 2013)
Pal	*Streptococcus pneumoniae*	Yes	Yes	(Stoffels 2015)
ϕ11	*Staphylococcus aureus*	Yes	No	(Stoffels 2015)
SPN9CC	*Salmonella typhimurium*	Yes	Not tested	(Young and Purton 2015)
LoGT-023 (Artilysin®)	*Acinetobacter baumannii*	Yes	Yes	Juliana de Costa Ramos (unpublished)
LysAB2	*Acinetobacter baumannii*	Yes	Yes	Juliana de Costa Ramos (unpublished)
CD27L$_{1\text{-}179}$	*Clostridium difficile*	Yes	Yes	This study

Table 5.1 The full list of endolysin expression attempts by the Purton Group. SPN9CC, LoGT-023 and LysAB2 target Gram-negative bacteria, while the others all target Gram-positive bacteria.

5.1.2 The need for more endolysins

As discussed in 1.2.2.4, Gram-positive endolysins have a very narrow spectrum of lytic activity. This property is highly advantageous in some respects, yet highly problematic in others. The advantage of a narrow spectrum antibiotic is that only the

target pathogenic bacteria are killed, while the commensal bacteria remain unharmed. This property limits the development of resistance and the opportunity for pathogenic re-colonisation. The disadvantage is that a precise identification of the pathogen causing the disease must be made before the correct endolysin can be prescribed by a clinician, and a distinct endolysin therapeutic must be produced for every common bacterial infection.

Unlike conventional antibiotics, where a single new drug discovery (such as teixobactin 1.1.2.2) has the potential to treat hundreds of different bacterial infections, endolysin therapy requires the development of a large arsenal of very specific protein pharmaceuticals. As shown in Table 5.1, the ability to produce active endolysins in *C. reinhardtii* is far from routine, with only a ~56% success rate so far. It is important, therefore, to continue to attempt to express new endolysins as well as to try to understand the factors affecting expression. This chapter demonstrates the successful expression of CD27L in the *C. reinhardtii* chloroplast, as well as documenting an attempt to achieve expression of Gp20, which has failed to express in previous efforts.

5.2 Aims and objectives

1. To produce a transgenic line of *C. reinhardtii* in which the endolysin CD27L accumulates.
2. To demonstrate the activity of CD27L on *C. difficile in vitro*.
3. To produce a transgenic line of *C. reinhardtii* in which Gp20 accumulates.
4. To demonstrate the activity of Gp20 on *P. acnes in vitro*.

5.3 Results and discussion

5.3.1 *Clostridium difficile*, ΦCD27 and $CD27L_{1-179}$

Clostridium difficile is a Gram-positive bacterium of particular importance due to its prevalence as a common hospital-acquired infection. *C. difficile* infection (CDI) is a toxin-mediated infection, and is the major cause of pseudomembranous colitis (Best et al. 2012). The ability of *C. difficile* to form spores, which can withstand harsh conditions and antibiotics, means that it is often the first bacterium to re-colonise the human gut after treatment with broad-spectrum antibiotics. While resistance to antibiotics has been reported in *C. difficile*, vancomycin remains the first line of treatment and until 2014 resistance to vancomycin was rarely observed (Hargreaves and Clokie 2014). However, recently there have been reports of reduced vancomycin susceptibility (Spigaglia 2016), emphasizing the need for new treatments.

5.3.2 *C. difficile* endolysin CD27L

The bacteriophage ΦCD27 infects *C. difficile,* and its endolysin, CD27L, has been identified and demonstrated to cause *C. difficile* lysis *in vitro* when expressed as a recombinant protein in *E. coli* (Mayer et al. 2008). CD27L has a similar conformation to many endolysins – a C-terminal cell binding domain, and an N-terminal catalytic domain. Removal of the C-terminal domain, to create $CD27L_{1-179}$ was found to improve enzyme activity and marginally extend the lytic range of the endolysin – enabling it to lyse more bacterial species (Mayer et al. 2008).

The *C. reinhardtii* chloroplast has been successfully used to express two active endolysins to date, Cpl-1 and Pal, both of which are specific to *S. pneumoniae*. To develop the *C. reinhardtii* chloroplast as a platform to produce endolysins, and given the specific nature of endolysins, it is important to show production of endolysins that are active against a range of bacteria. To this end, the truncated endolysin $CD27L_{1-179}$, which has a partially resolved crystal structure (PDB code 4CU5 and 3QAY) and has been shown to be active against a small range of bacteria when produced in *E. coli*, was chosen as a novel candidate for *C. reinhardtii* chloroplast expression.

5.3.2.1 Production of a *C. reinhardtii* line expressing CD27L

From hence forth, CD27L will refer to the truncated version of the endolysin, previously referred to as $CD27L_{1-179}$.

5.3.2.1.1 Gene design and codon optimization

The CD27L protein is accessible by PDB codes 3QAY (catalytic domain) and 4CU5 (Cell binding domain). The 179-amino acid sequence (catalytic domain) was obtained from RCSB PDB (a histidine amino acid was removed from the N-terminus, which remained from the His-tag) and the gene sequence was codon optimized for the *C. reinhardtii* chloroplast as described in 2.3.2. An HA tag was added to the C-terminus of the protein coding sequence for western blot detection, and SapI and SphI restriction sites were included for cloning purposes. The full-length gene was synthesized by Integrated DNA Technologies as a gBlock® Gene Fragment (Appendix J).

5.3.2.1.2 Expression cassette production

Given that CD27L has previously been expressed in *E. coli* without detrimental effects to *E. coli* growth, there were no concerns about its toxicity during cloning, so the pSRSapI transformation vector could be safely used instead of the variant vector pWUCA2 that allows UGA stop codons to be introduced into the coding region (Young and Purton 2015). The synthesized DNA was digested with SapI and SphI restriction enzymes and ligated into the pSRSapI vector and the insertion was confirmed by DNA sequencing.

5.3.2.1.3 *C. reinhardtii* transformation

The recipient line TN72 was transformed by the glass bead method using the pSRSapI_*cd27l*-HA plasmid. The resulting transformed line is represented in Figure 5.1. Photosystem II functionality is restored by *psbH* and the resulting transformants were capable of growing phototrophically on minimal media. Transformants were shown to reach a homoplasmic state by PCR (Appendix K) as described in 2.3.17.

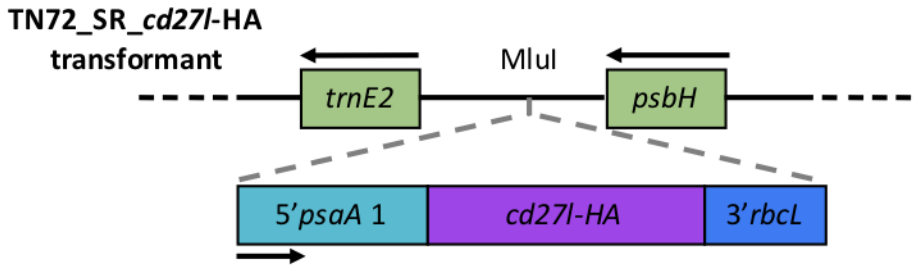


Figure 5.1 TN72 recipient strain successfully transformed to carry *cd27l*-HA under the *psaA* exon 1 promoter/5'UTR in the chloroplast, and with *psbH* functionality restored.

5.3.2.1.4 Western blot confirmation

Whole cell preparations of three TN72_SR_*cd27l-HA* transformants (T1-3), as well as TN72_SR_*cpl1*-HA and TN72_SR_Control (as positive and negative controls, respectively) were prepared and separated by SDS-PAGE. After blotting to a nitrocellulose membrane and blocking as described in 2.4.5, the membrane was probed with anti-HA primary antibody, then incubated with an IRDye® secondary antibody. Fluorescence was observed using the LiCor Odyssey® CLx system. The resulting image is shown in Figure 5.2.

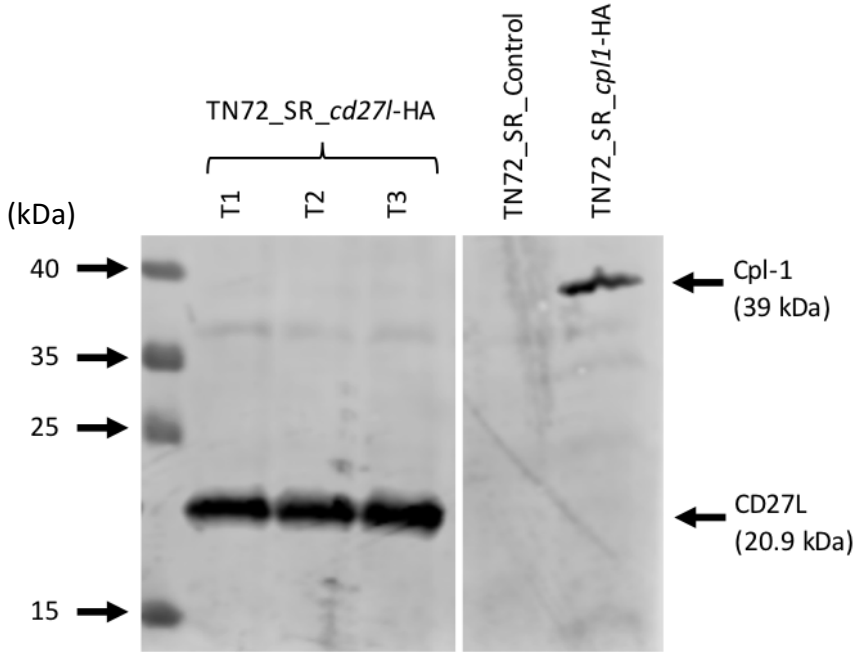


Figure 5.2 Western blot analysis shows that a ~21 kDa protein is detected by anti-HA antibodies in the three TN72_SR_*cd27l-HA* extracts, and not in the negative control. This is likely to be CD27L protein.

CD27L protein accumulates to easily detectable levels in all three transformant clones, and is of the expected size.

5.3.2.2 Activity of CD27L crude extract against 'stripped' *E. coli*

Since it was not feasible to set up *C. difficile* cultivation facilities in our lab, an assay using 'stripped' *E. coli* was used since CD27L had been shown to be active against this bacterium (Mayer et al. 2011). The *E. coli* had to be stripped of their outer-membrane to enable access to the cell wall by the endolysin as described in 2.7.2.2 (Nakimbugwe et al. 2006).

Crude extracts of TN72_SR_*cd27l*-HA and TN72_SR_Control were prepared as described in 2.5.1, and a Bradford assay was performed to measure the total protein in each sample. 5 µg of each crude extract, as well as 5 µg of BSA to act as an osmotic pressure control, were added to a 96-well plate, containing 200 µl of stripped *E. coli* in each well. The change in OD_{595} was measured by a FLUOstar Omega plate reader (BMG Labtech) over a period of 200 minutes at 37 °C. The resulting change in OD_{595} is displayed in Figure 5.3.

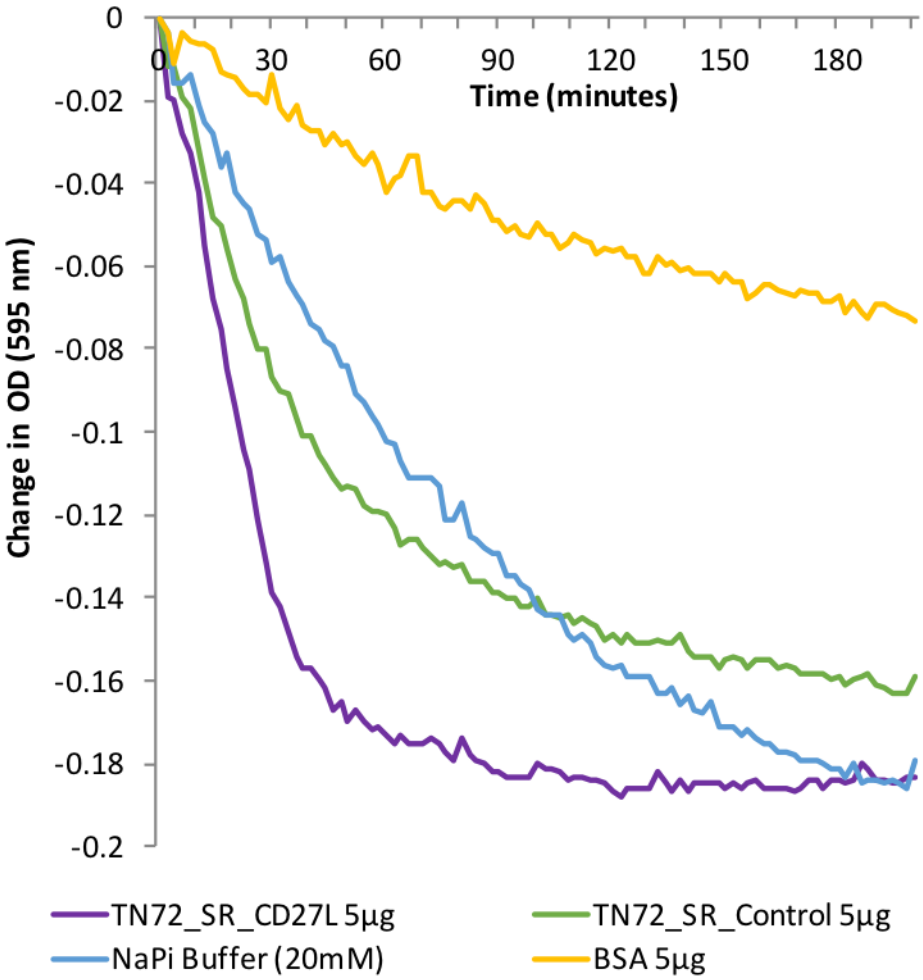


Figure 5.3 A turbidity reduction assay, measuring the reduction in OD (595 nm) in wells containing *E. coli* stripped of their outer-membrane and challenged with crude *C. reinhardtii* extracts. The BSA is used as an osmotic pressure control, while the NaPi buffer shows the rate of cell autolysis when added to the cells in the absence of osmotic pressure.

The crude extract containing CD27L appears to show activity against the stripped *E. coli* cells, as there is a much greater drop in OD than in the BSA control. However, there does appear to be some activity from the crude extract of the control strain also. The shape of the curve for the two *C. reinhardtii* crude extracts looks more like the characteristic kill curve expected, than the linear rate observed for the addition of NaPi Buffer.

5.3.2.3 Activity of CD27L crude extract against *C. difficile*

In order to test the endolysin against *C. difficile*, a category two biohazard for which our lab was not equipped, a collaboration was established with Dr Melinda Mayer, whose group at the Quadram Institute Bioscience (formerly the Norwich Institute of

Food Research) were the first to identify CD27L (Mayer et al. 2008). This enabled the testing of the CD27L-containing crude *C. reinhardtii* extract directly on *C. difficile* using the Mayer group's established protocol.

A turbidity reduction assay was performed as described in 2.7.2.3. Two concentrations of crude extract, in NaPi buffer, were used: 10 µg and 50 µg. The purified CD27L was provided by Dr Mayer and used as a positive control. It was an old (2008) NiNTA-purified CD27L sample produced in *E. coli*, so the concentration was deemed irrelevant due to its age as much of it was thought to have degraded, however, it still serves as a useful positive control.

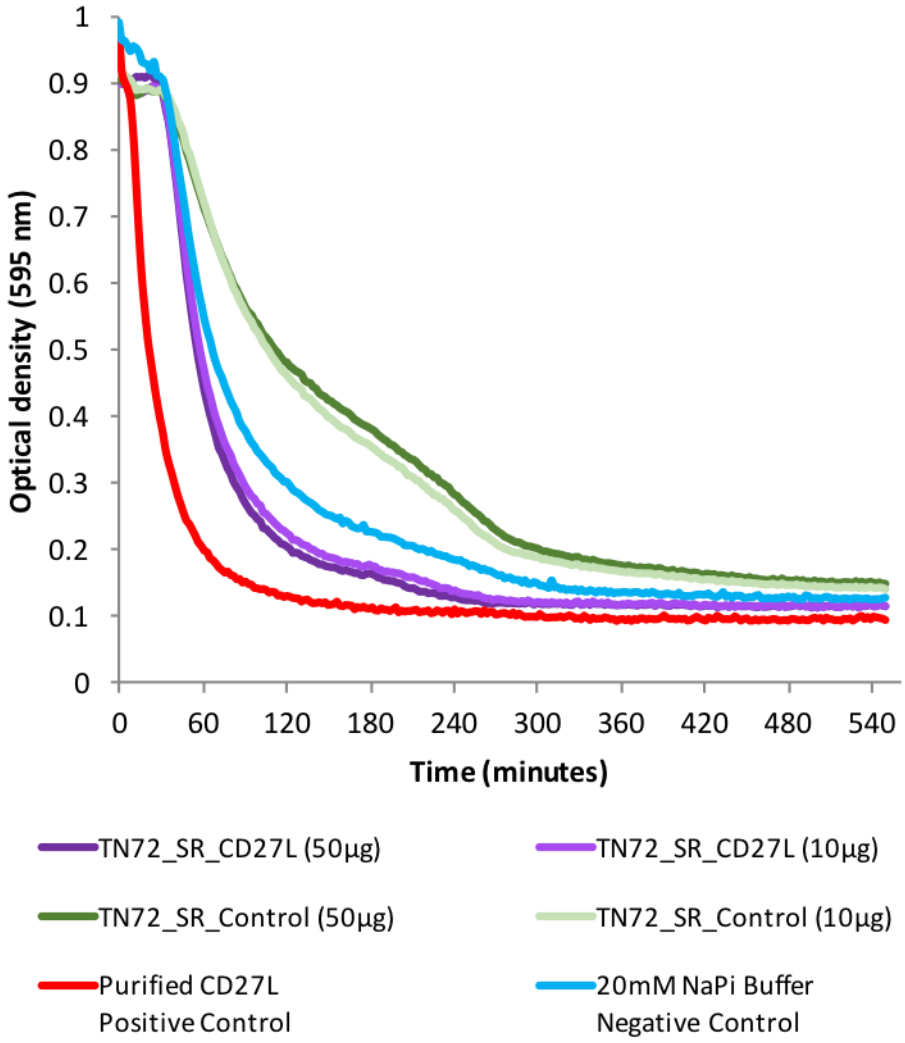


Figure 5.4 *Clostridium difficile* turbidity reduction assay. The purified CD27L was provided by Dr Mayer as a positive control and TN72_SR_Control as a negative control. Performed by Dr Melinda Mayer at Quadram Institute Bioscience, Norwich, UK.

Figure 5.4 shows that the TN72_SR_Control, which was used as a negative control, appears to have an inhibitory effect upon cell lysis when compared to the NaPi buffer control. This is believed to be due to the effect of a higher osmotic pressure on the *C. difficile* cells due to the presence of other proteins in the crude extract, and has been reported previously (Mayer et al. 2010).

5.3.2.4 Enriched CD27L crude extract

The osmotic pressure effects seen in the *C. difficile* turbidity reduction assay (Figure 5.4), and the apparent activity of the control extract observed in Figure 5.3 against stripped *E. coli* cells, hindered our ability to satisfactorily measure the activity of the *C. reinhardtii* produced endolysin. Consequently, it was decided to change the HA epitope tag on the C-terminus of the CD27L protein to a StrepII tag to better enable enrichment or purification. This was performed to enable a better comparison of endolysin activities without the effects of the *C. reinhardtii* crude extract.

5.3.2.4.1 Expression of CD27L-StrepII in *C. reinhardtii*

PCR was used to change the epitope tag on the *cd27l* gene. The reverse primer was designed to have an overhang with the StrepII tag sequence attached, thus removing the HA tag and adding the StrepII tag. The StrepII tag is shown in lower case, and the start and stop codons are underlined on the primers below:

cd27l.F: TGGTATTGCTCTTCAATGAAAATTTGTATTACT

cd27l_StrepII.R: ACTGACGCATGCTTATTAttttcgaattgtgggtgagacca GTTAATGTTTTTGTTTAAAACACCTTCAAC

The PCR product was digested with SphI and SapI restriction enzymes and ligated into pSRSapI as described in 2.3.10 and 2.3.11. Sequencing confirmed the addition of the correct encoding sequence for the StrepII tag.

The recipient line TN72 was transformed by the glass bead method using the pSRSapI_*cd27l*-StrepII plasmid. The resulting transformed line is represented in Figure 5.5. Photosystem II functionality was restored by *psbH* and the resulting transformants were capable of growing phototrophically on minimal media.

Transformants were shown to reach a homoplasmic state (Appendix L) by PCR as described in 2.3.17.

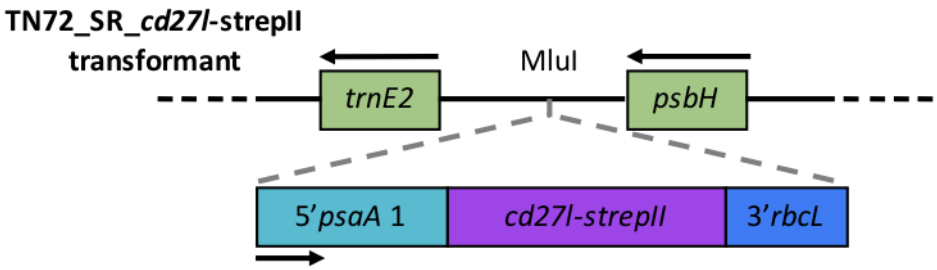


Figure 5.5 TN72 recipient strain successfully transformed to carry *cd27l*-StrepII under the *psaA* exon 1 promoter/5'UTR in the chloroplast, and with *psbH* functionality restored.

5.3.2.4.2 Confirmation by Western blotting

Whole cell preparations of TN72_SR_*cd27l*-StrepII, as well as TN72_SR_*cd27l*-HA, and TN72_SR_Control (as positive and negative controls respectively) were prepared and separated by SDS-PAGE. After blotting to nitrocellulose membrane and blocking as described in 2.4.5, the membrane was probed with anti-StrepII primary antibody, then incubated with IRDye® secondary antibody. Fluorescence was observed using the LiCor Odyssey® CLx system. The membrane was then re-probed with anti-HA primary antibody and IRDye® secondary antibody. The two images are shown in Figure 5.6.

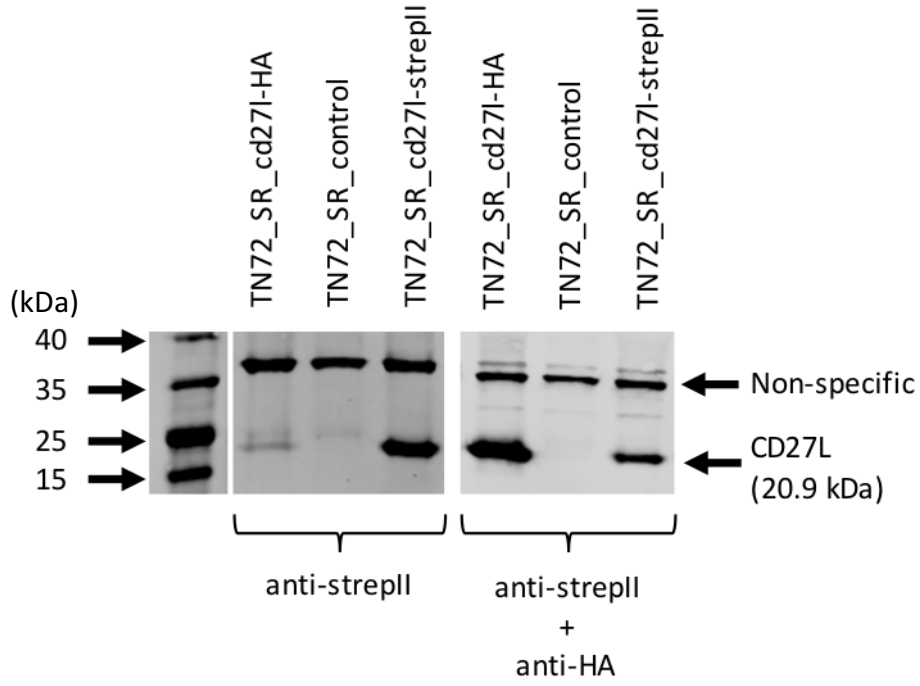


Figure 5.6 Western blot analysis of the membrane probed firstly with anti-StrepII primary antibodies (left), then additionally with anti-HA antibodies (right). CD27L accumulation was detected in the expected cell lines i.e. CD27L-StrepII was detected with the anti-StrepII antibody and the use of the anti-HA antibody also revealed the presence of CD27L-HA. The non-specific band confirms that a similar quantity of extract was loaded on to the gel for all three strains.

5.3.2.4.3 Relative protein accumulation levels

The use of the protein standard, described in 3.3.1.2.5, enabled the relative quantification of CD27L protein accumulation in the two *C. reinhardtii* lines. Figure 5.7 shows comparisons of the two whole protein cell preparations used in 5.3.2.4.2 and a 10 µg protein standard is included. The standard is tagged with HA and StrepII epitopes, enabling an approximate comparison of the protein accumulation seen in each sample.

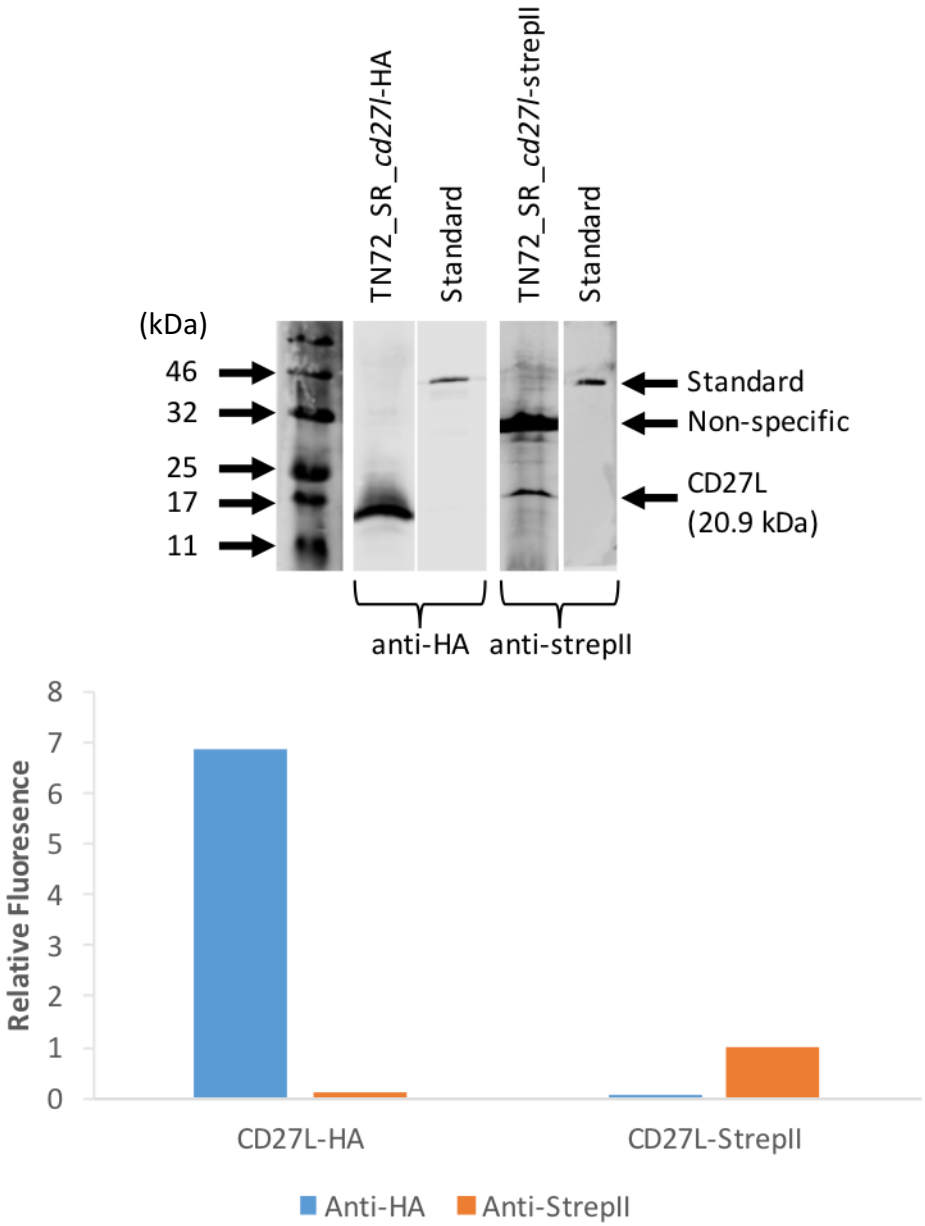


Figure 5.7 Western blot analysis of whole cell extracts, first normalised by culture OD_{740} and compared to a 10 µg protein standard. Quantification was performed using the LiCor Odyssey® CLx system, after which the standard was used to normalise the two blots and the signal from the 21 kDa bands compared in a bar chart.

The protein accumulation level of the HA-tagged protein appears to be approximately six times greater than the StrepII-tagged protein. However, as discussed in 3.3.2.2.4.3, the difference observed is possibly due to the different accessibility of the epitope tags on the Cpl_1-HA-StrepII protein standard. This is supported by that fact that this pattern also appears between the Cpl1-HA and Cpl1-StrepII accumulation levels and it seems unlikely that the StrepII tag is having such a consistently detrimental effect upon accumulation levels.

5.3.2.4.4 Enrichment of CD27L-StrepII using HiTrap StrepII Column

The inclusion of a StrepII tag on CD27L enables enrichment of the protein and will thus allow a clearer investigation into its antibacterial properties. *C. reinhardtii* cell extracts were prepared as described in 2.5.1, and ultracentrifuged at 100,000 g for 1 hour to remove cell debris. The resulting ultracentrifuged (UC) supernatant was then enriched for StrepII-tagged proteins using the commercially available HiTrap StrepII Column, as described in 2.5.3. Western blot analysis of the resulting fractions is shown in Figure 5.8.

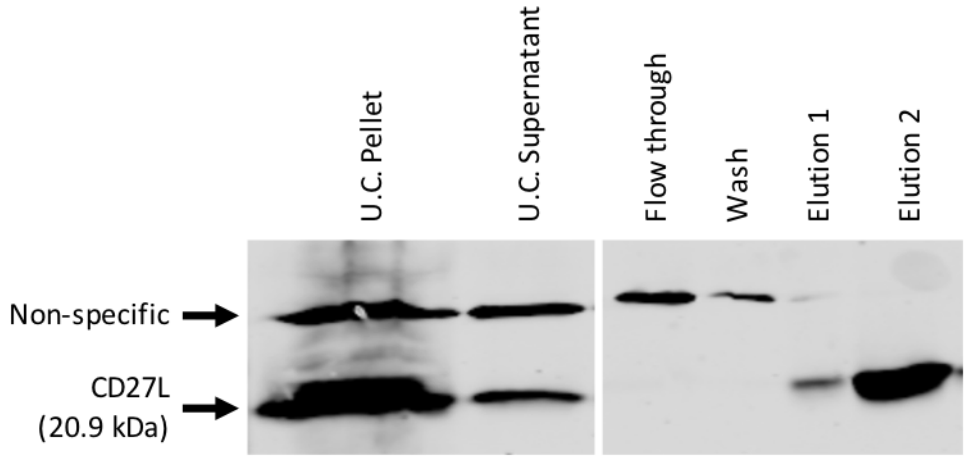


Figure 5.8 Western blot analysis using anti-StrepII primary antibody and IRDye® secondary antibody. Fractions from sample preparation and the HiTrap StrepII column were collected and separated by SDS-PAGE. The eluted fraction has no non-specific band, and 'Elution 2' was used for further experiments.

It is clear from Figure 5.8 that enrichment has been successful, and that the eluted fractions contain high quantities of CD27L and low quantities of the non-specifically recognised protein. It is also noticeable that a large proportion of the CD27L protein is lost in the ultracentrifuged pellet, but for the purposes of this investigation, this is not important. In the future, optimisation of this protocol may result in a greater yield.

5.3.2.4.5 Activity of enriched CD27L crude extract against stripped *E. coli*

The enriched CD27L extract was prepared as above (5.3.2.4.4) and an enriched Cpl-1 extract was used as a negative control. A Bradford assay was performed to measure the total protein in each sample.

The enriched extract was initially tested for activity using the stripped *E. coli* assay. 0.4 µg of each enriched extract, as well as 5 units of egg white lysozyme to act as a positive control, were added to a 96-well plate with stripped *E. coli*. The change in OD_{595} was measured over a period of 90 minutes at 37 °C. The resulting change in OD_{595} is displayed in Figure 5.3.

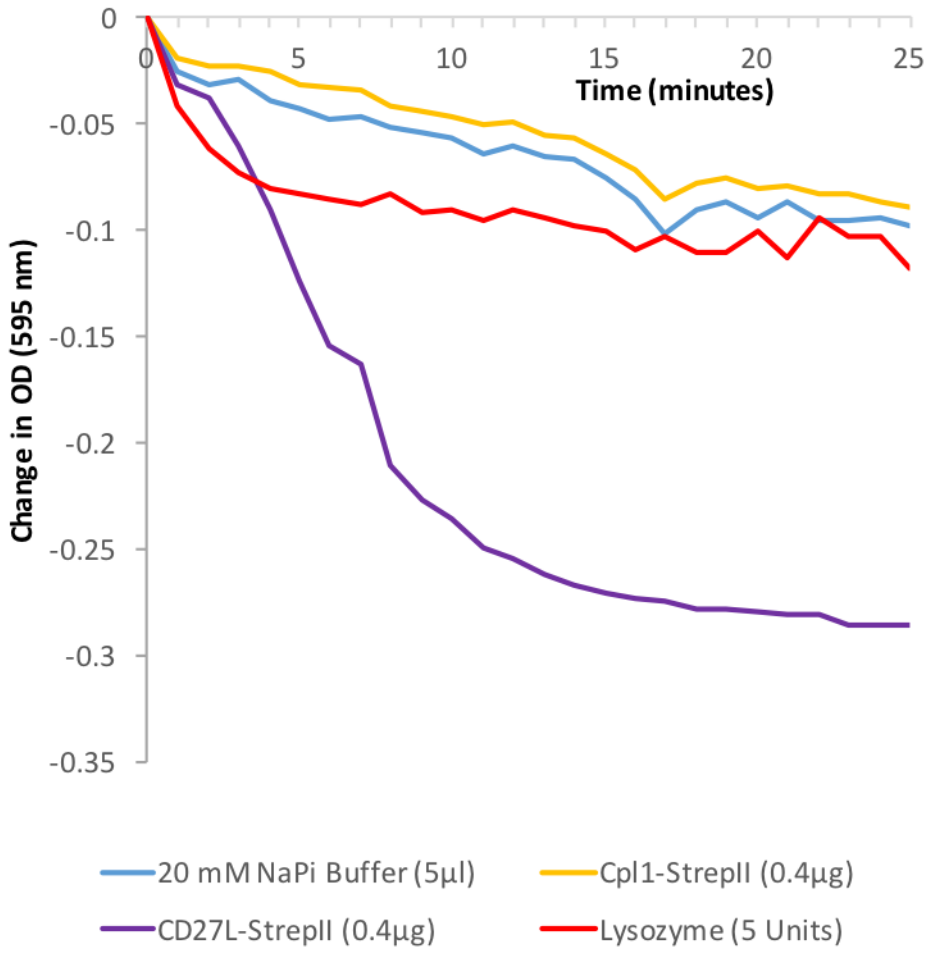


Figure 5.9 A turbidity reduction assay, measuring the reduction in OD_{595} in wells containing *E. coli* stripped of their outer-membrane and treated with enriched *C. reinhardtii* extracts. The Cpl-1 extract acts as a negative control. Only the first 25 minutes are shown, as the OD remains stable beyond this.

The enriched CD27L appears to be highly active, with the OD_{595} drop plateauing after approximately 15 minutes. The Cpl-1 enriched extract demonstrates the specificity of endolysins, and acts as a negative control in this instance. The 20mM NaPi buffer does not appear to produce the same level of cell lysis as observed in Figure 5.3,

probably because smaller sample volumes were used in this instance – around 5 µl as opposed to ~30 µl. There is little difference observed between the Cpl-1 and NaPi buffer, suggesting that this enrichment has removed many of the other proteins contributing to the osmotic pressure effect described previously. The lysozyme, which was included to act as a positive control, does cause some level of cell lysis but plateaus with an OD_{595} drop of approximately 0.1.

5.3.2.5 Activity of enriched CD27L crude extract against *C. difficile*

Given the success of enriched CD27L against *E. coli* shown in Figure 5.9, similar fresh samples were sent to Norwich for *in vitro* testing upon *C. difficile*. The turbidity reduction assay was performed as described in 2.7.2.3.

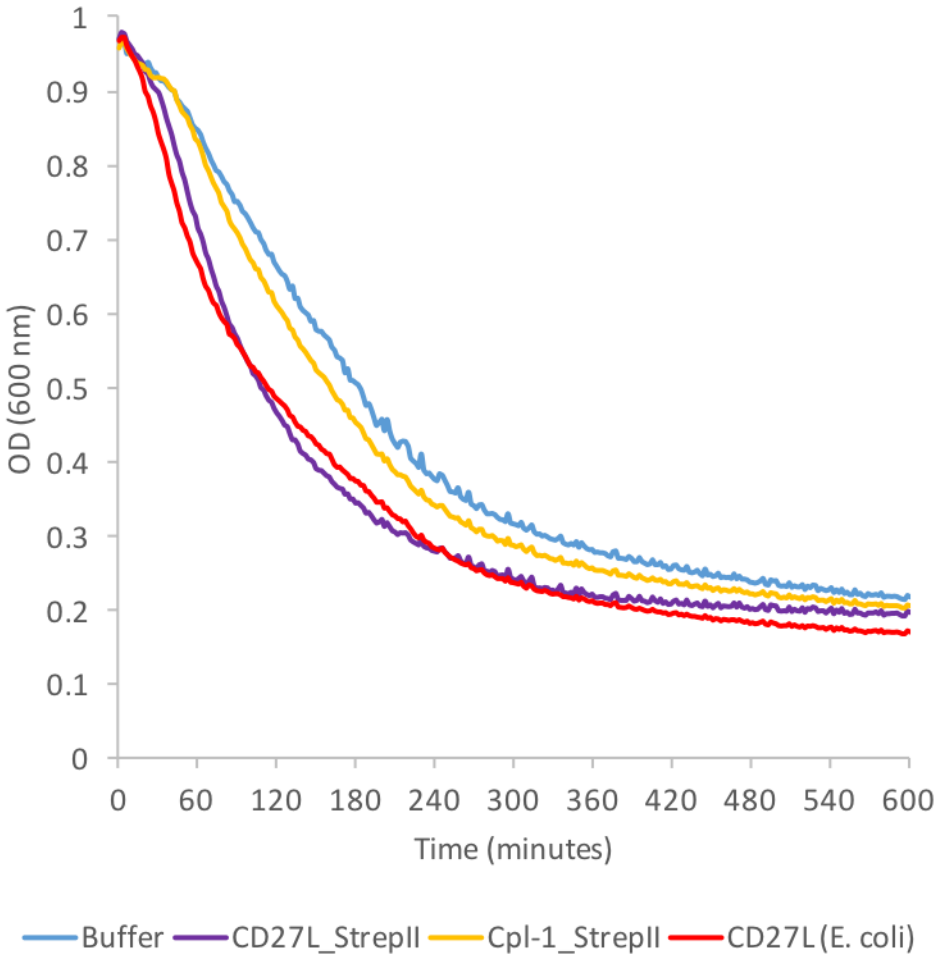


Figure 5.10 A turbidity reduction assay, measuring the reduction in OD_{600} in wells containing *C. difficile* challenged with enriched *C. reinhardtii* extracts. The Cpl-1 extract acts as a negative control. The buffer used is the Strep-Tactin® elution buffer provided by the HiTrap StrepII column kit, containing desthiobiotin (a biotin analogue). 0.3 µg of each enriched extract were used. The CD27L (*E. coli*) refers to a purified CD27L recombinant protein produced by Dr Mayer in *E. coli,* to act here as a positive control. This turbidity reduction assay was performed by Dr Melinda Mayer at Quadram Institute Bioscience, Norwich, UK.

The data show convincing lysis of *C. difficile* cells *in vitro*, despite the fact that autolysis appears to be relatively high, as is characteristic of *C. difficile* when incubated for extended periods of time in oxygenated conditions, as *C. difficile* is an obligate anaerobe (Mayer et al. 2011). The samples were, unfortunately, much more dilute than Dr Mayer commonly uses for turbidity reduction assays, and in the future a technique such as ammonium sulphate precipitation may enable more endolysin (~10 µg rather than 0.3 µg) to be added to each well.

The conclusions of all of these CD27L experiments is that another biologically active and stable endolysin can be produced in the *C. reinhardtii* chloroplast. The addition of a StrepII tag to the recombinant protein enables the fast and relatively cheap enrichment of crude extracts, which in turn has allowed us to observe more clearly the activity of the endolysin.

5.3.3 *Propionibacterium acnes,* bacteriophage PA6 and its endolysin Gp20

Propionibacterium acnes is a prevalent, Gram-positive, commensal bacterium known best for its role in the common skin condition acne vulgaris. Acne vulgaris is believed to affect over 80 % of adolescents and as such, antibiotics to treat it have been over-used and resistance has become commonplace (Leyden et al. 1983; Bojar and Holland 2004; Taylor et al. 2014). Some broad-spectrum antimicrobial agents, such as benzoyl peroxide, while still effective, have side effects that make their prolonged use unadvisable (Fulton et al. 1974; Bojar et al. 1995).

The treatment of acne vulgaris offers a large market and unmet need, which could possibly be served by endolysin treatment. As acne vulgaris is a topical affliction, it also presents a plausible target for treatment with a *C. reinhardtii* crude extract.

The bacteriophage PA6 is known to infect *P. acnes* and in 2007 its genome was sequenced (Farrar et al. 2007). The same study showed that PA6 was able to infect 32 different *P. acnes* isolates including antibiotic resistant strains. PA6 belongs to the *Siphoviridae*, the same virus family as Dp-1 (from which the endolysin Pal was isolated) and Φ11 (from which a *Staphylococcal* endolysin was isolated). The resulting genome analysis identified a gene encoding a putative amidase, named *gp20*. The identification of *gp20* as encoding an endolysin is further supported by the fact that the gene immediately downstream of it, *gp21*, appears to encode a holin: a classic endolysin-holin gene arrangement.

To date, there are no reports in the literature describing the successful expression of Gp20 as a recombinant protein, or a demonstration of peptidoglycan cleavage of *P. acnes* or cell lysis. However, a US patent (US_2012_0195872_A1) was filed in 2010 by Lysando (an aforementioned endolysin company) describing the fusion of Gp20 to an 18 amino acid peptide sequence, which was shown to be active against two *P. acnes* strains, presumably working in a similar manner to the Artilysins®, also discussed earlier (1.2.2.1).

5.3.3.1 Previous attempts to express Gp20 in the *C. reinhardtii* chloroplast

Several previous attempts to express Gp20 in the *C. reinhardtii* chloroplast have been made in the Purton lab. These have included insertion into the chloroplast in different expression cassettes, with different promoter/5'UTRs, and even as chimeric genes, but detection has not been achieved (Taunt 2013).

5.3.3.2 New attempt to produce a *C. reinhardtii* line expressing *gp20*

One variable that has not been altered however is the C-terminal HA tag. Epitope tags, such as the HA tag, are an extremely useful tool for detecting proteins for which no antibody is available, such as new proteins. Traditionally epitope tags are added to a protein terminus in the hope that they will not interfere with protein function. However, as seen in this thesis and elsewhere in the literature (e.g. Brault et al. (1999)), epitope choice and position can alter protein accumulation levels and function. A terminal tag remained a sensible choice, given that the structure of the Gp20 protein is unknown, so a simple switch from the C-terminus to the N-terminus of the protein was proposed.

5.3.3.2.1 Gene design and codon optimization of *gp20N*

The success of the pWUCA2 expression vector (which allows stop codons to be inserted into a gene-of-interest, thereby circumventing toxicity issues when cloning in *E. coli* (Young and Purton 2015)) in achieving similar levels of protein accumulation in the *C. reinhardtii* chloroplast, and the unknown toxicity of the Gp20 endolysin in *E. coli* led us to use the pWUCA2 plasmid as a precaution. *Codon Usage Optimizer Beta 0.92* (Kong 2013) was used to optimize the *gp20* gene for *C. reinhardtii* chloroplast expression. Three tryptophan codons spaced throughout the gene were changed from UGG to UGA. The HA tag coding sequence was included at the N terminus. SapI and SphI restriction sites were included for cloning into the pWUCA2 vector. The full-length *gp20N*-HA (the 'N' indicates that the HA tag is placed on the N terminus) gene was synthesized by Integrated DNA Technologies (Appendix M).

5.3.3.2.2 Expression cassette production

SapI and SphI restriction enzymes were used to digest and ligate the gene into pWUCA2 in frame, producing the plasmid pWUCA2_*gp20N*-HA. DH5α competent *E.*

coli cells were transformed with the resulting plasmid and sequencing confirmed the correct insertion.

5.3.3.2.3 *C. reinhardtii* transformation

The pWUCA2_*gp20N*-HA plasmid was used to transform recipient *C. reinhardtii* strain TN72 via the glass bead method as described 2.3.13. Figure 5.11 illustrates the resulting *C. reinhardtii* cell line, pWUCA2_*gp20N*-HA.

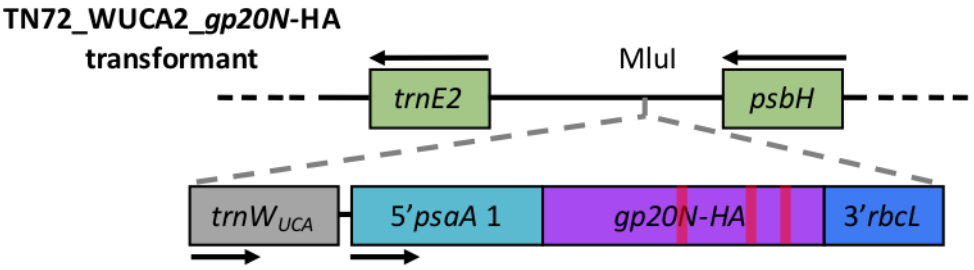


Figure 5.11 TN72 recipient strain successfully transformed to carry *gp20N*-HA under the *psaA* exon 1 promoter/5'UTR in the chloroplast, and with *psbH* functionality restored. A modified *trnW* gene has been co-introduced to translate the introduced UGA codons (indicated as red bars) in the *gp20N* gene as tryptophans.

The resulting transformants were confirmed for homoplasmicity by PCR as described in 2.3.17 (Appendix N). Finally, the correct insertion was confirmed by DNA sequencing as in 2.3.18.

5.3.3.2.4 Western blotting unable to confirm accumulation

To test for the accumulation of Gp20 in *C. reinhardtii*, whole cell preparations of TN72_WUCA2_*gp20N*-HA, TN72_A_*gp20-HA* (Taunt 2013), and TN72_SR_Control (as two negative controls) were prepared and separated by SDS-PAGE. After blotting to nitrocellulose membrane and blocking as described in 2.4.5, the membrane was probed with anti-HA primary antibody, then incubated with IRDye® secondary antibody. Fluorescence was observed using the LiCor Odyssey® CLx system. The resulting image is shown in Figure 5.12.

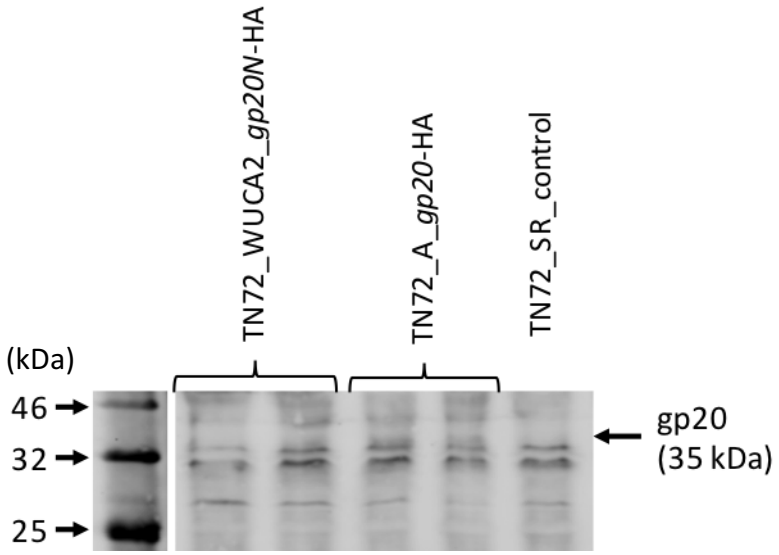


Figure 5.12 Western blot analysis of whole cell extracts. TN72_SR_Control acts as a negative control, and TN72_A_*gp20*-HA demonstrates the previous attempt made by Taunt (2013) to express *gp20* in the *C. reinhardtii* chloroplast.

No protein can be seen in Figure 5.12 at 35 kDa, the expected size for Gp20. The possible reasons for which are discussed in the following section.

5.3.3.3 Possible causes for *gp20* expression failure

As alluded to earlier in this thesis, failure of recombinant protein expression is quite commonplace in the *C. reinhardtii* chloroplast as the platform is still in the relatively early stages of development. Two possible reasons for failure to detect Gp20 are either the failure of translation or instability of the protein. This same conclusion was drawn by Taunt (2013), when *gp20* failed to express in *C. reinhardtii* on a previous occasion, with transcription issues and protein degradation being considered as unlikely due to previous literature suggesting that transcripts are usually present in failed recombinant protein expression attempts (Rochaix 2001), and endolysins having evolved to remain stable in a prokaryotic environment (Oey et al. 2009). Significant efforts were made by Taunt (2013) to investigate and ameliorate failures of translation. These efforts were focussed upon the ribosome binding site. Given that the upstream *atpA* promoter/5' UTR used by the pASapI expression cassette are known to be functional (See Cpl-1 work 3.3.2.2.1.4), it was supposed that the downstream box was playing an important role and in the instance of *gp20*, was hindering ribosome binding. This was investigated by including the first 34 codons of

the endogenous *atpA* gene followed by the GOI, to create a fusion (separated by a stromal processing peptidase). The plasmid used to perform this was named pASap2 and was created by Dr Chloe Economou (Taunt 2013). However, this strategy also failed to produce any visible level of protein accumulation.

5.3.3.4 Production of an *E. coli* line expressing Gp20

The failure to detect Gp20 accumulation in the *C. reinhardtii* chloroplast after attempting several strategies led us to investigate whether Gp20 was stable and active when expressed as a recombinant protein in *E. coli*. A previous investigation by Taunt (2013) failed to observe expression of Gp20 in *E. coli* lines transformed with pASapI containing the *gp20* gene, believing a lack of optimisation of the *atpA* promoter/5'UTR for *E. coli* expression to be responsible (Wannathong et al. 2016). We investigate this further here, as if protein accumulation can be achieved in *E. coli*, then it suggests that the protein is likely to be stable in *C. reinhardtii* and that translation is likely to be the key issue in the failure to detect the recombinant protein.

5.3.3.4.1 Expression cassette production

The *gp20*N-HA gene used previously could not be reused for expression in *E. coli* owing to the presence of stop codons in the gene, as it was originally designed for use with the pWUCA2 plasmid. A *gp20* gene with a C-terminal HA tag, codon optimized for *Synechocystis* expression in another project and synthesized by Integrated DNA technologies, was used as a template. Primers were designed to introduce a NcoI restriction site upstream of the ATG. A BamHI site already existed below the stop codons so did not need modifying. The introduced mutations are shown in lowercase, while the start and stop codons are underlined.

gp20_NcoI.F: CGTATACcc<u>ATG</u>GTGCGTTA

gp20.R: GATTGCGGATCC<u>TTATTA</u>GGC

The PCR product was digested and ligated into the high expression, IPTG inducible, plasmid pProEX HTb (Figure 5.13) and transformed into *E. coli* strain DH5α.

Figure Removed

Figure 5.13 Structural plasmid map of pProEX HTb and details of the multiple cloning site. Reproduced from http://www.biofeng.com/zaiti/dachang/pProEXHTb.html.

5.3.3.4.2 Coomassie stain confirmation of expression

Synthesis of the recombinant protein was induced using IPTG as described in 2.6.1. *E. coli* lysates were then created using BugBuster™ as described in 2.6.2 for both induced and uninduced cultures and prepared for protein analysis. The preparations were separated by SDS-PAGE and stained with Coomassie Brilliant Blue as described in 2.4.4. The stained gel was analysed using the LiCor Odyssey® CLx system and the resulting image is displayed in Figure 5.14.

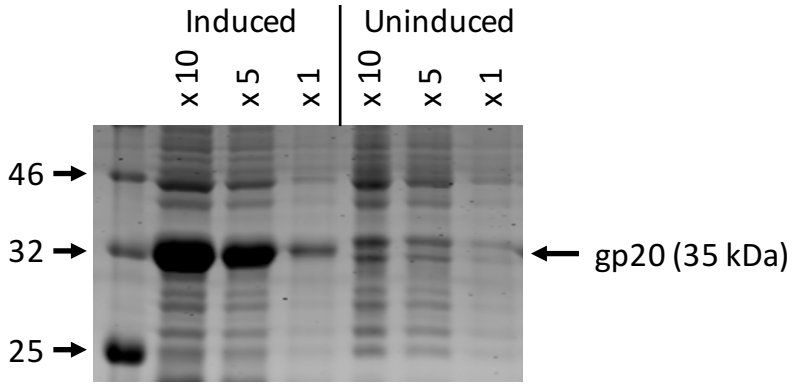


Figure 5.14 Whole cell preparations of induced and uninduced *E. coli* transformed with pProEX_*gp20*-HA were separated by SDS-PAGE and total protein staining was performed using a Coomassie stain.

High accumulation is observed in the induced *E. coli* at the expected size for Gp20, which is ~35 kDa. There is no protein seen in the uninduced samples which gives us confidence that the observed protein is indeed Gp20.

5.3.3.5 Activity of Gp20

A crude lysate of the induced *E. coli* expressing *gp20* was produced using a Stansted 'Pressure cell' homogeniser (2.6.3), and centrifuged at 5,000 g for 20 minutes to remove cell debris. The lysate was then filter sterilized using a 0.22 µm filter to remove any remaining viable *E. coli* cells. The same procedure was applied to a culture of untransformed DH5α cells as a negative control. Untransformed DH5α were used instead of uninduced DH5α_ProEX_*gp20*-HA to avoid issues around leaky expression of *gp20* in the uninduced control.

5.3.3.5.1 Crude lysate for activity assay

The pellet, lysate and filter sterilized lysate were prepared and separated by SDS-PAGE. After blotting to nitrocellulose membrane and blocking as described in 2.4.5, the membrane was probed with anti-HA primary antibody, then incubated with IRDye® secondary antibody. Fluorescence was observed using the LiCor Odyssey® CLx system. The resulting image is shown in Figure 5.15.

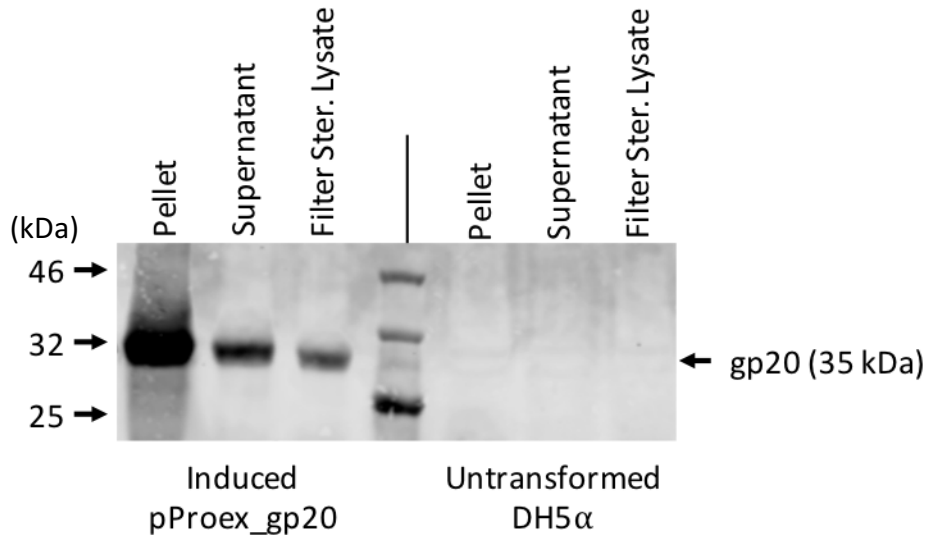


Figure 5.15 Western blot analysis shows HA-tagged Gp20 in all three lanes of the induced DH5α_ProEX_gp20-HA lysate, and not in the negative control lanes.

Although it appears that a large proportion of Gp20 protein is lost in the pellet, there is still a detectable quantity in the filter sterilised lysate. This is an encouraging result, as it suggests that recombinant Gp20 can exist in a stable form. This finding reduces the likelihood that Gp20 instability is the major cause of the lack of detectable protein the *C. reinhardtii* chloroplast, and requires us to look once again at translation elongation as the most likely issue preventing accumulation.

5.3.3.5.2 Colony forming unit assay against *P. acnes*

In order to test whether the recombinant Gp20 is active, a colony forming unit assay against *P. acnes* was performed using the filter sterilized lysates as shown in Figure 5.15.

Cultures of *P. acnes* were grown anaerobically as described in 2.1.5.3. The cultures were centrifuged at 5,000 g for 20 minutes and resuspended in 20 mM NaPi buffer. 50 µl of bacterial suspension and 50 µl of filter sterilized lysate or buffer were incubated at 37 °C for 0, 30, 60, 90, 120 or 150 minutes. Three 10 µl samples were taken after the stated incubation time from each combination and spotted onto blood agar plates. After anaerobic incubation at 37 °C for one week, growth was visible and the plates were photographed, shown in Figure 5.16.

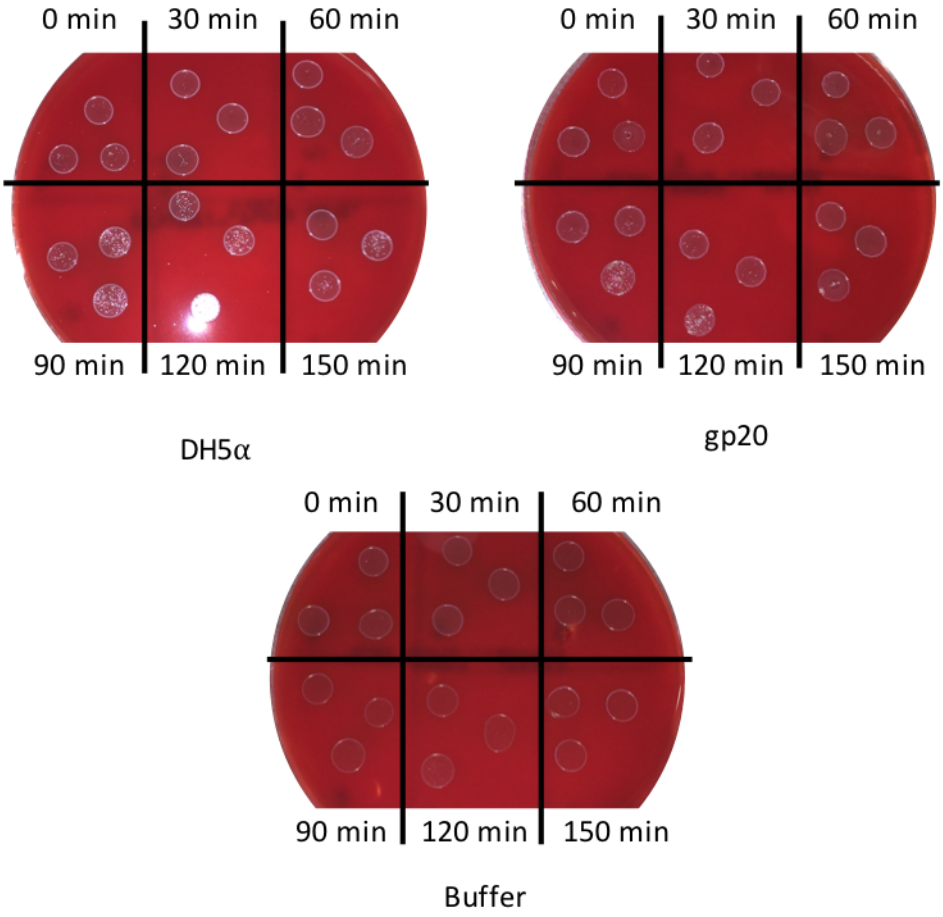


Figure 5.16 *P. acnes* growth on blood agar plates after incubation with crude *E. coli* lysates with and without Gp20, or just with 20 mM NaPi buffer. As is often the case with *P. acnes*, individual colonies tend not to be visible and a small lawn of cells appears instead (Stoffels 2015).

The presence of a lawn, rather than distinct colonies of *P. acnes* makes quantification of the antimicrobial effect of Gp20 difficult to measure. However, visually there appeared to be no difference in growth between any of the spots regardless of substrate and incubation time. In fact, the buffer alone appeared to reduce the lawn

the most, perhaps due to the osmotic pressure effects described previously in this thesis. Unfortunately, this experiment lacks a positive control, without which we cannot be sure that Gp20 is not working as, for example, the visible bacteria may be something other than *P. acnes* or the dosage might not be high enough.

This result tentatively suggests that Gp20, when produced as a recombinant protein in *E. coli,* does not display strong antibacterial properties towards *P. acnes*, however, as mentioned, the assay requires improvements for this conclusion to be more convincing. Since Taunt (2013) initially investigated *gp20* expression in *C. reinhardtii*, 13 other *P. acnes* targeting endolysins have been identified from other bacteriophages (Marinelli et al. 2012). These putative endolysins display remarkable sequence similarity, as do the whole bacteriophage genomes themselves. This is unusual for bacteriophage and unexpected, as usually a high level of genetic diversity exists amongst bacteriophage (Pride et al. 2012). The differences that do occur between endolysin sequences are mostly in the flexible linker region separating the two domains. This significant level of similarity led Marinelli et al. (2012) to suggest that any of the 14 endolysins chosen for recombinant production and endolysin therapy would be successful.

5.4 Conclusion and future work

This chapter has attempted to express two endolysins in the *C. reinhardtii* chloroplast. The first, CD27L, has been shown in the literature to be active when produced in *E. coli* and tested *in vitro* against *C. difficile* (Mayer et al. 2008). Here we have shown that CD27L can be expressed in *C. reinhardtii* to a relatively high level and is active *in vitro* against both stripped *E. coli* and *C. difficile*. In contrast, the second endolysin, Gp20, has no literature demonstrating its recombinant synthesis or activity *in vitro* or otherwise, and previous attempts to express *gp20* in *E. coli* and *C. reinhardtii* have failed, for unknown reasons (Taunt 2013). After failing to achieve expression in *C. reinhardtii* with a different epitope tag position, the focus shifted to achieving expression in *E. coli* and assaying for activity against *P. acnes*. Expression in *E. coli* was successful but activity was not satisfactorily assayed, as will be discussed further below.

5.4.1 CD27L endolysin

The investigation into expressing CD27L in *C. reinhardtii* has been relatively successful to date. Expression has been achieved and the use of the StrepII tag has enabled enrichment of CD27L from crude extracts, allowing more precise activity assays to be performed. Activity has been confirmed *in vitro* against both *C. difficile*, its natural target, and *E. coli* cells when stripped of their outer-membrane, which acts as a useful assay when *C. difficile* facilities are unavailable.

As with Cpl-1 in Chapter 3, attempts to quantify CD27L accumulation have produced varied results. The StrepII-tagged CD27L protein appears to accumulate to significantly lower levels than the HA-tagged version – a pattern also seen with Cpl-1. The finding that StrepII-tagged proteins consistently appear to accumulate to a lower level than their HA-tagged counterparts has led us to question the validity of the double-tagged protein standard produced in 3.3.1.2.6, as has been discussed in more detail in 3.3.2.2.4.3. For both this CD27L work and the Cpl-1 work, it is important to address this issue and find a suitable method of quantifying protein

accumulation, such as raising antibodies to CD27L itself and using the purified protein as a standard.

While the results shown in Figure 5.10 appear credible, demonstrating the activity of *C. reinhardtii*-produced CD27L against *C. difficile* requires a few minor modifications to the assay to make it more convincing. The first is to perform the assay with higher concentrations of endolysin. Dr Mayer usually includes 10 µg of endolysin per well, while we were only able to use 0.3 µg per well due to the diluted nature of the extract. This can be easily remedied by performing an ammonium sulphate precipitation upon the extract to concentrate the protein. It is expected that this improvement will widen the gap between the positive and negative controls, thus making the assay more convincing. Secondly, a distinction must be made between cell lysis and cell viability. While it is reasonable to assume that the two are related, a minor reduction in turbidity does not necessarily mean that cell viability is equally unaffected. For example, an endolysin may cleave the peptidoglycan in only a few locations, enough to lyse the cell, but leaving the wall otherwise mostly intact. This will register as a low level of turbidity reduction, but has reduced cell viability to the same level as an endolysin that breaks the wall into many pieces. In order to confirm the efficacy of an endolysin, it is therefore important to include a direct measurement of cell viability, such as a CFU assay. By plating the contents of the wells from the turbidity reduction assay onto nutrient agar it would be possible to quantify the remaining number of viable bacteria.

A recently discovered property of some endolysins, such as CD27L and CTP1L (Dunne et al. 2016) and Lys170 (Proença et al. 2015), is their encoding of a second, internal, translation start site. This second translation start site results in the additional synthesis of the C-terminal domain i.e. a free cell wall binding domain. This free CBD forms a multimeric endolysin with the full-length endolysin to enable it to become fully active. It was noted by Dunne et al. (2016) that the recombinant endolysin CTP1L produced from a codon-optimised gene displayed significantly lower activity than that produced from the wild-type nucleotide sequence. The codon optimised version produced much lower levels of free CBD, and the addition of recombinantly synthesised free CBD rescued activity levels.

The *cd27l* gene expressed here in *C. reinhardtii* is the truncated version of the full-length gene, as discussed in 5.3.2.1, and as such does not contain the CBD or the putative internal ribosome binding site identified by Dunne et al. (2016). Dunne et al. (2016) noted that the free CBD of CD27L was not necessary for activity, but postulated that it may play a role when the endolysin is under physiological conditions which have yet to be investigated.

It would be interesting to express the full-length CD27L endolysin in *C. reinhardtii*, using a codon optimised version but ensuring that the Shine Dalgarno (SD) and RBS of the putative internal translation site remain unchanged. This would answer the question of whether the *C. reinhardtii* chloroplast machinery is able to recognise this internal translation site. It seems unlikely that the internal site will be recognised by the chloroplast machinery, as although the chloroplast system does share many features with its prokaryotic ancestors, translation initiation is less well understood in the *C. reinhardtii* chloroplast. For example, although chloroplast 16S rRNA molecules display a highly conserved anti-SD sequence near their 3' end, the complementary SD sequences on chloroplast mRNA UTRs are hypervariable, if present at all (Rochaix et al. 1998). A C-terminal StrepII tag on the full-length CD27L protein will enable enrichment and detection of both the full-length protein and any product synthesised from the internal translation site. If the internal translation site fails to produce a free CBD, then it would also be of interest to express the CBD independently, under an alternative promoter/5'UTR.

The investigation by Dunne et al. (2016) concluded that the free CBD oligomerised with the full-length CD27L, but for this endolysin it did not appear to greatly affect activity. Therefore, it may be more relevant to look at CTP1L, an endolysin that targets *Clostridium tyrobutyricum,* and was shown to require the free CBD for activity. The potential for CTP1L as a future candidate for *C. reinhardtii* expression will be discussed later in this chapter.

Usually, at this stage of an endolysin project, most further work would involve the use of an animal model for *in vivo* testing. However, in the case of *C. difficile* models, the most helpful next stage is probably more advanced *in vitro* testing, using, for

example, a triple-stage chemostat model of the human gut. The reason for this is that the most well-used animal model for CDI, the Syrian Hamster, is not fully standardised and presents several important and problematic differences in pathophysiology to human infection (Best et al. 2012). The triple-stage human gut model was designed by Macfarlane et al. (1998) and has since been improved and updated, but essentially consists of three connected, pH controlled, anaerobically-maintained vessels, with growth media fed into the top at a specific flow rate. The model emulates the characteristics (spatial, temporal, nutritional and physicochemical) of the proximal to the distal bowel (Best et al. 2012). The triple-stage chemostat human gut model has been used extensively for testing *C. difficile* treatment efficacy (Freeman et al. 2007; Baines et al. 2008) and will enable various questions around the CD27L endolysin to be asked and answered, such as: how quickly does the endolysin degrade/denature in the human gut environment? How much endolysin is required to achieve a reduction in *C. difficile* bacterial load and how often must doses be given? What delivery mechanism is likely to enable the delivery of CD27L to the key *C. difficile* infection sites? How does CD27L compare to vancomycin treatment? Can CD27L act synergistically with vancomycin treatment? However, major issues with even an advanced *in vitro* model still exist, such as the absence of immunological influence. To address this, eventually some form of *in vivo* model will doubtless be required, but the use of advanced *in vitro* testing should reduce the reliance upon these and inform more precisely which tests need to be carried out.

An interesting final consideration of *C. reinhardtii*-produced CD27L is the delivery mechanism. To date, *C. reinhardtii*-produced therapeutics have been considered most viable for topical applications due to the potential cost savings made by using a crude extract, on the assumption that a crude extract would not achieve FDA approval for systemic use. However, *C. difficile* has recently emerged as an increasingly prevalent pathogen in food animals (Songer 2004; Keel et al. 2007), such as pigs, calves and ostriches as well as pets such as cats, dogs and horses (Gould and Limbago 2010). These offer a potential market that may be less tightly regulated and enable *C. reinhardtii* produced therapeutics to show a proof-of-concept. Work by Dr

Priscilla Rajakumar in the Purton Lab investigated the administration of oral vaccines produced in *C. reinhardtii* to chickens and fish, and discussed the use of the *C. reinhardtii* cell as a delivery capsule (Rajakumar 2016). While this concept has not been practically tested, Rajakumar (2016) suggests that the cell wall and the chloroplast of *C. reinhardtii* protect the therapeutic recombinant proteins from degradation through the upper levels of the digestive system, enabling them to reach the gut. This will, of course, depend upon which animal is being treated, as for example the chickens studied by Rajakumar (2016) have a very short intestinal tract, while the same strategy may not work in cattle.

5.4.2 Gp20 endolysin

The Gp20 endolysin has been problematic since recombinant expression was first attempted by Taunt (2013). As an endolysin targeting a topical and prevalent bacterial species, it would be an ideal candidate for production in *C. reinhardtii* as a treatment for acne vulgaris. However, previous attempts to achieve *gp20* expression in the *C. reinhardtii* chloroplast, as well as new attempts described in this chapter, have all failed to produce detectable levels of expression. To test whether the Gp20 protein is stable and active outside the *P. acnes* intracellular environment, the endolysin was expressed in *E. coli*. In *E. coli* Gp20 accumulated to high levels, but colony forming unit assays using a crude lysate against *P. acnes* failed to show any activity. Given high levels of interest in *P. acnes*, and the identification of at least 14 relevant putative endolysins, it does seem remarkable that no reports of a recombinantly produced endolysin targeting *P. acnes* exist, other than the aforementioned patent application by Lysando. Here we will discuss in more detail the possible reasons for issues faced in expression and activity, and potential future lines of investigation.

After failed initial attempts to express *gp20* in the *C. reinhardtii* chloroplast, Taunt (2013) investigated whether translation initiation was the issue by creating a *cpl1*:*gp20* fusion, but again expression failed, while a positive control *cpl1*:*pal* succeeded. This suggested that the ribosome was stalling at some point within the *gp20* gene, an issue most commonly associated with codon usage. However, further

codon optimisation failed to solve the problem. Taunt (2013) was also unable to detect *gp20* expression in *E. coli*, leading to a suggestion that rapid protein turnover may be to blame. However, strong *gp20* expression in *E. coli* was described earlier in this chapter, making this a less likely suspect, although admittedly the *E. coli* and chloroplast environments are not identical.

As mentioned earlier, the remarkable similarity between the 14 *P. acnes* targeting endolysins described so far means that testing the expression of a different endolysin is unlikely to make any difference. The potential demand for solutions to *P. acnes* infection from both the cosmetic and pharmaceutical industries means that this endolysin remains worthy of investigation.

The alignment performed by Marinelli et al. (2012) (Figure 5.17) provides a useful indication of the two endolysin domains, as the area of least homology is likely to be the glycine rich linker domain (circa position 200). This is further supported by database comparisons predicting that the N-terminal domain is associated with N-acetylmuramoyl-1-alanine amidase activity (Farrar et al. 2007). Given that endolysin modules generally remain stable when separated (Mayer et al. 2008; Dunne et al. 2016), it would be interesting to attempt to express the two separate domains in the *C. reinhardtii* chloroplast. This would give an insight into where the ribosome is stalling, and could potentially be subjected to further codon optimisation to minimise this. Furthermore, as demonstrated with CD27L, occasionally the endolysin catalytic domain alone is sufficient for cell lysis which could be tested here if expression can be achieved.

Figure Removed

Figure 5.17 Protein sequence alignment of 14 putative endolysins from *P. acnes* bacteriophages. The alignment shows that the sequence is highly conserved, with the majority of differences located in the glycine-rich linker at around position 200. The C-terminus cell wall binding module is less highly conserved than the catalytic N-terminus module. Gp20 from the PA6 bacteriophage is the 4th from the bottom, labelled PA6_rev_gp20, as it is a revised annotation of the PA6 genome. Reproduced from Marinelli et al. (2012).

The next issue to resolve is the apparent lack of activity of the *E. coli*-produced Gp20 protein against *P. acnes*. The first question to address in this case is the CFU assay itself, which was not fully satisfying and was not repeated due to time constraints. The investigation requires a suitable positive control, the obvious choice being an antibiotic such as erythromycin, the first line of treatment in a clinical setting (Zaenglein et al. 2016). However, it would be preferable to have a more specific antibacterial, which will partially confirm that the bacteria seen are indeed *P. acnes* and not a contaminating species. The peptide adrenomedullin is secreted by human epithelial cells and has been demonstrated to have an antimicrobial effect upon several microorganisms found on the skin and respiratory tract, but is particularly effective against *P. acnes*, *M. luteus, P. gingivalis* and *H. pylori* (Allaker et al. 1999). It therefore represents a narrower-spectrum antimicrobial, when used at an optimal concentration, than erythromycin and may provide a useful positive control. To further confirm that the bacterium in question is *P. acnes*, Gram stain (+), catalase (+) and indole (+) tests, followed by sequencing of the 16S rRNA gene for species-level identification can be performed (Butler-Wu et al. 2011).

Assuming that the above assay is improved to an acceptable standard but activity is still not observed, several issues could be at play. The first is that *P. acnes* is known to secrete many protective enzymes which allow it to survive in the harsh human skin environment, including hydrolases and proteases (Holland et al. 2010), as well as propionic and acetic acid which inhibit the growth of competing species (Wilson 2005). It is plausible that some combination of these factors is causing the endolysin to denature or be degraded before it can bind and lyse the cell wall. As the endolysins have evolved to be active from within the bacterial cell, it is possible that such extrinsic barriers are blocking its action. Another possibility is that the holin, which is normally expressed alongside the endolysin, is required for cell lysis, although this seems unlikely and has not been reported before, to the author's knowledge.

As seen earlier in this chapter, while most endolysins consist of a simple modular monomer, others such as PlyC are multimeric, and still others such as CTP1L require oligomerisation with their own truncated C-terminal domain. It is possible that the lack of activity observed with Gp20 is due to the requirement to oligomerise like

CTP1L for full activity. Figure 5.17 shows a conserved methionine residue at approximately position 193 which could be an internal start codon. Also at position 210 in Figure 5.17 the alternative start codon, GTG, codes for a valine. This residue is conserved as either a methionine or a valine in the other putative endolysins, however an analysis of the nucleotide sequence is required to identify whether this is a potential internal start codon. This location is similar to that observed in CTP1L and CD27L and would similarly result in the synthesis of a free CBD. However, no Shine Dalgarno consensus can be seen upstream of either the methionine-encoding ATG or valine-encoding GTG when the nucleotide sequence of *gp20* is examined, and analysis using the NCBI Open Reading Frame Finder (https://www.ncbi.nlm.nih.gov/orffinder/) software did not identify either as a potential ORF (Figure 5.18).

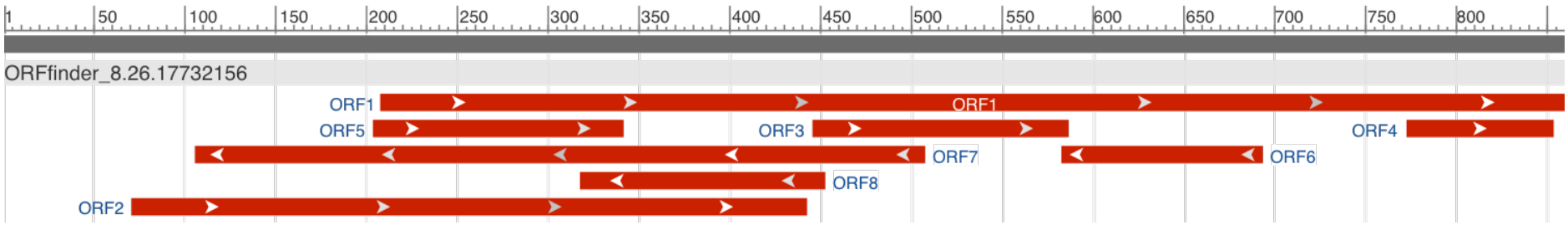


Figure 5.18 Results from https://www.ncbi.nlm.nih.gov/orffinder/ with the putative endolysin sequence gene ID 5247034. Only ORF1 is in frame and in the same direction as the full-length endolysin. ORF1 is 654 nucleotides in length, producing a 217 amino acid protein.

The internal ribosome binding site and ORF identified in Figure 5.18 would result in the synthesis of a free CBD, like CTP1L, but includes a much larger portion of the catalytic domain than would be expected. Nevertheless, it would be interesting to attempt to express *gp20* in *E. coli* using the wild-type nucleotide sequence, which may identify whether an internal ribosome binding site is active. Such a finding may result in the improvement of activity against *P. acnes*.

5.4.3 Future endolysins

As discussed earlier in this chapter, the problems surrounding Gp20 cannot be solved by merely attempting to express an alternative *P. acnes* targeting endolysin, as the level of homology between the proteins is so high that it is extremely unlikely that this will make any difference. In this case the strategy needs to be to better understand why the endolysin is failing to express and attempting to solve any problems with activity.

There are several other endolysins which would be interesting to express in the *C. reinhardtii* chloroplast in the future. CTP1L is an interesting endolysin due to its requirement to oligomerise with its free CBD, but is also a commercially relevant endolysin, as it targets *Clostridium tyrobutyricum. Clostridium tyrobutyricum* contamination can lead to extensive costs in the cheese production industry through 'late-blowing', a phenomenon affecting high-pH cheeses such as Edammer and Gouda, and resulting in excess bubbles and an abnormal flavour. An application of CTP1L is therefore in food safety and production, and given that *C. reinhardtii* has GRAS status, this may well be a very well suited market for *C. reinhardtii*-produced endolysins.

Work by Jado et al. (2003) demonstrated that two *S. pneumoniae* endolysins, Pal and Cpl-1, are able to act synergistically. There are numerous advantages to exploiting drug synergies therapeutically, such as lower dosages, greater efficacy, and minimising the opportunity for resistance to occur. It would be interesting to investigate whether a similar synergy can be observed with endolysins targeting other bacteria, such as *C. difficile*. Wang et al. (2015) identified another *C. difficile* endolysin, named PlyCD, which was demonstrated to be active against *C. difficile in*

vitro and, similarly to CD27L, more active in its truncated form, $PlyCD_{1\text{-}174}$. Successful expression of PlyCD in the *C. reinhardtii* chloroplast would allow testing of both PlyCD as a stand-alone endolysin, as well as enabling an investigation into whether a synergistic effect can be observed between PlyCD and CD27L.

As a final note on selecting future endolysins, one method might be termed a 'disease-orientated' approach. This approach looks at the bacterial species associated with certain identifiable diseases and designing 'endolysin cocktails' to kill all of those species. For example, the skin disease impetigo is typically treated with a broad-spectrum antibiotic cream (George and Rubin 2003), despite the fact that it is known to be caused by one of only two bacterial species: *Staphylococcus aureus* or *Streptococcus pyogenes.* It is therefore conceivable that a cocktail of two or more endolysins targeting *S. aureus* and *S. pyogenes* would be a viable treatment. This particular example, as a topical application, would again be well suited to *C. reinhardtii* production. Two potential endolysins, in this case, could be LysH5 (targeting *S. aureus*) (Gutiérrez et al. 2014) and PlyC (targeting *S. pyogenes*) (Shen et al. 2013).

Overall, this chapter has demonstrated that while the *C. reinhardtii* chloroplast can be a very well suited platform for endolysin production, there remain several issues and unexplored areas which leave recombinant endolysin expression far from routine. However, non-expressing endolysin genes such as *gp20* offer an opportunity to understand more about factors affecting chloroplast expression, such as codon optimisation and translation elongation. The suggested future work laid out in this section may help us to progress the *C. reinhardtii* chloroplast platform towards achieving more reliable protein production.

Chapter 6 Final discussion and future prospects

6.1 Summary

6.1.1 Summary of results

The research described in this thesis sought to address two overarching issues. The first is the improvement of the genetic engineering technology underpinning the development of the *C. reinhardtii* chloroplast as a novel platform for recombinant protein production. The second is the advancement of endolysins as an alternative to conventional antibiotics.

Chapter 3 investigates the issue of relatively low levels of transgene expression seen in *C. reinhardtii* and demonstrates techniques available to improve this. As demand grows for protein therapeutics, such as monoclonal antibodies, hormones (e.g. human insulin), and vaccines, so does the requirement for suitable platforms for their recombinant production. As discussed in 1.3.1, there are various platforms currently used for therapeutic protein production, each with different advantages and limitations. The green alga *C. reinhardtii* offers a novel alternative cell factory, with important advantages over several more widely used platforms. However, several limitations of *C. reinhardtii*, including low levels of expression, need to be addressed before it can realistically compete with the established platforms. Initially, methods used to compare protein accumulation levels between different transgenic *C. reinhardtii* lines were investigated. The chapter highlights the limitations of comparing protein accumulation levels whilst clonal variation appears to be having such a large effect, as well as proposing and testing improved methods of comparison. As such, attempts to investigate the effects of codon pair optimisation remain inconclusive. However, the chapter does demonstrate that the introduction of two expression cassettes into the *C. reinhardtii* chloroplast plastome results in a robust two-fold increase in protein accumulation compared with a single expression cassette.

The growing threat of antibiotic resistance has encouraged research into developing alternatives to conventional antibiotics. Endolysins are of increasing interest and the first endolysin based therapies are now available (1.3.3.3). Chapter 4 investigates whether the activity of these endolysins can be improved against their target bacterium, *S. pneumoniae*, either by direct mutation of the protein or through synergistic effects with holins. Site directed mutagenesis was used to disrupt the cell wall binding domain of Cpl-1 and expression was achieved for all six mutants in *C. reinhardtii*. However, such modifications appeared only to decrease the activity of the endolysin, demonstrating that the functionality of the two active choline binding sites is essential for the activity of the endolysin in this instance. Dimerisation of Cpl-1 via a disulphide bridge was also investigated, but could not be achieved in the *C. reinhardtii* chloroplast. Several reasons for the failure of Cpl-1 to dimerise were suggested, along with possible solutions, mostly revolving around the steric hindrance of the C-terminal epitope tag. Finally, the holin protein Cph-1 was expressed in the *C. reinhardtii* chloroplast. However, expression quickly ceased and the possible reasons for this were discussed.

Chapter 5 describes attempts at increasing the suite of endolysins that can be produced in the *C. reinhardtii* chloroplast. The *C. difficile* targeting endolysin, CD27L, was successfully expressed in the *C. reinhardtii* chloroplast and its activity against *C. difficile* was demonstrated *in vitro*. Meanwhile, the *P. acnes* targeting endolysin, Gp20, failed to express in transformed *C. reinhardtii* lines. However, it was successfully expressed in *E. coli* under an inducible promoter, but activity could not be observed against *P. acnes*.

6.1.2 Summary of short-term future work

Future work has been roughly divided into two categories in this thesis: ‘short-term’ future work and ‘long-term’ future work. While both of these terms are relative, the goal of this system is to distinguish between work that will enable the progression or conclusion of current projects (short-term), and work that looks further to the future and would serve as a stand-alone project in its own right (long-term). Relevant short-

term future work has been suggested at the end of each results chapter, much of which is very specific. However, the most interesting area that each results chapter highlighted for further investigation is summarised briefly below. Potential paths for long-term future projects are discussed in 6.2 below.

The issue of clonal variation highlighted in Chapter 3 limits our ability to accurately investigate strategies for improving recombinant protein accumulation in the *C. reinhardtii* chloroplast, but also offers an opportunity to understand what leads to a very high or very low expressing 'clone'. In order to enable the investigation into subtler methods of improving recombinant protein accumulation, in 3.4.2 we suggest that the use of a reporter gene, such as luciferase, and the use of high-throughput techniques will enable 'comparison-of-means' analysis. Advancements in whole genome sequencing, and importantly its falling cost, may eventually enable a methodical whole genome analysis of high-expressing clones to identify mutations or insertions that are resulting in high levels of protein accumulation. These mutations might be affecting nuclear genes for chloroplast regulatory factors, proteases, or chaperones. These changes can then be purposefully introduced alongside future transgenes, thus improving *C. reinhardtii* as a protein production platform.

The endolysin Cpl-1 was the main focus of investigation in Chapter 4, and as such much of the proposed future work was specific to this endolysin. However, the unexpected and unexplained result that Cph-1 expression was achieved and then ceased is worthy of further investigation. Silencing mechanisms, while common in nuclear transformations, have not previously been described in the chloroplast (Daniell et al. 2005). It is therefore important to understand how the termination of expression has occurred, the findings of which may help prevent this from reoccurring, but also could become a useful molecular switch.

Chapter 5 demonstrated the successful expression and activity of a synthetic gene encoding CD27L, and also documented the unsuccessful attempts to express *gp20* in *C. reinhardtii*. However, it is the *gp20* result that remains interesting in terms of gaining a greater understanding of *C. reinhardtii* expression and its suitability as a platform for producing endolysins. Despite repeated attempts and failures to achieve

gp20 expression in the *C. reinhardtii* chloroplast, there is still some way to go before we can deduce the precise reason for this and attempt to ameliorate it. Codon optimisation, translation initiation, and epitope tag location have all been investigated and do not appear to be causing failure of expression. Attempting to express the two domains of the endolysin separately may shed some light on to the problem, and aid us in a broader understanding of expression failure in the *C. reinhardtii* chloroplast.

6.2 Discussion and long-term future prospects

6.2.1 Alternative strategies to improving expression levels in *C. reinhardtii* chloroplast

6.2.1.1 Synthetic promoters and 5'UTRs

Synthetic promoters and 5'UTRs are becoming widely used in synthetic biology and in recombinant protein production, due to their frequent ability to provide more consistent and higher final protein yields than endogenous regulatory elements (Gilman and Love 2016). As discussed previously in this thesis (3.1.2), chloroplast gene expression appears to be mostly controlled at the level of translation in *C. reinhardtii*, rather than transcription, and as such the improvement of 5'UTRs is of greater interest than that of promoters at this stage. To date, the most consistent levels of gene expression in the *C. reinhardtii* chloroplast have been achieved using endogenous promoter/5'UTR sequences from highly expressed proteins, and in some cases higher levels have been achieved through combinations with nuclear mutants (Michelet et al. 2011) and altered CES mechanisms (Manuell et al. 2007) which will be discussed is the next section (6.2.1.2).

Research by Specht & Mayfield (2013) used a luciferase reporter gene (Figure 6.1) and high-throughput analysis to compile an unbiased dataset of mutated *psaA* and *psbD* 5'UTRs, from which the first successful synthetic *C. reinhardtii* chloroplast 5'UTR was designed. The resulting synthetic 5'UTR improved protein accumulation by approximately 66 %, compared to endogenous 5'UTRs.

Figure Removed

Figure 6.1 *C. reinhardtii* transformed with expression cassettes containing mutated 5'UTRs controlling luciferase expression in the chloroplast. This false colour image illustrates the differences in luciferase expression under different 5'UTRs. Techniques such as this enable the processing and expression quantification of hundreds of *C. reinhardtii* colonies in a very small amount of time. The positive control (wild type UTR) and negative control (no-luciferase) are indicated on the bottom left of the plate. Reproduced from Specht & Mayfield (2013).

As high-throughput techniques become more accessible and defined in *C. reinhardtii*, the potential for producing large unbiased datasets improves. The availability of such datasets may enable the identification of consensus motifs in both the promoter and 5'UTRs of endogenous genes and pave the way towards producing synthetic promoters and 5'UTRs that dramatically improve protein yields as well as providing consistent levels of gene expression. The work by Specht & Mayfield (2013) is the first step in achieving this, but their use of particle bombardment to transform *C. reinhardtii* recipient lines remains subject to the issue of clonal variation, and as such very precise deductions about 5'UTR consensus sequences are difficult to draw and an even larger dataset is likely to be required. Recent advancements in precision gene editing may provide the solution to this issue, as will be discussed in the following section.

6.2.1.2 The potential of CRISPR/Cas9 gene editing

As an alternative to modifying the 5'UTR which is regulated by various elements in the *C. reinhardtii* chloroplast (a *cis*-acting element), one could also modify the

regulatory elements themselves (*trans*-acting factors). Such an experiment was performed by Michelet et al. (2011), where chloroplast transgenes were shown to be upregulated when expressed in a cellular background that included a nuclear mutation. The mutation in this instance was within a *trans*-splicing factor that was imported to the chloroplast from the nucleus, thus preventing *trans*-splicing in the chloroplast. However, this mutation also led to a loss of photoautotrophy meaning that the *C. reinhardtii* cells must be grown in the dark and require acetate as a carbon source. The discovery of CRISPR/Cas9 gene editing, and the recent application of this technology to the *C. reinhardtii* nucleus (Baek et al. 2016; Shin et al. 2016), will result in experiments such as that described by Michelet et al. (2011) becoming far easier to perform since targeted knock-outs of nuclear genes would be possible. Rather than knocking out *trans*-acting elements such as splicing factors which, as described, result in other undesirable phenotypes, CRISPR/Cas9 could be used to knock-out other factors that indirectly control protein accumulation, such as chloroplast-targeted proteases. The ease with which such mutations could be introduced should quickly enable researchers to manipulate the chloroplast environment into one that is more amenable to recombinant protein accumulation without having large detrimental effects on cell physiology.

6.2.1.3 Expression of novel chaperones

As discussed earlier, the primary suspect for causing low rates of successful recombinant protein expression in the *C. reinhardtii* chloroplast has been translation initiation and elongation. However, poor protein stability or high proteolysis could also be playing a role in this. Misfolded proteins and those that have not reached their native state are targets of proteases. While knocking out proteases may be part of the solution to improving protein accumulation, the quality of protein being produced is equally important, and disruption of proteases may have other detrimental effects upon the cell (Ramundo and Rochaix 2014).

While recombinant protein yields in *E. coli* have been steadily improved over the past four decades, through a process of largely trial and error, we must attempt to reproduce these advances in *C. reinhardtii* by analysing those improvements and

specifically introducing. Many of the successful improvements in *E. coli* were due to the co-expression of chaperones (Makino et al. 2011), and therefore this is worthy of investigation. Many chaperones are known to be expressed in the *C. reinhardtii* nucleus and targeted to the chloroplast to aid protein folding (Schroda 2004). However, some novel chaperones may have specific abilities to aid in the folding of certain recombinant proteins, and studies in *E. coli* have found that co-overproduction of certain chaperones in the *E. coli* cytoplasm have aided in protein solubility and yields (de Marco et al. 2007). It would, therefore, be of interest to identify relevant chaperones, such as Heat Shock Protein 70 (Willmund et al. 2008), and attempt to express them alongside endolysins in the *C. reinhardtii* chloroplast with the goal of improving protein accumulation and expression success rates.

6.2.2 Other microalgal hosts for scale-up

C. reinhardtii has been an important model organism for the study of photosynthesis, chloroplast biogenesis, and flagellar function for decades (Rochaix 1995). Given the years of research behind it and its well-defined genetics, *C. reinhardtii* was the obvious initial candidate for proving the concept of microalgae as a recombinant protein production platform. However, when it comes to scaling up microalgae production for commercial exploitation, other species may be more suitable, and research performed on *C. reinhardtii* should be readily transferable to these species. For example, some species such as those of *Spirulina*, *Chlorella*, *Haematococcus,* and *Dunaliella* have all been grown on commercial scales for some time (Figure 6.2) and are well suited to cultivation (Rosenberg et al. 2008).

Figure Removed

Figure 6.2 *Spirulina* growing in open ponds (raceways) for commercial production at Cyanotech's facilities in Hawaii[2].

The advantage to using one of the above microalgae is that the technology already exists for their large scale production, as well as data on growth optimisation and commercial viability. While yields of therapeutic proteins from *C. reinhardtii* are very low, hovering around 5-8 mg L^{-1} culture volume (1.3.3.3), the potential for cheap cultivation and high volumes goes some way towards negating this issue. In contrast CHO cells, as an example, must be grown in highly controlled environments in extremely expensive media, it is understandable their yield per litre must be very high to make this economically viable. It is therefore plausible to expect that moving to a new algal platform could be economically feasible without requiring vast increases in protein yield.

Other microalgae of interest include extremophiles, because their culture conditions generally minimise contamination, which is a big issue when growing microalgae at the commercial scales and open raceways shown in Figure 6.2. For example, *Cyanidioschyzon merolae* is a microalga that inhabits hot (40 – 56 °C) and acidic (pH 0.2-4) environments, and its chloroplast has recently been transformed for the first time (Zienkiewicz et al. 2017). Attempting to express recombinant proteins, such as Cpl-1, in *C. merolae* and similar organisms will be an important step in determining

[2] https://www.cyanotech.com/company/facility.html

the best option for commercial scale-up. However, use of high temperatures and extremophiles will require investigation into the thermostability of the proteins looking to be produced, and will not be suitable for all products. Furthermore, the cost of maintaining cultures at such temperatures may negate or minimize the savings made by adopting such a system.

6.3 Concluding remarks

6.3.1 Views on the future of antibiotics

As with many emerging technologies, the adoption of endolysins as antibiotics depends heavily upon the status of the existing technology: conventional antibiotics. The recent discovery of a new antibiotic, teixobactin, and the novel manner in which it was discovered (see 1.1.2.2) may yet extend the ruling era of small molecule antibiotics. While the discovery of new antibiotics is clearly a good thing, it is vitally important that complacency does not limit the research into endolysins and other antibiotic alternatives. Conventional antibiotics have clear disadvantages when compared to endolysins – namely the huge selective pressure they exert upon all bacteria to form resistance, thus making us increasingly dependent upon new discoveries. It is of great importance, therefore, that research continues into the development of more advanced antibiotics so that we can spread our reliance across several strategies rather than continuing to depend upon the discovery of new conventional antibiotics. With approximately 10^{31} bacteriophage in existence on the planet, there is a vast, and constantly evolving, supply of potential antibiotics in the form of phage-therapy and endolysins, offering a much more stable and sustainable source of future therapies.

For full adoption of endolysins and replacement of conventional antibiotics, both much improved point-of-care diagnostics, and much more research into tackling Gram-negative bacteria are required. Both of these are stumbling blocks that will take time and resources to overcome, but the final result will be a far superior solution to tackling bacterial infections. Furthermore, if endolysins, or any other technology, do become the first-line treatment for bacterial infections, research

should constantly strive to find alternatives, as the 'arms-race' with bacteria is never-ending, and resistance to any new technologies will eventually emerge.

6.3.2 Views on the future of microalgae-produced therapeutics

Hypothetical conversations with drug regulatory bodies and thought-experiments around the advantages of algae-based therapeutics are clearly important initial stages of developing a new recombinant protein platform, and currently more immediate issues such as low recombinant protein yields are preventing advancement of the platform. However, to perform a cost comparison of microalgae and other platforms, and to establish the viability of the platform, it is vitally important to understand the extent to which drug regulatory bodies will look favourably upon proteins produced in GRAS organisms such as *C. reinhardtii*. If it is the case that crude or enriched algal extracts can be approved, then this will be enormously encouraging for the future of the platform. However, if this is not the case, then the use of established platforms will almost always be more financially favourable. In this situation, the future of microalgae as recombinant protein production platforms must shift towards alternative markets, such as cosmetics, veterinary medicine, and food safety, and look to other proteins that cannot be expressed in other systems or have marked synergies with algal crude extracts. Other protein production platforms with GRAS status, such as spinach (Schulz et al. 2015), and possibly even the bacterial system *Lactococcus lactis* (Cano-Garrido et al. 2014), are likely to reach the point of commercial viability sooner than microalgae and this will offer valuable insights into how these proteins are received and regulated by the relevant regulatory bodies.

References

Adam, D. 2002. Global antibiotic resistance in Streptococcus pneumoniae. The Journal of antimicrobial chemotherapy 50:1–5.

Adam, Z., A. Rudella & K. J. van Wijk. 2006. Recent advances in the study of Clp, FtsH and other proteases located in chloroplasts. Current Opinion in Plant Biology 9:234–240.

Ahmad, V., M. S. Khan, Q. M. S. Jamal, M. A. Alzohairy, M. A. Al Karaawi & M. U. Siddiqui. 2017. Antimicrobial potential of bacteriocins: in therapy, agriculture and food preservation. International Journal of Antimicrobial Agents 49:1–11.

Ajuebor, J., O. McAuliffe, J. O'Mahony, R. P. Ross, C. Hill & A. Coffey. 2016. Bacteriophage endolysins and their applications. Science Progress 99:183–199.

Alkatib, S., L. B. Scharff, M. Rogalski, T. T. Fleischmann, A. Matthes, S. Seeger, M. A. Schöttler, S. Ruf & R. Bock. 2012. The Contributions of Wobbling and Superwobbling to the Reading of the Genetic Code. PLoS Genetics 8:11.

Allaker, R. P., C. Zihni & S. Kapas. 1999. An investigation into the antimicrobial effects of adrenomedullin on members of the skin, oral, respiratory tract and gut microflora. FEMS Immunology and Medical Microbiology 23:289–293.

Allen, H. K., J. Trachsel, T. Looft & T. A. Casey. 2014. Finding alternatives to antibiotics. Annals of the New York Academy of Sciences 1323:91–100.

Almaraz-Delgado, A. L., J. Flores-Uribe, V. H. Pérez-España, E. Salgado-Manjarrez & J. A. Badillo-Corona. 2014. Production of therapeutic proteins in the chloroplast of Chlamydomonas reinhardtii. AMB Express 4:57.

Aminov, R. I. 2010. A brief history of the antibiotic era: Lessons learned and challenges for the future. Frontiers in Microbiology 1:1–7.

Appelbaum, P. C. 1992. Antimicrobial Resistance in Streptococcus pneumoniae: An Overview. Clinical Infectious Diseases 15:77–83.

Arias, C. A. & B. E. Murray. 2015. clinical implications of basic research A New Antibiotic and the Evolution of Resistance:1168–1170.

Aw, R., G. R. Barton & D. J. Leak. 2017. Insights into the prevalence and underlying causes of clonal variation through transcriptomic analysis in Pichia pastoris. Applied Microbiology and Biotechnology 101:5045–5058.

Baek, K., D. H. Kim, J. Jeong, S. J. Sim, A. Melis, J.-S. Kim, E. Jin & S. Bae. 2016. DNA-free two-gene knockout in Chlamydomonas reinhardtii via CRISPR-Cas9 ribonucleoproteins. Scientific Reports 6:30620.

Baines, S. D., R. O'Connor, K. Saxton, J. Freeman & M. H. Wilcox. 2008. Comparison

of oritavancin versus vancomycin as treatments for clindamycin-induced Clostridium difficile PCR ribotype 027 infection in a human gut model. Journal of Antimicrobial Chemotherapy 62:1078–1085.

Barahimipour, R., J. Neupert & R. Bock. 2016. Efficient expression of nuclear transgenes in the green alga Chlamydomonas: synthesis of an HIV antigen and development of a new selectable marker. Plant Molecular Biology 90:403–418.

Bax, R. 1997. P. Antibiotic resistance: a view from the pharmaceutical industry. Clinical Infectious Diseases 24:5151–5153.

Berkmen, M. 2012. Production of disulfide-bonded proteins in Escherichia coli. Protein Expression and Purification 82:240–251.

Berry, A. M., R. A. Lock, D. Hansman & J. C. Paton. 1989. Contribution of autolysin to virulence of Streptococcus pneumoniae. Infection and Immunity 57:2324–2330.

Bertani, G. 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. Journal of bacteriology 62:293–300.

Best, E. L., J. Freeman & M. H. Wilcox. 2012. Models for the study of Clostridium difficile infection. Gut Microbes 3:145–167.

Bhatia, A., J.-F. Maisonneuve & D. H. Persing. 2004. Propionibacterium Acnes and Chronic Diseases. The infectious etiology of chronic diseases: defining the relationship, enhancing the research, and mitigating the effects: workshop summary.

Blowers, A. D., G. S. Ellmore, U. Klein & L. Bogorad. 1990. Transcriptional analysis of endogenous and foreign genes in chloroplast transformants of Chlamydomonas. Plant Cell 2:1059–1070.

Bogaert, D., R. De Groot & P. W. M. Hermans. 2004. Streptococcus pneumoniae colonisation: The key to pneumococcal disease. Lancet Infectious Diseases 4:144–154.

Bojar, R. A., W. J. Cunliffe & K. T. Holland. 1995. The short-term treatment of acne vulgaris with benzoyl peroxide: effects on the surface and follicular cutaneous microflora. British Journal of Dermatology 132:204–208.

Bojar, R. A. & K. T. Holland. 2004. Acne and propionibacterium acnes. Clinics in Dermatology 22:375–379.

Bowdish, D. M. E., D. J. Davidson & R. E. W. Hancock. 2005. A re-evaluation of the role of host defence peptides in mammalian immunity. Current protein & peptide science 6:35–51.

Brault, V., U. Sauder, M. C. Reedy, U. Aebi & C. a Schoenenberger. 1999. Differential epitope tagging of actin in transformed Drosophila produces distinct effects on myofibril assembly and function of the indirect flight muscle. Molecular biology of the cell 10:135–49.

Briers, Y. & R. Lavigne. 2015. Breaking barriers: expansion of the use of endolysins as novel antibacterials against Gram-negative bacteria. Future microbiology 10:377–90.

Briers, Y., M. Schmelcher, M. J. Loessner, J. Hendrix, Y. Engelborghs, G. Volckaert & R. Lavigne. 2009. The high-affinity peptidoglycan binding domain of Pseudomonas phage endolysin KZ144. Biochemical and Biophysical Research Communications 383:187–191.

Briers, Y., M. Walmagh, V. Van Puyenbroeck, A. Cornelissen, W. Cenens, A. Aertsen, H. Oliveira, J. Azeredo, G. Verween, J. P. Pirnay, S. Miller, G. Volckaert & R. Lavigne. 2014. Engineered endolysin-based "Artilysins" to combat multidrug-resistant gram-negative pathogens. American Society for Microbiology 5.

Brooker, R. J. 2010. Genetic Principals. The McGraw-Hill Companies, S.r.l. - Publishing Group Italia.

Brown, L., J. M. Wolf, R. Prados-Rosales & A. Casadevall. 2015. Through the wall: extracellular vesicles in Gram-positive bacteria, mycobacteria and fungi. Nature reviews. Microbiology 13:620–30.

Buey, R. M., B. Monterroso, M. Menéndez, G. Diakun, P. Chacón, J. A. Hermoso & J. F. Díaz. 2007. Insights into Molecular Plasticity of Choline Binding Proteins (Pneumococcal Surface Proteins) by SAXS. Journal of Molecular Biology 365:411–424.

Butler-Wu, S. M., E. M. Burns, P. S. Pottinger, A. S. Magaret, J. L. Rakeman, F. A. Matsen & B. T. Cookson. 2011. Optimization of periprosthetic culture for diagnosis of Propionibacterium acnes prosthetic joint infection. Journal of Clinical Microbiology 49:2490–2495.

Cabello, F. C. 2006. Heavy use of prophylactic antibiotics in aquaculture: A growing problem for human and animal health and for the environment. Environmental Microbiology 8:1137–1144.

Cano-Garrido, O., F. L. Rueda, L. Sànchez-García, L. Ruiz-Ávila, R. Bosser, A. Villaverde & E. García-Fruitós. 2014. Expanding the recombinant protein quality in Lactococcus lactis. Microbial Cell Factories 13:167.

Catalão, M. J., F. Gil, J. Moniz-Pereira & M. Pimentel. 2011. Functional analysis of the Holin-Like proteins of mycobacteriophage Ms6. Journal of Bacteriology 193:2793–2803.

Cha, T. S., W. Yee & A. Aziz. 2012. Assessment of factors affecting Agrobacterium-

mediated genetic transformation of the unicellular green alga, Chlorella vulgaris. World Journal of Microbiology and Biotechnology 28:1771–1779.

Chapot-Chartier, M.-P. P. 2014. Interactions of the cell-wall glycopolymers of lactic acid bacteria with their bacteriophages. Frontiers in Microbiology 5:1–10.

Cheng, G., H. Hao, S. Xie, X. Wang, M. Dai, L. Huang & Z. Yuan. 2014. Antibiotic alternatives: The substitution of antibiotics in animal husbandry? Frontiers in Microbiology 5:1–15.

Cheng, Q. & V. A. Fischetti. 2007. Mutagenesis of a bacteriophage lytic enzyme PlyGBS significantly increases its antibacterial activity against group B streptococci. Applied Microbiology and Biotechnology 74:1284–1291.

Cheng, X., X. Zhang, J. W. Pflugrath & F. W. Studier. 1994. The structure of bacteriophage T7 lysozyme, a zinc amidase and an inhibitor of T7 RNA polymerase. Proceedings of the National Academy of Sciences of the United States of America 91:4034–4038.

Chisti, Y. 2008. Biodiesel from microalgae beats bioethanol. Trends in Biotechnology 26:126–131.

Choquet, Y., D. B. Stern, K. Wostrikoff, R. Kuras, J. Girard-Bascou & F. A. Wollman. 1998. Translation of cytochrome f is autoregulated through the 5′ untranslated region of petA mRNA in Chlamydomonas chloroplasts. Proceedings of the National Academy of Sciences 95:4380–4385.

Chung, P. Y. & R. Khanum. 2017. Antimicrobial peptides as potential anti-biofilm agents against multi-drug resistant bacteria. Journal of Microbiology, Immunology and Infection 50:405-410.

Clatworthy, A. E., E. Pierson & D. T. Hung. 2007. Targeting virulence: a new paradigm for antimicrobial therapy. Nature chemical biology 3:541–548.

ContraFect. no date. Product Pipeline. ContraFect Corporation (CFRX). <http://www.contrafect.com/pipeline/overview> (15 July 2017).

Coragliotti, A. T., M. V. Beligni, S. E. Franklin & S. P. Mayfield. 2011. Molecular factors affecting the accumulation of recombinant proteins in the chlamydomonas reinhardtii chloroplast. Molecular Biotechnology 48:60–75.

Crutchfield, A., K. Diller & J. Brand. 1999. Cryopreservation of *Chlamydomonas reinhardtii* (Chlorophyta). European Journal of Phycology 34:43–52.

Cullis, C. A., B. J. Vorster, C. Van Der Vyver & K. J. Kunert. 2009. Transfer of genetic material between the chloroplast and nucleus: How is it related to stress in plants? Annals of Botany 103:625–633.

Daniel, A., C. Euler, M. Collin, P. Chahales, K. J. Gorelick & V. a Fischetti. 2010. Synergism between a novel chimeric lysin and oxacillin protects against infection by methicillin-resistant Staphylococcus aureus. Antimicrobial agents and chemotherapy 54:1603–12.

Daniell, H., S. Kumar & N. Dufourmantel. 2005. Breakthrough in chloroplast genetic engineering of agronomically important crops. Trends in biotechnology 23:238–45.

Datar, R. V, T. Cartwright & C. G. Rosen. 1993. Process economics of animal cell and bacterial fermentations: a case study analysis of tissue plasminogen activator. Bio/technology (Nature Publishing Company) 11:349–357.

Davies, J. 2006. Are antibiotics naturally antibiotics? Journal of Industrial Microbiology and Biotechnology 33:496–499.

Demain, A. L. & P. Vaishnav. 2011. Production of Recombinant Proteins by Microbes and Higher Organisms. Comprehensive Biotechnology, Second Edition 3:333–345.

Derrien, B. 2013. Strep Tag II purification of C. reinhardtii chloroplast complex. Bio-Protocol 3:2–6.

Derrien, B., W. Majeran, G. Effantin, J. Ebenezer, G. Friso, K. J. van Wijk, A. C. Steven, M. R. Maurizi & O. Vallon. 2012. The purification of the Chlamydomonas reinhardtii chloroplast ClpP complex: Additional subunits and structural features. Plant Molecular Biology 80:189–202.

Dessinioti, C. & A. Katsambas. 2017. Propionibacterium acnes and antimicrobial resistance in acne. Clinics in Dermatology 35:163–167.

Doehn, J. M., K. Fischer, K. Reppe, B. Gutbier, T. Tschernig, A. C. Hocke, V. A. Fischetti, J. Löffler, N. Suttorp, S. Hippenstiel & M. Witzenrath. 2013. Delivery of the endolysin Cpl-1 by inhalation rescues mice with fatal pneumococcal pneumonia. The Journal of antimicrobial chemotherapy 68:2111–7.

Dreesen, I. A. J., G. C. E. Hamri & M. Fussenegger. 2010. Heat-stable oral alga-based vaccine protects mice from Staphylococcus aureus infection. Journal of Biotechnology 145:273–280.

Drugbank. "Benzylpenicillin." <https://www.drugbank.ca/drugs/DB01053> (accessed 31 July 2017).

Dumont, J., D. Euwart, B. Mei, S. Estes & R. Kshirsagar. 2015. Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. Critical reviews in Biotechnology 1:1–13.

Dunne, M., S. Leicht, B. Krichel, H. D. T. Mertens, A. Thompson, J. Krijgsveld, D. I. Svergun, N. Gómez-Torres, S. Garde, C. Uetrecht, A. Narbad, M. J. Mayer & R. Meijers. 2016. Crystal structure of the CTP1L endolysin reveals how its activity is regulated by a secondary translation product. Journal of Biological Chemistry 291:4882–4893.

Dunne, M., H. D. T. Mertens, V. Garefalaki, C. M. Jeffries, A. Thompson, E. A. Lemke, D. I. Svergun, M. J. Mayer, A. Narbad & R. Meijers. 2014. The CD27L and CTP1L Endolysins Targeting Clostridia Contain a Built-in Trigger and Release Factor. PLoS Pathogens 10(7).

Dutcher, S. K., L. Li, H. Lin, L. Meyer, T. H. Giddings, a. L. Kwan, B. L. Lewis & J. Berman. 2012. Whole-Genome Sequencing to Identify Mutants and Polymorphisms in Chlamydomonas reinhardtii. Genes|Genomes|Genetics 2:15–22.

Dwidar, M., A. K. Monnappa & R. J. Mitchell. 2012. The dual probiotic and antibiotic nature of Bdellovibrio bacteriovorus. BMB Reports 45:71–78.

Dworkin, M. & S. Falkow. 2006. The prokaryotes: a handbook on the biology of bacteria. Chapter 10.1 The Family Chloroflexceae. New York: Springer-Verlag.

Entenza, J. M., J. M. Loeffler, D. Grandgirard, V. A. Fischetti & P. Moreillon. 2005. Therapeutic effects of bacteriophage Cpl-1 lysin against Streptococcus pneumoniae endocarditis in rats. Antimicrobial Agents and Chemotherapy 49:4789–4792.

Van Essche, M., M. Quirynen, I. Sliepen, G. Loozen, N. Boon, J. Van Eldere & W. Teughels. 2011. Killing of anaerobic pathogens by predatory bacteria. Molecular Oral Microbiology 26:52–61.

Farrar, M. D., K. M. Howson, R. a Bojar, D. West, J. C. Towler, J. Parry, K. Pelton & K. T. Holland. 2007. Genome sequence and analysis of a Propionibacterium acnes bacteriophage. Journal of bacteriology 189:4161–7.

Feiner, R., T. Argov, L. Rabinovich, N. Sigal, I. Borovok & A. A. Herskovits. 2015. A new perspective on lysogeny: prophages as active regulatory switches of bacteria. Nature Reviews Microbiology 13:641–650.

Fernandes, R., P. Amador & C. Prudêncio. 2013. β-Lactams: chemical structure, mode of action and mechanisms of resistance. Reviews in Medical Microbiology 24:7–17.

Fernández-Tornero, C., E. García, R. López, G. Giménez-Gallego & A. 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 321:163–173.

Fernebro, J. 2011. Fighting bacterial infections-future treatment options. Drug resistance updates : reviews and commentaries in antimicrobial and anticancer chemotherapy 14:125–39.

Ferreira-Camargo, L. S., M. Tran, J. Beld, M. D. Burkart & S. P. Mayfield. 2015. Selenocystamine improves protein accumulation in chloroplasts of eukaryotic green algae. AMB Express 5:126.

Fischetti, V. a. 2008. Bacteriophage lysins as effective antibacterials. Current opinion in microbiology 11:393–400.

Franklin, S. E. & S. P. Mayfield. 2004. Prospects for molecular farming in the green alga Chlamydomonas reinhardtii. Current Opinion in Plant Biology 7:159–165.

Franklin, S., B. Ngo, E. Efuet & S. P. Mayfield. 2002. Development of a GFP reporter gene for Chlamydomonas reinhardtii chloroplast. The Plant journal : for cell and molecular biology 30:733–44.

Freeman, J., S. D. Baines, K. Saxton & M. H. Wilcox. 2007. Effect of metronidazole on growth and toxin production by epidemic Clostridium difficile PCR ribotypes 001 and 027 in a human gut model. Journal of Antimicrobial Chemotherapy 60:83–91.

Fuhrmann, M., W. Oertel & P. Hegemann. 1999. A synthetic gene coding for the green fluorescent protein (GFP) is a versatile reporter in Chlamydomonas reinhardtii. The Plant journal : for cell and molecular biology 19:353–361.

Fulton, J. E., A. Farzad-Bakshandeh & S. Bradley. 1974. Studies on the Mechanism of Action of Topical Benzoyl Peroxide and Vitamin A Acid in Acne Vulgaris. Journal of Cutaneous Pathology 1:191–200.

García, J. L., E. García, a Arrarás, P. García, C. Ronda & R. López. 1987. Cloning, purification, and biochemical characterization of the pneumococcal bacteriophage Cp-1 lysin. Journal of virology 61:2573–80.

García, P., M. P. González, E. García, J. L. García & R. López. 1999. The molecular characterization of the first autolytic lysozyme of Streptococcus pneumoniae reveals evolutionary mobile domains. Molecular Microbiology 33:128–138.

Garrett, J., C. Bruno & R. Young. 1990. Lysis protein S of phage lambda functions in Saccharomyces cerevisiae. Journal of Bacteriology 172:7275–7277.

Gaskins, H. R., C. T. Collier & D. B. Anderson. 2002. Antibiotics As Growth Promotants:Mode of Action. Animal Biotechnology 13:29–42.

Gaspar, P., G. Moura, M. A. S. Santos & J. L. Oliveira. 2013. MRNA secondary structure optimization using a correlated stem-loop prediction. Nucleic Acids Research 41:3–7.

Gehre, F., S. L. Leib, D. Grandgirard, J. Kummer, A. Bhlmann, F. Simon, R. Gumann, A. S. Kharat, M. G. Tuber & A. Tomasz. 2008. Essential role of choline for pneumococcal virulence in an experimental model of meningitis. Journal of Internal Medicine 264:143–154.

George, A. & G. Rubin. 2003. A systematic review and meta-analysis of treatments for impetigo. British Journal of General Practice 53:480–487.

Gerding, D. N., C. A. Muto & R. C. Owens. 2008. Measures to control and prevent Clostridium difficile infection. Clinical Infectious Diseases 46:S43-49.

Gilman, J. & J. Love. 2016. Synthetic promoter design for new microbial chassis. Biochemical Society Transactions 44:731–737.

Gimpel, J. A., J. S. Hyun, N. G. Schoepp & S. P. Mayfield. 2015. Production of recombinant proteins in microalgae at pilot greenhouse scale. Biotechnology and Bioengineering 112:339–345.

Gould, L. H. & B. Limbago. 2010. *Clostridium difficile* in Food and Domestic Animals: A New Foodborne Pathogen? Clinical Infectious Diseases 51:577–582.

Govea-Alonso, D. O., M. A. Tello-Olea, J. Beltrán-López, E. Monreal-Escalante, J. A. Salazar-Gonzalez, B. Bañuelos-Hernández & S. Rosales-Mendoza. 2017. Assessment of Carrot Callus as Biofactories of an Atherosclerosis Oral Vaccine Prototype. Molecular Biotechnology 59:482–489.

Grandgirard, D., J. M. Loeffler, V. A. Fischetti & S. L. Leib. 2008. Phage Lytic Enzyme Cpl-1 for Antibacterial Therapy in Experimental Pneumococcal Meningitis 197:1519–1522.

Grabowski, G. A., M. Golembo & Y. Shaaltiel. 2014. Taliglucerase alfa: An enzyme replacement therapy using plant cell expression technology. *Molecular Genetics and Metabolism* 1:1-8.

Gregory, J. A., F. Li, L. M. Tomosada, C. J. Cox, A. B. Topol, J. M. Vinetz & S. Mayfield. 2012. Algae-produced pfs25 elicits antibodies that inhibit malaria transmission. PLoS ONE 7(5).

Gregory, J. A., A. Shepley-McTaggart, M. Umpierrez, B. K. Hurlburt, S. J. Maleki, H. A. Sampson, S. P. Mayfield & M. C. Berin. 2016. Immunotherapy using algal-produced Ara h 1 core domain suppresses peanut allergy in mice. Plant Biotechnology Journal 14:1541–1550.

Gregory, J. A., A. B. Topol, D. Z. Doerner & S. Mayfield. 2013. Alga-produced cholera toxin-Pfs25 fusion proteins as oral vaccines. Applied and Environmental Microbiology 79:3917–3925.

Grossman, A. R., E. E. Harris, C. Hauser, P. A. Lefebvre, D. Martinez, D. Rokhsar, J. Shrager, C. D. Sil, D. Stern, O. Vallon & Z. Zhang. 2003. *Chlamydomonas reinhardtii* at the Crossroads of Genomics. Society 2:1137–1150.

Gu, J., Y. Feng, X. Feng, C. Sun, L. Lei, W. Ding, F. Niu, L. Jiao, M. Yang, Y. Li, X. Liu, J. Song, Z. Cui, D. Han, C. Du, Y. Yang, S. Ouyang, Z. J. Liu & W. Han. 2014. Structural and Biochemical Characterization Reveals LysGH15 as an Unprecedented "EF-Hand-Like" Calcium-Binding Phage Lysin. PLoS Pathogens 10.

Gupta, S. K. & P. Shukla. 2015. Advanced technologies for improved expression of recombinant proteins in bacteria: perspectives and applications. Critical Reviews in Biotechnology 0:1–10.

Gutiérrez, D., P. Ruas-Madiedo, B. Martínez, A. Rodríguez & P. García. 2014. Effective Removal of Staphylococcal Biofilms by the Endolysin LysH5. PloS ONE 9:e107307.

Gutman, G. A. & G. W. Hatfield. 1989. Nonrandom utilization of codon pairs in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America 86:3699–703.

Hafiz, S. & C. L. Oakley. 1976. Clostridium difficile: isolation and characteristics. Journal of medical microbiology 9:129–36.

Hall, I. & E. O'Toole. 1935. Intestinal flora in new-born infants: With a description of a new pathogenic anaerobe, Bacillus difficilis. American Journal of Diseases of Children 49:390–402.

Hanahan, D. 1983. Studies on transformation of Escherichia coli with plasmids. Journal of Molecular Biology 166:557–580.

Hao, H., G. Cheng, Z. Iqbal, X. Ai, H. I. Hussain, L. Huang, M. Dai, Y. Wang, Z. Liu & Z. Yuan. 2014. Benefits and risks of antimicrobial use in food-producing animals. Frontiers in Microbiology 5:1–11.

Hargreaves, K. R. & M. R. J. Clokie. 2014. Clostridium difficile phages: Still difficult? Frontiers in Microbiology 5:1–14.

Harris, E. H. 2009. The Chlamydomonas sourcebook. Volume 1. 2nd edition. San Diego, CA: Academic Press.

He, D. M., K. X. Qian, G. F. Shen, Z. F. Zhang, Y. N. LI, Z. L. Su & H. B. Shao. 2007. Recombination and expression of classical swine fever virus (CSFV) structural protein E2 gene in Chlamydomonas reinhardtii chroloplasts. Colloids and Surfaces B: Biointerfaces 55:26–30.

Hermoso, J. A., B. Monterroso, A. Albert, B. Galán, O. Ahrazem, P. García, M. Martínez-Ripoll, J. L. García & M. Menéndez. 2003. Structural basis for selective recognition of pneumococcal cell wall by modular endolysin from phage Cp-1. Structure 11:1239–1249.

Hidaka, Y., M. Ohno, B. Hemmasi, O. Hill, W. G. Forssmann & Y. Shimonishi. 1998. In vitro disulfide-coupled folding of guanylyl cyclase-activating peptide and its precursor protein. Biochemistry 37:8498–8507.

Holland, C., T. N. Mak, U. Zimny-Arndt, M. Schmid, T. F. Meyer, P. R. Jungblut & H. Brüggemann. 2010. Proteomic identification of secreted proteins of Propionibacterium acnes. BMC microbiology 10:230.

Hoopes, J. T., C. J. Stark, H. A. Kim, D. J. Sussman, D. M. Donovan & D. C. Nelson. 2009. Use of a bacteriophage lysin, PlyC, as an enzyme disinfectant against Streptococcus equi. Applied and Environmental Microbiology 75:1388–1394.

Hosler, J. P., E. A. Wurtz, E. H. Harris, N. W. Gillham & J. E. Boynton. 1989. Relationship between gene dosage and gene expression in the chloroplast of Chlamydomonas reinhardtii. Plant Physiology 91:648–655.

Howe, C. J. & S. Purton. 2007. The Little Genome of Apicomplexan Plastids: its raison d'etre and a Possible Explanation for the "Delayed Death" Phenomenon. Protist 158:121–133.

Huang, H., A. Weintraub, H. Fang & C. E. Nord. 2009. Antimicrobial resistance in Clostridium difficile. International Journal of Antimicrobial Agents 34:516–522.

Hughes, C. & G. G. Meynell. 1974. High frequency of antibiotic-resistant enterobacteria in the River Stour, Kent. Lancet (London, England) 2:451–3.

Hyams, C., E. Camberlein, J. M. Cohen, K. Bax & J. S. Brown. 2010. The Streptococcus pneumoniae capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms. Infection and Immunity 78:704–715.

Hyvonen, M. 2002. Refolding of disulfide rich proteins. <http://camelot.bioc.cam.ac.uk/~marko/methods/refold.pdf>.

Irwin, B., J. D. Heck & G. W. Hatfield. 1995. Codon Pair Utilization Biases Influence Translational Elongation Step Times. Biochemistry 270:22801–22806.

Jado, I., R. López, E. García, A. Fenoll, J. Casal & P. García. 2003. Phage lytic enzymes as therapy for antibiotic-resistant Streptococcus pneumoniae infection in a murine sepsis model. The Journal of antimicrobial chemotherapy 52:967–73.

Jeffrey, G. A. 1997. An Introduction to Hydrogen Bonding. Topics in Physical Chemistry. New York: Oxford Univeristy Press.

Jervis, E. J., C. A. Haynes, G. Kilburn, E. J. Jervis, C. A. Haynes & D. G. Kilburn. 1997. ENZYMOLOGY : Surface Diffusion of Cellulases and Their Isolated Binding Domains on Cellulose Surface Diffusion of Cellulases and Their Isolated Binding Domains on Cellulose. The Journal of Biological Chemistry 272:24016–24023.

Jinkerson, R. E. & M. C. Jonikas. 2015. Molecular techniques to interrogate and edit the Chlamydomonas nuclear genome. Plant Journal 82:393–412.

Joshi, S. R., R. M. Parikh & A. K. Das. 2007. Insulin--history, biochemistry, physiology and pharmacology. The Journal of the Association of Physicians of India 55 Suppl:19–25.

Kaiser, G. E. 2011. BIOL 230 Lecture Guide - Mode of Action of Penicillins. <http://faculty.ccbcmd.edu/courses/bio141/lecguide/unit2/control/penresillus.html> (31 July 2017).

Keel, K., J. S. Brazier, K. W. Post, S. Weese & J. G. Songer. 2007. Prevalence of PCR ribotypes among Clostridium difficile isolates from pigs, calves, and other species. Journal of clinical microbiology 45:1963–4.

Kim, J. 1997. Protein Disulfide Isomerase as a Regulator of Chloroplast Translational Activation. Science 278:1954–1957.

Kim, J. Y., Y. G. G. Kim & G. M. Lee. 2012. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. Applied Microbiology and Biotechnology 93:917–930.

Kindle, K. L., K. L. Richards & D. B. Stern. 1991. Engineering the chloroplast genome: techniques and capabilities for chloroplast transformation in Chlamydomonas reinhardtii. Proceedings of the National Academy of Sciences of the United States of America 88:1721–1725.

Kong, K. 2013. Codon Usage Optimizer (CUO). GitHub. <https://github.com/khai-/CUO>.

Korndörfer, I. P., J. Danzer, M. Schmelcher, M. Zimmer, A. Skerra & M. 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 364:678–89.

Kristensen, C., T. Kjeldsen, F. C. Wiberg, L. Schäffer, M. Hach, S. Havelund, J. Bass, D. F. Steiner & A. S. Andersen. 1997. Alanine scanning mutagenesis of insulin. The Journal of biological chemistry 272:12978–83.

Kropat, J., A. Hong-Hermesdorf, D. Casero, P. Ent, M. Castruita, M. Pellegrini, S. S. Merchant & D. Malasarn. 2011. A revised mineral nutrient supplement increases biomass and growth rate in Chlamydomonas reinhardtii. The Plant journal for cell and molecular biology 66:770–80.

LaFee, S. & H. Buschman. 2017. Novel Phage Therapy Saves Patient with Multidrug-Resistant Bacterial Infection. *UC San Diego Health*. <https://health.ucsd.edu/news/releases/Pages/2017-04-25-novel-phage-therapy-saves-patient-with-multidrug-resistant-bacterial-infection.aspx> (14 July 2017).

Lai, M. J., N. T. Lin, A. Hu, P. C. Soo, L. K. Chen, L. H. Chen & K. C. Chang. 2011. Antibacterial activity of Acinetobacter baumannii phage PhiAB2 endolysin (LysAB2) against both Gram-positive and Gram-negative bacteria. Applied Microbiology and Biotechnology 90:529–539.

Lam, S. J., N. M. O'Brien-Simpson, N. Pantarat, A. Sulistio, E. H. H. Wong, Y.-Y. Chen, J. C. Lenzo, J. A. Holden, A. Blencowe, E. C. Reynolds & G. G. Qiao. 2016. Combating multidrug-resistant Gram-negative bacteria with structurally nanoengineered antimicrobial peptide polymers. Nature Microbiology 1:16162.

Lauersen, K. J., O. Kruse & J. H. Mussgnug. 2015. Targeted expression of nuclear transgenes in Chlamydomonas reinhardtii with a versatile, modular vector toolkit. Applied Microbiology and Biotechnology 99:3491–3503.

Lella, M., S. Kamilla, V. Jain & R. Mahalakshmi. 2016. Molecular Mechanism of Holin Transmembrane Domain i in Pore Formation and Bacterial Cell Death. ACS Chemical Biology 11:910–920.

Lewis, K. 2013. Platforms for antibiotic discovery. Nature Reviews Drug Discovery 12:371–387.

Leyden, J. J., K. J. McGinley, S. Cavalieri, G. F. Webster, O. H. Mills & A. M. Kligman. 1983. Propionibacterium acnes resistance to antibiotics in acne patients. Journal of the American Academy of Dermatology 8:41–45.

Lim, J.-A., H. Shin, S. Heu & S. Ryu. 2014. Exogenous lytic activity of SPN9CC endolysin against gram-negative bacteria. Journal of microbiology and biotechnology 24:803–11.

Ling, L. L., T. Schneider, A. J. Peoples, A. L. Spoering, I. Engels, B. P. Conlon, A. Mueller, D. E. Hughes, S. Epstein, M. Jones, L. Lazarides, V. a Steadman, D. R. Cohen, C. R. Felix, K. A. Fetterman, W. P. Millett, A. G. Nitti, A. M. Zullo, C. Chen et al. 2015. A new antibiotic kills pathogens without detectable resistance. Nature 517:455–459.

Livermore, D. M. 2011. Discovery research: The scientific challenge of finding new antibiotics. Journal of Antimicrobial Chemotherapy 66:1941–1944.

Loeffler, J. M., S. Djurkovic & V. A. Fischetti. 2003. Phage Lytic Enzyme Cpl-1 as a Novel Antimicrobial for Pneumococcal Bacteremia. Infection and Immunity 71:6199–6204.

Loeffler, J. M., D. Nelson & V. a Fischetti. 2001. Rapid killing of Streptococcus pneumoniae with a bacteriophage cell wall hydrolase. Science 294:2170–2172.

Loessner, M. J. 2005. Bacteriophage endolysins--current state of research and applications. Current opinion in microbiology 8:480–7.

Loessner, M. J., K. Kramer, F. Ebel & S. 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 44:335–49.

Lucks, J. B., D. R. Nelson, G. R. Kudla & J. B. Plotkin. 2008. Genome landscapes and bacteriophage codon usage. PLoS Computational Biology 4.

Macfarlane, G. T., S. Macfarlane & G. R. Gibson. 1998. Validation of a Three-Stage Compound Continuous Culture System for Investigating the Effect of Retention Time on the Ecology and Metabolism of Bacteria in the Human Colon. Microbial Ecology 35:180–187.

Madigan, M. T. & J. M. Martinko. 2006. Brock Biology of Microorganisms, 11th edn. International Microbiology. Upper Saddle River, NJ ; London : Pearson Prentice Hall.

Maestro, B. & J. M. Sanz. 2016. Choline Binding Proteins from Streptococcus pneumoniae: A Dual Role as Enzybiotics and Targets for the Design of New Antimicrobials. Antibiotics 5.

Mailer, J. S. & B. Mason. 2001. Penicillin: Medicine's Wartime Wonder Drug and Its Production at Peoria, Illinois. *Illinois Periodicals Online*. <http://www.lib.niu.edu/2001/iht810139.html> (13 July 2017).

Makino, T., G. Skretas & G. Georgiou. 2011. Strain engineering for improved expression of recombinant proteins in bacteria. Microbial Cell Factories. 10:32-42.

Maliga, P. 2014. Chloroplast Biotechnology : Methods and Protocols. Totowa, NJ: Humana Press.

Manuell, A. L., M. V. Beligni, J. H. Elder, D. T. Siefker, M. Tran, A. Weber, T. L. McDonald & S. P. Mayfield. 2007. Robust expression of a bioactive mammalian protein in Chlamydomonas chloroplast. Plant Biotechnology Journal 5:402–412.

de Marco, A., E. Deuerling, A. Mogk, T. Tomoyasu & B. Bukau. 2007. Chaperone-based procedure to increase yields of soluble recombinant proteins produced in E. coli. BMC Biotechnology 7:32.

Marinelli, L. J., S. Fitz-gibbon & C. Hayes. 2012. Propionibacterium acnes Bacteriophages Display Limited Genetic 3:1–13.

Markey, P. M., S. E. Racine, C. N. Markey, C. J. Hopwood, P. K. Keel, S. A. Burt, M. C. Neale, C. L. Sisk, M. Steven, K. L. Klump, E. Lansing & E. Lansing. 2015. Algal chloroplast produced camelid VH H antitoxins are capable of neutralizing botulinum neurotoxin. Plant Biotechnology Journal 6:300–308.

Martín, a C., R. López & P. García. 1998. Functional analysis of the two-gene lysis system of the pneumococcal phage Cp-1 in homologous and heterologous host cells. Journal of bacteriology 180:210–7.

Maul, J. E., J. W. Lilly, L. Cui, C. W. DePamphilis, W. Miller, E. H. Harris & D. B. Stern. 2002. The Chlamydomonas reinhardtii plastid chromosome: islands of genes in a sea of repeats. The Plant cell 14:2659–2679.

Mayer, M. J., V. Garefalaki, R. Spoerl, A. Narbad & R. Meijers. 2011. Structure-based modification of a Clostridium difficile-targeting endolysin affects activity and host range. Journal of Bacteriology 193:5477–5486.

Mayer, M. J., A. Narbad & M. J. Gasson. 2008. Molecular characterization of a Clostridium difficile bacteriophage and its cloned biologically active endolysin. Journal of Bacteriology 190:6734–6740.

Mayer, M. J., J. Payne, M. J. Gasson & A. Narbad. 2010. Genomic sequence and characterization of the virulent bacteriophage CTP1 from Clostridium tyrobutyricum and heterologous expression of its endolysin. Applied and Environmental Microbiology 76:5415–5422.

Mayfield, S. P., S. E. Franklin & R. A. Lerner. 2003. Expression and assembly of a fully active antibody in algae. Proceedings of the National Academy of Sciences of the United States of America. 100:438–442.

Mayfield, S. P. & J. Schultz. 2004. Development of a luciferase reporter gene, luxCt, for Chlamydomonas reinhardtii chloroplast. Plant Journal 37:449–458.

McCullers, J. A., Å. Karlström, A. R. Iverson, J. M. Loeffler & V. A. Fischetti. 2007. Novel strategy to prevent otitis media caused by colonizing Streptococcus pneumoniae. PLoS Pathogens 3:1–3.

McDaniel, L. S., J. Yother, M. Vijayakumar, L. McGarry, W. R. Guild & D. E. Briles. 1987. Use of insertional inactivation to facilitate studies of biological properties of pneumococcal surface protein A (PspA). The Journal of experimental medicine 165:381–394.

Meek, R. W., H. Vyas & L. J. V. Piddock. 2015. Nonmedical Uses of Antibiotics: Time to Restrict Their Use? PLoS Biology 13:1–11.

Merchant, S. & L. Bogorad. 1986. Rapid degradation of apoplastocyanin in Cu(II)-deficient cells of Chlamydomonas reinhardtii. Journal of Biological Chemistry 261:15850–15853.

Merchant, S. S., S. E. Prochnik, O. Vallon, E. H. Harris, J. Karpowicz, G. B. Witman, A. Terry, A. Salamov, L. K. Fritz-laylin, L. Maréchal-Drouard, W. F. Marshall, L. Qu, D. R. Nelson, A. Sanderfoot, M. H. Spalding, V. V. Kapitonov, Q. Ren, P. Cardol, H. Cerutti et al. 2010. The Chlamydomonas Genome Reveals the Evolution of Key Animal and Plant Functions. Science 318:245–250.

Michelet, L., L. Lefebvre-Legendre, S. E. Burr, J. D. Rochaix & M. Goldschmidt-Clermont. 2011. Enhanced chloroplast transgene expression in a nuclear mutant of Chlamydomonas. Plant Biotechnology Journal 9:565–574.

Micreos. Staphefekt Product Fact Sheets. <http://www.micreos.com/content/human-health.aspx> (15 July 2017).

Miller, S. I. 2016. Antibiotic resistance and regulation of the gram-negative bacterial outer membrane barrier by host innate immune molecules. mBio 7:e01541-16.

Monterroso, B. A., J. L. Sáiz, P. García, J. L. García & M. Menéndez. 2008. Insights into the structure-function relationships of pneumococcal cell wall lysozymes, LytC and Cpl-1. Journal of Biological Chemistry 283:28618–28628.

Monterroso, B., C. López-Zumel, J. L. García, J. L. Sáiz, P. García, N. E. Campillo & M. Menéndez. 2005. Unravelling the structure of the pneumococcal autolytic lysozyme. Biochemical Journal 391:41–49.

Mudd, E. A., P. Madesis, E. M. Avila & A. Day. 2014. Excision of plastid marker genes using directly repeated DNA sequences. Methods in Molecular Biology 1132:107–123.

Nakimbugwe, D., B. Masschalck, D. Deckers, L. Callewaert, A. Aertsen & C. W. Michiels. 2006. Cell wall substrate specificity of six different lysozymes and lysozyme inhibitory activity of bacterial extracts. FEMS Microbiology Letters 259:41–46.

Nelson, D. 2016. Immunotherapeutics Derived from Bacteriophage Endolysins. The Lysin Meeting. New York, USA.

Nelson, D., L. Loomis & V. a Fischetti. 2001. Prevention and elimination of upper respiratory colonization of mice by group A streptococci by using a bacteriophage lytic enzyme. Proceedings of the National Academy of Sciences of the United States of America 98:4107–12.

Nielsen, J. 2013. Production of biopharmaceutical proteins by yeast. Bioengineered 4:207–211.

O'Neill, J., C. By, J. I. M. O. Neill & J. O'Neill. 2015. Securing New Drugs for Future Generations: The Pipeline of Antibiotics. The Review on Antimicrobial Resistance:42.

Obaro, S. & R. Adegbola. 2002. The pneumococcus: Carriage, disease and conjugate vaccines. Journal of Medical Microbiology 51:98–104.

Oey, M., M. Lohse, B. Kreikemeyer & R. Bock. 2009. Exhaustion of the chloroplast protein synthesis capacity by massive expression of a highly stable protein antibiotic. Plant Journal 57:436–445.

Oliveira, H., L. D. R. Melo, S. B. Santos, F. L. Nobrega, E. C. Ferreira, N. Cerca, J. Azeredo & L. D. Kluskens. 2013. Molecular Aspects and Comparative Genomics of Bacteriophage Endolysins. Journal of Virology 87:4558–4570.

Paik, J., I. Kern, R. Lurz & R. Hakenbeck. 1999. Mutational analysis of the Streptococcus pneumoniae bimodular class A penicillin-binding proteins. Journal of Bacteriology 181:3852–3856.

Pang, T., T. C. Fleming, K. Pogliano & R. Young. 2013. Visualization of pinholin lesions in vivo. Proceedings of the National Academy of Sciences of the United States of America 110:E2054-63.

Park, K. Y. & S. J. Wi. 2016. Potential of plants to produce recombinant protein products. Journal of Plant Biology 59:559–568.

Parmley, S. 2014. Lysin in wait. Science-Business eXchange 7.

Pastagia, M., R. Schuch, V. a Fischetti & D. B. Huang. 2013. Lysins: the arrival of pathogen-directed anti-infectives. Journal of medical microbiology 62:1506–16.

Payne, D. J., M. N. Gwynn, D. J. Holmes & D. L. Pompliano. 2007. Drugs for bad bugs: confronting the challenges of antibacterial discovery. Nature Reviews. Drug Discovery 6:29–40.

Pirofski, L. A. & A. Casadevall. 2008. The damage-response framework of microbial pathogenesis and infectious diseases. Advances in Experimental Medicine and Biology 635:135–146.

Portillo, M. E., S. Corvec, O. Borens & A. Trampuz. 2013. Propionibacterium acnes: An underestimated pathogen in implant-associated infections. *BioMed Research International*.

Pride, D. T., J. Salzman, M. Haynes, F. Rohwer, C. Davis-Long, R. A. White, P. Loomer, G. C. Armitage & D. A. Relman. 2012. Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome. The ISME Journal 6:915–926.

Proença, D., C. Velours, C. Leandro, M. Garcia, M. Pimentel & C. São-José. 2015. A two-component, multimeric endolysin encoded by a single gene. Molecular Microbiology 95:739–753.

Purton, S. 2007. Tools and techniques for chloroplast transformation of Chlamydomonas. Advances in experimental medicine and biology 616:34–45.

Rajakumar, P. D. 2016. The chloroplast of Chlamydomonas reinhardtii as a platform for recombinant vaccine production. PhD thesis. London: University College London.

Ramundo, S. & J. D. Rochaix. 2014. Chloroplast unfolded protein response, a new plastid stress signaling pathway? Plant Signaling & Behavior 9:e972874.

Rasala, B. A. & S. P. Mayfield. 2015. Photosynthetic biomanufacturing in green algae; production of recombinant proteins for industrial, nutritional, and medical uses. Photosynthesis research 123:227–239.

Rasala, B. A., M. Muto, P. A. Lee, M. Jager, R. M. F. Cardoso, C. A. Behnke, P. Kirk, C. A. Hokanson, R. Crea, M. Mendez & S. P. Mayfield. 2009. Production of therapeutic proteins in algae, analysis of expression of seven human proteins in the chloroplast of Chlamydomonas reinhardtii. Plant biotechnology journal 8:719–33.

Reddy, B. L. & M. H. Saier. 2013. Topological and phylogenetic analyses of bacterial holin families and superfamilies. Biochimica et Biophysica Acta - Biomembranes 1828:2654–2671.

Reindel, R. & C. R. Fiore. 2017. Phage Therapy: Considerations and Challenges for Development. Clinical Infectious Diseases 64:1589–1590.

Resch, G., P. Moreillon & V. A. Fischetti. 2011a. A stable phage lysin (Cpl-1) dimer with increased antipneumococcal activity and decreased plasma clearance. International Journal of Antimicrobial Agents 38:516–21.

Resch, G., P. Moreillon & V. A. Fischetti. 2011b. PEGylating a bacteriophage endolysin inhibits its bactericidal activity. AMB Express 1:29.

Rice, L. B. 2008. Federal Funding for the Study of Antimicrobial Resistance in Nosocomial Pathogens: No ESKAPE. The Journal of Infectious Diseases 197:1079–1081.

Rochaix, J.-D. 1995. Chlamydomonas reinhardtii as the Photosynthetic Yeast. Annual review of genetics 29:209–230.

Rochaix, J.-D. 2001. Assembly, function, and dynamics of the photosynthetic machinery in Chlamydomonas reinhardtii. Plant physiology 127:1394–8.

Rochaix, J.-D., M. Goldschmidt-Clermont & S. Merchant. 1998. Advances in Photosynthesis and Respiration: The molecular biology of chloroplasts and mitochondria in Chlamydomonas. Netherlands: Springer.

Ronda, C., R. Lopez & E. Garcia. 1981. Isolation and characterization of a new bacteriophage, Cp-1, infecting Streptococcus pneumoniae. Journal of Virology 40:551–559.

Rosenberg, J. N., G. A. Oyler, L. Wilkinson & M. J. Betenbaugh. 2008. A green light for engineered algae: redirecting metabolism to fuel a biotechnology revolution. Current Opinion in Biotechnology 19:430–436.

Sagermann, M. & B. W. Matthews. 2002. Crystal structures of a T4-lysozyme duplication-extension mutant demonstrate that the highly conserved beta-sheet region has low intrinsic folding propensity. Journal of molecular biology 316:931–940.

Santillan-Jimenez, E., R. Pace, S. Marques, T. Morgan, C. McKelphin, J. Mobley & M. Crocker. 2016. Extraction, characterization, purification and catalytic upgrading of algae lipids to fuel-like hydrocarbons. Fuel 180:668–678.

Sanz, J. M., E. Díaz & J. L. García. 1992. Studies on the structure and function of the N-terminal domain of the pneumococcal murein hydrolases. Molecular Microbiology 6:921–931.

Savva, C. G., J. S. Dewey, J. Deaton, R. L. White, D. K. Struck, A. Holzenburg & R. Young. 2008. The holin of bacteriophage lambda forms rings with large diameter. Molecular microbiology 69:784–793.

Scaife, M. A., G. T. D. T. Nguyen, J. Rico, D. Lambert, K. E. Helliwell & A. G. Smith. 2015. Establishing Chlamydomonas reinhardtii as an industrial biotechnology host. Plant Journal 82:532–546.

Schmelcher, M., D. M. Donovan & M. J. Loessner. 2012. Bacteriophage endolysins as novel antimicrobials. Future Microbiology 7:1147–1171.

Schroda, M. 2004. The Chlamydomonas genome reveals its secrets: Chaperone genes and the potential roles of their gene products in the chloroplast. Photosynthesis Research 82:221–240.

Schulz, S., A. Stephan, S. Hahn, L. Bortesi, F. Jarczowski, U. Bettmann, A.-K. Paschke, D. Tusé, C. H. Stahl, A. Giritch & Y. Gleba. 2015. Broad and efficient control of major foodborne pathogenic strains of Escherichia coli by mixtures of plant-produced colicins. Proceedings of the National Academy of Sciences of the United States of America 112:e5454-60.

Scranton, M. A., J. T. Ostrand, F. J. Fields & S. P. Mayfield. 2015. Chlamydomonas as a model for biofuels and bio-products production. Plant Journal 82:523–531.

Scranton, M. A., J. T. Ostrand, D. R. Georgianna, S. M. Lofgren, D. Li, R. C. Ellis, D. N. Carruthers, A. Dräger, D. L. Masica & S. P. Mayfield. 2016. Synthetic promoters capable of driving robust nuclear gene expression in the green alga Chlamydomonas reinhardtii. Algal Research 15:135–142.

Sebaihia, M., B. W. Wren, P. Mullany, N. F. Fairweather, N. Minton, R. Stabler, N. R. Thomson, A. P. Roberts, A. M. Cerdeño-Tárraga, H. Wang, M. T. G. Holden, A. Wright, C. Churcher, M. a Quail, S. Baker, N. Bason, K. Brooks, T. Chillingworth, A. Cronin et al. 2006. The multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome. Nature genetics 38:779–786.

Sengupta, S., M. K. Chattopadhyay & H. P. Grossart. 2013. The multifaceted roles of antibiotics and antibiotic resistance in nature. Frontiers in Microbiology 4:1–13.

Shatzkes, K., E. Singleton, C. Tang, M. Zuena, S. Shukla, S. Gupta, S. Dharani, O. Onyile, J. Rinaggio, N. D. Connell & D. E. Kadouri. 2016. Predatory bacteria attenuate Klebsiella pneumoniae burden in rat lungs. American Society for Microbiology 7:1–9.

Shen, Y., M. Barros, T. Vennemann, D. T. Gallagher, Y. Yin, S. B. Linden, R. D. Heselpoth, D. J. Spencer, D. M. Donovan, J. Moult, V. A. Fischetti, F. Heinrich, M. Lische & D. C. Nelson. 2016. A bacteriophage endolysin that eliminates intracellular streptococci. eLife 5:1–26.

Shen, Y., T. Köller, B. Kreikemeyer & D. C. Nelson. 2013. Rapid degradation of Streptococcus pyogenes biofilms by PlyC, a bacteriophage-encoded endolysin. Journal of Antimicrobial Chemotherapy 68:1818–1824.

Sherman, M. Y. & S. B. Qian. 2013. Less is more: Improving proteostasis by translation slow down. Trends in Biochemical Sciences 38:585–591.

Shi, N. 2013. Cross section of a Chlamydomonas reinhardtii algae cell, a 3D representation. <https://commons.wikimedia.org/wiki/File:Cross_section_of_a_Chlamydomonas_reinhardtii_algae_cell,_a_3D_representation.jpg> (16 July 2017).

Shi, Y., N. Li, Y. Yan, H. Wang, Y. Li, C. Lu & J. Sun. 2012. Combined Antibacterial activity of phage lytic proteins holin and lysin from streptococcus suis bacteriophage SMP. Current Microbiology 65:28–34.

Shin, S.-E., J.-M. Lim, H. G. Koh, E. K. Kim, N. K. Kang, S. Jeon, S. Kwon, W.-S. Shin, B. Lee, K. Hwangbo, J. Kim, S. H. Ye, J.-Y. Yun, H. Seo, H.-M. Oh, K.-J. Kim, J.-S. Kim, W.-J. Jeong, Y. K. Chang et al. 2016. CRISPR/Cas9-induced knockout and knock-in mutations in Chlamydomonas reinhardtii. Scientific Reports 6:27810.

Silver, L. L. 2011. Challenges of antibacterial discovery. Clinical Microbiology Reviews 24:71–109.

Silver, L. L. & K. A. Bostian. 1993. Discovery and Development of New Antibiotics: the Problem of Antibiotic Resistance. Antimicrobial Agents and Chemotherapy 37:377–383.

Smith, D. L., D. K. Struck, J. M. Scholtz & R. Young. 1998. Purification and biochemical characterization of the lambda holin. Journal of bacteriology 180:2531–40.

Smith, E., L. Howard & E. Dymek. 1996. Scanning electron microscope image, showing an example of green algae (Chlorophyta). <http://remf.dartmouth.edu/images/algaeSEM/source/4.html> (16 July 2017).

Sockett, R. E. & C. Lambert. 2004. Opinion: Bdellovibrio as therapeutic agents: a predatory renaissance? Nature Reviews Microbiology 2:669–675.

Song, J., F. Xia, H. Jiang, X. Li, L. Hu, P. Gong, L. Lei, X. Feng, C. Sun, J. Gu & W. Han. 2016. Identification and characterization of HolGH15: The Holin of Staphylococcus Aureus bacteriophage GH15. Journal of General Virology 97:1272–1281.

Songer, J. G. 2004. The emergence of Clostridium difficile as a pathogen of food animals. Animal Health Research Reviews 5:321–326.

Specht, E. A. & S. P. Mayfield. 2013. Synthetic oligonucleotide libraries reveal novel regulatory elements in chlamydomonas chloroplast mRNAs. ACS Synthetic Biology 2:34–46.

Specht, E., S. Miyake-Stoner & S. Mayfield. 2010. Micro-algae come of age as a platform for recombinant protein production. Biotechnology Letters 32:1373–1383.

Spellberg, B. 2014. The future of antibiotics. Critical Care 18:228.

Spicer, A. & S. Purton. 2016. Genetic Engineering of Microalgae. Current Status and Future Prospects. Pp. 139–163 in Microalgal Production for Biomass and High-Value Products (S. P. Slocombe & J. R. Benemann, eds.). CRC Press.

Spigaglia, P. 2016. Recent advances in the understanding of antibiotic resistance in Clostridium difficile infection. Therapeutic Advances in Infectious Disease 3:23–42.

Spolaore, P., C. Joannis-Cassan, E. Duran & A. Isambert. 2006. Commercial applications of microalgae. Journal of Bioscience and Bioengineering 101:87–96.

Stanek, R. J., M. B. Maher, N. B. Norton & M. A. Mufson. 2011. Emergence of a unique penicillin-resistant Streptococcus pneumoniae serogroup 35 strain. Journal of Clinical Microbiology 49:400–404.

Stegemann, S., S. Hartmann, S. Ruf & R. Bock. 2003. High-frequency gene transfer from the chloroplast genome to the nucleus. Proceedings of the National Academy of Sciences of the United States of America 100:8828–8833.

Steinbrenner, J. & G. Sandmann. 2006. Transformation of the green alga Haematococcus pluvialis with a phytoene desaturase for accelerated astaxanthin biosynthesis. Applied and Environmental Microbiology 72:7477–7484.

Stoffels, L. 2015. Microalgae as a novel production platform for antibacterial proteins. PhD thesis. London: University College London.

Stoffels, L., H. N. Taunt, B. Charalambous & S. Purton. 2017. Synthesis of bacteriophage lytic proteins against Streptococcus pneumoniae in the chloroplast of Chlamydomonas reinhardtii. Plant Biotechnology Journal 15:1130–1140.

Sun, M., K. Qian, N. Su, H. Chang, J. Liu & G. Chen. 2003. Foot-and-mouth disease virus VP1 protein fused with cholera toxin B subunit expressed in Chlamydomonas reinhardtii chloroplast. Biotechnology Letters 25:1087–1092.

Sun, X. & J. C. Linden. 1999. Shear stress effects on plant cell suspension cultures in a rotating wall vessel bioreactor. Journal of Industrial Microbiology & Biotechnology 22:44–47.

Surzycki, R., K. Greenham, K. Kitayama, F. Dibal, R. Wagner, J. D. Rochaix, T. Ajam & S. Surzycki. 2009. Factors effecting expression of vaccines in microalgae. Biologicals 37:133–138.

Tängdén, T., R. A. Hickman, P. Forsberg, P. Lagerbäck, C. G. Giske & O. Cars. 2014. Combination antibiotic therapy for multidrug-resistant Gram-negative bacteria. Antimicrobial agents and chemotherapy 9734:1–5.

Taunt, H. N. 2013. The Synthesis of Novel Antibacterial Proteins in the Chlamydomonas reinhardtii Chloroplast. PhD thesis. London: University College London.

Taylor, E. J. M., Y. Yu, J. Champer & J. Kim. 2014. Resveratrol Demonstrates Antimicrobial Effects Against Propionibacterium acnes In Vitro. Dermatology and Therapy 4:249–257.

Tolba, M., M. U. Ahmed, C. Tlili, F. Eichenseher, M. J. Loessner, M. Zourob, L. Yang, R. Bashir, V. Ramaswamy, V. M. Cresence, J. S. Rejitha, M. U. Lekshmi, K. S. Dharsana, S. P. Prasad, H. M. Vijila, P. Mook, S. J. O'Brien, I. A. Gillespie, C. Genigeorgis et al. 2012. A bacteriophage endolysin-based electrochemical impedance biosensor for the rapid detection of Listeria cells. The Analyst 137:5749.

Tran, M., C. Van, D. J. Barrera, P. L. Pettersson, C. D. Peinado, J. Bui & S. P. Mayfield. 2013. Production of unique immunotoxin cancer therapeutics in algal chloroplasts. Proceedings of the National Academy of Sciences of the United States of America 110:E15-22.

Tran, M., B. Zhou, P. L. Pettersson, M. J. Gonzalez & S. P. Mayfield. 2009. Synthesis and assembly of a full-length human monoclonal antibody in algal chloroplasts. Biotechnology and Bioengineering 104:663–673.

US. Department of Health and Human Services 2013. Antibiotic resistance threats in the United States, 2013. *Centres for Disease Control and Prevention*. <http://www.cdc.gov/drugresistance/threat-report-2013/index.html>.

Vahrenholz, C., G. Riemen, E. Pratje, B. Dujon & G. Michaelis. 1993. Mitochondrial DNA of Chlamydomonas reinhardtii: the structure of the ends of the linear 15.8-kb genome suggests mechanisms for DNA replication. Current Genetics 24:241–247.

Varon, M. 1979. Selection of predation-resistant bacteria in continuous culture. Nature 277:386–388.

Ventola, C. L. 2015. The antibiotic resistance crisis: part 1: causes and threats. P & T : A peer-reviewed journal for formulary management 40:277–83.

Walker, T. L., S. Purton, D. K. Becker & C. Collet. 2005. Microalgae as bioreactors. Plant cell reports 24:629–41.

Walsh, C. T. & T. A. Wencewicz. 2014. Prospects for new antibiotics: a molecule-centered perspective. The Journal of Antibiotics 67:7–22.

Wang, I.-N. N., D. L. Smith & R. Young. 2000. Holins: the protein clocks of bacteriophage infections. Annual review of microbiology 54:799–825.

Wang, Q., C. W. Euler, A. Delaune & V. A. Fischetti. 2015. Using a novel lysin to help control *Clostridium difficile* infections. Antimicrobial Agents and Chemotherapy 59:AAC.01357-15.

Wang, X., M. Brandsma, R. Tremblay, D. Maxwell, A. M. Jevnikar, N. Huner & S. Ma. 2008. A novel expression platform for the production of diabetes-associated autoantigen human glutamic acid decarboxylase (hGAD65). BMC Biotechnology 8:87.

Wannathong, T., J. C. Waterhouse, R. E. B. Young, C. K. Economou & S. Purton. 2016. New tools for chloroplast genetic engineering allow the synthesis of human growth hormone in the green alga Chlamydomonas reinhardtii. Applied Microbiology and Biotechnology 100:5467–5477.

Werth, B. (2014). *The billion-dollar molecule*. New York: Simon & Schuster, p.122.

White, R., S. Chiba, T. Pang, J. S. Dewey, C. G. Savva, A. Holzenburg, K. Pogliano & R. Young. 2011. Holin triggering in real time. Proceedings of the National Academy of Sciences of the United States of America 108:798–803.

WHO. 2007. Pneumococcal conjugate vaccine for childhood immunization – WHO position paper. Weekly epidemiological record 82:93–104.

WHO. 2016. Antimicrobial resistance Fact sheet. World Health Organization. <http://www.who.int/mediacentre/factsheets/fs194/en/> (14 July 2017).

Willmund, F., K. V. Dorn, M. Schulz-Raffelt & M. Schroda. 2008. The Chloroplast DnaJ Homolog CDJ1 of Chlamydomonas reinhardtii Is Part of a Multichaperone Complex Containing HSP70B, CGE1, and HSP90C. Plant Physiology 148:2070–2082.

Wilson, M. 2005. Microbial inhabitants of humans : their ecology and role in health and disease. Cambridge, UK: Cambridge University Press.

Witzenrath, M., B. Schmeck, J. M. Doehn, T. Tschernig, J. Zahlten, J. M. Loeffler, M. Zemlin, H. Muller, B. Gutbier, H. Schutte, S. Hippenstiel, V. a Fischetti, N. Suttorp & S. Rosseau. 2009. Systemic use of the endolysin Cpl-1 rescues mice with fatal pneumococcal pneumonia. Critical Care Medicine 37:642–649.

Wurm, F. M. 2004. Production of recombinant protein therapeutics in cultivated mammalian cells. Nature Biotechnology 22:1393–1398.

Xu, J. & N. Zhang. 2014. On the way to commercializing plant cell culture platform for biopharmaceuticals: present status and prospect. Pharmaceutical Bioprocessing 2:499–518.

Yang, S.-C., C.-H. Lin, C. T. Sung & J.-Y. Fang. 2014. Antibacterial activities of bacteriocins: application in foods and pharmaceuticals. Frontiers in Microbiology 5:241.

Yang, Z., Y. Li, F. Chen, D. Li, Z. Zhang, Y. Liu, D. Zheng, Y. Wang & G. Shen. 2006. Expression of human soluble TRAIL in Chlamydomonas reinhardtii chloroplast. Chinese Science Bulletin 51:1703–1709.

Yao, J., Y. Weng, A. Dickey & K. Y. Wang. 2015. Plants as factories for human pharmaceuticals: Applications and challenges. International Journal of Molecular Sciences 16:28549–28565.

Young, R. 1992. Bacteriophage lysis: mechanism and regulation. Microbiological reviews 56:430–81.

Young, R. E. B. & S. Purton. 2014. Cytosine deaminase as a negative selectable marker for the microalgal chloroplast: A strategy for the isolation of nuclear mutations that affect chloroplast gene expression. Plant Journal 80:915–925.

Young, R. E. B. & S. Purton. 2015. Codon reassignment to facilitate genetic engineering and biocontainment in the chloroplast of Chlamydomonas reinhardtii. Plant Biotechnology Journal:1–10.

Young, R. E. B. & S. Purton. 2016. Codon reassignment to facilitate genetic engineering and biocontainment in the chloroplast of Chlamydomonas reinhardtii. Plant Biotechnology Journal 14:1251–1260.

Zaenglein, A. L., A. L. Pathy, B. J. Schlosser, A. Alikhan, H. E. Baldwin, D. S. Berson, W. P. Bowe, E. M. Graber, J. C. Harper, S. Kang, J. E. Keri, J. J. Leyden, R. V. Reynolds, N. B. Silverberg, L. F. Stein Gold, M. M. Tollefson, J. S. Weiss, N. C. Dolan, A. A. Sagan et al. 2016. Guidelines of care for the management of acne vulgaris. Journal of the American Academy of Dermatology 74:945–973e33.

Zaffiri, L., J. Gardner & L. H. Toledo-Pereyra. 2012. History of antibiotics. From salvarsan to cephalosporins. Journal of Investigative Surgery 25:67–77.

Zedler, J. A. Z. Z., D. Gangl, B. B. Hamberger, S. Purton & C. Robinson. 2015. Stable expression of a bifunctional diterpene synthase in the chloroplast of Chlamydomonas reinhardtii. Journal of Applied Phycology 27:2271–2277.

Zienkiewicz, M., T. Krupnik, A. Droak, A. Golke & E. Romanowska. 2017. Transformation of the Cyanidioschyzon merolae chloroplast genome: prospects for understanding chloroplast function in extreme environments. Plant Molecular Biology 93:171–183.

Chapter 7 Appendices

7.1 Results Chapter 3 Appendices

7.1.1 Appendix A

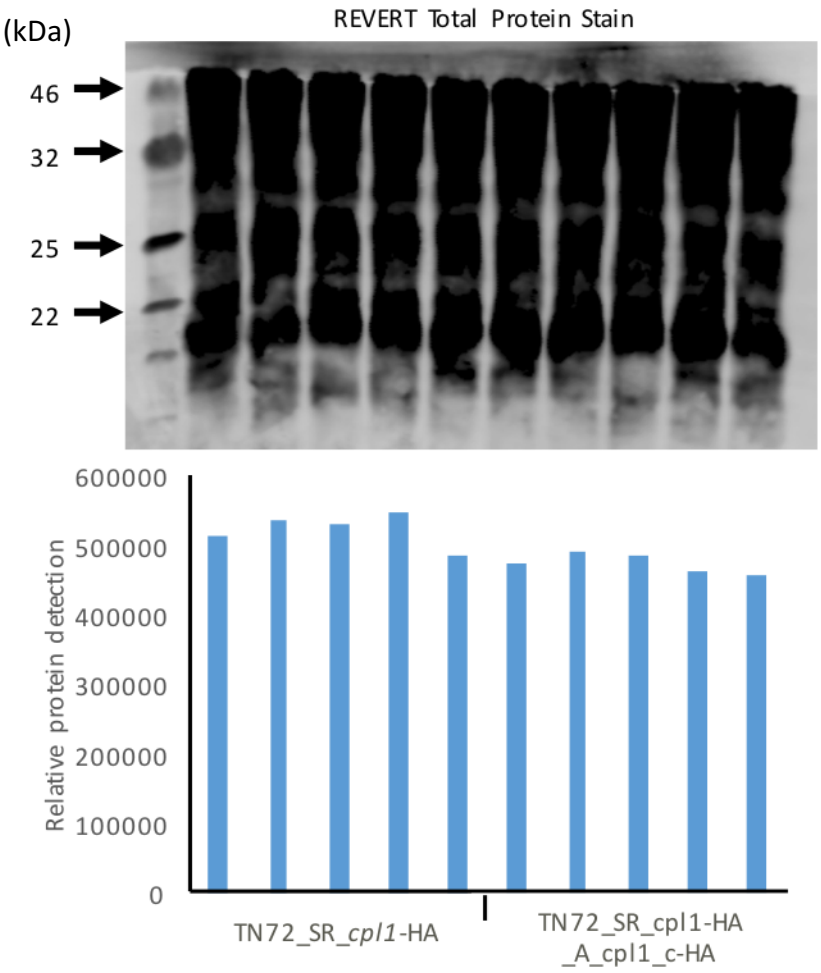


Figure 7.1 Total protein staining of five technical repeats of TN72_SR_*cpl1*-HA and TN72_SR_*cpl1*-HA_A_*cpl1*_*c*-HA. These data were used for normalisation of the anti-HA signal to account for loading error.

7.1.2 Appendix B

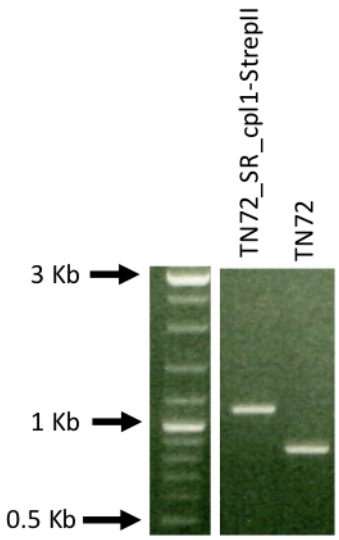


Figure 7.2 Demonstration of homoplasmicity of TN72_SR_*cpl1*-StrepII. Primers used in this instance were Flank1, psa.R, RbcL.F which provide an 850 bp band for the recipient strain and a 1135 bp band for a homoplasmic strain.

7.1.3 Appendix C

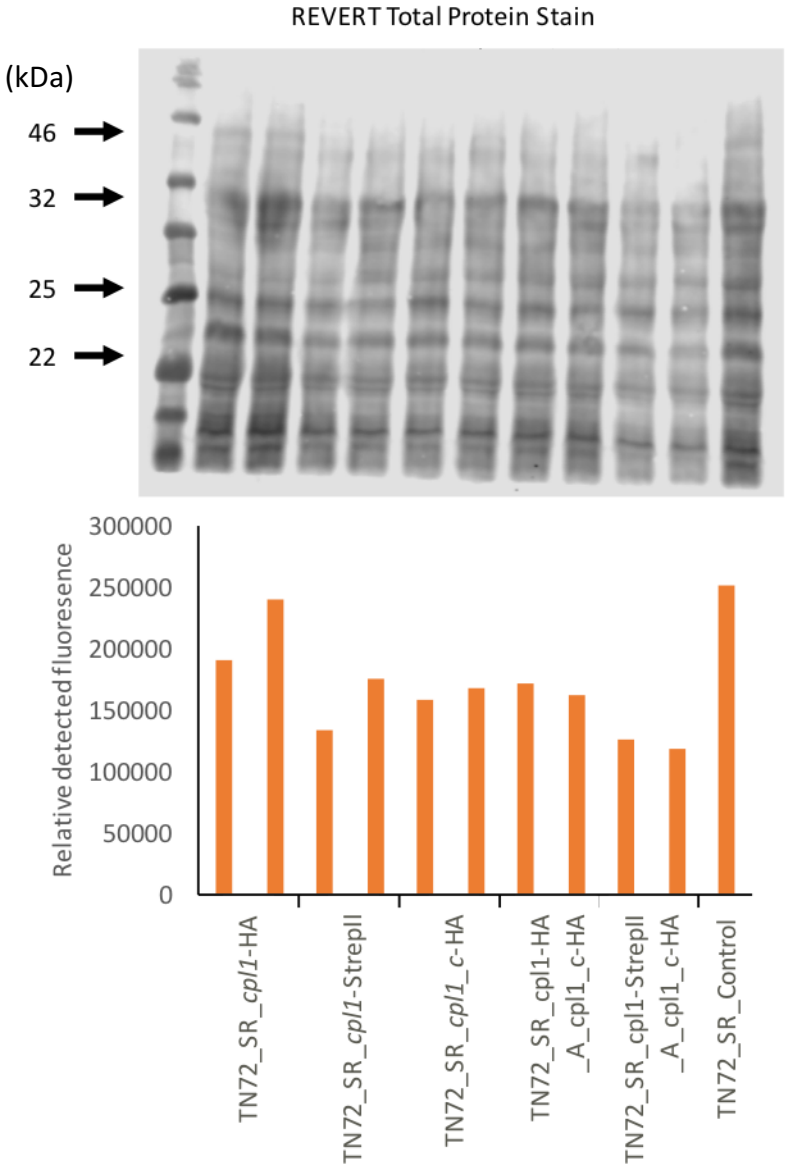


Figure 7.3 Total protein staining used to normalise for loading errors when comparing total protein accumulation between the above *C. reinhardtii* lines.

7.1.4 Appendix D

Sequence length: 21279 Number of protein genes: 32 Total codons: 7093

Amino acid	Codon	Number	Frequency per thousand												
Phe	UUU	171	24.1	Ser	UCU	147	20.7	Tyr	UAU	87	12.3	Cys	UGU	61	8.6
	UUC	285	40.2		UCC	0	0.0		UAC	152	21.4		UGC	3	0.4
Leu	UUA	544	76.7		UCA	144	20.3	STOP	UAA	30	4.2	STOP	UGA	0	0.0
	UUG	5	0.7		UCG	9	1.3		UAG	2	0.3	Trp	UGG	170	24.0
Leu	CUU	152	21.4	Pro	CCU	119	16.8	His	CAU	23	3.2	Arg	CGU	245	34.5
	CUC	0	0.0		CCC	2	0.3		CAC	148	20.9		CGC	10	1.4
	CUA	53	7.5		CCA	200	28.2	Gln	CAA	218	30.7		CGA	2	0.3
	CUG	9	1.3		CCG	13	1.8		CAG	10	1.4		CGG	0	0.0
ile	AUU	343	48.4	Thr	ACU	210	29.6	Asn	AAU	83	11.7	Ser	AGU	80	11.3
	AUC	102	14.4		ACC	3	0.4		AAC	198	27.9		AGC	30	4.2
	AUA	4	0.6		ACA	193	27.2	Lys	AAA	240	33.8	Arg	AGA	8	1.1
Met	AUG	182	25.7		ACG	12	1.7		AAG	9	1.3		AGG	2	0.3
Val	GUU	258	36.4	Ala	GCU	433	61.0	Asp	GAU	135	19.0	Gly	GGU	586	82.6
	GUC	0	0.0		GCC	9	1.3		GAC	115	16.2		GGC	24	3.4
	GUA	261	36.8		GCA	173	24.4	Glu	GAA	311	43.8		GGA	19	2.7
	GUG	6	0.8		GCG	23	3.2		GAG	20	2.8		GGG	7	1.0

Figure 7.4 Codon frequency chart generated for *C. reinhardtii* chloroplast transgene codon optimization, from *Codon Usage Optimizer Beta 0.92* (Kong 2013).

7.1.5 Appendix E

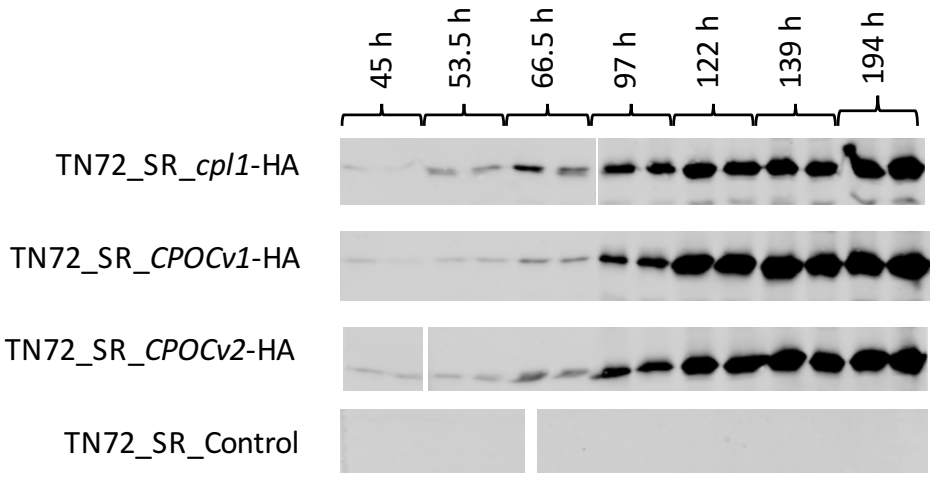


Figure 7.5 Western blot analysis demonstrating the varying levels of HA-tagged protein accumulation over a 194 hour culture period.

7.1.6 Appendix F

A one-way ANOVA test was carried out at each time point e.g. t = 45 for samples taken at 45 hours into the experiment. In this experiment, four time times were analysed, with 3 replicates of each on the western blot (Figure 3.37).

Results at t= 45

Summary of Data

	Treatments			
	CPOCv1	CPOCv2	cpl1-HA	Total
N	3	3	3	9
∑X	0.115	0.4351	0.1253	0.6754
Mean	0.0383	0.145	0.0418	0.075
$\sum X^2$	0.0044	0.0632	0.0054	0.073
Std.Dev.	0.0035	0.0066	0.0075	0.0528

Result Details

Source	***SS***	***df***	***MS***	
Between-treatments	0.0221	2	0.011	*F* = 293.57558
Within-treatments	0.0002	6	0	
Total	0.0223	8		

The *f*-ratio value is 293.57558. The *p*-value is < .00001. The result is significant at $p < .05$.

Results at t= 58

Summary of Data

	Treatments			
	CPOCv1	CPOCv2	cpl1-HA	Total
N	3	3	3	9
∑X	0.3091	1.5043	0.2813	2.0948
Mean	0.103	0.5014	0.0938	0.2328
$\sum X^2$	0.0319	0.7607	0.0274	0.82
Std.Dev.	0.0038	0.0565	0.0221	0.2038

Result Details

Source	***SS***	***df***	***MS***	
Between-treatments	0.325	2	0.1625	*F* = 131.77838
Within-treatments	0.0074	6	0.0012	
Total	0.3324	8		

The *f*-ratio value is 131.77838. The *p*-value is .000011. The result is significant at $p < .05$.

Results at t= 77

Summary of Data

	Treatments			
1	CPOCv1	CPOCv2	cpl1-HA	Total
N	3	3	3	9
∑X	0.9648	1.4821	0.663	3.1099
Mean	0.3216	0.494	0.221	0.3455
$\sum X^2$	0.3137	0.8476	0.1472	1.3085
Std.Dev.	0.0411	0.2403	0.0178	0.171

Result Details

Source	***SS***	***df***	***MS***	
Between-treatments	0.1144	2	0.0572	*F* = 2.87208
Within-treatments	0.1195	6	0.0199	
Total	0.2339	8		

The *f*-ratio value is 2.87208. The *p*-value is .133348.
The result is *not* significant at $p < .05$.

Results at t= 119

Summary of Data

	Treatments			
1	CPOCv1	CPOCv2	cpl1-HA	Total
N	2	2	2	6
∑X	0.894	0.9605	0.8009	2.6554
Mean	0.447	0.4803	0.4004	0.4426
$\sum X^2$	0.4022	0.4775	0.3281	1.2078
Std.Dev.	0.0505	0.1271	0.086	0.0807

Result Details

Source	***SS***	***df***	***MS***	
Between-treatments	0.0064	2	0.0032	*F* = 0.36948
Within-treatments	0.0261	3	0.0087	
Total	0.0325	5		

The *f*-ratio value is 0.36948. The *p*-value is .718715.
The result is *not* significant at $p < .05$.

7.1.7 Appendix G

A one-way ANOVA test was carried out at each time point e.g. t = 45 for samples taken at 45 hours into the experiment. In this second running of the experiment, seven time points were analysed but with only two replicates of each (Figure 7.5).

Results at t = 45

Summary of Data

	Treatments			
	CPOCv1	CPOCv2	Cpl-1	Total
N	2	2	2	6
∑X	0.0837	0.2317	0.0405	0.3559
Mean	0.0418	0.1158	0.0202	0.0593
$\sum X^2$	0.0039	0.0268	0.0009	0.0316
Std.Dev.	0.02	0.0006	0.0059	0.0458

Result Details

Source	***SS***	***df***	***MS***	
Between-treatments	0.0101	2	0.005	*F* = 34.67033
Within-treatments	0.0004	3	0.0001	
Total	0.0105	5		

The *f*-ratio value is 34.67033. The *p*-value is .008445. The result is significant at $p < .05$.

Results at t = 53.5

Summary of Data

	Treatments			
	CPOCv1	CPOCv2	Cpl-1	Total
N	2	2	2	6
∑X	0.0713	0.1784	0.1145	0.3642
Mean	0.0357	0.0892	0.0573	0.0607
$\sum X^2$	0.0026	0.0159	0.0066	0.0251
Std.Dev.	0.0062	0.0057	0.0008	0.0244

Result Details

Source	***SS***	***df***	***MS***	
Between-treatments	0.0029	2	0.0015	*F* = 61.0499
Within-treatments	0.0001	3	0	
Total	0.003	5		

The *f*-ratio value is 61.0499. The *p*-value is .003714. The result is significant at $p < .05$.

Results at t = 66.5

Summary of Data

	Treatments			
	CPOCv1	CPOCv2	Cpl-1	Total
N	2	2	2	6
∑X	0.1001	0.2606	0.1458	0.5065
Mean	0.05	0.1303	0.0729	0.0844
$\sum X^2$	0.005	0.0356	0.011	0.0516
Std.Dev.	0.0003	0.0405	0.0193	0.0421

Result Details

Source	***SS***	***df***	***MS***	
Between-treatments	0.0068	2	0.0034	*F* = 5.09397
Within-treatments	0.002	3	0.0007	
Total	0.0089	5		

The *f*-ratio value is 5.09397. The *p*-value is .108497. The result is *not* significant at $p < .05$.

Results at t = 97

Summary of Data

	Treatments			
	CPOCv1	CPOCv2	Cpl-1	Total
N	2	2	2	6
∑X	0.1803	0.2498	0.1553	0.5854
Mean	0.0901	0.1249	0.0776	0.0976
$\sum X^2$	0.0168	0.0313	0.0121	0.0603
Std.Dev.	0.0235	0.0102	0.0095	0.0251

Result Details

Source	***SS***	***df***	***MS***	
Between-treatments	0.0024	2	0.0012	*F* = 4.82318
Within-treatments	0.0007	3	0.0002	
Total	0.0031	5		

The *f*-ratio value is 4.82318. The *p*-value is .11554. The result is *not* significant at $p < .05$.

Results at t = 122

Summary of Data

	Treatments			
	CPOCv1	CPOCv2	Cpl-1	Total
N	2	2	2	6
∑X	0.9233	0.8594	0.3558	2.1385
Mean	0.4617	0.4297	0.1779	0.3564
$\sum X^2$	0.4288	0.3714	0.0634	0.8636
Std.Dev.	0.0505	0.0458	0.0095	0.1424

Result Details

Source	***SS***	***df***	***MS***	
Between-treatments	0.0966	2	0.0483	*F* = 30.59972
Within-treatments	0.0047	3	0.0016	
Total	0.1014	5		

The *f*-ratio value is 30.59972. The *p*-value is .010101. The result is significant at $p < .05$.

Results at t = 139

Summary of Data

	Treatments			
	CPOCv1	CPOCv2	Cpl-1	Total
N	2	2	2	6
∑X	1.1899	1.1952	0.3528	2.7379
Mean	0.595	0.5976	0.1764	0.4563
$\sum X^2$	0.7086	0.7301	0.0634	1.5021
Std.Dev.	0.0245	0.126	0.0342	0.2248

Result Details

Source	***SS***	***df***	***MS***	
Between-treatments	0.2351	2	0.1175	*F* = 20.00184
Within-treatments	0.0176	3	0.0059	
Total	0.2527	5		

The *f*-ratio value is 20.00184. The *p*-value is .018426. The result is significant at $p < .05$.

Results at t = 194

Summary of Data

	Treatments			
	CPOCv1	CPOCv2	Cpl-1	Total
N	2	2	2	6
∑X	1.4863	1.4019	0.6651	3.5533
Mean	0.7431	0.7009	0.3326	0.5922
$\sum X^2$	1.1065	0.9986	0.222	2.327
Std.Dev.	0.0445	0.1261	0.0284	0.211

Result Details

Source	***SS***	***df***	***MS***	
Between-treatments	0.204	2	0.102	*F* = 16.38062
Within-treatments	0.0187	3	0.0062	
Total	0.2227	5		

The *f*-ratio value is 16.38062. The *p*-value is .024298. The result is significant at $p < .05$.

7.2 Results Chapter 4 Appendices

7.2.1 Appendix H

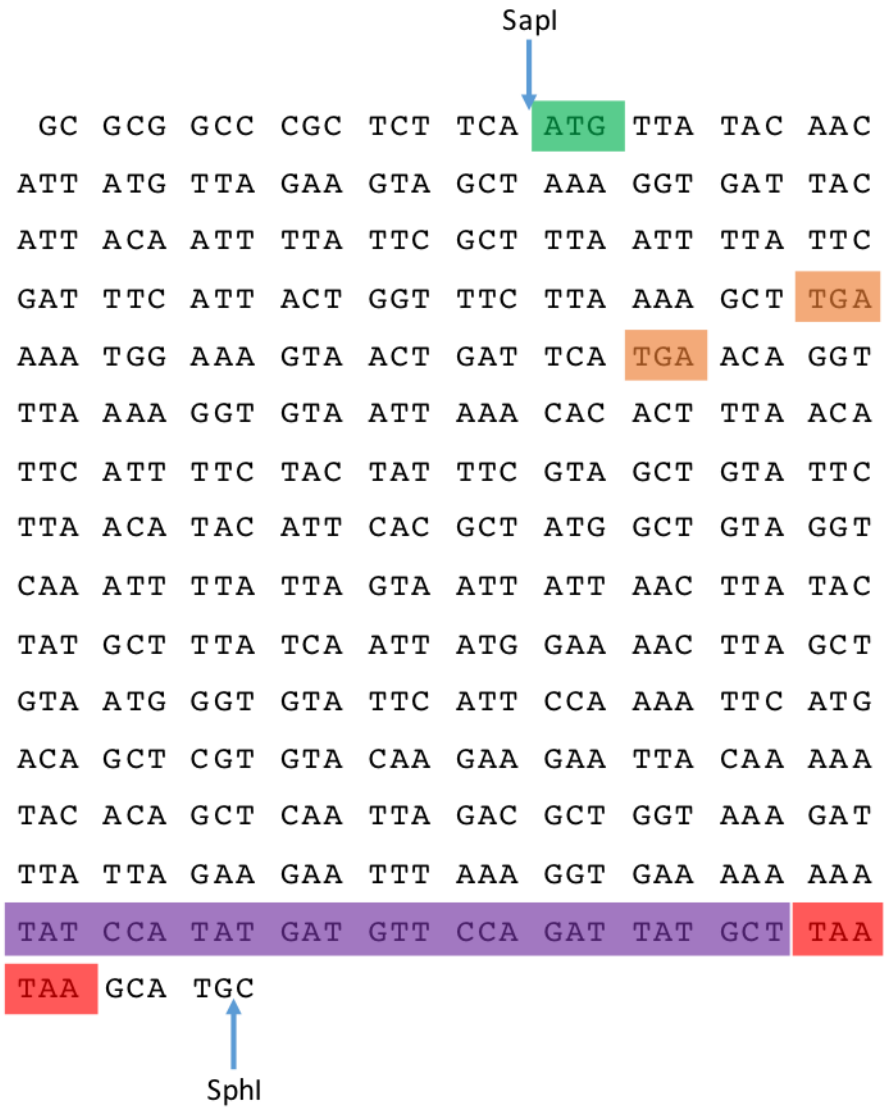


Figure 7.6 *cph1* sequence as ordered from Integrated DNA Technologies. The start codon is highlighted in green. The introduced TGA codons for tryptophan are highlighted in orange. The stop codons are highlighted in red. The HA tag is highlighted in purple. SapI and SphI restriction sites are indicated by blue arrows.

7.2.2 Appendix I

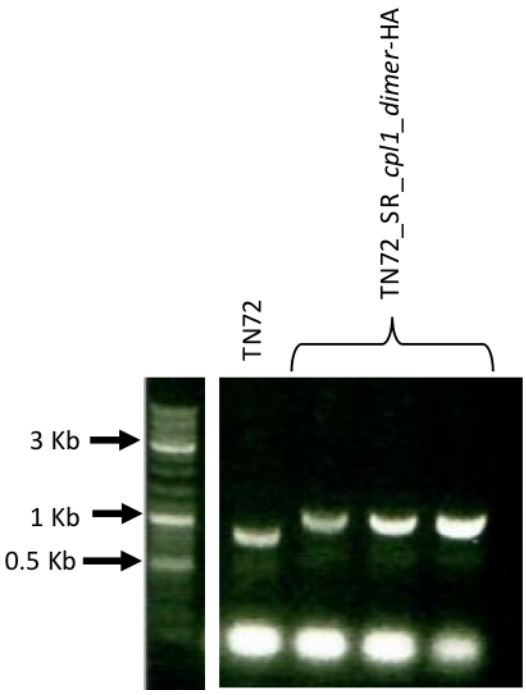


Figure 7.7 Demonstration of homoplasmicity of TN72_SR_*cpl1_dimer*-HA. Primers used in this instance were Flank1, psa.R, RbcL.F which produce an 850 bp band for the recipient strain and an 1135 bp band for a homoplasmic strain.

7.3 Results Chapter 5 Appendices

7.3.1 Appendix J

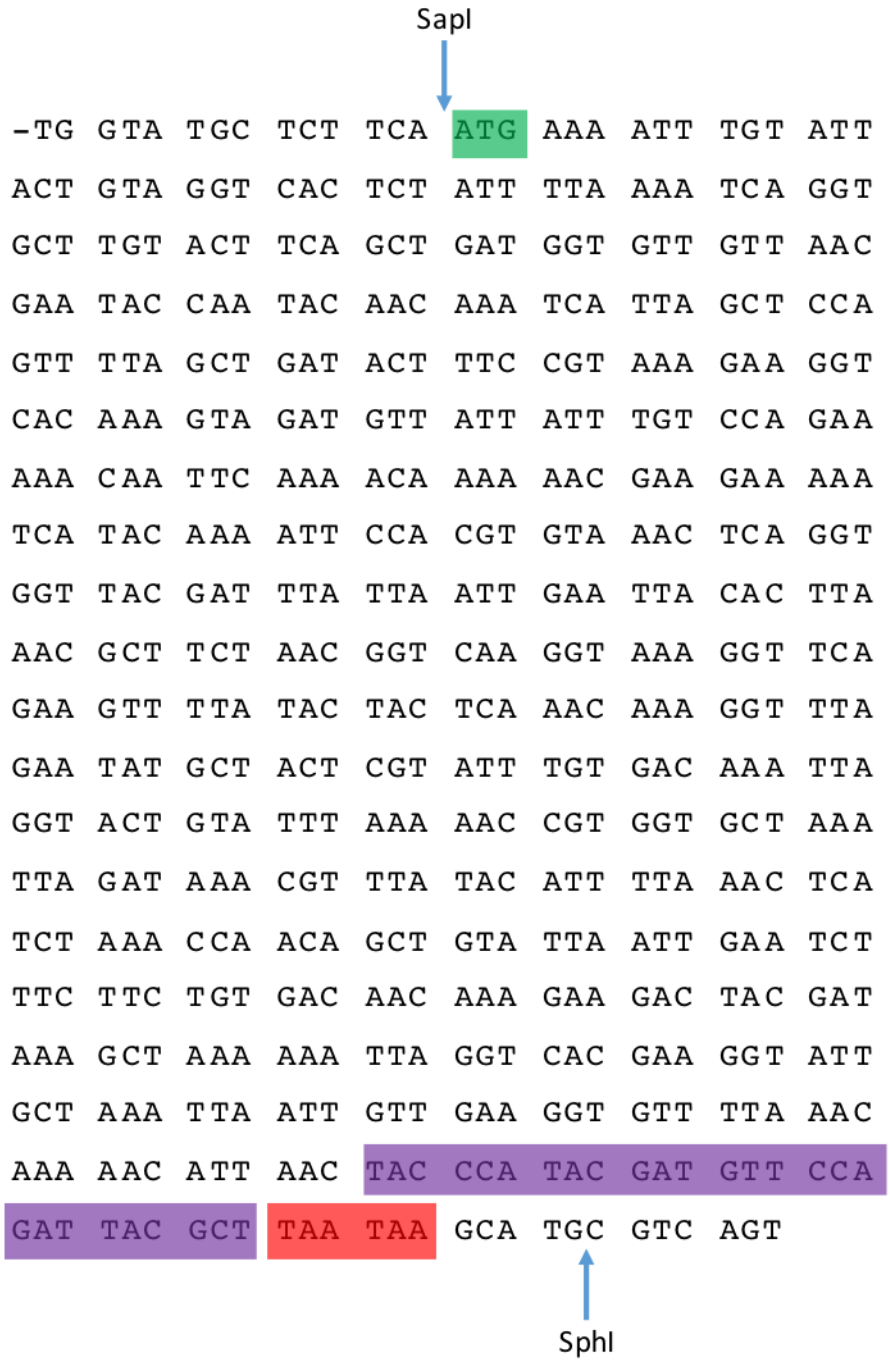


Figure 7.8 *cd27l* sequence as ordered from Integrated DNA Technologies. The start codon is highlighted in green. The stop codons are highlighted in red. The HA tag is highlighted in purple. SapI and SphI restriction sites are indicated by blue arrows.

7.3.2 Appendix K

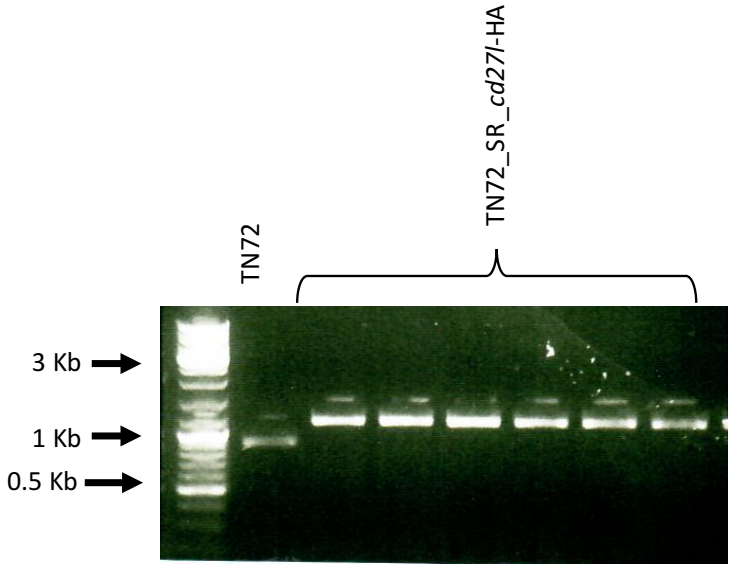


Figure 7.9 Demonstration of homoplasmicity of TN72_SR_*cd27l*-HA. Primers used in this instance were Flank1, psa.R, RbcL.F which produce an 850 bp band for the recipient strain and an 1135 bp band for a homoplasmic strain. Six transformants were tested.

7.3.3 Appendix L

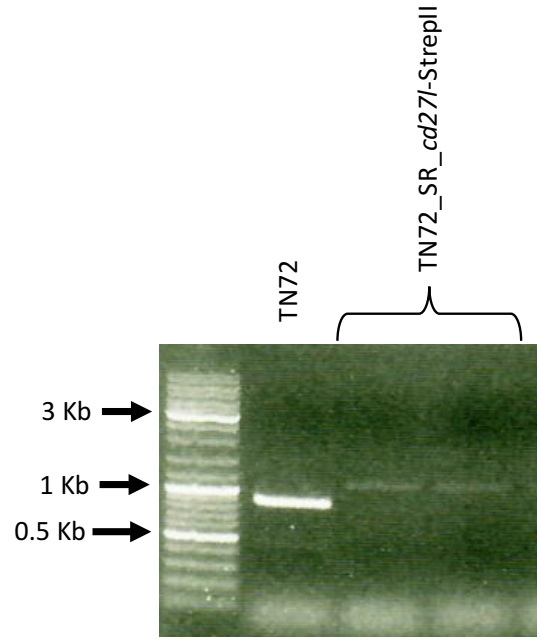


Figure 7.10 Demonstration of homoplasmicity of TN72_SR_*cd27l*-StrepII. Primers used in this instance were Flank1, psa.R, RbcL.F which produce an 850 bp band for the recipient strain and an 1135 bp band for a homoplasmic strain. Two transformants were tested.

7.3.4 Appendix M

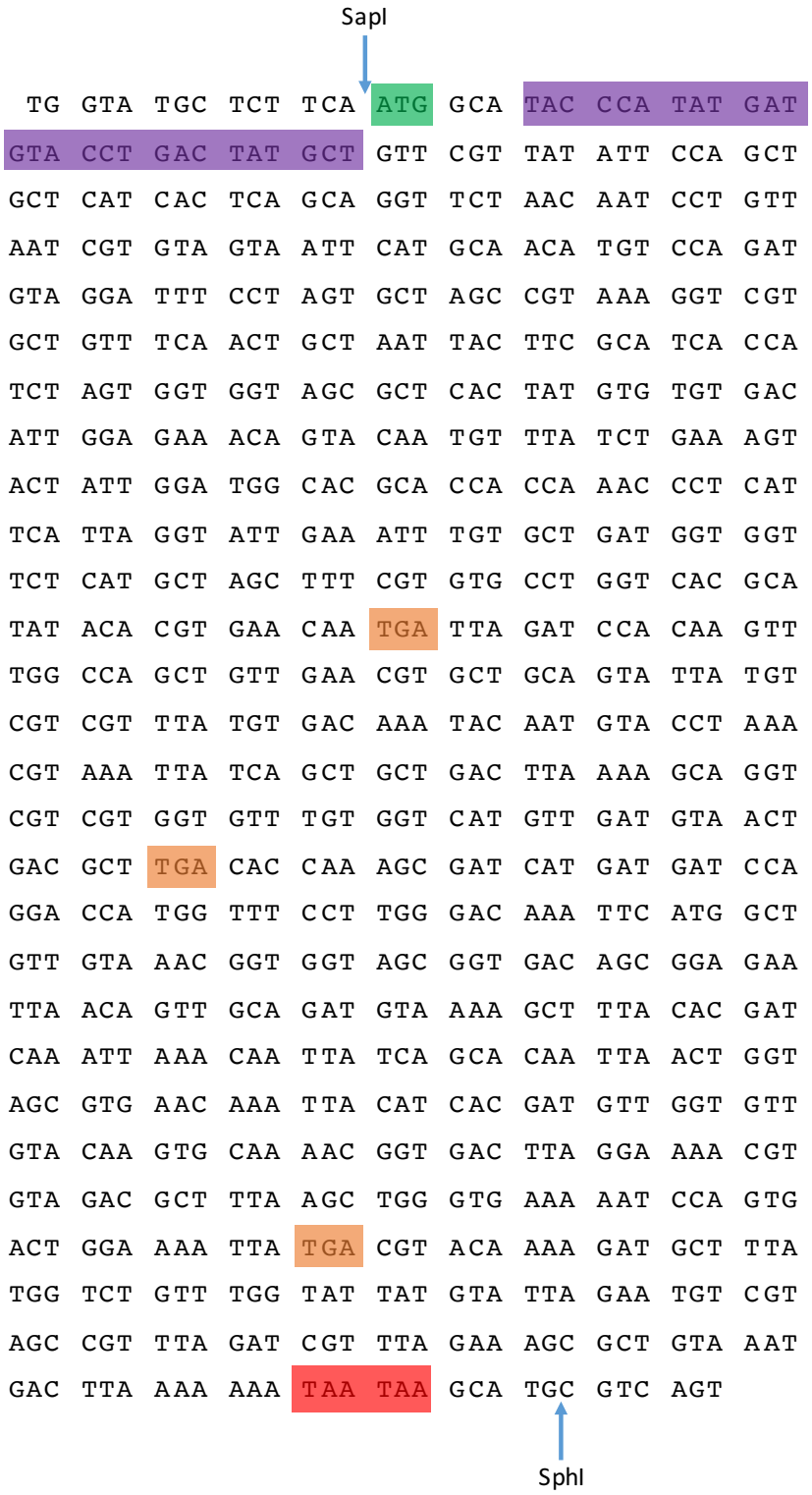


Figure 7.11 *gp20N* sequence as ordered from Integrated DNA Technologies. The start codon is highlighted in green. The introduced TGA codons for tryptophan are highlighted in orange. The stop codons are highlighted in red. The HA tag is highlighted in purple. SapI and SphI restriction sites are indicated by blue arrows.

7.3.5 Appendix N

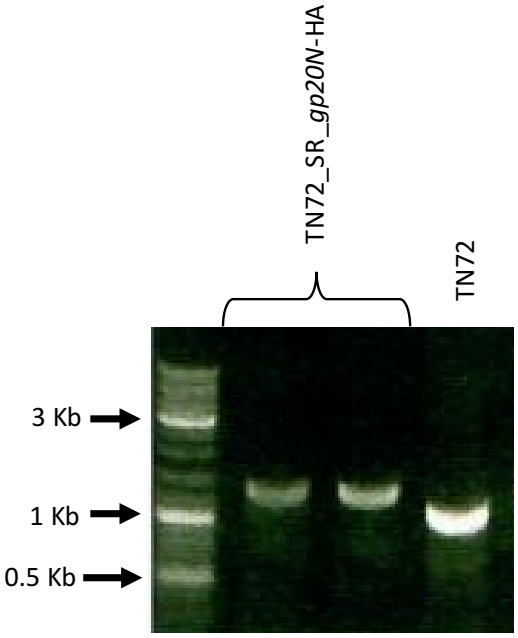


Figure 7.12 Demonstration of homoplasmicity of TN72_SR_*gp20N*-HA. Primers used in this instance were Flank1, psa.R, RbcL.F which produce an 850 bp band for the recipient strain and an 1135 bp band for a homoplasmic strain. Two transformants were tested.

7.3.6 Appendix O: List of Primers

Primers in order of appearance in this thesis.

NAME	SEQUENCE	NOTES
Flank1	GTCATTGCGAAAATACTGGTGC	Primers for general sequencing, PCR and homoplasmicity testing
RbcL.F	CGGATGTAACTCAATCGGTAG	
Rbcl.R	CAAACTTCACATGCAGCAGC	
Psa.R	CATGGATTTCTCCTTATAATAAC	
atpA.R	ACGTCCACAGGCGTCGT	
123a	GGAGGGAAGTAGGCAGTAGC	
psbH.R	GCAACAGGAACTTCTAAAGC	
psaSEQ	AACTATTTGTCTAATTTAATAACC	

cpl1.F	GCTCTTCAATGGTTAAAAAAAATGATTTATTC	Primers for the site-directed mutageneis of the cell wall binding domain
cpl1.R	GCATGCTTATTAAGCATAATCTGGAAC	
W202A_SDM.F	GTACAGCTAAACAAGATTCAAAAG	
W202A_SDM.R	CTTTTGAATCTTGTTTAGCTGTAC	
W209A_SDM.F	CAAAAGGTGCTTGGTTTC	
W209A_SDM.R	GAAACCAAGCACCTTTTG	
W223A_SDM.F	CCATATAATAAAGCTGAAAAAATTG	
W223A_SDM.R	CAATTTTTTCAGCTTTATTATATGG	
W230A_SDM.F	GTGGTGTTGCTTATTATTTCG	
W230A_SDM.R	CGAAATAATAAGCAACACCAC	
Cpl1_BssHII.F	CAGGCAATTTGCTTACACCTTT	Primers to introduce a BssHII restriction site into Cpl-1
Cpl1_BssHII.R	TTAAGATAAGCGCGCTACATCC	
cpl1_C45S.F	ACATATTTAAATCCATCATTATCAGCTC	Primers to remove and introduce cysteine residues to create Cpl1_dimer
cpl1_C45S.R	GAGCTGATAATGATGGATTTAAATATGT	
cpl1_D324C.F	GTGAATTAGCATGTAATCCTTCATTT	
cpl1_D324C.R	AAATGAAGGATTACATGCTAATTCAC	
cd27l.F	TGGTATTGCTCTTCAATGAAAATTTGTATTACT	Primers for the replacement of the HA tag with a StrepII tag on *cd27l*
cd27l_StrepII.R	ACTGACGCATGCTTATTATTTTTCGAATTG TGGGTGAGACCAGTTAATGTTTTTGTTTAA AACACCTTCAAC	
gp20_NcoI.F	CGTATACCCATGGTGCGTTA	Primers to introduce an NcoI restriction site to *gp20*
gp20.R	GATTGCGGATCCTTATTAGGC	