Cardiff University

School of Biosciences

Design and Optimisation of Expression Vectors for the Synthesis of Eukaryotic Proteins in Plants

Ryan Jamie Coates

December 2023

Cardiff School of Biosciences

The Sir Martin Evans Building

Museum Avenue

CF10 3AX

A thesis submitted for the degree of Doctor of Philosophy for Cardiff University

(~48,500 words)

Abstract

Recombinant proteins are vital in sectors like medicine, industry, and agriculture. Conventional expression systems face challenges such as the inability of prokaryotic systems to perform eukaryotic post-translational modifications and the high costs of scaling up eukaryotic cell cultures. Plants are scalable and can produce eukaryotic modifications but suffer from low yields and underdeveloped downstream processing methods. This thesis aimed to enhance plant-based production of recombinant proteins, focussing on developing and optimizing plant expression vectors and protein processing procedures to maximize yields. A range of eukaryotic proteins, including those from plants and mammals, varying in size and sub-cellular localisation were expressed to understand how protein complexity affects yield. The goal was to express full-length rat and human P2X7 receptors and two truncated versions of isolated cytoplasmic domains. Using Golden Gate and conventional cloning, several expression constructs were created. These constructs significantly increased eGFP production in plants, up to 199-fold more than initial constructs and 2.2-fold more than the previously published pJL-TRBO vector. By comparing different protein extraction and purification methods an effective downstream processing pipeline was developed, enabling eGFP quantification of up to 20% total soluble protein. Compared to starting constructs, eGFP production was increased 254-fold at the protein level using the best construct when measured by western blot. With methods optimised for eGFP, other target proteins were expressed, extracted, purified, and quantified. This led to the development of several useful expression constructs for plants, enabling detectable production of the 53 kDa isolated cytoplasmic domain of P2X7 in a transient *N. benthamiana* expression system, although full-length transmembrane protein expression was not detected. Overall, this research successfully developed optimized expression vectors and protein extraction and purification techniques for high yields of poly(His)-tagged cytoplasmic proteins and paves the way for future large-scale research to express more diverse protein targets.

Acknowledgements

I would like to express my deepest gratitude to Dr. Mark Young and Dr. Simon Scofield, my PhD supervisors, for their invaluable support, guidance, and knowledge. My sincere thanks also go to several of my lab colleagues and friends, namely Dr. Tamara Lechon Gomez, Dr. Lorenz Fuchs, and Alistair Holland for their technical advice and assistance with various aspects throughout this PhD. I would also like to thank Dr. Patrick Hardinge, Dr. Walter Dewitte and my assessor, Professor Colin Berry, for their support and advice, particularly in the later stages of this research. I further extend my gratitude to all those who were involved in any capacity in this research or the completion of this thesis. Your contributions are deeply appreciated. I am also thankful to my friends and family for their unwavering support and understanding throughout my academic pursuits. Lastly, I'd like to extend my appreciation to the BBSRC and South West Biosciences Doctoral Training Partnership for providing financial support during this research, which of course was vital to the completion of this thesis.

Table of Contents

List of Figures and Tables

List of Abbreviations

Abbreviation	Meaning
UTR	Untranslated region
MAR	Matrix attachment region
TMV	Tobacco Mosaic Virus
pJL-TRBO	Plasmid made by John Lindbo that is Tobacco mosaic virus derived and enables RNA Based Overexpression
pNS	Simple expression cassette with a promoter and terminator in non-replicating Golden Gate expression vector
pNC	Complex expression cassette with a promoter, 5' UTR, 3' UTR, double terminator and MAR in non-replicating Golden Gate expression vector
pRS	Simple expression cassette with a promoter and terminator in the replicating vector pJL-TRBO
pRC	Complex expression cassette with a promoter, 5' UTR, 3' UTR, double terminator and MAR in the replicating vector pJL-TRBO
pRI	Analogous to pJL-TRBO (with a different coding sequence)
pRU	pJL-TRBO vector with both additional 5' and 3' UTRs
pR5	pJL-TRBO vector with only the additional 5' UTR
pR3	pJL-TRBO vector with only the additional 3' UTR
PCP	Plant Cell Pack
FPLC	Fast Protein Liquid Chromatography
HPB	Purification tag with a HA epitope tag, a Prescission® protease recognition site, and the biotinylation site of the *A. thaliana* 3-methylcrotonyl-CoA carboxylase
CBD	Cellulose binding domain purification tag
eGFP	Enhanced green fluorescent protein
hf	Human full length P2X7
hc	Human cysteine rich domain P2X7
hb	Human ballast domain P2X7
rf	Rat full length P2X7
rc	Rat cysteine rich domain P2X7
rb	Rat ballast domain P2X7
DORN1	Does not respond to nucleotides 1 / P2K1 receptor

HMA4	Heavy Metal ATPase 4
rFC	Recombinant Factor C
RANGAP1	Ran GTPase activating protein 1

Table of Constructs

Construct name	Description	Vector	Promoter	5' UTR	CDS	3' UTR	Terminator	MAR
pNS-eGFP	eGFP expressed in a non-replicating vector with minimal DNA regulatory elements	pICH47742	CaMV 35S	N/A	eGFP	N/A	Nopaline Synthase	N/A
pNC-eGFP	eGFP expressed in a non-replicating vector with a complex combination of DNA regulatory elements	pICH47742	CaMV 35S	Synthetic CPMV	eGFP	CPMV	*N. tabacum* extensin and *N. benthamiana* Actin	RB7
pRS-eGFP	eGFP expressed in a replicating vector with minimal DNA regulatory elements	pJL-TRBO	CaMV 35S	N/A	eGFP	N/A	Nopaline Synthase	N/A
pRC-eGFP	eGFP expressed in a replicating vector with a complex combination of DNA regulatory elements	pJL-TRBO	CaMV 35S	Synthetic CPMV	eGFP	CPMV	*N. tabacum* extensin and *N. benthamiana* Actin	RB7
pRI-eGFP	eGFP expressed in a replicating vector with no additional DNA elements	pJL-TRBO	CaMV 35S	N/A	eGFP	N/A	N/A	N/A
pRU-eGFP	eGFP expressed in a replicating vector with a 5' and 3' UTR as additional DNA elements	pJL-TRBO	CaMV 35S	Synthetic CPMV	eGFP	CPMV	N/A	N/A
pR5-eGFP	eGFP expressed in a replicating vector with only a 5' UTR as an additional DNA element	pJL-TRBO	CaMV 35S	Synthetic CPMV	eGFP	N/A	N/A	N/A
pR3-eGFP	eGFP expressed in a replicating vector with only a 3' UTR as an additional DNA element	pJL-TRBO	CaMV 35S	N/A	eGFP	CPMV	N/A	N/A
pNS-hfP2X7	Human full-length P2X7 with a C-terminal eGFP fusion expressed in a non-replicating vector with minimal DNA regulatory elements	pICH47742	CaMV 35S	N/A	Human full-length P2X7-eGFP	N/A	Nopaline Synthase	N/A
pNS-hcP2X7	Human cysteine-rich P2X7 (truncated) with a C-terminal eGFP fusion expressed	pICH47742	CaMV 35S	N/A	Human cysteine-	N/A	Nopaline Synthase	N/A

	in a non-replicating vector with minimal DNA regulatory elements				rich P2X7-eGFP			
pNS-hbP2X7	Human ballast P2X7 (truncated) with a C-terminal eGFP fusion expressed in a non-replicating vector with minimal DNA regulatory elements	pICH47742	CaMV 35S	N/A	Human ballast P2X7-eGFP	N/A	Nopaline Synthase	N/A
pNS-rfP2X7	Rat full-length P2X7 with a C-terminal eGFP fusion expressed in a non-replicating vector with minimal DNA regulatory elements	pICH47742	CaMV 35S	N/A	Rat full-length P2X7-eGFP	N/A	Nopaline Synthase	N/A
pNS-rcP2X7	Rat cysteine-rich P2X7 (truncated) with a C-terminal eGFP fusion expressed in a non-replicating vector with minimal DNA regulatory elements	pICH47742	CaMV 35S	N/A	Rat cysteine-rich P2X7-eGFP	N/A	Nopaline Synthase	N/A
pNS-rbP2X7	Rat ballast P2X7 (truncated) with a C-terminal eGFP fusion expressed in a non-replicating vector with minimal DNA regulatory elements	pICH47742	CaMV 35S	N/A	Rat ballast P2X7-eGFP	N/A	Nopaline Synthase	N/A
pNC-DORN1	DORN1 with a C-terminal eGFP fusion expressed in a non-replicating vector with a complex combination of DNA regulatory elements	pICH47742	CaMV 35S	Synthetic CPMV	DORN1-eGFP	CPMV	*N. tabacum* extensin and *N. benthamiana* Actin	RB7
pNC-HMA4	HMA4 with a C-terminal eGFP fusion expressed in a non-replicating vector with a complex combination of DNA regulatory elements	pICH47742	CaMV 35S	Synthetic CPMV	HMA4-eGFP	CPMV	*N. tabacum* extensin and *N. benthamiana* Actin	RB7
pNC-rFC	Factor C with a C-terminal eGFP fusion expressed in a non-replicating vector with a complex combination of DNA regulatory elements	pICH47742	CaMV 35S	Synthetic CPMV	Factor C-eGFP	CPMV	*N. tabacum* extensin and *N. benthamiana* Actin	RB7
pNC-RANGAP1	RANGAP1 with a C-terminal eGFP fusion expressed in a non-replicating vector	pICH47742	CaMV 35S	Synthetic CPMV	RANGAP1-eGFP	CPMV	*N. tabacum* extensin and *N.*	RB7

	with a complex combination of DNA regulatory elements						*benthamiana* Actin	
pNC-hfP2X7	Human full-length P2X7 with a C-terminal eGFP fusion expressed in a non-replicating vector with a complex combination of DNA regulatory elements	pICH47742	CaMV 35S	Synthetic CPMV	Human full-length P2X7-eGFP	CPMV	*N. tabacum* extensin and *N. benthamiana* Actin	RB7
pNC-hcP2X7	Human cysteine-rich P2X7 (truncated) with a C-terminal eGFP fusion expressed in a non-replicating vector with a complex combination of DNA regulatory elements	pICH47742	CaMV 35S	Synthetic CPMV	Human cysteine-rich P2X7-eGFP	CPMV	*N. tabacum* extensin and *N. benthamiana* Actin	RB7
pNC-hbP2X7	Human ballast P2X7 (truncated) with a C-terminal eGFP fusion expressed in a non-replicating vector with a complex combination of DNA regulatory elements	pICH47742	CaMV 35S	Synthetic CPMV	Human ballast P2X7-eGFP	CPMV	*N. tabacum* extensin and *N. benthamiana* Actin	RB7
pNC-rfP2X7	Rat full-length P2X7 with a C-terminal eGFP fusion expressed in a non-replicating vector with a complex combination of DNA regulatory elements	pICH47742	CaMV 35S	Synthetic CPMV	Rat full-length P2X7-eGFP	CPMV	*N. tabacum* extensin and *N. benthamiana* Actin	RB7
pNC-rcP2X7	Rat cysteine-rich P2X7 (truncated) with a C-terminal eGFP fusion expressed in a non-replicating vector with a complex combination of DNA regulatory elements	pICH47742	CaMV 35S	Synthetic CPMV	Rat cysteine-rich P2X7-eGFP	CPMV	*N. tabacum* extensin and *N. benthamiana* Actin	RB7
pNC-rbP2X7	Rat ballast P2X7 (truncated) with a C-terminal eGFP fusion expressed in a non-replicating vector with a complex combination of DNA regulatory elements	pICH47742	CaMV 35S	Synthetic CPMV	Rat ballast P2X7-eGFP	CPMV	*N. tabacum* extensin and *N. benthamiana* Actin	RB7

pRC-DORN1	DORN1 with a C-terminal eGFP fusion expressed in a replicating vector with a complex combination of DNA regulatory elements	pJL-TRBO	CaMV 35S	Synthetic CPMV	DORN1-eGFP	CPMV	*N. tabacum* extensin and *N. benthamiana* Actin	RB7
pRC-HMA4	HMA4 with a C-terminal eGFP fusion expressed in a replicating vector with a complex combination of DNA regulatory elements	pJL-TRBO	CaMV 35S	Synthetic CPMV	HMA4-eGFP	CPMV	*N. tabacum* extensin and *N. benthamiana* Actin	RB7
pRC-rFC	Factor C with a C-terminal eGFP fusion expressed in a replicating vector with a complex combination of DNA regulatory elements	pJL-TRBO	CaMV 35S	Synthetic CPMV	Factor C-eGFP	CPMV	*N. tabacum* extensin and *N. benthamiana* Actin	RB7
pRC-RANGAP1	RANGAP1 with a C-terminal eGFP fusion expressed in a replicating vector with a complex combination of DNA regulatory elements	pJL-TRBO	CaMV 35S	Synthetic CPMV	RANGAP1-eGFP	CPMV	*N. tabacum* extensin and *N. benthamiana* Actin	RB7
pRC-hfP2X7	Human full-length P2X7 with a C-terminal eGFP fusion expressed in a replicating vector with a complex combination of DNA regulatory elements	pJL-TRBO	CaMV 35S	Synthetic CPMV	Human full-length P2X7-eGFP	CPMV	*N. tabacum* extensin and *N. benthamiana* Actin	RB7
pRC-hcP2X7	Human cysteine-rich P2X7 (truncated) with a C-terminal eGFP fusion expressed in a replicating vector with a complex combination of DNA regulatory elements	pJL-TRBO	CaMV 35S	Synthetic CPMV	Human cysteine-rich P2X7-eGFP	CPMV	*N. tabacum* extensin and *N. benthamiana* Actin	RB7
pRC-hbP2X7	Human ballast P2X7 (truncated) with a C-terminal eGFP fusion expressed in a replicating vector with a complex combination of DNA regulatory elements	pJL-TRBO	CaMV 35S	Synthetic CPMV	Human ballast P2X7-eGFP	CPMV	*N. tabacum* extensin and *N. benthamiana* Actin	RB7
pR5-DORN1	DORN1 with a C-terminal eGFP fusion expressed in a replicating vector with	pJL-TRBO	CaMV 35S	Synthetic CPMV	DORN1-eGFP	N/A	N/A	N/A

	only a 5’ UTR as an additional DNA element							
pR5-HMA4	HMA4 with a C-terminal eGFP fusion expressed in a replicating vector with only a 5’ UTR as an additional DNA element	pJL-TRBO	CaMV 35S	Synthetic CPMV	HMA4-eGFP	N/A	N/A	N/A
pR5-rFC	Factor C with a C-terminal eGFP fusion expressed in a replicating vector with only a 5’ UTR as an additional DNA element	pJL-TRBO	CaMV 35S	Synthetic CPMV	Factor C-eGFP	N/A	N/A	N/A
pR5-RANGAP1	RANGAP1 with a C-terminal eGFP fusion expressed in a replicating vector with only a 5’ UTR as an additional DNA element	pJL-TRBO	CaMV 35S	Synthetic CPMV	RANGAP1-eGFP	N/A	N/A	N/A
pR5-hfP2X7	Human full-length P2X7 with a C-terminal eGFP fusion expressed in a replicating vector with only a 5’ UTR as an additional DNA element	pJL-TRBO	CaMV 35S	Synthetic CPMV	Human full-length P2X7-eGFP	N/A	N/A	N/A
pR5-hcP2X7	Human cysteine-rich P2X7 (truncated) with a C-terminal eGFP fusion expressed in a replicating vector with only a 5’ UTR as an additional DNA element	pJL-TRBO	CaMV 35S	Synthetic CPMV	Human cysteine-rich P2X7-eGFP	N/A	N/A	N/A
pR5-hbP2X7	Human ballast P2X7 (truncated) with a C-terminal eGFP fusion expressed in a replicating vector with only a 5’ UTR as an additional DNA element	pJL-TRBO	CaMV 35S	Synthetic CPMV	Human ballast P2X7-eGFP	N/A	N/A	N/A
pR5-rfP2X7	Rat full-length P2X7 with a C-terminal eGFP fusion expressed in a replicating vector with only a 5’ UTR as an additional DNA element	pJL-TRBO	CaMV 35S	Synthetic CPMV	Rat full-length P2X7-eGFP	N/A	N/A	N/A
pR5-rcP2X7	Rat cysteine-rich P2X7 (truncated) with a C-terminal eGFP fusion expressed in a	pJL-TRBO	CaMV 35S	Synthetic CPMV	Rat cysteine-	N/A	N/A	N/A

	replicating vector with only a 5’ UTR as an additional DNA element				rich P2X7-eGFP			
pR5-rbP2X7	Rat ballast P2X7 (truncated) with a C-terminal eGFP fusion expressed in a replicating vector with only a 5’ UTR as an additional DNA element	pJL-TRBO	CaMV 35S	Synthetic CPMV	Rat ballast P2X7-eGFP	N/A	N/A	N/A

Chapter 1 – Introduction

1.1 – An overview of recombinant protein production

1.1.1 History and relevance of recombinant proteins

Recombinant proteins are proteins produced in heterologous expression systems, different to the native host organism usually with the aim of increasing yields or reducing production costs. The technology is closely linked with the invention of genetic engineering in the mid-late 20th century following the deciphering of the genetic codon (Nirenberg et al. 1965), identification of the first gene (Shapiro et al. 1969), the creation of recombinant DNA and the first recombinant organism (Jackson et al. 1972; Cohen et al. 1973), and finally the isolation of DNA ligase and restriction enzymes (Lehman 1974). These sequential technological developments were key for the field of recombinant protein production.

Recombinant proteins are important in many fields with medical, industrial, and agricultural applications. The first recombinant protein to be marketed was recombinant insulin (Quianzon and Cheikh 2012), but today they are particularly prevalent in the pharmaceutical industry for the production of vaccines, antibodies, and diagnostic tools. In fact, the biopharmaceutical sector has been the largest sector of the pharmaceutical market (Owczarek et al. 2019), with protein drugs representing 10% of the drug market in 2017 (Usmani et al. 2017). In an industrial setting, common uses of recombinant enzymes include in laundry detergents and biofuel production (Rabert et al. 2013; Puetz and Wurm 2019), and agricultural biotechnology expresses recombinant proteins in genetically modified organisms to confer beneficial traits such as enhanced nutritional value and resistance to abiotic and biotic stresses (Aziz et al. 2022) through the expression of pathogenesis-related proteins (Santos and Franco 2023), insecticidal and herbicidal proteins (Gupta et al. 2021; Dong et al. 2021), nutritional proteins (Le et al. 2016), stress resistance proteins (Parmar et al. 2017), and proteins that improve shelf life (Aziz et al. 2022). Importantly, these proteins are typically expressed within the plant to enhance host properties, where the *plant* is the valuable product, not the *protein* itself. This creates an important distinction of recombinant proteins that are expressed *within* a product, that are never extracted from the plant, compared to *as* a product when they are extracted from the plant. This research primarily focusses on the latter, where recombinant proteins are made as the product with protein extraction and purification from plant tissue being key. Recombinant proteins are produced using ‘expression systems’, a term which used to describe all the components of a recombinant protein production system, including the host organism, transformation method, constructs used, and protein extraction and purification methods.

1.1.2 Conventional expression systems and associated challenges

Conventional expression systems employed for recombinant protein production mostly include prokaryotic hosts like *E. coli*, and eukaryotic cell cultures using yeast, insect, and mammalian cells (Gecchele et al. 2015). Typical challenges include optimising protein yields, quality and function, whilst ensuring proper protein solubility, folding, stability and preservation of biological activity. However, each expression system has its own unique advantages and disadvantages. Prokaryotic expression systems are cheapest, scalable, and produce high yields of proteins with simple engineering methods and rapid growth times, but cannot produce complex post-translational modifications, tend to form inclusion bodies which hinders downstream processing (DSP) and can produce endotoxins (Mamat et al. 2015). Yeast expression hosts, such as *Saccharomyces cerevisiae*, combine the benefits of unicellular organisms with the ability to perform some eukaryotic post-translational modifications, making them suitable for producing complex proteins that prokaryotic systems cannot (Gomes et al. 2018). Insect cell cultures can produce higher complexity proteins, often with high quality, but are much more costly, and are expensive to scale (Gecchele et al. 2015). Mammalian expression systems, often utilising Chinese Hamster Ovary (CHO) or Human Embryonic Kidney (HEK) cells, can produce all complex post-translational modifications and are widely utilized in the production of therapeutic proteins. However, these are also very costly to scale and are sensitive to environmental changes so require tightly controlled cultivation conditions with specialist equipment (O'Flaherty et al. 2020). Furthermore, differences in the characteristics of target proteins result in some proteins being more challenging to produce than others. Factors that influence production typically include size, post-translational modifications, and intracellular localisation, with larger, modified, or membranous proteins being more difficult to produce than smaller, unmodified, or cytosolic proteins. As a result of these factors, there is not a 'one size fits all' expression system for all recombinant proteins with each expression system occupying a different niche for different proteins. Currently there is no optimal expression system for modified membranous proteins, as these require complex post-translational modifications that only insect or mammalian cell cultures offer, whilst being produced at low yields requiring upstream scalability that only prokaryotic expression systems offer. Fortunately, there is potential for this problem to be addressed using plant expression systems.

1.2 – Recombinant protein production in plants

Plants represent promising recombinant protein expression systems as they are able to produce complex post-translational modifications with very high levels of upstream scalability, requiring little to no specialist equipment and at low costs. Plants have been used to make recombinant proteins as early as 1989, where they were used to express antibodies (Hiatt et al. 1989), and have had

increasing success in recent years, particularly with vaccines (Malaquias et al. 2021; Zhang et al. 2014).

Plants offer several advantages including low upstream costs (Avesani et al. 2013; Chen and Davis 2016), high capacity for scalability (Burnett and Burnett 2019), potential for edible vaccines (Merlin et al. 2014; Kurup and Thomas 2019), lack of endotoxin production (Fischer and Emans 2000; Merlin et al. 2014), rapid protein production times that make them particularly suitable for vaccine production (Zhang et al. 2014; Chen and Davis 2016; Rattanapisit et al. 2020; Makatsa et al. 2021), and the ability to produce complex post-translational modifications (Gomord and Faye 2004; Millar et al. 2019). In addition, a wide range of plant species have been used for recombinant protein production, including tobacco species (Ma et al. 2003; Chen and Lai 2015; Burnett and Burnett 2019; Moon et al. 2019), legumes (Burnett and Burnett 2019), fruit and vegetable crops (Merlin et al. 2014; Kurup and Thomas 2019; Moon et al. 2019; Shanmugaraj et al. 2021), cereals (Mirzaee et al. 2022), aquatic plants (Khvatkov et al. 2018; Chanroj et al. 2021), algae (Ahmad et al. 2020), and mosses (Baur et al. 2005; Gitzinger et al. 2009; Niederkrüger et al. 2014; Reski et al. 2018). These can utilise whole plants or cell-cultures, dependent on the application.

The success of plant expression systems has largely been hindered by several limitations, including the presence of a carbohydrate rich cell wall, which must be processed to extract proteins. Other reported limitations include low yields, particularly in plant cell cultures (Schillberg and Finnern 2021), lengthy transformation times, non-human glycosylation of proteins (Castilho and Steinkellner 2012), and poorly optimised methods for producing recombinant proteins in many plant species (Khvatkov et al. 2018; Chanroj et al. 2021). Fortunately, recent efforts have been made to address these challenges, including the maximising of yields through engineering efforts, minimising of production times through continually improving transient transformation methods (Norkunas et al. 2018), humanising protein glycosylation (Castilho and Steinkellner 2012), optimising and automating downstream processing (Kumar et al. 2021), and the use of transient transformation to enhance the speed of protein production. Consequently, there are many factors to consider when producing recombinant proteins in plants.

The following subsections are the scope of this thesis and will explore these factors splitting them into upstream processing, including selection of host species, transformation methods, and construct optimisation, and downstream processing, including protein extraction and purification. This thesis has three core data chapters. Chapter 3 explores the optimisation of upstream processing, which refers to all factors affecting initial production and cultivation of host plants expressing target proteins. This includes the development and optimization of expression vectors,

transformation methods, and the cultivation and scaling up of plant expression hosts. Chapter 4 explores the optimisation of downstream processing, which encapsulates the steps following target protein production in the plant host. This includes extraction and purification of target proteins, exploring a variety of methods aimed to achieve maximum yields and purities of target proteins. The final data chapter, Chapter 5, explores the impact of protein complexity on yields in plants using quantitative comparative methods developed in Chapters 3 and 4. In this research, the term protein 'complexity' refers to factors such as size, post-translational modifications, and intracellular localisation. Chapter 5 also explores the effects of different plant chassis for expressing target proteins, and develops assays to test the function of P2X7, a major target protein in this research. The major pipeline and chapters of this research is shown diagrammatically in Figure 1.1.

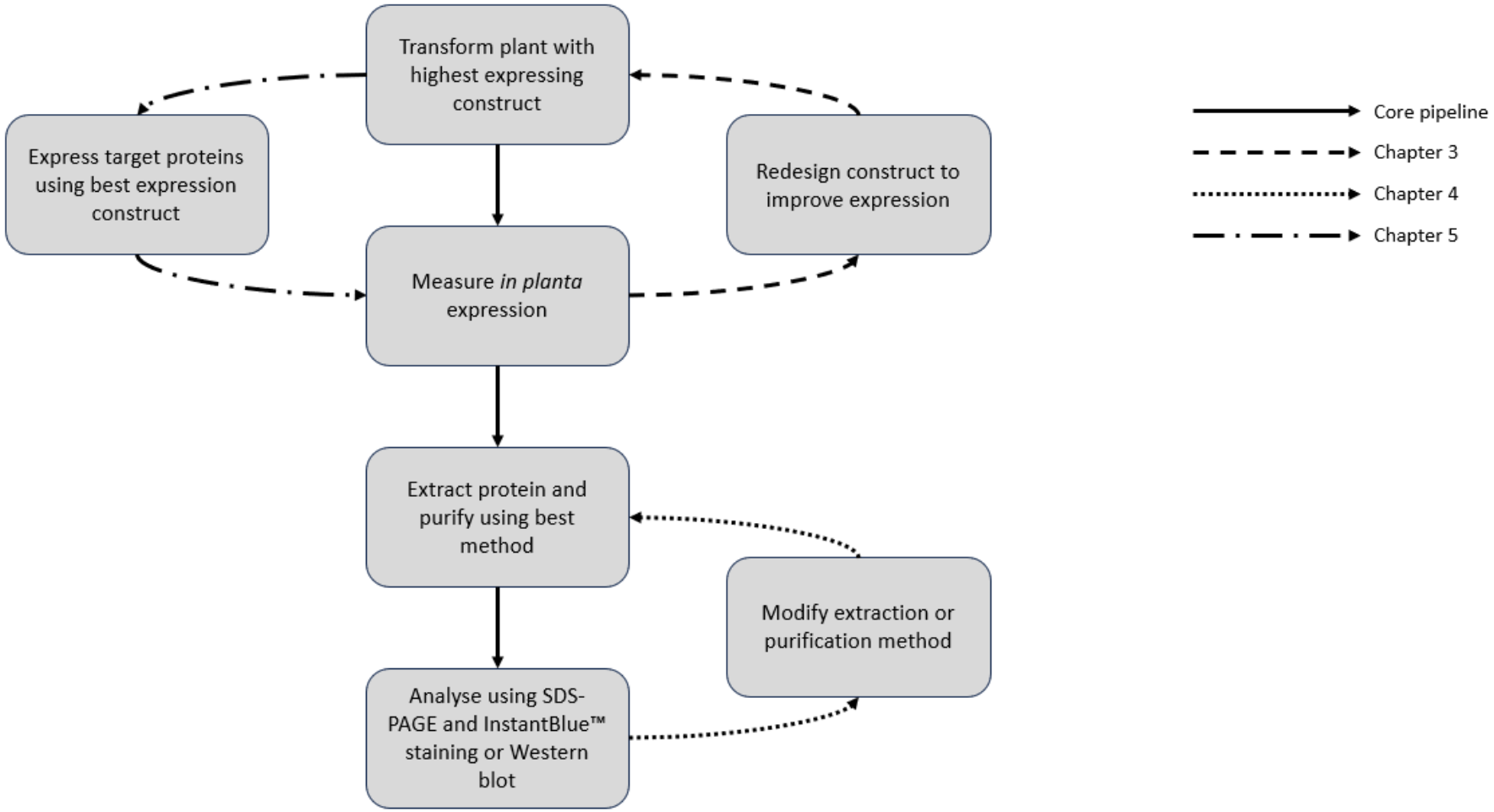


Figure 1.1: The major pipeline in this research. Solid arrows show the core pipeline used. Dashed lines show the cycle used in Chapter 3 to develop improved expression constructs, dotted lines show the cycle used in Chapter 4 to develop suitable extraction and purification methods, and dot-dash lines show the focus of Chapter 5 to test the existing constructs and methods using different target proteins.

1.3 – Upstream processing of plant recombinant proteins

Upstream processing refers to all factors affecting initial production and cultivation of host plants expressing target proteins. This includes factors such as the selection of a suitable plant host, the type of transgenesis, the design and utilisation of genetic expression constructs, and the co-expression of other proteins that may enable improved target protein yields. Each of these are discussed in this section and Table 1.1 summarises the information in the following sub-sections, with factors utilised in this research shown in bold, alongside the justification of their use.

Table 1.1: Upstream processing factors investigated in this research. Examples of factors are shown, with specific examples that are used in this research shown in **bold**, alongside the justification for using them.

Category	Examples (factors directly used in this research are shown in bold)	Justification
Plant host species	***Nicotiana* spp.** ***A. thaliana*** Legumes Fruit and vegetables Cereals Mosses Aquatic plants and algae	Chosen species are model organisms with well-established research tools and transformation methods (Burnett and Burnett 2019; Chen and Lai 2015).
Type of transgenesis	**Transient** **Stable (nuclear)** Stable (chloroplastic)	Not all target proteins of this research can be produced using chloroplast transformation but can using transient and stable transformations (Lehtimäki et al. 2015).
Genetic elements	**Promoters** **Terminators** **UTRs** **Matrix attachment regions** Enhancers Introns	Chosen DNA elements can be used in every construct, assembled using modular cloning and have been extensively researched. The specific elements used have been characterised to show high levels of gene expression (Ali and Kim 2019; Peyret et al. 2019; Diamos and Mason 2018). Intron use depends on CDS.
Expression vectors	**Non-replicating** **Replicating** (**Tobamoviruses**, Potexviruses, Comoviruses or Geminiviruses)	Tobamoviruses are well researched in *Nicotiana* spp. and show high expression levels (Hefferon 2017). Comparing non-replicating and replicating vectors is useful.
Co-expressed proteins	**Silencing suppressors** (**p19**, p24, HC-Pro, p22, P0, P17 and NSs) **Cell-cycle proteins** **Stress response proteins** Protease inhibitors	Many different silencing suppressors available, p19 is best characterised and widely used, shown to be highly effective (Peyret et al. 2019). Cell-cycle and stress response proteins found to be successful in similar research (Norkunas et al. 2018).

1.3.1 Plant species

The most popular plants used to produce recombinant proteins, and the plants mainly utilised in this research, are *Nicotiana* plants, usually *N. benthamiana* on small scales, and *N. tabacum* on larger scales. These plants are useful as they produce high yields with large leaves, are scalable due to their rapid growth cycles and established infiltration methods, and produce large numbers of seeds (Burnett and Burnett 2019). Moreover, transient expression methods are optimised for *Nicotiana*, including agroinfiltration on small scales, and vacuum infiltration on larger scales, which can reduce transgenesis time by 30-fold (Chen and Lai 2015). Several replicating vectors, explored later, are available for use in *Nicotiana* species and have had widespread success at driving high levels of protein expression (Yamamoto et al. 2018; Mardanova et al. 2017; Marillonnet et al. 2005; Huang et al. 2009). A final advantage is that recombinant proteins produced in tobacco plants are usually harvested before flowering occurs, reducing the risk of transgene escape, despite them being cross-pollinators (Moon et al. 2019). Disadvantages of *Nicotiana* species include the high level of phenolic and alkaloid compounds, which makes downstream purification more difficult, and the instability of recombinant proteins in planta, meaning other storage methods must be used (Burnett and Burnett 2019). Furthermore, although tobacco BY-2 cells are the most commonly used plant cell culture, plant cells are larger than mammalian cells, producing lower yields of target protein per litre of culture, and lack any significant advantages over other eukaryotic cell cultures (Schillberg et al. 2019). Despite these drawbacks, engineering efforts have been made to improve BY-2 cell culture expression systems, including cell lines with reduced protease activity or humanised glycosylation (Karki et al. 2021). Finally, researchers have developed an expression system called plant cell packs, derived from BY-2 cells, that enable the rapid multiplex screening of constructs in 96-well plates (Rademacher et al. 2019; Gengenbach et al. 2020), which are explored in this research. Thus, for rapid transient expression of recombinant proteins which will be promptly extracted and have a high demand and consequent turnover rate, *Nicotiana* plants are an ideal host.

The other expression species used in this research is *Arabidopsis thaliana*. These plants can be used to express recombinant proteins, though are less widely used. Transient transformation methods are not well established in *A. thaliana*, so stable transformants are usually made. However, this method results in long growth times and is hindered by low leaf biomass resulting in many plants being necessary to obtain high yields of protein. Jeong et al. (2018) has successfully produced and extracted a multi-subunit plant membrane protein in stably transformed *A. thaliana*, and mCherry has been expressed in *A. thaliana* at concentrations up to 400 μg/g (Von Schaewen et al. 2018), though this host rarely confers major benefits over others in this section due to less leaf biomass than *Nicotiana* plants, lower protein contents than cereal seeds and leguminous plants, a lack of

edible organs, and slower growth rates than aquatic plants. Many of these plants were not used in this research but their properties, advantages and disadvantages are summarised below.

While legumes have lower leaf biomass than tobacco, they have a high protein content in their seeds, and have cheaper cultivation costs due to lower nitrogen requirements (Burnett and Burnett 2019). One study used the legume *Mucuna bracteate* to transiently express recombinant proteins and showed expression levels 2-fold higher than *N. benthamiana* (Abd-Aziz et al. 2020). Simple syringe agroinfiltration does not work for this species as the leaves are too brittle, but vacuum infiltration, suitable for large scale expression, works. Unfortunately, the researchers found that the recombinant protein, in this case IgG, had more structural heterogeneity, lacking glycan groups than are produced in *N. benthamiana*, which may suggest that this legume is less consistent. Recent efforts have been made to improve recombinant protein expression in legumes, where peas, lentils, and fava beans were successfully engineered to produce GFP localised to the cytoplasm or the apoplast (Debler, Henares and Lee 2021). However, the authors noted that the GFP expression in the cytoplasm was around 8 times lower than in *N. benthamiana*, and 111 times lower in the apoplast. Nonetheless, legumes may be an alternate option for recombinant protein production, with the potential for higher expression than *Nicotiana* plants in some cases.

Several fruit and vegetable crops have been used to produce recombinant proteins, including lettuce and Solanaceae species. These usually have low protein contents relative to seeds, but recombinant proteins have been expressed at moderate levels, with viral particles expressed at up to 8% of total soluble protein (TSP) in tomato fruits, and 0.4% TSP in potato tubers (Merlin et al. 2014). Note however, that this corresponds to 160 µg/g and 120 µg/g respectively, demonstrating the higher total protein content in potato tubers (Merlin et al. 2014). The main advantage of recombinant protein expression in fruits and vegetables is that they are edible and can be stored or freeze-dried. When these tissues express immunogenic proteins, they elicit an immune response, largely through mucosal immunity, without the need for purification, negating major costs (Kurup and Thomas 2019) though variable yields between plants can cause problems with dosage control (Moon et al. 2019). Potatoes, tomatoes, carrots, bananas, and lettuce, among other non-fruit and vegetable crops, have been successfully engineered and employed as edible vaccines for a range of diseases (Kurup and Thomas 2019). Importantly, ingested plant tissue is exposed to a range of gastrointestinal environments with different pH, protease activity, and absorption capabilities which must each be considered for effective applications, although the presence of a plant cell wall has been shown to provide natural bio-encapsulation and protection of antigens throughout the gastrointestinal tract (Kurup and Thomas 2019). Furthermore, these expression hosts benefit from a wide variety of known tissue-specific promoters, allowing expression specifically in the fruits, tubers, or roots (Lim

et al. 2012; Ding et al. 2022). Thus, if dosages can be made consistent, expressing recombinant proteins in fruits and vegetables can be an ideal option for therapeutic proteins for consumption.

Similarly, cereal crops have been made to express recombinant proteins, primarily in their seeds due to their high protein concentrations and abundance. Cereal seeds have intracellular storage vesicles where proteases are downregulated resulting in improved recombinant protein stability and storage long term (Burnett and Burnett 2019). The other dominant plant cell culture utilises rice, reviewed for pharmaceutical protein production (Kuo et al. 2013). One study reports high recombinant lysozyme yields in rice, expressed at 45% of TSP, or 0.6% of dry weight (Huang et al. 2002). However, whether these yields are competitive compared to other eukaryotic cell-cultures is debated (Schillberg et al. 2019).

Mosses such as *Physcomitrella patens* have shown success for recombinant protein production (Baur et al. 2005; Gitzinger et al. 2009; Niederkrüger et al. 2014; Reski et al. 2018), particularly for glycosylated biopharmaceuticals (Huether et al. 2005; Decker & Reski 2007; Reski et al. 2015), where proteins can be secreted into the culture medium which dramatically reduces downstream purification costs (Schaaf et al. 2005). In fact, α-galactosidase A produced in moss has been shown to be more effective in treating enzymatic deficiencies in mice than the more conventional agalsidase alfa which is produced from mammalian cells with a mannose-6-phosphate modification that the moss version lacks (Shen et al. 2015). This demonstrates the potential superiority of the system.

Lastly, aquatic plants can also be used to produce recombinant proteins. Many of these species have extremely rapid growth rates and high protein concentrations. In fact, duckweed can have a doubling rate as low as 24 hours and its dry weight can be up to 45% protein (Khvatkov et al. 2018). Additionally, these almost only propagate vegetatively, reducing the risk of transgene escape and introgression should these be grown outside of contained facilities. Khvatkov et al., (2018) expressed recombinant human granulocyte colony-stimulating factor in *Wolffia arrhiza* at just under 0.2% of TSP. Whilst this is a low yield, the extremely rapid growth time of this species can result in high yields, provided the large amounts of biomass can be processed. The other more widely used aquatic plant is *Lemna minor*, which has also been used for the expression of several different recombinant proteins (Khvatkov et al. 2018). Lastly, algal species have been successfully transformed to express recombinant proteins at low yields, though replicating vectors have been successfully employed in these hosts, which could lead to improved yields (Ahmad et al. 2020; Malla et al. 2021). Thus, aquatic plants are promising expression hosts due to their rapid biomass accumulation, though increases in target protein yield may be necessary to make them a competitive plant expression system.

1.3.2 Type of transgenesis

Directly relevant to the host species is the type of transgenesis, as some species can only be transformed in certain ways. This research utilises both stable nuclear and transient transformation, with a focus on the latter, though it should be stated that other options for transgenesis exist which are also discussed below.

Nuclear transformation is the most common stable transgenesis technique. Using stably transformed cereals, recombinant proteins have been produced in moderate yields (Liu and Timko 2022). Nuclear expression has some advantages and disadvantages. The main advantages are that transgenesis is stable meaning plants only need to be propagated for further generations rather than retransformed for each batch, and that tissue-specific gene expression is available, allowing tight control of protein production. Disadvantages include the lengthy process of creating stable lines, involving the generation of homozygous transformants and screening plant transformants for those with the highest expression, which often produce lower yields than transient transformants (Maclean et al. 2007). Additionally, nuclear transformants experience transcriptional gene silencing through DNA methylation and transgene silencing (Gallego-Bartolomé 2020), in addition to post-transcriptional gene silencing. Unless targeted insertions are performed (Buyel et al. 2021), transformation methods integrate the transgene into the genome randomly, so position effects are an issue with variable expression depending on where the transgene integrates (Burnett and Burnett 2019). Researchers investigated position effects and found that expression levels varied by up to 10-fold, demonstrating the importance of genome position (Day et al. 2000). Fortunately, there are additional DNA elements that can be incorporated into the transgene, notably including scaffold/matrix attachment regions (S/MARS) which affect the surrounding chromatin structure and increase transcriptional efficacy (Diamos and Mason 2018). MARs have been used in stable rice transformants to improve expression by 3-fold (Egelkrout et al. 2012). Stably transformed *A. thaliana* and *N. benthamiana* plants are investigated in this research, though to a lesser extent than transient transformants.

Alternatively, chloroplastic transgene expression can also be stable and presents some unique advantages. These organelles are abundant in photosynthetic cells of plants, and consequently, chloroplast transgenesis can result in protein expression levels of up to 70% of TSP within leaves (Oey et al. 2009; Ruhlman et al. 2010). Other reported advantages of chloroplastic expression include the ability to express multiple protein coding sequences as polycistronic units, and transgene stability over several generations, without the spread of the transgene in pollen as chloroplasts are almost always maternally inherited (Egelkrout et al. 2012; Moon et al. 2019). The major downside of this transformation method is the inability to perform some protein post-translational modifications

(Lehtimäki et al. 2015), removing one of the advantages that plant expression systems have over prokaryotic systems.

Transient protein expression is the most used by industry for several reasons. Firstly, this transformation method is faster than any others, with transgenic plants being made, and target proteins extracted within weeks or less (Chen and Davis 2016). Secondly, the methods are very well developed, with techniques for large scale transformation being 30-fold faster than bench scale transformation (Chen and Lai 2015), and viral vectors being continually developed for maximal protein expression, used in conjunction with silencing inhibitors. Furthermore, proteins with complex PTMs can still be produced and still with excellent yields, and there is a vast array of DNA sequences that have been well characterised in transient transformants. The main drawbacks are the lack of optimised transient transformation techniques for non-*Nicotiana* plants, and that transgenesis always needs to be performed on every generation of plants. Fortunately, several different methods can be used to transiently transform plants including small scale syringe infiltration, to large scale vacuum infiltration with the appropriate equipment (Chuang et al. 2022). Several improvements in transformation efficiency in agroinfiltrated *N. benthamiana* plants have been reported in a study by Norkunas et al., (2018) who used a modified infiltration buffer, a heat shock of transformed plants, and co-expression of several different proteins to increase recombinant protein yields. By combining different improvements, these researchers reported a 3.5-fold increase in GUS expression when compared to the already efficacious pEAQ-HT system. As *N. benthamiana* is the primary plant target used in this research, transient transformation using agroinfiltration is the dominant method of transgenesis utilised in this research.

1.3.3 Genetic elements

Expression construct design plays an important role in influencing recombinant protein yields, influencing several factors including transcription, translation, and degradation of products. The engineering of expression constructs has paramount importance in this research and utilises various classes of DNA elements to exert several effects. A diagram of genetic elements discussed in this section can be seen in Figure 1.2

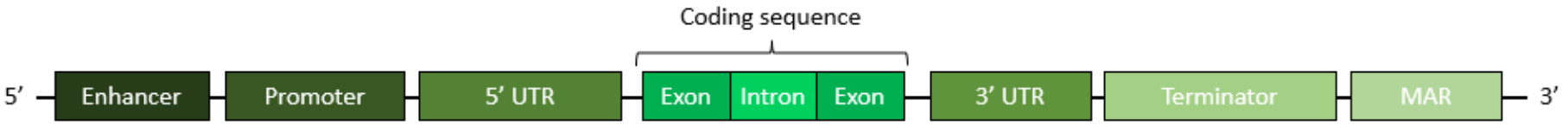


Figure 1.2: Visual schematic of DNA elements discussed. This research utilised every DNA element shown except enhancers and introns.

1.3.3.1 Influencing transcription

Transcription can be affected using promoters, terminators, enhancers, introns, and matrix attachment regions. Generally, minimal expression constructs utilise a promoter to initiate transcription and a terminator to end it. Promoter sequences exist on a spectrum of strengths, with some promoters driving high levels of transcription and others driving lower levels, influencing the number of mRNA molecules produced. They also can be tissue specific or inducible, enabling spatial and temporal control over gene expression (Beringer et al. 2017), but these have not been investigated in this research. Generally, high recombinant protein yields are desirable so typically constitutive, global, often virally derived promoters such as the 35S Cauliflower Mosaic Virus (CaMV) promoter are used. This is arguably the best studied plant promoter and is commonly used in expression constructs, including in this research. However, other promoters are worthy targets including synthetic promoters, reviewed by Ali and Kim (2019), which can drive higher levels of transcription in some plants and tissues than CaMV, with one reporting up to 25-fold higher expression in transgenic tobacco with over 57-fold higher in certain tissue types (Kumar et al. 2011). Terminators that are commonly used include the CaMV 35S, nopaline synthase (NOS), octopine synthase (OCS) and Heat Shock Protein (HSP) terminators (Nagaya et al. 2010; Limkul et al. 2015; Diamos and Mason 2018). Diamos and Mason (2018) characterised the use of terminators in tandem in tobacco and lettuce plants, demonstrating improved gene expression when double terminators were stacked together, and showed that different pairs of terminators worked with varying efficacies depending on the order of the DNA elements and the host species in which they are employed. In fact, these researchers paired double terminators with matrix attachment regions (MARs) and showed that GFP production could be enhanced by over 60-fold using this arrangement, compared to constructs with a single NOS terminator and no MAR. MARs, whilst not essential to a functioning transcriptional unit, influence transcription by binding the nuclear matrix to create an active chromatin hub (Allen 2008; Diamos and Mason 2018). One study showed a 3-fold increase in gene expression when using an MAR in stably transformed rice (Vain et al. 1999). Importantly, Diamos and Mason (2018) showed that MARs can improve protein expression in transiently transformed plants. Consequently, the work by Diamos and Mason is a cornerstone of the research presented here as the terminators and MAR employed are based on it.

Lastly, enhancers and introns can also improve expression but were not utilised in this research. Like promoters, enhancers recruit trans-acting proteins to create an active-chromatin hub that promotes transcription (Tippens et al. 2018). These can influence tissue-specific gene expression and act more distally to the transcriptional start site than promoters. Similarly, introns are non-coding DNA

elements within the coding sequence, but also have been shown to improve gene expression by a process known as intron-mediated enhancement, reportedly increasing mRNA accumulation by up to 10-fold (Bourdon et al. 2001; Rose 2008). However, several different coding sequences are utilised in this research and as intron positioning is unique for each coding sequence they were not investigated here.

1.3.3.2 Influencing translation

Translation can be affected by incorporating untranslated regions into constructs, which can influence translation efficacy and mRNA stability. The most powerful of these sequences are often virally derived. 5' UTRs typically improve translation efficiency by affecting ribosome binding, with some UTRs increasing expression more than others. For example, the 5' UTR from alfalfa mosaic virus has been shown to increase target protein expression by 4-fold (Mardanova et al. 2017), and the commercially viable HyperTrans system utilised a modified 5' UTR from cowpea mosaic virus (CPMV), with upstream start codons removed to reduce translation of non-target products, enabling protein accumulation up to 20% of TSP (Sainsbury and Lomonossoff 2008; Sainsbury et al. 2009). Further engineering efforts have been performed on this UTR to generate synthetic variants that improve this already high expression by up to 2-fold (Peyret et al. 2019). This synthetic 5' UTR was chosen for incorporation into expression constructs in this research. The same publication tested synthetic variants of the CPMV 3' UTR but failed to create a variant that improves expression. The native CPMV 3' UTR contains highly structured regions that influence expression by improving RNA stability (Meshcheriakova et al. 2014) and was used in this research.

Additionally, codon usage affects the efficiency of translation. Some codons are more commonly used in host plants than others, resulting in tRNA populations for rarely used codons being less abundant in a cell than tRNA for more commonly used codons (Webster et al. 2016; Buyel et al. 2021; Plotkin and Kudla 2010). This means that coding sequences containing rare codons may incur a rate limiting step on protein production that can be avoided by utilising more commonly used codon. In fact, one study expressed a Bt toxin, Cry1A, and found that codon optimisation improved expression by 100-fold (Perlak et al. 1991). Similarly, one study expressed miraculin in plants and found that codon optimisation improved expression by 1.5-fold (Hiwasa-Tanase et al. 2010), and another study expressing the yeast enzyme phytase found that expression in rice could be improved 120-fold through codon optimisation and the addition of a signal sequence (Hamada et al. 2004) though the relative effects of each are unknown. Codon optimisation is an important element in construct design and has been carefully considered in this research. Where possible, coding sequences were designed and codon optimised, with few or no rare codons used. However, some of

the coding sequences utilised here were derived from pre-existing vectors for mammalian cell expression and are not codon optimised.

1.3.4 Expression vectors

Directly relevant to the two previous subsections are the use of different expression vectors. Importantly, all vectors replicate within their host organisms as controlled by origins of replication, however, some vectors replicate target RNA molecules using viral encoded replicase proteins. In this research, non-replicating vectors are defined as vectors that do not replicate target RNA molecules, and replicating vectors are defined as those that do. Non-replicating vectors, such as those used in the Golden Gate expression system, are typically used for cloning of transgenes and transformation into the host but provide no additional influence on gene expression following transformation. In contrast, replicating vectors enable a further increase in protein expression through the replication of target RNAs where replicase and movement proteins bind virally derived UTRs to exert their effects. There are many replicating vectors available for use in plants which have been recently reviewed (Mahmood et al. 2023; Abrahamian et al. 2020; Hefferon 2017). Naturally, replicating vectors are derived from viruses and contain the whole viral genome. These are typically engineered to express target protein alongside the viral coat protein, either as a fusion or using an additional sub-genomic promoter. The major downsides of using full viral vectors are biocontainment issues, as the virus contains all necessary components for infection, limitations on transgene size (Hefferon 2017), and the potential need to cleave viral coat proteins from target recombinant proteins if expressed as a chimaeric fusion (Lico et al. 2008) through the use of proteases, inteins, or other cleavage mechanisms. Instead, deconstructed vectors can be used which do not contain the whole viral genome, lacking at least one element, inhibiting viral function. Often, the coat protein is removed to prevent systemic spread of the virus, meaning an additional method is needed to insert the viral vector into plant cells. However, deconstructed viral vectors do not have the size limits and biocontainment issues of full viral vectors (Hefferon 2017). Several classes of deconstructed replicating vectors are widely used in plant biotechnology and include tobamoviruses, potexviruses, comoviruses, and geminiviruses.

Tobamoviruses are rod shaped single stranded RNA viruses which include tobacco mosaic virus (TMV). Engineered vectors derived from TMV include the magnICON (Marillonnet et al. 2004; Marillonnet et al. 2005) and TMV RNA-based overexpression (TRBO; Lindbo 2007b) expression systems. TMV infects a broad range of plants and is well-studied, particularly in *Nicotiana* plants, where it has been used to express several recombinant proteins (Hefferon 2017). Furthermore, the TRBO system has been shown to drive high levels of protein expression of 3 to 5 mg/g fresh weight and lacks the TMV coat protein so there is no risk of transgene escape with this system (Lindbo

2007b). As a result of these factors, it was chosen as the replicating vector for use in this research, however other viral vectors exist and are viable alternatives for further research.

Briefly, potexviruses are flexuous, filamentous single stranded RNA viruses and include potato virus X (PVX). PVX has also been used to express several proteins, with target protein yields of up to 1 mg/g or 30% TSP (Mardanova et al. 2015). However, it has been reported that larger transgenes result in vector instability and reduced expression (Avesani et al. 2007). Comoviruses form icosahedral particles and have two separate single stranded RNA molecules. A widely used example is CPMV. A well-known engineered version is the pEAQ vector which is a non-replicating overexpression system containing modified 5' and 3' UTRs derived from CPMV that drive increased transgene expression (Sainsbury et al. 2009). Further developments to PVX vectors involved the combined use of a 5' UTR from alfalfa mosaic virus, which improved gene expression by 3 to 4-fold (Mardanova et al. 2009). A later publication combined the advantages of both pEAQ and PVX vectors to generate pEff, a replicating vector that drives target protein expression at 30% TSP (Mardanova et al. 2017). Lastly, geminiviruses form twinned icosahedral particles with single stranded DNA and have broad host ranges. Many deconstructed vectors derived from geminiviruses have been created and are reviewed by Bhattacharjee and Hallan (2022). Bean yellow dwarf virus (BeYMV) is an example which has been successfully engineered to produce antibodies to Ebola, Zika and West Nile viruses (Huang et al. 2009). Each of these viral vectors could drive high levels of recombinant protein expression, but only the TMV-derived TRBO vector was explored in this research.

1.3.5 Co-expression and suppression of proteins

The final way to improve upstream processing in recombinant protein expression in through the co-expression or suppression of other proteins in transformed tissue. The most commonly co-expressed proteins are suppressors of gene silencing. Post-transcriptional gene silencing, or RNA interference, is an anti-viral plant defence mechanism that occurs when siRNAs bind mRNA transcripts and cause degradation (Zhang et al. 2015), reducing gene expression. However, several viruses have evolved to produce silencing-suppressor proteins that work through binding dsRNA or siRNAs, preventing proper assembly of silencing complexes, degradation of siRNAs or argonaute proteins of the silencing complex, or inhibition of a methyltransferase, HEN1, that stabilises siRNAs (Abrahamian et al. 2020). Commonly used silencing suppressors include p19 from tomato bush stunt virus, p24 from grapevine leafroll-associated virus-2 (Mardanova et al. 2017), HC-Pro from potyvirus (Abrahamian et al. 2020), p22 from crinivirus (Abrahamian et al. 2020), P0 from Polerovirus, P17 from Aureusvirus, and NSs from Tospovirus (Peyret et al. 2019). Research from Peyret et al. (2019) assessed the effectiveness of several and found p19 to be the most effective silencer. As a result, some experiments in this research involved the co-expression p19, though others were not investigated.

Another publication by Norkunas et al. (2018) co-expressed several other proteins alongside p19, including Bcl-2 associated athanogene 4 (BAG4) from *A. thaliana*, a stress response protein, cucumber mosaic virus (CMV) 2b, another suppressor of gene silencing, and two replicase associated proteins (Rep) from tobacco yellow dwarf virus, Rep and RepA, involved in cell cycle regulation. In this study, the co-expression of BAG4, p19, CMV2b, and Rep/RepA increased expression of GUS by up to 2-fold, 2.5-fold, 2.5-fold, and 3-fold, in agroinfiltrated *N. benthamiana* plants, respectively. As a result, these proteins were also investigated during this research. Finally, co-expressing protease suppressors (Clemente et al. 2019) or transforming plant lines that have reduced protease activity has been shown to improve protein expression (Kuo et al. 2013), however these two factors were not investigated during this research.

1.4 – Downstream processing of plant recombinant proteins

1.4.1 Extraction methods

Downstream processing represents the major costs of recombinant protein production in plant expression chassis, comparable to other systems (Kuo et al. 2013; Merlin et al. 2014; Chen and Davis 2016; Dong et al. 2019). Initial extraction aims to liberate target proteins from host cells, discarding tissue that does not express the protein. Various methods are available for this which are often dependent on the nature of the target protein. Broadly, these methods include isolation of plant biomass that expresses the protein. Next, the cell wall is disrupted using either mechanical or chemical methods. Following this, the protein is extracted from the cells using an extraction buffer, of which several are investigated here, with a careful balance of reagents to maximise protein extraction and stability without compromising the ability to purify the protein further downstream. Alternatively, aqueous two-phase separation (ATPS) techniques can be used to extract protein from ground tissue. There are several variations of ATPS (Iqbal et al. 2016), one of which is investigated in this research to a minor extent. In this research, the general extraction process involved the isolation of transformed leaves, grinding of tissue in liquid nitrogen, and protein extraction using an extraction buffer with centrifugation and filtration steps to remove debris.

During extraction, there are several factors to consider. Firstly, protein stability is temperature dependent, so temperatures are kept low to prevent denaturation. This is not an issue when grinding tissue in liquid nitrogen, but larger scale extractions using manual grinding are not feasible, so automated methods are preferable for scaled up experiments. Non-specialist equipment can be used including food processors, tissue homogenisers, and screw presses, where tissue is sheared typically whilst suspended in extraction buffer in a temperature-controlled environment (Wilken and Nikolov 2012; Buyel et al. 2016; Hansen et al. 2022). Screw presses should only be used on a case-by-

case basis for the extraction of cytosolic proteins, as membranous organelles and corresponding proteins are typically not retained with much of the protein, at least 50%, remaining in the pulp (Dotsenko and Lange 2016; Hansen et al. 2022). Secondly, proteins have unique environmental requirements for proper folding and function, so ionic strength needs to be controlled for, typically with the use of a buffering agent and salts. It has been reported that a pH below 5 precipitates the most abundant leaf proteins, pigments and debris, and the presence of secondary metabolites within plants, including alkaloids and phenolic compounds can interfere with target protein stability and purification effectiveness (Wilken and Nikolov 2012). Lastly, the presence of plant proteases within extracts means that inhibitors are necessary, which are often expensive and, in this research, represent the vast majority of extraction buffer costs.

1.4.2 Protein purification

Once total protein has been extracted, it is often necessary to purify target protein for downstream applications. The main purpose of this research was to develop optimised expression constructs to synthesise eukaryotic proteins in plants. However, recombinant proteins must be extracted from the expression chassis and often purified from native host proteins for use in downstream applications. As such, effective production systems should be compatible with purification methods.

There are many available options for purifying proteins, including chromatographic and non-chromatographic techniques, both of which are experimented with in this research. Non-chromatographic techniques that have been used in plants include flocculation (Elmer et al. 2009; Buyel and Fischer 2013; Buyel and Fischer 2014), precipitation by heat, salts, or pH (Menzel et al. 2016; Opdensteinen et al. 2021) and ultrafiltration and diafiltration (Buyel and Fischer 2013; Opdensteinen et al. 2019). Briefly, flocculation involves the addition of a flocculant that causes aggregation of cellular components, increasing their mass and allowing removal through sedimentation or centrifugation. Precipitation methods function based on the unique properties of proteins, where some precipitate at different temperatures and pH. Both flocculation and precipitation are often paired with a filtration technique. Ultrafiltration uses semi-permeable membranes to separate components based on their size and shape, enabling smaller molecules to filter through the membrane whilst larger molecules are retained. This technique is used later in this research as a means for concentrating protein samples. Diafiltration involves the addition of buffer to maintain a constant volume, removing low molecular weight molecules and enabling buffer exchange to purify and concentrate proteins.

Alternatively, chromatographic techniques can be used to separate proteins in a mixture by their properties. Such methods include size exclusion chromatography, reverse phase and hydrophobic

interaction chromatography, ion exchange chromatography, and affinity chromatography. These methods separate proteins based on their size, hydrophobicity, charge, and affinity to a ligand, respectively. Chromatographic techniques can also be automated using fast protein liquid chromatography (FPLC) systems to fine tune purification processes, saving time and resources whilst minimising human error. FPLC methods are tested in this research, using several different purification columns and FPLC systems.

Importantly, the biochemical properties of proteins can be modified using affinity tags. The addition of these tags then enables purification methods to be used that would be unsuitable for the target protein lacking the tag. Protein tags have been developed that enable non-chromatographic purification from plant extracts, including fusions of oleosins, elastin-like proteins, and g-Zein-derived peptides, which enable purification of proteins by affecting hydrophobicity (Wilken and Nikolov 2012; Gerszberg and Hnatuszko-Konka 2022; Parmenter et al. 1995), temperature-dependent solubility (Rigano et al. 2013), and intracellular localisation (Torrent et al. 2009; Joseph et al. 2012), respectively. However, affinity tags coupled with chromatographic techniques are more commonly used, including in this research. Many examples exist and have been extensively reviewed (Coates et al. 2022; Pina et al. 2014; Kimple et al. 2013), including tags that grant affinity to antibodies (Jeong et al. 2018; Daniell et al. 2009), metals allowing immobilised metal affinity chromatography (IMAC; Sainsbury et al. 2016; Wilken and Nikolov 2012), macromolecules (Pina et al. 2014; Islam et al. 2018), or to other proteins (Jeong et al. 2018; Qi and Katagiri 2009). In this research, several affinity tags that have shown success in plant expression systems were tested, including tags that bind streptavidin (Qi and Katagiri 2009), cellulose (Islam et al. 2018), antibodies, and nickel beads. IMAC using nickel beads is the dominant purification method used in this research, and is further investigated in several different forms, assessing the suitability of the method using free and immobilised ligand, with manual and automated purification methods. The protein extraction and purification pipelines tested in this research can be seen diagrammatically in Figure 1.3.

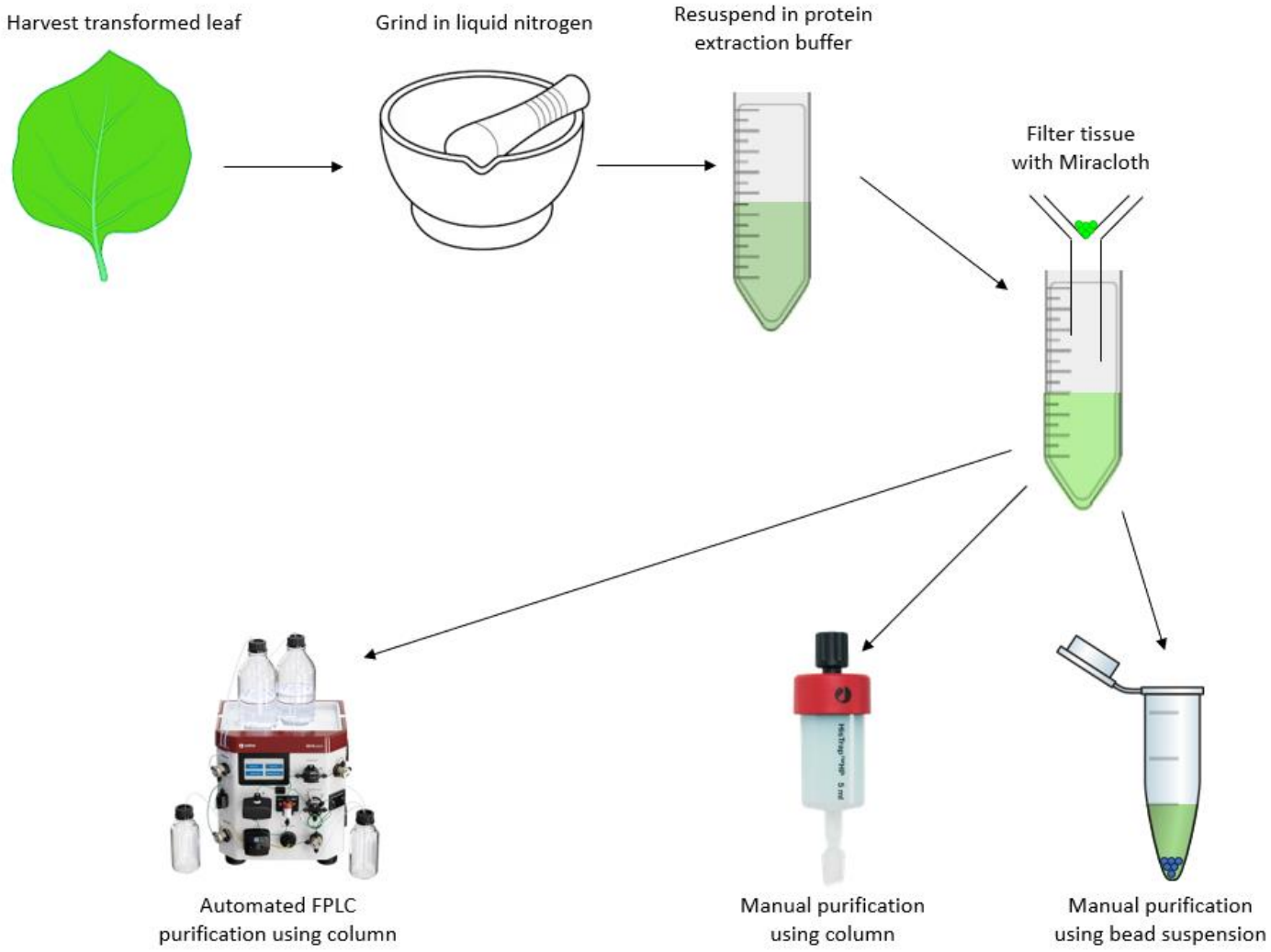


Figure 1.3: Downstream processing pipeline used in this research. Proteins are extracted from transformed tissue by grinding harvested leaves in liquid nitrogen, followed by suspension in a protein extraction buffer. Successful protein extracts then have target proteins purified using either manual or automated methods.

1.5 – The limits of recombinant protein complexity in plants

The final factor assessed during this research is the difference in expression levels between proteins of varying complexity, defined here as factors including protein size, post-translational modifications, and intracellular localisation. It is well established that different recombinant proteins are produced in varying yields, with a range of literature showing this (Lindbo 2007b; Mardanova et al. 2017). However, most of these studies have expressed simple proteins that localise to the cytoplasm and have no or few post-translational modifications. To my knowledge, no study has actively examined the effects of increasing protein complexity on yield, but several factors contribute to yield variation.

Firstly, protein size influences yield, with larger proteins being more difficult to produce than smaller ones due to increased metabolic impact on the cell. Larger proteins require more complex folding, have reduced stability, and are more enegetically demanding to translocate (Farías-Rico et al. 2018). Secondly, the presence of post-translational modification impacts protein yields, with modified proteins requiring further processing within organelles where modifying enzymes are localised. This results in further trafficking of proteins through the cell, particularly the endoplasmic reticulum and Golgi apparatus but also in other organelles (Pizarro and Norambuena 2014). Whilst the presence of these modifications has not been directly measured in this research, some protein targets are naturally modified in their endogenous host organism and are consequently unable to be produced in prokaryotic expression systems. Lastly, the intracellular localisation of proteins impacts yields, partly due to increased trafficking and processing being required to transport proteins through the cell, and partly because some cellular components have lower volumes within a cell. This is particularly true for plasma membrane proteins which are typically hydrophobic, require strict environments to properly fold, must be transported effectively, and finally integrated into the membrane which comprises a much lower proportion of cellular volume than the cytosol (Carpenter et al. 2008; Pizarro and Norambuena 2014).

As a result, several protein targets are investigated in this research which are chosen for their varying sizes, intracellular localisations, and endogenous PTMs, although PTMs were not directly measured in this research. These proteins are summarised in Table 1.2, and are discussed in detail in Chapter 5. However, it is important to note that the selection of these proteins is not exhaustive but instead is a compromise between the number of protein targets and protein diversity. These include cytosolic and plasma membrane proteins, with varying post-translational modifications and sizes, from a range of host species including plant hosts to act as controls. The main target protein of this

research was initially P2X7, a mammalian transmembrane cation channel with extensive PTMs, but this was later expanded to include other protein targets and truncated P2X7 variants lacking key domains to reduce complexity. However, as a mammalian protein is being expressed in plants, functional assays were also performed to indicate whether the target protein is folding and functioning correctly. To understand this, functional P2X7 assays in mammalian hosts must be discussed.

Table 1.2: Details and relevance of target proteins expressed in this research.

Protein (UniProt entry accession code)	**Size (kDa / AA)**	**Modifications**	**Intracellular Localisation**	**Host organism and role**	**Importance**
eGFP (derived from P42212)	27 kDa / 238 AA	None	Cytosolic	*Aequorea victoria* - fluorescent protein	Reporter protein to measure construct expression.
DORN1 (Q9LSR8)	103 kDa / 766 AA	Glycosylation	Plasma membrane	*A. thaliana* - receptor for extracellular ATP, playing a role in plant immune responses (Chen et al. 2017; Kumar et al. 2020).	Native to plants so can be produced functionally and acts primarily as a proof of concept in this research.
Putative cadmium/zinc-transporting ATPase HMA4 (O64474)	128.7 kDa / 1172 AA	None	Plasma membrane	*A. thaliana* - detoxification and root-to-shoot translocation of zinc and cadmium (Laurent et al. 2016; Zimmermann et al. 2009)	An unmodified integral membrane protein native to plants so can be produced functionally and acts primarily as a proof of concept in this research.

Factor C (P28175)	113.8 kDa / 1019 AA	Glycosylation and disulfide bonds	Cytosolic	*Tachypleus tridentatus* - plays a key role in its blood clotting system and is sensitive to bacterial endotoxins (Dubczak et al. 2021)	A modified cytosolic protein of industrial and medical relevance for its role in sterility testing.
RAN GTPase activating protein 1 (RANGAP1; Q9LE82)	60.3 kDa / 535 AA	None	Cytosolic	*A. thaliana* - involved in nuclear transport and is essential for cell division and growth (Xu et al. 2008)	An unmodified cytosolic protein native to plants so can be produced functionally and acts primarily as a proof of concept in this research.
Human and rat full-length P2X7 (Q99572 and Q64663, respectively)	Human - 68.6 kDa / 595 AA Rat - 68.3 kDa / 595 AA	Glycosylation, disulfide bonds, and palmitoylation	Plasma membrane	*Homo sapiens and Rattus norvegicus* - ATP-gated integral membrane protein channel involved in a range of physiological responses including inflammation and cell death (Tewari and Seth 2015; Lara et al. 2020)	A medically relevant mammalian protein involved in a range of pathophysiological disorders. A nearly full structure of rat P2X7 is available but the structure of human P2X7 is not.

Human and rat cysteine-rich P2X7 (truncated versions of Q99572 and Q64663, respectively)	Human - 27.7 kDa / 239 AA Rat - 27.7 kDa / 239 AA	Disulfide bonds, and palmitoylation	Cytosolic but palmitoylation causes membrane association	Truncated version of P2X7 lacking the transmembrane domain but containing a palmitoylated cysteine-rich region and the C-terminal signalling domain	This truncated version of P2X7 lacks the transmembrane domain, so should be cytosolic but contains a heavily palmitoylated region that could cause association to membranes.
Human and rat ballast P2X7 (truncated versions of Q99572 and Q64663, respectively)	Human - 23.1 kDa / 201 AA Rat – 23.2 kDa / 201 AA	None	Cytosolic	Truncated version of P2X7 lacking the transmembrane domain and palmitoylated cysteine-rich region, containing only the C-terminal signalling domain	This truncated version of P2X7 lacks both the transmembrane domain and palmitoylated region, containing only the C-terminal signalling domain considered to be cytosolic. This C-terminal signalling domain is involved in P2X7 function and signalling, though specific roles are poorly understood.

1.6 – P2X7 and functional assays

Despite testing a range of different protein targets, P2X7 was the main target protein of this project because at the start of this work no 3D-structure of full-length P2X7 was available, and the human protein is an important potential target for anti-inflammatory drugs. P2X proteins are a family of ATP-gated ion channels, involved in a range of physiological processes including inflammation, pain perception, and neurotransmission (Burnstock et al. 2014). P2X7 is different from other P2X receptors, in that it has a larger intracellular C-terminal signalling domain, a lower ATP-sensitivity, requiring high micromolar concentrations of ATP to open the ion channel, compared to other P2X receptors which are activated by low micromolar to nanomolar ATP concentrations. P2X7 has been linked to several pathologies including neurodegenerative diseases, psychiatric disorders, chronic pain, inflammation, and cancers (Tewari and Seth 2015; Lara et al. 2020). P2X7 is unique among the family of P2X purinergic receptors for its dual functionality. Initially, upon binding extracellular ATP, P2X7 acts like a typical ion channel, allowing influx of cations into the cell. This early ion flux can modulate various cellular activities, including signal transduction and gene expression (Karasawa et al. 2017). However, what sets P2X7 apart is its ability to transition from a selective cation channel to forming a large, nonselective pore upon prolonged ATP stimulation. This pore formation allows molecules up to 900 Da to pass through, significantly altering the cell's ionic balance (Karasawa et al. 2017). This leads to a cascade of downstream effects including the activation of inflammatory pathways, the release of cytokines, and cell death (Oliveira-Giacomelli et al. 2021). As functional information can be obtained using structural studies, producing P2X7 recombinantly is vital for understanding the precise mechanisms governing its pore formation and the subsequent biological effects. It allows for the analysis of its structure-function relationship, identifying conformational changes that occur during pore formation and cellular signalling cascades. Whilst a cryoEM structure for full-length rat P2X7 has been published (Figure 1.4, Panel A; McCarthy et al. 2019), the structure of full-length human P2X7 is still not known, and there are significant differences between the rat and human P2X7 receptors, including differential agonist sensitivity (Donnelly-Roberts et al. 2009), that may be explained through obtaining the structure of human P2X7. Despite efforts, full-length human P2X7 has not been successfully produced in conventional expression systems in sufficient yields for structural studies. Commercially, human P2X7 can be purchased as HEK293 cells expressing the protein (abm, catalogue number: T6316), or as purified protein from a wheat germ cell-free system (abcam, catalogue number: ab159042). Importantly, these expression systems are not cheaply scalable and the purified protein product from wheat germ cells has a molecular weight of 94.9 kDa with an additional N-terminal GST tag, so may not have the correct folding, modifications, or function as native P2X7. However, whole plants have the potential capacity to produce the

protein and have not been tested as expression hosts for P2X7 to date. Furthermore, the C-terminal signalling domain of P2X7 is thought to contribute to the many of these diseases, so truncated versions of P2X7 (shown in Figure 1.4, Panel B) comprising the soluble cytoplasmic C-terminus alone are also viable targets to produce recombinantly, as if the full-length structure cannot be produced in plants, functional insights may be obtained through studying these truncated versions. This rationale was used in this research, where N-terminal truncations of P2X7 were made to lack key domains, including the transmembrane domain and a heavily palmitoylated region. It was expected that if full-length P2X7 could be produced in plants the yields would likely be low as it is a post-translationally modified intermembrane protein. However, even if yields are low, it is possible that functional assays may be able to detect P2X7 activity in transformed plants.

There are several existing assays to detect P2X7 activity. These include the measurement and quantification of cation uptake using dyes such as Yo-Pro-1 and ethidium bromide, which each fluoresce when the dyes bind nucleic acid, or Fluo-4 which measures calcium influx (Pelegrín 2011; Karasawa et al. 2017). Alternatively, patch-clamp assays can be used which directly measure the flow of ions across P2X7 channels (Karasawa et al. 2017), or cell death assays can measure lactate dehydrogenase release (Bidula et al. 2019) following P2X7 activation by ATP. In this research, dye uptake methods were used to analyse P2X7 function as they were relatively straightforward to set up, and routinely used in the Young lab to assay mammalian P2X7 receptor function. P2X7 activity can be convincingly demonstrated if any activity is abolished in the presence of antagonist molecules, such as JNJ-47965567 (Karasawa and Kawate 2016), employed here.

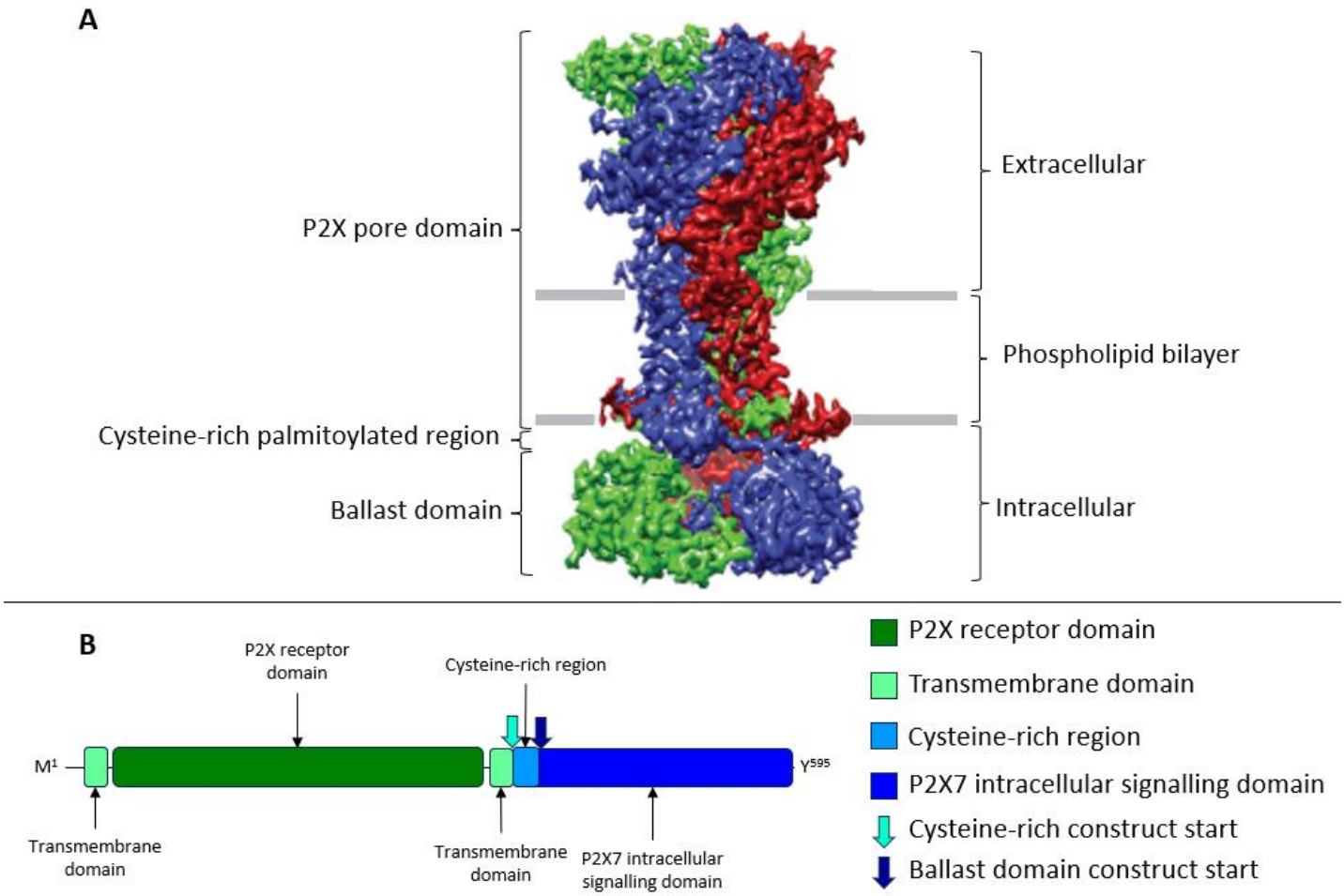


Figure 1.4: The structure of P2X7. A – Published full-length rat P2X7 CryoEM structure by McCarthy et al. (2019), modified to show the pore domain, cysteine-rich region, and ballast domain, alongside localisation of each region. **B** – line diagram of domains of P2X7 and sites of importance, showing truncation points for the two C-terminal isoforms used in this research (blue arrows).

1.7 – Aims, objectives and structure

Overall, plants are a promising expression platform for recombinant proteins, but efforts need to be made to improve *in planta* yields, develop cost effective downstream purification processes, and assess the effects of protein complexity on yields. These points are the scope of this research and are the framework of this thesis.

The aim of this research is to develop a plant-based expression system for complex recombinant proteins, with a focus on P2X7. This is done through a series of objectives. Firstly, *in planta* expression is increased using genetic engineering approaches, testing different regulatory regions, transformation methods, and expression vectors in *N. benthamiana*, using eGFP as a reporter to measure this. Secondly, protein extraction and purification methods are tested, and their effectiveness compared using SDS-PAGE, InstantBlue™ staining, and western blotting, to develop an effective downstream processing pipeline. Thirdly, once protein expression and extraction are improved, different proteins of varying complexity are expressed to identify the relationship between protein complexity and yield. Finally, the production of P2X7 is measured in several expression systems, and experiments are attempted to determine protein functionality.

This thesis first covers the materials and methods employed during this research in Chapter 2. Chapter 3 focusses on improving *in planta* yields, primarily using *N. benthamiana* as the expression host, experimenting with different genetic components and transformation methods. Quantification of expression is performed primarily using eGFP as a reporter with fluorescence measured using photon imaging and confocal microscopy. Chapter 4 develops an effective extraction and purification method to isolate target proteins from transformed tissue, assessing the effects of different extraction methods and buffers, and purification tags and methods. Metrics to measure the effectiveness primarily use SDS-PAGE, InstantBlue™ staining and western blotting. Chapter 5 then takes the best expression constructs and extraction methods and tests these with different proteins, varying in complexity. Detection of target proteins is primarily performed using SDS-PAGE and western blotting, and involves measurements of total extracted protein, purified protein and concentrated protein. Here, there is also some investigation of other plant expression systems with a focus on P2X7 specifically. Finally, the overall findings are discussed in greater context with the wider literature where limitations are explained, and avenues for future work are considered to address these limitations and answer any questions raised by this research.

Collectively, this research enhances our understanding of regulatory elements used in expression construct design, with later constructs showing clear improvements in protein yields than initial constructs. It also develops our scientific understanding of the relationship between protein

complexity and yield, and sheds light on the varying capacities of different expression hosts and systems for producing different proteins. It also raises further questions that could form the basis for future work, including the use of other plant species as expression hosts, different replicating vectors and purification methods, and increasing the overall scale of experiments.

Chapter 2 - Materials and Methods

2.1 – Gene synthesis

DNA sequences were ordered from GeneWiz® using their PriorityGene Service following codon optimisation using the Benchling® codon optimisation tool.

2.2 – DNA purification

2.2.1 Plasmid DNA extraction

Plasmid DNA was extracted using plasmid DNA extraction kits following QIAprep Spin Miniprep Kit instructions (Qiagen, catalogue number: 27106) as follows: 5 mL of bacterial culture was centrifuged at 4,000 *x g* for 10 minutes. The supernatant was discarded, and the pellet was resuspended in 250 µL of resuspension buffer (50 mM Tris·HCl, 10 mM EDTA, 100 µg/mL RNase A, pH 8). 250 µL of lysis buffer (200 mM NaOH 200 mM and 1% SDS (w/v)) was added and the solution gently inverted 10 times, and the reaction incubated at room temperature for 2-3 minutes. 350 µL of neutralization buffer (4.2 M Gu-HCl, 0.9 M KOAc, pH 4.2) was added and the tube again inverted to mix. Samples were then centrifuged for 10 minutes at 20,000 *x g*. 800 µL of supernatant was transferred into a spin column and collection tube, centrifuged for 1 minute and the flow through discarded. Following this, 750 µL of wash buffer (10 mM Tris·HCl, 80% ethanol (v/v)) was added to the spin column, centrifuged for 1 minute, and the flow through discarded. A second spin was performed to remove residual buffer. The spin column containing the DNA was transferred to a sterile Eppendorf tube and 30-50 µL of 65°C Milli-Q H_2O was added and incubated at room temperature for 1 minute. The DNA was then eluted with a 1-minute centrifugation.

2.2.2 DNA extraction from an agarose gel

DNA extractions from agarose gels were carried out using a Zymoclean Gel DNA recovery kit (Zymoresearch, catalogue number: D4001/D4002) according to the manufacturer's instructions. This involved carefully excising the DNA band from the agarose gel and resuspending in 3 volumes (w/v) of ADB buffer followed by heating at 50°C for 5-10 minutes. The sample was centrifuged at 20,000 *x g* for 1 minute using the provided spin columns and the flow through was discarded with the DNA bound to the membrane. This was then washed twice in 200 µL of wash buffer (10 mM Tris·HCl, 80% ethanol (v/v)) with a 1-minute centrifugation between washes and the flow through discarded. The spin column containing the DNA was transferred to a sterile Eppendorf tube and 10-30 µL of 65°C Milli-Q H_2O was added and incubated at room temperature for 1 minute. The DNA was then eluted with a 1-minute centrifugation.

2.3 – Cloning

2.3.1 Q5 Amplification of DNA fragments

PCR amplification of DNA fragments using Q5 DNA polymerase was carried out according to manufacturer's instructions (NEB, catalogue number: M0491S), with slight modifications. This was as follows: 0.25 µL of Q5 high-fidelity DNA polymerase, 4 µL of 5 x buffer, 1 µL 10 mM dNTPs mix (NEB, catalogue number: N0447S), 1 µL of each 10 µM forward and reverse primers, 100 ng of template DNA, and Milli-Q H_2O to a final volume of 20 µL was added to a PCR tube. For difficult to amplify fragments, these volumes were scaled up proportionately to a 50 µL reaction volume, keeping the amount of template DNA the same and replacing the excess with water instead. Thermocycler conditions used were typically an initial denaturation of 95°C for 2 minutes, followed by 30 cycles of 95°C denaturation for 10 seconds, 55-60°C annealing for 10 seconds, and 72°C amplification for 30 seconds per kilobase of target amplicon, followed by a final extension step of 72°C for 10 minutes, and a final hold of 4°C. For difficult to amplify fragments, a variant called a touch-up PCR was used, where the first 10 cycles of annealing would start at 48°C and gradually increase to the target annealing temperature over these 10 cycles.

2.3.2 General Golden Gate cloning

Golden Gate cloning was achieved by mixing 1.5 µL of T4 DNA ligase buffer, 1 µL Type IIS restriction endonuclease (either BsaI (Thermofisher, catalogue number: ER0291) or BpiI (Thermofisher, catalogue number: ER1012)), 1 µL of T4 DNA ligase (NEB, catalogue number: M0202S), vector: insert in a 1:3 molar ratio, and Milli-Q H_2O to a total volume of 15 µL. For constructs with <3 inserts, these were placed in a thermocycler for 10 cycles of 37°C for 5 minutes, then 16°C for 5 minutes, followed by a final 37°C cutting step for 10 minutes, then a 55°C followed 80°C denaturation step, each for 10 minutes, with a final hold of 4°C. For longer and more difficult constructs with >3 inserts, 100 cycles were used instead. 5 µL of the reaction mix was used to transform *E. coli* cells.

2.3.3 General conventional cloning

Conventional restriction endonuclease cloning was carried out by mixing 1 x NEB rCutSmart™ Buffer with 0.5 µL of appropriate restriction enzyme (AvrII (NEB, catalogue number: R0174S) or PacI (NEB, catalogue number: R0547S)), 500 ng of DNA and Milli-Q H_2O to a total volume of 10 µL and incubated at the optimal temperature of the restriction endonuclease for 3 hours or overnight. The digest reaction was separated on an agarose gel and the appropriate band extracted from the gel. Ligation reactions were carried out using an approximate 3:1 Insert: vector molar ratio at maximum

concentration, 1 x T4 DNA ligase buffer, 0.5 µL DNA ligase and Milli-Q H_2O to a final volume of 10 µL and left overnight at 16°C. 5 µL of this ligation reaction was used to transform *E. coli*.

2.3.4 Blue-White Screening of Colonies

E. coli colonies containing the Golden Gate vector with the desired insert were screened for using blue-white selection on agar plates containing appropriate antibiotic, 100 µM IPTG and 40 µg/mL X-gal. Blue colonies expressing the LacZ screening marker were ignored, and white colonies were screened by colony PCR.

2.3.5 Screening of colonies by colony PCR

Colony PCRs for screening were carried out using 1 x PCRBIO Taq Mix Red (PCRBiosystems, catalogue number: PB10.13-02), 0.5 µM forward and reverse primer, and Milli-Q H_2O were added to a total volume 10 µL. The target colony was carefully picked into this solution and microwaved for 10 seconds and placed in a thermocycler. Thermocycling conditions used were 95°C for 2 minutes, then 35 cycles of 95°C denaturation for 15 seconds, 55°C annealing for 15 seconds, and 72°C amplification for 20 seconds per kilobase of amplicon DNA, followed by a final amplification stage of 72°C for 2 minutes.

2.3.6 Agarose gel electrophoresis

Agarose gel electrophoresis was carried out using a 0.8-2% agarose TAE gel containing SafeView at 50 nL/mL, in 1 x TAE buffer, at 100 volts for suitable running time (20-60 minutes, depending on desired product separation). Samples were mixed with 1 x NEB purple loading dye if necessary. A DNA size marker was loaded to outer lanes (1Kb plus; NEB, catalogue number: N0469S; SmartLadder, Eurogentec, catalogue number: MW-1700-02).

2.3.7 Creation of Phase 1 expression constructs

To create the simple (S) and complex (C) expression cassettes used in this research, each DNA element needed to be designed in a way that they could ultimately be combined into the final expression construct. For the creation of the simple (S) non-replicating (N) construct, pNS, the viral CaMV 35S promoter was already obtained by the Scofield group in the MoClo components kit, and was amplified by PCR using Q5 polymerase to incorporate the PacI restriction site at the 5’ end of the promoter, with the BpiI restriction site at the terminal ends and the appropriate sticky ends, 5’-

GGAG and AATG-3', enabling digestion and ligation into the Pro + 5'UTR Level 0 module, pICH41295. This was then transformed into *E. coli*, white colonies were screened by colony PCR and positive clones were picked overnight into broth. The plasmid DNA was extracted by miniprep and confirmed by DNA sequencing. The eGFP sequence, already used within the Young lab, was also amplified by Q5 PCR using overhangs that produce the sticky ends 5'-AATG and TTCG-3', cloned into the Level 0 module pICH41308 which was transformed into *E. coli*. White colonies were screened by colony PCR and positive clones picked overnight into broth, with the plasmid DNA extracted by miniprep and confirmed by DNA sequencing. Finally, the nopaline synthase terminator was amplified to contain a 3' AvrII recognition site, with the terminal overhangs of 5'-TTCG and CGCT-3'. This part was digested and ligated into pICH41276 which was transformed into *E. coli*, white colonies screened and positive clones picked overnight into broth. The plasmid DNA was extracted by miniprep and confirmed by DNA sequencing. Following this, the three parts could be combined into the Level 1 acceptor plasmid, pICH47742, shown in Figure 2.1, using a Golden Gate cloning reaction with the BsaI restriction enzyme. This was transformed into *E. coli,* screened by colony PCR, and positive clones grown in broth overnight, miniprepped, and sequenced, creating the first expression construct, pNS-eGFP.

The cloning of the complex (C) non-replicating (N) expression construct, pNC, required the redesigning of every DNA element. The MoClo components kit has existing Level 0 acceptor plasmids with set sticky ends, each designed for a different DNA element to be inserted into. Briefly, there are three pre-designed slots for DNA elements to be inserted 5' of the coding sequence. These are the promoter, 5' UTR, and an N-terminal tag as options (5' → 3'). Downstream of the coding sequence module, there are another three pre-designed slots, a C-terminal tag, 3' UTR, and terminator module. The complex expression construct required four DNA elements 3' of the coding sequence; these were the 3' UTR, *two* terminators, and a matrix attachment region. As a result, each DNA element was designed to fit into a new order using the existing sticky ends, with the 3' UTR occupying the acceptor for the C-terminal tag, the two terminators being constructed together as a single unit to occupy the 3' UTR module, and the terminal matrix attachment region designed to be ligated into the terminator module. A diagram of this cloning method, along with sticky ends, can be seen in Figure 2.2.

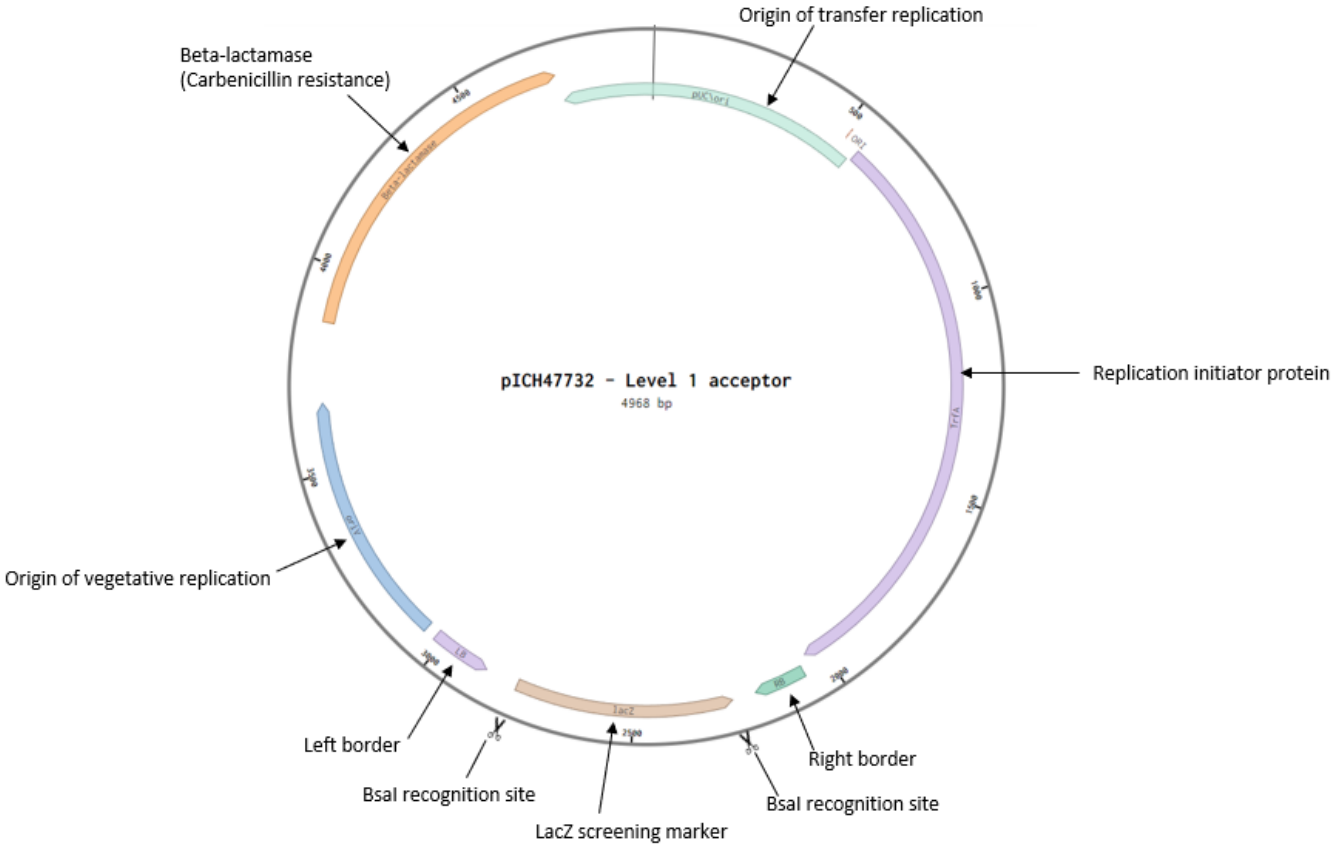


Figure 2.1: Vector map of the Golden Gate Level 1 expression vector, pICH47742. This is a Golden Gate vector that enables expression of target proteins in plants and does not replicate target RNA molecules. It contains a beta-lactamase gene for carbenicillin selection in bacteria, and a LacZ gene for the expression of beta-galactosidase enabling blue-white screening, flanked by the BsaI restriction enzymes for Golden Gate cloning, enabling blue-white selection of bacterial clones. Flanking these are the left and right borders that enable *Agrobacterium* mediated transfer of expression constructs into the nuclei of plant cells. The two origins of replication enable replication of the vector in both *E. coli* and *Agrobacterium*.

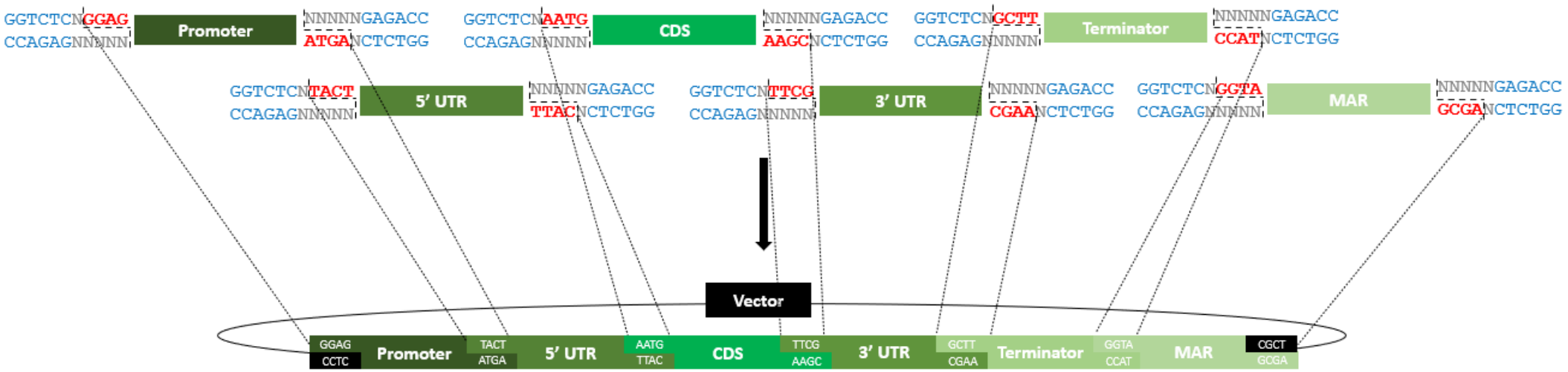


Figure 2.2: Schematic showing the assembly of the complex expression cassette. Different level 0 modules hold different DNA elements, including promoters, untranslated regions (UTRs), coding sequences (CDS), terminators, and matrix attachment regions (MARs). These are contained in Level 0 vectors, which are not shown for simplicity. Each level 0 vector can be digested at BsaI recognition sites (blue) to create different 4 bp sticky ends (red) that enable construction into a level 1 transcriptional unit/expression cassette. This diagram displays the order of assembled level 0 parts used to generate the level 1 construct, pNC, alongside the relevant sticky ends for each sequence. Constructs shown are 5' to 3', left to right.

The constitutive viral CaMV 35S promoter used for the complex constructs was identical to that used for the simple expression constructs, with a PacI site inserted at the 5' end of the promoter by PCR amplification using primers with overlaps (see Table S1 containing primer sequences) enabling downstream excision and ligation in the replicating vector pJL-TRBO.

Both the synthetic 5' UTR and the CPMV 3' UTRs were synthesised from GeneWiz as linear DNA fragments with the appropriate BpiI restriction recognition sites and sticky ends incorporated to enable direct Golden Gate cloning into their Level 0 acceptor vectors. The 5' UTR was ligated into the pICH41246 vector, and the 3' UTR into the pAGM1301 vector. Each was then transformed into *E. coli* and white colonies were screened by colony PCR. Positive clones were picked into broth overnight, miniprepped and confirmed by DNA sequencing.

Because of the sticky end compatibility of the 3' UTR module, the eGFP coding sequence needed to have its 3' sticky end redesigned so that it could ligate to the 3' UTR. This means that the eGFP was instead cloned into the vector pAGM1287, using Q5 PCR amplification to create the correct sticky end. The linear product was analysed by gel electrophoresis and excised, then digested and ligated into the Level 0 vector using BpiI. This was transformed into *E. coli* and white colonies were screened, cultured in liquid media, miniprepped and the DNA confirmed by sequencing.

The terminator elements, *N. tabacum* extensin (EU) and *N. benthamiana* Actin (Act), have a little more complexity than the other parts, as this required the combining of two independent DNA elements into one Level 0 part. Theoretically, these could have also been synthesised as one functional unit ready to ligate into the Level 0 acceptor, but the initial design was going to incorporate a third CaMV 35S terminator to be utilised as double terminator. This meant that there were going to be two double terminators created and tested as each was shown to be promising in previous research (Diamos and Mason 2018). These were going to be the 35S CaMV::Act double terminator, and the EU::Act terminator. As such, the parts were designed with modularity in mind so that they could be cloned in the appropriate orders, which required a lower level of Golden Gate vectors, Level -1. Typically these vectors are used to remove internal Type IIS restriction endonuclease recognition sites by creating silent mutations, in a process called domestication. However, for the creation of the complex expression cassette, the Level -1 system was utilised to combine the double terminators in tandem. Due to difficulties amplifying the CaMV 35S terminator, the CaMV 35S::Act double terminator was scrapped, resulting in the creation of only the EU::Act terminator. However, because of the initial design goals, the DNA sequences of these were initially synthesised and cloned into the Level -1 expression vector, pAGM1311, with custom sticky ends

created so that the two terminators could scarlessly be ligated together. The Level -1 Type IIS restriction endonuclease is BsaI, so these overhangs were designed at each end of both terminators. The two terminators were independently PCR amplified with their appropriate overhangs and sticky ends, and cloned into the Level -1 acceptors, which were transformed into *E. coli*. White colonies were screened by colony PCR and positive clones were grown up overnight in broth, had the plasmid DNA isolated by miniprep, and the DNA sequenced. Once both sequences were successfully cloned into the Level -1 vectors they were then combined into the Level 0 acceptor module. This construct was transformed into *E. coli*, and white colonies were screened by colony PCR. Successful colonies were picked into broth, the DNA miniprepped and confirmed by DNA sequencing. Importantly, whilst it is favourable to domesticate DNA elements and remove internal Type IIS recognition sites, an internal site was *not* domesticated within the *N. benthamiana* Actin terminator as it is risky to mutate regulatory DNA elements as this could adversely affect the secondary structure and function of the terminator. This meant that the double terminator Level 0 part contained an internal BsaI cut-site that would *not* be removed during the Level 1 construct creation process, reducing cloning efficiency to preserve terminator function.

Finally, the RB7 matrix attachment region was synthesised by GeneWiz but with an additional overhang containing the 3' AvrII recognition site to enable downstream cloning into the replicating vector. The synthesised DNA fragment also contained terminal BpiI overhangs that enable direct cloning into the appropriate Level 0 acceptor vector. *E. coli* was transformed with this construct and white colonies screened by colony PCR. Positive clones were grown in broth overnight, miniprepped and the DNA sequenced.

Following this design and cloning, all the Level 0 DNA elements utilised in this research had been created and could be cloned into the Level 1 non-replicating (N) acceptor vector, pICH47742, shown in Figure 2.1, using a Golden Gate reaction with BsaI as the restriction enzyme. This was transformed into *E. coli* and white colonies were screened by colony PCR. Positive clones were growth in broth and the DNA miniprepped and sequenced, creating the complex eGFP expression construct, pNC-eGFP.

2.3.8 Creation of Phase 2 and 3 expression constructs

pJL-TRBO, shown in Figure 2.3, requires conventional cloning to input a coding sequence, utilising either PacI, AvrII, or NotI as the appropriate restriction endonucleases. The Phase 2 and 3 constructs utilised this replicating vector and were amplified with a PacI recognition site at the 5' end and an AvrII site at the 3' end.

For the pRS-eGFP construct, the pNS-eGFP construct was PCR amplified with primers that generate the PacI and AvrII recognition sites. The product was separated using gel electrophoresis and the appropriate band corresponding to the insert was gel extracted. The vector and insert were digested with PacI and AvrII, and the digested vector was isolated using a QIAquick PCR Purification Kit according to manufacturer instructions. The linear vector and inserts were ligated together overnight. This method was also used for the Phase 3 constructs using the pNC-eGFP construct as a template for PCR amplification to generate the PacI and AvrII overhangs in the desired locations. For the Phase 2 construct pRC-eGFP, pNC-eGFP was digested and ligated into the pJL-TRBO as the PacI and AvrII recognition sites were already present.

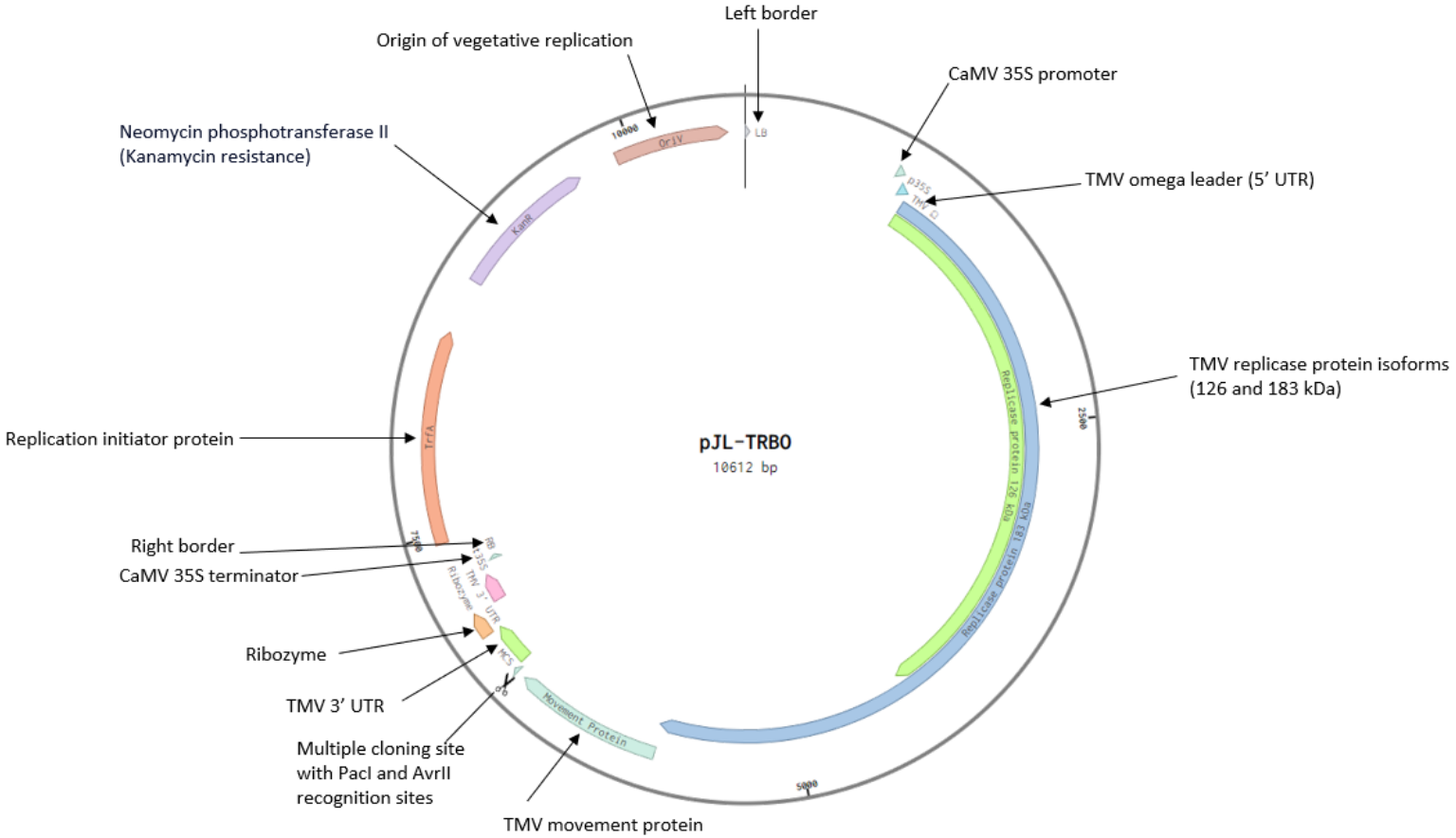


Figure 2.3: Vector map of the replicating vector, pJL-TRBO. This is a deconstructed vector derived from tobacco mosaic virus that enables expression of target proteins in plants and replicated target RNA molecules. It contains a neomycin phosphotransferase II gene for kanamycin selection in bacteria, and a multiple cloning site with the PacI and AvrII restriction recognition sites. In addition, the vector contains the TMV movement and replicase proteins that enable RNA replication and spread through leaf plasmodesmata, enabling neighbouring cells not directly transformed by the *Agrobacterium* to express proteins. Downstream of the multiple cloning site is a ribozyme sequence and CaMV 35S terminator which each ensure efficient transcript termination. Flanking the viral proteins and cloning site are the TMV 5' and 3' UTRs, enabling replication and movement of RNA in between the UTRs. The left and right borders enable *Agrobacterium* mediated transfer of expression constructs into the nuclei of plant cells.

2.3.9 Development of the 'Improving Infiltration' Level 2 construct

The 'Improving Infiltration' Level 2 construct contained multiple transcriptional units, expressing Rep, RepA, CMV2b, BAG4, and p19 proteins to further improve the yields of recombinant proteins in transient plant transformants. Each protein coding sequence was codon optimised according to the *A. thaliana* codon usage template using the Benchling codon optimisation tool, and synthesised by GeneWiz (see Table S1 for synthesised DNA) to contain BpiI overhangs enabling ligation into the Level 0 acceptor plasmid, pICH41308. Each insert was digested and ligated into the vector and transformed into *E. coli*. White colonies were screened by colony PCR and positive clones selected, grown in broth, miniprepped and the DNA sequenced. Following this, Level 1 transcriptional units were made using these coding sequences and regulatory elements existing in the Golden Gate toolkit, obtained by the Scofield laboratory prior to this research. Level 1 transcriptional units were created with the Rep and RepA proteins under control of a 5' Mas promoter from *A. tumefaciens* with a 3' nopaline synthase terminator. The BAG4 and CMV2b coding sequences were combined with a 5' RbcS2 promoter from *Solanum lycopersicum*, also with the 3' nopaline synthase UTR. The p19 protein was expressed under control of the CaMV 35S promoter at the 5' end, and a nopaline synthase terminator at the 3' end. As a Level 2 construct was being created from these components, each had to be ligated into a different Level 1 acceptor with sticky ends that enable the combination of several level 1 cassettes. To achieve this, the CaMV 35S::p19::NosT construct was ligated into the Level 1 acceptor pICH47732. The Mas::Rep::NosT construct was ligated into the Level 1 acceptor, pICH47742 and the Mas::RepA::NosT construct was ligated into pICH47751. The RbcS2::CMV2b::NosT construct was ligated into pICH47761, and the RbcS2::BAG4::NosT construct was cloned into pICH47772. This design enables the ligation of the constructs in that order from 5' to 3', with an end-linker at the 3' most end that enables the Level 2 acceptor to close. Once each Level 1 construct had been assembled using a BsaI Golden Gate reaction, *E. coli* were transformed, white colonies screened by colony PCR, and positive clones grown in broth overnight. These then had DNA extracted by miniprep and were confirmed by DNA sequencing. Following this, each Level 1 construct was combined into the Level 2 acceptor pAGM4673 creating the expression cassette shown in Figure 3.14. This Level 2 construct was transformed into *E. coli*. Due to the size of this construct, a colony PCR was not suitable for screening. Instead, several colonies were picked into broth overnight and miniprepped the following day, and the constructs were screened by restriction digest and analysis by gel electrophoresis.

2.4 – Bacterial cells

2.4.1 Antibiotic concentrations

Antibiotics were used for bacterial selection at the following concentrations: Kanamycin (50 μg/ml), Spectinomycin (100 μg/ml), Carbenicillin (100 μg/ml), Rifampicin (30 μg/ml), Gentamycin (20 μg/ml)

2.4.2 Growth of bacteria

Bacteria were grown on LB agar plates (LB Agar, 32 g/L, pH 7.2 (Melford; catalogue number: L24033)) or liquid media culture (Luria Broth, 20 g/L, pH 7.2 (Melford; catalogue number: L24060)) containing appropriate antibiotic. Level -1, 2 and pJL-TRBO clones were grown on kanamycin, level 0 clones on spectinomycin, and level 1 clones on kanamycin. Agrobacterium was also grown in the presence of rifampicin and gentamycin.

2.4.3 Transformation of *E. coli*

E. coli (strain DH5α, Zymoresearch, catalogue number: T3007; or strain DH10 Beta, NEB, catalogue number: C3019H) were transformed according to manufacturer's instructions, with some slight modifications as follows: Cells were thawed on ice for 10 minutes, and had either 1 µL DNA (purified DNA) or 5 µL for a ligation mix added to 50 µL of competent cells. These were left on ice for 15-30 minutes. For DH10 Beta cells only, a heat shock was performed by suspending the cells in a 42°C water bath for 30 seconds and were immediately transferred to ice. Cells then had 250 µL of outgrowth medium added and were incubated at 37°C at 100 rpm for 1 hour. Cells were then plated onto agar plates containing appropriate antibiotics, and IPTG and X-gal if blue-white screening was used.

2.4.4 Transformation of *A. tumefaciens*

50 µL aliquots of *Agrobacterium tumefaciens* (Str. GV3101; VWR, catalogue number: 103753-234) were thawed on ice for approximately 10 minutes. 150 ng of DNA to transform was added to each aliquot. The solution was gently pipetted into a BIO-RAD Gene Pulser® Cuvette (catalogue number: 165-2086), and the cells were electroporated at 2500 volts until completed, indicated by the BIO-RAD MicroPulser™ machine. 1 mL of LB was added, and the solution was transferred to an Eppendorf tube, and incubated at 28°C for 2 hours. Cells were then plated onto agar plates containing gentamycin, rifampicin and the appropriate expression plasmid antibiotic.

2.5 – Plants

2.5.1 Antibiotic concentrations

Antibiotics were used for plant selection at the following concentrations: Kanamycin (50 µg/ml), Cefotaxime (100 µg/ml)

2.5.2 Growth of *N. benthamiana*

Nicotiana plants were grown at 25°C in soil and sand in a 3:1 ratio (Levington advance professional growing media M2 grade; Melcourt horticultural sharp sand). The soil was heat treated prior to potting at 70°C for at least 2 hours.

2.5.3 Growth of *A. thaliana*

A. thaliana seeds (Col-0 ecotype) were washed in 1 mL 70% ethanol. Seeds were briefly vortexed and centrifuged at 2,000 *x g* using a 5 second pulse in a benchtop centrifuge. The ethanol was removed, and the seeds rinsed with 1 mL Elix H_2O. Seeds were again vortexed and centrifuged using a short pulse, and the water removed. They were then rinsed in 1 mL of 10% bleach solution containing 1% triton X-100 (SigmaAldrich; catalogue number: T8787), and incubated at room temperature for 10 minutes, vortexing occasionally. Following this, seeds were then centrifuged using a short pulse and had the solution removed in a sterile fume hood using sterile pipette tips. 1 mL Elix H_2O was used to rinse the seeds, vortexed, centrifuged using a short pulse, and removed again. This water rinse was repeated twice. Seeds were then resuspended in a final 1 mL Elix H_2O, and carefully pipetted into rows onto micro agar plates (Duchefa Biochemie; catalogue number: M1002) containing appropriate antibiotic. These were then incubated in growth cabinets at 25°C in continuous light, before being transferred to soil once four leaves had formed.

2.5.4 Growth and maintenance of BY-2 tobacco cell cultures

Tobacco BY-2 cell suspensions were grown in 95 mL mLS medium (4.3 g/L MS salt mix, 30 g/L sucrose 200 mg/L KH_2PO_4, 1 mg/L Thiamine·HCl, 100 mg/L *myo*-Inositol, and 0.236 mg/L 2,4-D, pH 5.8) in a 300 mL conical flask, sub-cultured every 7 days by transferring 1.2 mL of cell suspension to fresh mLS media. Cell suspensions were grown at 27°C at 130 RPM in darkness.

2.5.5 Stable transformation of *A. thaliana* by floral dipping

Agrobacterium transformed with the plasmid of interest were grown up in 50 mL LB with appropriate antibiotic for 48 hours. Following this, they were diluted in 450 mL of fresh LB without antibiotic and grown for 6 hours. 12.5 g of sucrose was added to each solution, and they were left to grow for another 1 hour and 30 minutes. Solutions were transferred to a beaker and 250 µL of silwet L-77 was added immediately before dipping. Four-week-old *A. thaliana* plants were transformed by floral dipping in the Agrobacterium solution by gently dipping and slowly trailing flowers in the

solution. These were then gently dabbed onto a paper towel to remove excess fluid and contained in a loosely tied autoclave bag away from direct light.

2.5.6 Transient transformation of *A. thaliana* protoplasts

Transformation of *A. thaliana* protoplasts was performed according to Wu et al. (2009). Leaves were removed from five-week-old plants and sandwiched between a piece of time tape (SLS; catalogue number: DD140250J) and magic tape (Scotch; catalogue number: 7100214399), on the upper and lower epidermis of the leaf, respectively. Following this, the tape was carefully peeled apart to expose the mesophyll cells. The upper epidermis, still attached to the time tape, was left to shake at 40 rpm in enzyme solution (1% cellulase, 0.25% macerozyme, 0.4 M mannitol, 10 mM $CaCl_2$, 20 mM KCl, 0.1% BSA and 20 mM MES, pH 5.7) for 1 hour. Following this, the isolated protoplasts were centrifuged for 3 minutes at 100 × g, washed twice with 25 mL of 4°C modified W5 solution (154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, 5 mM glucose, and 2 mM MES, pH 5.7) and incubated on ice for 30 min. The protoplasts were then again centrifuged for 3 minutes at 100 *x g* and resuspended in modified MMg solution (0.4 M mannitol, 15 mM $MgCl_2$, and 4 mM MES, pH 5.7). Then, the protoplasts were mixed in 200 µL of MMg with around 30 µg of plasmid DNA at room temperature. An equal volume of a PEG solution (40% (w/v) PEG, 0.1 M $CaCl_2$, 0.2 M mannitol) was added, and the mixture incubated at room temperature for 5 minutes. Following this, 3 mL of W5 solution was gently added, the solution mixed, and protoplasts were pelleted by centrifugation at 100 *x g* for 1 min. This wash step was repeated twice, and the protoplasts were again gently resuspended in 1 mL of W5 solution. The protoplasts were then transferred to 6-well plates, covered with 1% BSA, and incubated at room temperature for 16 hours in light.

2.5.7 Stable transformation of *N. benthamiana*

Stable transformation of *N. benthamiana* plants was performed as previously decribed with slight modifications (Forestier et al. 2021). Leaves from six-week old *N. benthamiana* plants were removed and sequentially sterilised in 10% bleach, Elix H_2O, 70% ethanol, and Elix H_2O again. Leaf discs of approximately 1 cm in diameter were then cut from the leaves using a potato borer. Discs were then dipped into solutions of *Agrobacterium* transformed with the plasmid of interest, and transferred to shooting media (MS medium, 1 mg/L 6-Benzylaminopurine (BAP), 0.1 mg/L Naphthalene acetic acid (NAA), 50 mg/L Kanamycin, 100 mg/L Cefotaxime (Claforan, Duchefa C0111) micro agar plates. These were transferred to fresh shooting media fortnightly, and onto rooting media (MS medium, 50 mg/L Kanamycin, 100 mg/L Cefotaxime (Claforan, Duchefa C0111) after approximately 3 months.

Calli were then selected and transferred to fresh rooting media until plantlets developed. These plantlets were then transferred to soil to propagate.

2.5.8 Transient transformation of *N. benthamiana*

Six-week-old *N. benthamiana* plants were used for transient transformation. *Agrobacterium* containing the expression construct of interest were grown in 50 mL LB containing appropriate antibiotic in a sterile volumetric flask for 48 hours at 28°C with shaking at 200 RPM allowing aeration. Cells were transferred to a 50 mL falcon tube and pelleted by centrifugation at 10,000 *x g*, at 4°C for 20 minutes. The supernatant was discarded, and the cells were then diluted in activation buffer (10 mM $MgCl_2$, 10 mM MES 200 µM acetosyringone, pH 5.6) to an OD600 of 1.0. The activated *Agrobacterium* were left at room temperature for two hours and were then hand infiltrated into the underside of leaves using a needleless 10 mL syringe as previously described by Lindbo (2007b). Briefly, the *Agrobacterium* solution was pulled into the syringe and gently held on the abaxial surface of the leaf, maintaining gentle pressure using a digit on the adaxial surface of the leaf to form a seal. The solution was gently injected into the air spaces which can be visually monitored by watching the solution spread throughout the leaf. Where possible, leaves were saturated using this method, typically requiring at least two infiltration sites per leaf. To maximise the infiltrated area, hand-infiltrated leaves were then vacuum infiltrated by inverting the plant in activated *Agrobacterium* solution, applying a vacuum 2 minutes followed by releasing the vacuum in a controlled manner over 5-10 seconds. Leaves were analysed at 7 days post-infiltration.

2.5.9 Preparation and transformation of plant cell packs derived from BY-2 cell cultures

Plant cell packs were generated and transformed as previously described (Gengenbach et al. 2020), with some modifications. *Agrobacterium* solutions containing the plasmid to be transformed were grown in broth for two days as described above. *N. tabacum* BY-2 cell suspension cultures between 4-7 days old were concentrated by sedimentation at 1 *x g* for 30 minutes, and excess culture media was carefully removed by pipetting. Plant cell packs were produced by pipetting 300 µL of concentrated cell suspension into 96-well receiver plates using sterile 1 mL pipette tips with approximately 1 cm of the tip removed. The plate was then centrifuged at 500 *x g* for 2 minutes to remove excess media, generating the plant cell packs, porous disks of BY-2 cells. Following, 1 mL of *Agrobacterium* culture was centrifuged at 20,8000 *x g* for 1 minute, and the cell cultures resuspended in PCP infiltration buffer (0.5 g L^{-1} Murashige-Skoog type medium 146 mM sucrose, 10 mM glucose monohydrate 200 µM acetosyringone, 15 mM MES, pH 5.6) to an OD600 of 0.05-0.5.

100 µL of activated *Agrobacterium* culture was gently pipetted onto target PCPs, and the plates left at room temperature for 1 hour without an additional liquid removal step. Finally, the 96-well plates containing the transformed PCPs were inverted over a container containing 100 mL of deionised water, sealed using autoclave tape, incubated at 27°C in darkness for 4 days, and had resulting fluorescence analysed using photon imaging.

2.5.10 Functional assays in plant cell packs

Plant cell packs transformed with P2X7 constructs were incubated with either 5 µM Yo-Pro-1 or ethidium bromide ± 10 µM JNJ-47965567 for 20 minutes at room temperature. An initial image was taken in a PhotonIMAGER Optima using settings for each cationic dye (ethidium bromide excitation = 520-540 nm, emission = 590-610 nm; Yo-Pro-1 excitation = 450-470 nm, emission = 500-520 nm). Following incubation, 30 µM Bz-ATP was added to transformed PCPs to activate expressed P2X7 receptors. An image was taken immediately after activation at T0. An image was acquired every minute for 10 minutes, with a final image acquired at 20 minutes once any reactions would have terminated. Data were analysed using the M3Vision software.

2.6 – RNA and protein extractions

2.6.1 RNA extraction

RNA was extracted from single leaves of *N. benthamiana* plants at 8 weeks old, with transformed tissue collected 1 week after agroinfiltration. The leaf tissue was finely ground in liquid nitrogen, and RNA was extracted according using TRI Reagent (Zymo Research, catalogue Number R2050-1-200) using the protocol for TRIzol™ Reagent (Invitrogen, catalogue Number 15596026) as follows: 1 mL of TRI reagent was added per 50-100 mg of ground tissue in a microcentrifuge tube and the sample centrifuged for 5 minutes at 12,000 *x g* and 4°C. Following, the supernatant was transferred to a fresh microcentrifuge tube and incubated for 5 minutes at room temperature. 200 µL chloroform was added per 1 mL of TRI reagent used and the solution incubated for 2-3 minutes at room temperature, followed by a 12,000 *x g* centrifugation step for 15 minutes at 4°C. The resulting upper aqueous phase was very carefully transferred to a fresh microcentrifuge tube, without disturbing the lower or interphase layers. Next, 500 µL isopropanol was added per 1 mL of TRI reagent used in the lysis step, and incubated at room temperature for 10 minutes, followed by a 12,000 *x g* centrifugation step for 10 minutes at 4°C. The supernatant was carefully removed and the resulting RNA pellet was washed in 1 mL of 75% ethanol per 1 mL of TRI reagent used, briefly vortexed, and centrifuged at 7,500 *x g* for 5 minutes at 4°C. The supernatant was removed and the pellet air dried for 10 minutes, before being resuspended in 50 µL of RNase-free water at 65°C.

2.6.2 Total protein extraction

Protein Extraction Buffer (PEB) 1.0

N. benthamiana tissue was finely ground in liquid nitrogen and the total protein extracted in 2 volumes of PEB-1.0 (Phosphate-buffered saline and 1 x Roche cOmplete protease inhibitor) by incubation for 30 minutes at 4°C. A total sample was taken. The protein sample was centrifuged at 4,500 *x g* for 10 minutes to remove cellular debris, the supernatant taken and the pellet resuspended in 1 mL PEB-1.0 with a sample taken. Following, the pellet was resuspended in PEB-1.0 containing 1% FC-12, incubated on ice for 30 minutes and then centrifuged at 20,800 *x g* for 10 minutes. The supernatant was retained and the pellet resuspended in 1 mL PEB-1.0 containing 1% FC-12.

PEB 2.0

Benchtop method: *N. benthamiana* tissue was finely ground in liquid nitrogen and the total protein extracted in 4 volumes PEB-2.0 containing 50 mM HEPES-KOH, pH 7.6, 150 mM NaCl, 10% v/v glycerol, 0.25% FC-12, 1 mM PMSF, 1 x cOmplete protease inhibitor by incubating at 4°C for 30 minutes. A total sample was taken. The protein sample was centrifuged at 4,000 *x g* for 10 minutes at 4°C to remove cellular debris, and the pellet resuspended in 1 mL PEB 2.0 and retained for analysis. The supernatant was taken and filtered through double-layer Miracloth, then centrifuged at 22,000 *x g* for 10 minutes at 4°C. The supernatant containing the cytosolic proteins was retained, and the pellet was resuspended in 1 mL PEB-2.0 containing 1% FC-12, left on ice for 30 minutes and then centrifuged at 22,000 *x g* for 10 minutes. The supernatant of this solution was retained and is thought to contain the membranous proteins. The pellet was resuspended in extraction buffer and retained.

Ultracentrifuge method: *N. benthamiana* tissue was finely ground in liquid nitrogen and the total protein extracted in 4 volumes PEB-2.0 containing 50 mM HEPES-KOH, pH 7.6, 150 mM NaCl, 10% v/v glycerol, 0.25% FC-12, 1 mM PMSF, 1 x cOmplete protease inhibitor by incubating at 4°C for 30 minutes. A total sample was taken. The protein sample was centrifuged at 18,000 *x g* for 10 minutes at 4°C to remove cellular debris, and the pellet resuspended in 1 mL PEB 2.0 and retained for analysis. The supernatant was taken and filtered through double-layer Miracloth, then centrifuged at 100,000 *x g* for 1 hour at 4°C. The supernatant containing the cytosolic proteins was retained, and the pellet was resuspended in 1 mL PEB-2.0 containing 1% FC-12, left on ice for 30 minutes and then centrifuged at 100,000 *x g* for 30 minutes. The supernatant of this solution was retained and is thought to contain the membranous proteins. The pellet was resuspended in extraction buffer and retained.

PEB 3.0

N. benthamiana tissue was finely ground in liquid nitrogen and the total protein extracted in 4 volumes PEB-3.0 containing 50 mM Tris-HCl, 150 mM NaCl, and 10 mM EDTA, pH 8.0 by incubating at 4°C for 30 minutes. The extract solution had debris removed by centrifugation at 12,000 *x g* for 15 minutes, followed by the addition of 600 µL of 5 M NaCl and 4.66 mL of saturated ammonium sulphate. Anhydrous ethanol was added to the solution in a 1:3 v/v ratio and mixed. The sample was centrifuged at 3,000 *x g* for 5 minutes and the upper phase retained. N-butanol was added to the ethanol extract in a 1:4 v/v ratio, mixed and the centrifugation step repeated with the lower aqueous phase containing the extracted protein retained.

PEB 4.1

N. benthamiana tissue was finely ground in liquid nitrogen and the total protein extracted in 4 volumes PEB-4.1 containing 50 mM Tris-HCl, pH 7.4, 330 mM sucrose, 5 mM EDTA, 5 mM DTT and 1 x sigma plant protease inhibitor cocktail by incubating at 4°C for 30 minutes. The protein sample was centrifuged at 2,000 *x g* for 30 minutes at 4°C to remove cellular debris, the supernatant taken and filtered through double-layer Miracloth, then centrifuged at 100,000 *x g* for 1 hour at 4°C. The supernatant is considered to contain the cytosolic proteins. The pellet was resuspended in 2 mL PEB-4.1 containing 1% FC-12, left on ice for 30 minutes and then centrifuged at 100,000 *x g* for 30 minutes. The supernatant of fractionation is thought to contain membranous proteins.

PEB 4.2

N. benthamiana tissue was finely ground in liquid nitrogen and the total protein extracted in 4 volumes PEB-4.2 containing 50 mM Tris-HCl, pH 7.4, 165 mM NaCl, 5 mM DTT and 1 x sigma plant protease inhibitor cocktail by incubating at 4°C for 30 minutes. The protein sample was centrifuged at 4,000 *x g* for 30 minutes at 4°C to remove cellular debris, the supernatant taken and filtered through double-layer Miracloth, then centrifuged at 100,000 *x g* for 1 hour at 4°C. The supernatant is considered to contain the cytosolic proteins. The pellet was resuspended in 2 mL PEB-4.2 containing 1% FC-12, left on ice for 30 minutes and then centrifuged at 100,000 *x g* for 30 minutes. The supernatant of fractionation is thought to contain membranous proteins.

PEB 4.3

N. benthamiana tissue was finely ground in liquid nitrogen and the total protein extracted in 4 volumes PEB-4.3 containing 50 mM phosphate, pH 8.0, 10 mM Tris, pH 8.0, 500 mM NaCl, 0.1% Tween 20, 0.1% FC-12, and 1 x sigma plant protease inhibitor by incubating at 4°C for 30 minutes. The protein sample was centrifuged at 12,000 *x g* for 20 minutes at 4°C to remove cellular debris.

The supernatant containing cytoplasmic and membranous proteins was taken and filtered through double-layer Miracloth.

PEB 5.0

N. benthamiana tissue was finely ground in liquid nitrogen and the total protein extracted in 4 volumes PEB-5.0 containing 20 mM sodium phosphate, pH 7.5, 150 mM NaCl, 1% (v/v) Triton X-100, 5 mM DTT, 1 mM PMSF and 1 x sigma plant protease inhibitor cocktail by incubating at 4°C for 30 minutes. The protein sample was centrifuged at 4,000 *x g* for 20 minutes at 4°C to remove cellular debris. The supernatant containing cytoplasmic and membranous proteins was taken and filtered through double-layer Miracloth.

2.7 – Expression analyses

2.7.1 Statistical analyses

Statistical analyses were carried out using R version 4.2.0 (2022-04-22 ucrt) -- "Vigorous Calisthenics", or using the built in statistical analysis on GraphPad Prism 10. All data was tested for a normal distribution using a Shapiro-Wilks test, and parametric tests were used where data was normally distributed. When measuring the difference between means, a one-way ANOVA was used. In cases that the data deviated substantially from a normal distribution and an ANOVA produced a large test statistic and low R^2 value a non-parametric Kruskal-Wallis test was used instead.

2.7.2 Analysis of in planta fluorescence by confocal microscopy

Unless stated, for *A. thaliana* plants and protoplasts, an Axio Observer.Z1 from Zeiss was used to screen for eGFP fluorescence using customised bandpass filters for eGFP. Z-stacks of 50 µm in depth were taken of transformed cells to analyse the intracellular distribution of fluorescence with variable gain between experiments and pinhole size of 1 Airy Unit using each magnification. In protoplasts, cells were also stained with the red membrane stain FM™ 4-64 (ThermoFisher, catalogue number: T3166), according to manufacturer's instructions, at a working concentration of 5 µg/mL, and visualised on confocal dishes (VWR, catalogue number: 734-2905). The excitation laser used for eGFP excitation was 488 nm. A magnification of x10 or x20 was used, and the images processed using the ImageJ software, using maximum intensity Z-stacks to generate images, and adding a scale bar. Where stated in a legend, using ImageJ, brightness and contrast have been equally enhanced across images to aid viewing.

For *N. benthamiana* leaves and some *A. thaliana* leaves when stated, a Zeiss LSM 880 confocal microscope with Airyscan was used. A 1 cm^2 section of interest was excised and placed on a glass slide with a droplet of water followed by a cover slip and visualised. Confocal images were taken

using Z-stacks of 30 µm using a pinhole size of 0.8 Airy Units to generate a 1.7 µm section on a 20-times magnification. Laser excitation of eGFP was achieved using a wavelength of 488 nm with a gain of 700. Dimensions for images were X = 1024, Y = 1024. Images were processed using FIJI and maximum intensity projections of Z-stacks are shown for qualitative images. For quantitative images, Z-stacks were not combined, and instead individual images were scanned to identify the centre of the nucleus, the brightest region within transformed cells. Regions of interest (ROIs) were manually selected following the cellular perimeter and measurements were obtained to gather the area, mean expression, and integrated density. This was performed for both the nuclei of cells and the surrounding cytoplasm within only the central plane in 2D. 3D images were not obtained which would allow quantification of cellular fluorescence across all planes. Full microscopy settings for using the Zeiss LSM 880 confocal microscope can be found in the supplementary material.

2.7.3 Analysis of in planta fluorescence using photon imaging

Whole tobacco leaves were analysed using a PhotonIMAGER Optima™ by Biospace Lab by placing leaves on a baseplate at an appropriate distance from the lens. Bandpass filters used for eGFP are as follows: GFP: Background = 430 nm, Excitation = 488 nm, Emission = 530 nm. White light and fluorescence images were obtained, overlayed, the leaves selected individually using the software M3 Vision, and the resulting data exported and analysed using Microsoft Excel. Quantification was performed by measuring the number of photons emitted per cm^2 per second per steradian (ph s^{-1} cm^{-2} sr^{-1}) across the whole area of each leaf or of only transformed areas using the M3 Vision Magic Wand tool.

2.7.4 Detection of transcripts using RT-PCR

Extracted RNA had any contaminating DNA removed using the DNA-*free*™ DNA Removal Kit (Invitrogen Catalogue Number AM1906) according to manufacturer's instructions. This involved adding 5 µL of buffer and 1 µL of rDNaseI to 44 µL of RNA sample (maximum concentration possible), mixing and incubation at 37°C for 30 minutes. Following, 5.5 µL of DNase inactivation reagent was added, the solution mixed and incubated at room temperature for 2 minutes. Finally, the sample was centrifuged at 10,000 *x g* for 90 seconds, and the supernatant transferred to a fresh microcentrifuge tube. The RNA quality was then assessed by running 2 µL on an agarose gel. The now DNA-free RNA was then reverse transcribed using a ProtoScript® II First Strand cDNA Synthesis Kit (NEB, catalogue number E6560S) according to manufacturer's instructions. This involved mixing 1

µg of RNA with 2 µL of $d(T)_{23}$ VN primers or random primers, 10 µL of reaction mix and 2 µL of enzyme mix, with nuclease free water added to a total volume of 20 µL. If using random primers, a preincubation at 25°C for 5 minutes was performed. This solution was then incubated at 42°C for 1 hour, followed by denaturation at 80°C for 5 minutes. The resulting cDNA was diluted to a total volume of 100 µL in nuclease free water. cDNA was analysed by PCR using 0.5 µM of each forward and reverse primer, 1 x PCRBIO Taq Mix Red (PCRBiosystems, catalogue number: PB10.13-02), 4.9µL of Milli-Q H_2O and 0.1 µL of cDNA, alongside a no template control. Thermocycling conditions used were 95°C for 2 minutes, then 35 cycles of 95°C denaturation for 10 seconds, 55°C annealing for 10 seconds, and 72°C amplification for 20 seconds per kilobase of amplicon DNA, followed by a final amplification stage of 72°C for 2 minutes. Amplicons were analysed by agarose gel electrophoresis using a 2.5% agarose gel for amplicons <500 bp in length, and a 0.8% agarose gel for amplicons >500 bp in length.

2.7.5 Quantification of protein concentration by Bradford assay and fluorescence

The concentration of protein in samples was analysed using a Bio-Rad protein assay. To do this, 15 µL of sample (either pure or diluted 1:10, sample: Milli-Q H_2O) or BSA standard (500, 250, 125, 67.5, and 31.25 µg/ml) were gently mixed with 200 µL of 1 x Bio-Rad Protein Assay solution (Bio-Rad, Catalogue number: 5000006) in a transparent 96-well plate. The plate then had absorbance read using a Clariostar Plate Reader. A BSA standard curve was generated, and the sample protein concentrations plotted against it. Statistical analyses were carried out to identify whether the difference in fluorescence was significant to the BSA standards.

2.7.6 SDS-PAGE

Hand-cast 8-12% acrylamide SDS-PAGE gels were used to separate protein samples, ran on 200V for 1-2 hours. Alternatively, NuPAGE™ 4 to 12%, Bis-Tris gels were used, and ran at 200V for 50 minutes. The samples were denatured prior to gel loading in 6x SDS loading buffer (0.375 M Tris pH 6.8, 12% SDS, 60% glycerol, 0.6 M DTT, 0.06% bromophenol blue; 1:5 v/v) by incubating at 100°C for 2 minutes.

2.7.7 Coomassie blue staining of protein gels

Total protein staining was carried out by adding 10 ml of InstantBlue™ into a square plastic tray containing the SDS-PAGE gel and gently rocked for 15 minutes. Following this, the gel was destained

with three 5-minute washes in Milli-Q H_2O and photographed under a transilluminator with white light.

2.7.8 Quantification of extracted GFP by western blot

To western blot, the SDS-PAGE gel was carefully transferred to a mini nitrocellulose membrane (ThermoFisher, catalogue number: LC2000) and put into a Trans-Blot® Turbo™ Transfer System (BIO-RAD, catalogue number 1704150) on the 7-minute turbo setting to transfer the proteins to the membrane. The membrane was then blocked in Intercept® (TBS) Blocking Buffer (LI-COR, catalogue number 927-60001) by rotating in 10 mL of the solution at room temperature for 1 hour. Following this, the membrane was left in primary antibody solution (10 mL Blocking buffer containing 2% Tween (v/v) and primary antibody (Anti-GFP (Plant Specific) Antibody; antibodies.com, catalogue number A50024) at a concentration 1/5000 v/v) overnight at 4°C. The next day, the membrane was washed 4 times in 10 mL TBS containing 2% v/v Tween, each for 5 minutes rocking at room temperature. Following, the secondary antibody solution was added (10 mL Blocking buffer containing 2% Tween (v/v) and secondary antibody (Goat anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ Plus 800; Invitrogen, catalogue number A32730) at a concentration of 1/13500 v/v), and incubated, rocking, for an hour at room temperature. The membrane was then washed four times again in TBS-tween solution and visualised using an Odyssey® DLx Imaging System.

2.8 – Protein purification

2.8.1 His-pulldown using Ni-Sepharose beads with PEB 4.2

Approximately 1 mL of Ni-Sepharose beads slurry (SigmaAldrich, Catalogue number: GE17-5268-01) was pelleted using a 10 second 2,000 *x g* centrifugation and the supernatant removed and replaced with 1 mL solubilisation buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 1% (w/v) dodecyl maltoside, 1 x Roche cOmplete protease inhibitor) and gently mixed. 100 µL of bead suspension was added to total protein extracts with 25 mM imidazole added and the sample rotated overnight at 4°C. Following, the beads within the protein extract were pelleted with a 2,000 *x g* centrifugation for 2 minutes at room temperature. The supernatant was either discarded or optionally retained for analysis. The bead pellet was washed six times through the addition of 500 µl wash buffer (Solubilisation buffer with 25 mM imidazole) followed by mixing through inversion and a 10 second 2,000 *x g* centrifugation, with the supernatant removed after each wash. Finally, bound protein was eluted by adding 100 µL of elution buffer (Solubilisation buffer with 500 mM imidazole)

and incubating for 5 minutes at room temperature. This elution was optionally performed twice. Eluted protein extract could then be used for downstream analysis by SDS-PAGE.

2.8.2 His-pulldown using Ni-Sepharose beads with PEB 5.0

Approximately 1 mL of Ni-Sepharose beads slurry was pelleted using a 10 second 2,000 *x g* centrifugation and the supernatant removed and replaced with 1 mL solubilisation buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 1% (v/v) Triton X-100) and gently mixed. Protein extracts were mixed 1:1 with 2 x binding buffer (20 mM sodium phosphate, pH 7.4, 925 mM NaCl, 40 mM imidazole). 100 µL of bead suspension was added to total protein extracts and the sample rotated overnight at 4°C. Following, the beads within the protein extract were pelleted with a 2,000 *x g* centrifugation for 2 minutes at room temperature. The supernatant was either discarded or optionally retained for analysis. The bead pellet was washed six times through the addition of 500 µl wash buffer (solubilisation buffer with 20 mM imidazole) followed by mixing through inversion and a 10 second 2,000 *x g* centrifugation, with the supernatant removed after each wash. Finally, bound protein was eluted by adding 100 µL of elution buffer (solubilisation buffer with 500 mM imidazole) and incubating for 5 minutes at room temperature. This elution was optionally performed twice. Eluted protein extract could then be used for downstream analysis by SDS-PAGE.

2.8.3 His-pulldown using Ni-NTA magnetic beads in PEB 5.0

Approximately 1 mL of HisPur™ Ni-NTA beads (ThermoFisher, Catalogue number: 88831) in slurry were pelleted using a magnetic stand and the supernatant was removed and replaced with 1 mL solubilisation buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 1% (v/v) Triton X-100) and gently mixed. Protein extracts were mixed 1:1 with 2 x binding buffer (20 mM sodium phosphate, pH 7.4, 925 mM NaCl, 40 mM imidazole). 100 µL of bead suspension was added to total protein extracts and the sample rotated overnight at 4°C. Following, the beads within the protein extract were pelleted with a 2,000 *x g* centrifugation for 2 minutes at room temperature. The supernatant was either discarded or optionally retained for analysis. The bead pellet was resuspended in 1 mL wash buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 20 mM imidazole) and transferred to a fresh microtube held in a magnetic stand. Three washes were performed by collecting the beads against the magnet, removing the supernatant and replacing it with 1 mL wash buffer, gently mixing through inversion, and repeating the process. After three washes the clear supernatant was removed, and the bound protein was eluted by adding 100 µL of elution buffer (20 mM sodium

phosphate, pH 7.4, 500 mM NaCl, 500 mM imidazole) and incubated for 10 minutes at room temperature. Eluted protein extract could then be used for downstream analysis by SDS-PAGE.

2.8.4 Manual HisTrapHP column purification

Total protein extracts were mixed 1:1 with 2 x binding buffer and sequentially filtered through a 0.45 µM then 0.2 µM membrane. Using a needless syringe, the HisTrapHP column (Cytiva, Catalogue number: 29051021) was equilibrated with 10 column volumes (CVs) of equilibration buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 20 mM imidazole). The sample was loaded onto the column and the column washed with 5 CVs equilibration buffer. Following, the protein was eluted in using 3 CVs of elution buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 500 mM imidazole) and collected into microtubes. The column was then reset with 5 CVs of elution buffer followed by 10 CVs of equilibration buffer. The column was then washed with 10 CVs 20% ethanol and stored in this solution.

2.8.5 AKTA HisTrapHP column purification

AKTA purification was performed using either an AKTA Prime or AKTA Start. In both cases, the flowrate used was either 1 or 4 column volumes (CVs) per minute. Column equilibration was performed using 5 CVs of equilibration buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 20 mM imidazole), followed by sample loading, and equilibration using 15 CVs of equilibration buffer. Following this, a gradient elution was performed using 20 CVs ranging from 0-100% elution buffer; 1 mL fractions were collected. The column was then equilibrated with 5 CVs of elution buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 500 mM imidazole) followed by 5 CVs of equilibration buffer. Finally, 5 CVs of 20% ethanol was passed through the column for cleaning and storage.

2.8.6 HPB-tag

Purification of the HPB-tagged eGFP protein was performed as previously described (Qi and Katagiri 2011) with some minor modifications as follows: Tissue expressing eGFP tagged with HPB was ground and had total protein extracted and fractionated using the PEB 4.2 method. 200 µL of Dynabeads® M-280 Streptavidin magnetic beads (ThermoFisher, Catalogue number: 11205D) were washed three times in PEB 4.2 into a final volume of 200 µL. 100 µL of the bead solution was then added to the total protein extract expressing the HPB-tagged eGFP and rotated at 30 RPM for 2 hours to allow binding. Following this, the beads were collected using a magnetic stand and the supernatant was removed. The beads were washed three times with RIPA buffer 1 (50 mM Tris-HCl,

150 mM NaCl, 1 mM EDTA, pH 7.4, 0.5% sodium-deoxycholate), with gentle vortexing for 10 seconds in between each wash followed by magnetic pulldown. Next, RIPA buffer 2 (50 mM Tris-HCl, 20 mM NaCl, 1 mM EDTA, pH 7.4, 0.5% sodium-deoxycholate) was used for three washes, and the beads resuspended with RIPA buffer 2 to a final volume of 40 µL. 40 µL of 2 x SDS sample buffer was added to the solution, mixed, and the solution heated at 100°C for 10 minutes to denature and liberate bound proteins. This solution was then analysed by SDS-PAGE and western blotting.

2.8.7 CBD-tag

Purification of the CBD-tagged eGFP variant was performed as previously described (Islam et al. 2018) with some slight modifications as follows: Tissue expressing eGFP tagged with CBD was ground and had total protein extracted and fractionated using the PEB 4.2 method. 500 mg of microcrystalline cellulose (SigmaAldrich, Catalogue number: 435236) was washed in four bead volumes of water and the fine particles removed following sedimentation. This wash was repeated five times until the supernatant was clear. The cellulose beads were mixed with water at 1:1 ratio (v/v) and added to the protein extract containing the CBD-tagged eGFP, and incubated for 1 hour with gentle rotation at 60 RPM at 4°C. Following this, beads were centrifuged at 2,000 *xg* for 2 minutes, the supernatant removed and the beads collected. The beads were washed in four bead volumes of wash buffer (Tris-HCl, pH 7.5) to remove loosely bound proteins. Finally, the protein was liberated through denaturation in 2 x SDS-PAGE buffer (100 mm Tris-HCl, pH 6.8, 4% (w/v) SDS, 0.1% (w/v) bromophenol blue, 20% (v/v) glycerol 200 mM β-mercaptoethanol) at 100°C for 10 minutes. This solution was then analysed by SDS-PAGE and western blotting.

2.8.8 FLAG-tag (Fab-Trap)

Purification of eGFP with a FLAG-tag was performed using a Fab-Trap according to manufacturer's instructions with some minor buffer modifications (DYKDDDDK Fab-Trap; ChromoTek, Catalogue number: AB_2894836). The Fab-Trap beads were resuspended by gentle pipetting and 25 µL of bead slurry was transferred into a 1.5 mL reaction tube, with 500 µL ice-cold dilution buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 80% v/v dH_2O added immediately prior to use) added. The beads were sedimented by centrifugation at 2,500 *x g* for 5 minutes at 4°C, and the supernatant discarded. Tissue expressing eGFP with the FLAG-tag was ground and had total protein extracted and fractionated using the PEB 4.2 method and was mixed with dilution buffer in a 2:3 v/v ratio. The 500 µL protein extract and beads were mixed, and the solution rotated at 4 °C for 1 hour to allow protein binding. Following incubation, the beads were sedimented by centrifugation at 2,500 *x g* for 5 min at

4°C, and the supernatant discarded. The beads containing bound protein were washed by resuspension in 500 µL wash buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05 % Sodium deoxycholate, 0.5 mM EDTA, 80% v/v dH_2O added prior to use), collected by centrifugation, and the supernatant discarded. This wash was repeated twice more, the beads transferred to a fresh tube on the final wash step and the remaining supernatant removed. Elution was achieved by resuspending the beads in 80 µL of 2x SDS-sample buffer (100 mM Tris-HCl, pH 6.8, 4% (w/v) SDS, 0.1% (w/v) bromophenol blue, 20% (v/v) glycerol 200 mM β-mercaptoethanol), boiling the beads for 5 min at 100°C followed by centrifugation at 2,500 *x g* for 2 min at 4°C. This supernatant was then analysed by SDS-PAGE and western blotting.

2.8.9 Anti-GFP antibody (GFP-Trap)

Purification of eGFP using a GFP-Trap was done according to manufacturer's instructions with some minor buffer modifications (GFP-Trap Magnetic Particles M-270; ChromoTek, Catalogue number: AB_2827592). The GFP-Trap beads were resuspended by gentle pipetting and 25 µL of bead slurry was transferred into a 1.5 mL reaction tube, with 500 µL ice-cold dilution buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 0.018% sodium azide) added. The beads were collected using a magnetic stand, and the supernatant discarded. Tissue expressing eGFP with the FLAG-tag was ground and had total protein extracted and fractionated using the PEB 4.2 method and was mixed with dilution buffer in a 2:3 v/v ratio. The 500 µL protein extract and beads were mixed, and the solution rotated at 4 °C for 1 hour to allow protein binding. Following incubation, the beads were collected using a magnetic stand, and the supernatant discarded. The beads containing bound protein were washed by resuspension in 500 µL wash buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05 % Sodium deoxycholate, 0.5 mM EDTA), collected using a magnetic stand, and the supernatant discarded. This wash was repeated twice more, the beads transferred to a fresh tube on the final wash step and the remaining supernatant removed. Elution was achieved by resuspending the beads in 80 µL of 2 x SDS-sample buffer (100 mM Tris-HCl, pH 6.8, 4% (w/v) SDS, 0.1% (w/v) bromophenol blue, 20% (v/v) glycerol 200 mM β-mercaptoethanol), boiling the beads for 5 min at 100°C followed by bead collection using a magnetic stand. The supernatant was then analysed by SDS-PAGE and western blotting.

2.9 Protein concentration using VivaSpin columns

Purified protein was concentrated further using Vivaspin 500 centrifugal concentrator columns (Merck, catalogue number: Z629367) with a MWCO of 3000 daltons. Centrifugation steps were performed at 15,000 *x g* until the sample volume reached 25-30 µL or up to 24 hours.

Chapter 3 – Improving Transgene Expression via Construct Design

3.1 – Chapter Summary

The aim of this chapter was to develop expression constructs that generate improved protein yields *in planta*, relative to minimal early constructs, containing only a promoter, coding sequence, and terminator. The work documented in this chapter details the creation of several enhanced plant expression constructs that produce high yields of desired protein. Using eGFP as a readout, enabling quantification of protein production *in planta*, and using primers specifically designed to quantify transcription, several different expression constructs were created and compared. These constructs were created and tested in three distinct phases, based on the results of the previous phase.

Phase 1 involved the creation of non-replicating constructs; a simple (S) construct containing only CaMV 35S promoter, eGFP coding sequence, and nopaline synthase terminator (35S::eGFP:NosT) in a Level 1 non-replicating Golden Gate vector (N), referred to as pNS-eGFP; and a complex (C) expression construct composed of the CaMV 35S promoter, a synthetic 5' UTR, eGFP coding sequence, CPMV 3' UTR, extensin and Actin terminator, and a RB7 matrix attachment region in a Level 1 non-replicating Golden Gate vector, referred to as pNC-eGFP.

In an attempt to further improve expression, Phase 2 involved conventional cloning these constructs into a TMV-based RNA replicating vector (R), pJL-TRBO (Lindbo 2007b), and testing this against the original. These are the same regulatory elements in the replicating vector and so are termed pRS and pRC for the simple and complex collection of DNA elements, respectively. The use of the replicating vector dramatically improved eGFP expression, particularly with pRC-eGFP.

Many of the regulatory elements incorporated into pRC affect transcription which should not affect the replicating RNA in TMV vectors. However, as the complex construct did function in the replicating vector, it was hypothesised that the UTRs were most likely responsible for the increase in expression as these function at the RNA-level, and so are the most likely DNA element to affect expression within the RNA-replicating vector. Consequently, Phase 3 constructs were made that incorporated different combinations of UTRs in an attempt to improve expression further and elucidate the relative effects of the UTRs compared to the other regulatory elements in pRC. These included expressing the isolated eGFP coding sequence (pRI-eGFP) with no additional DNA elements, with both 5' and 3' UTRs (pRU-eGFP), or with only the 5' (pR5-eGFP) or 3' (pR3-eGFP) UTRs.

Upon testing each of these constructs by measuring *in planta* eGFP fluorescence, the best expression construct was determined to be pRC-eGFP, with up to 199-fold higher fluorescence than the initial pNS-eGFP construct, and 2.2-fold the fluorescence of pRI-eGFP, analogous to the previously

published pJL-TRBO vector. Overall, this chapter documents a substantial genetic engineering and cloning effort to create a plant expression construct that results in increased recombinant protein expression relative to starting expression constructs.

3.2 – Introduction

Recombinant protein expression is typically carried out in prokaryotic expression systems, utilising *E. coli*-based expression where possible, due to the high yields, low cost and ease of use of the expression system (Tripathi and Shrivastava 2019). However, prokaryotes are unable to produce all post-translational modifications, meaning that they are unable to produce all recombinant proteins. When a target protein is subject to PTMs, eukaryotic expression systems are used, which are typically expensive to scale up (He et al. 2014). This is particularly a problem for proteins with a low intracellular abundance like membrane proteins, so there is currently no optimal expression system for modified membrane proteins, as prokaryotes cannot produce the PTMs, and eukaryotic cell cultures are too costly to scale. Here, plant expression systems are desirable as they can produce PTMs and are considered the cheapest expression system to scale (De Martinis et al. 2022). However, plant expression systems typically have low yields relative to conventional expression systems (Schillberg et al. 2019). A viable solution to this problem may be to generate expression constructs that maximise the yields of recombinant proteins.

Recently there has been increasing success in improving the yields of plant-made recombinant proteins through construct design, particularly using combinatorial or synthetic biology approaches to design novel DNA elements including promoters, untranslated regions (UTRs), terminators and matrix attachment regions (MARs), and the use of replicating vectors. However, there has not been a single publication that has combined all these factors, and typical plant expression constructs consist of only a promoter, coding sequence and terminator, far from maximising protein yields.

Here, using modular cloning, several expression constructs were created that combine various published genetic elements, testing the effectiveness of combining promoters, UTRs, terminators, and matrix attachment regions, and assessing their effectiveness within non-replicating and replicating vectors. These DNA elements can be simplified into groups. Promoters, terminators, and matrix attachment regions primarily affect transcription, influencing the number of RNA molecules produced from a transcriptional unit. Conversely, UTRs primarily affect translational efficiency and RNA stability. All of these are investigated in this research as optimising transcriptional and translational efficiency as well as transcript stability are essential for maximising target protein expression. Viral-derived replicating vectors also exist, which enable amplification of target RNA molecules and systemic spread between plant cells, maximising the number of affected cells and the

number of available transcripts. It is likely that DNA elements that influence transcription provide no benefit within replicated RNA molecules and may even be deleterious to expression levels. However, due to the modular cloning approach utilised in this research, some expression constructs contain promoters, terminators, and MARs within the replicated RNA. Consequently, several other expression constructs were created that lack these additional DNA elements to assess whether they are incompatible for use within replicating vectors.

Golden Gate cloning was chosen as the major method for construct creation as it allows for the assembly of multiple DNA fragments in a predefined linear order within a single reaction. Using Type IIS restriction enzymes, shown in Figure 3.1, which cut DNA outside of their recognition sites, and T4 DNA ligase, which ligates the DNA fragments, cloning reactions can be performed cyclically, with each cycle of digestion and ligation creating irreversibly ligated DNA that lacks the restriction enzyme recognition site. This method has a high efficiency and flexibility, enabling the creation of large constructs with customised sticky ends, designed by the researcher. It is rapid and scalable, with digestion and ligation performed in a single reaction tube, and the customised sticky ends enable scarless cloning where no additional nucleotides are left between assembled DNA fragments. Using this approach, large constructs can be created using genetic elements in the order seen in Figure 3.2. This also enables each DNA element to be put into standardised positions for use in Golden Gate cloning, in a process known as domestication. Domestication refers to the process of adapting a DNA sequence for compatibility with Golden Gate cloning and involves modifying the DNA sequence to remove internal Type IIS restriction enzyme recognition sites that could interfere with the intended assembly process, and the addition of appropriate restriction sites flanking the DNA fragment to be cloned, ensuring linear insertion into the destination vector, and the creation of expression constructs. Importantly, the native Golden Gate acceptor vectors contain a LacZ coding sequence which gets replaced by the designed DNA insert, enabling blue/white selection on agar plates containing IPTG and x-gal. This makes the screening process simple, as blue colonies are unsuccessful clones, and white colonies are worth investigating further as the LacZ coding sequence has been replaced by the DNA sequence of your choosing.

The replicating vector used in this research was not domesticated for Golden Gate cloning, so once a full expression cassette containing all the DNA elements was created, it was conventionally cloned using AvrII and PacI restriction enzymes into the replicating vector. This approach enables rapid construction of expression cassettes followed by simple digestion from the Golden Gate vectors and ligation into target replicating vectors. The result is several Golden Gate domesticated regulatory elements that can be used in future research, and the creation of several expression constructs combining a variety of these DNA elements which are tested in this chapter.

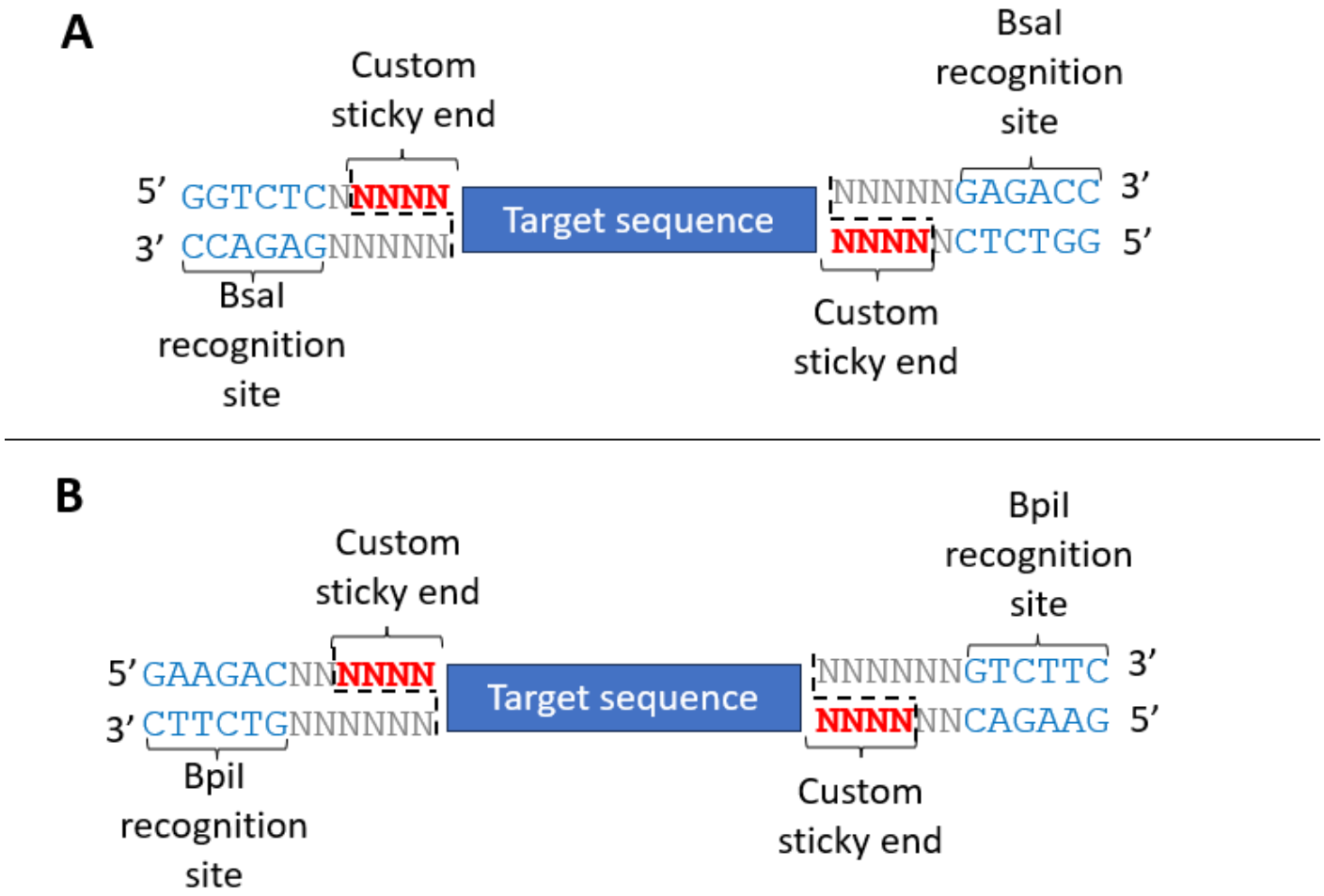


Figure 3.1: Type IIS restriction enzymes used for Golden Gate cloning. Both BsaI (A) and BpiI (B) are type IIS restriction enzymes that cut downstream of their recognition site (blue) creating a custom sticky end (red) that enables a single enzyme to be used to digest and later ligate multiple different sticky ends. The recognition site can be excluded from the ligated product, meaning cycles of digesting and ligating can be performed in a single reaction to maximise cloning efficiency.

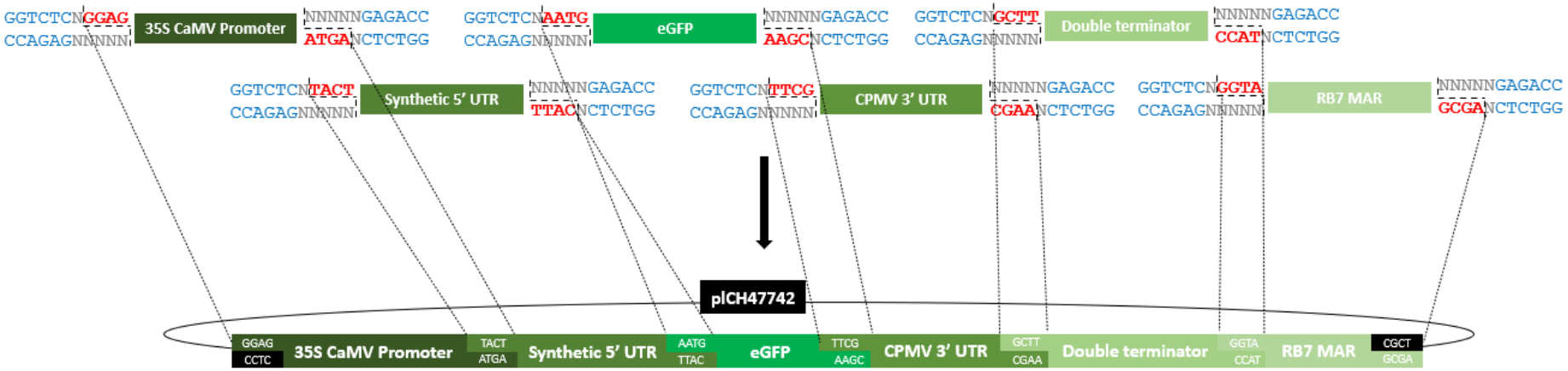


Figure 3.2: Schematic showing the Golden Gate assembly of the pNC expression vector. The type IIS restriction enzyme, BsaI, has a 4 bp cut site (red) downstream of its recognition site (blue), enabling the creation of custom sticky ends. This enabled the creation of the pNC construct containing (5' to 3') the 35S CaMV promoter, a synthetic 5' UTR (Peyret et al. 2019), eGFP coding sequence, the CPMV 3' UTR (Peyret et al. 2019), *N. tabacum* Extensin and *N. benthamiana* Actin double terminator (Diamos & Mason 2018), and RB7 matrix attachment region (Diamos & Mason 2018) in the level 1 Golden Gate vector, pICH47742.

The major DNA elements utilised here are the 35S CaMV promoter, a synthetic 5' UTR (Peyret et al. 2019), eGFP coding sequence, CPMV 3' UTR (Peyret et al. 2019), *N. benthamiana* Actin and *N. tabacum* extensin double terminator and RB7 MAR (Diamos and Mason 2018). The DNA element that influenced expression the most was found to be the 5' UTR, producing high levels of per cell protein expression when utilised in pR5-eGFP. However, the expression construct with the highest overall expression was found to be pRC, which contained all the regulatory elements within the replicating vector backbone, suggesting that the promoter, terminators, and MAR play a role within the replicating vector. Overall, this chapter shows that combining DNA elements can significantly improve recombinant protein expression in plants, particularly when combined with a replicating vector.

3.3 – Results

3.3.1 Design, cloning, screening, and experimental analyses of initial constructs (Phase 1)

The overall aim of this research was to produce a plant-based expression system for the production of complex mammalian membrane proteins, focussing on P2X7 as a target. As such, early expression constructs were created utilising a simple transcriptional unit to drive the expression of different C-terminally eGFP-tagged P2X7 isoforms from human and rats. Three isoforms from each mammal were created and tested, which are discussed in detail in Chapter 5; these were the full-length protein, a truncated variant missing the transmembrane domain but containing the cysteine-rich palmitoylated region, and a further truncated version missing both the membrane domain and the cysteine-rich motif. Specifically, these utilised minimal expression constructs with the CaMV 35S promoter, the P2X7 coding sequence, and a nopaline synthase terminator, chosen as the parts are well characterised in the literature. These expression constructs are explored later in Chapter 5 but were the main reason for designing more powerful genetic expression constructs, as it was clear that the expression from these constructs was too low for a viable plant protein expression system. Consequently, using eGFP as the coding sequence to measure expression, several expression constructs were made in an effort to improve transgene expression.

3.3.1.1 Selection of DNA elements

In order to design the enhanced expression constructs, DNA elements shown to produce high levels of protein expression in the literature were selected. For the promoter, it was decided that the constitutive CaMV 35S promoter would be used due to its extensive characterisation and high expression, although there was some evidence of improved promoter targets with less characterisation. For the UTRs, a synthetic 5' UTR, derived from CPMV-HT (Peyret et al. 2019), and the native CPMV RNA-2 3' UTR were the best candidates. For the terminators and MAR, one recent study tested several terminators and matrix attachment regions in *N. benthamiana* plants and identified the most powerful combinations of these to incorporate into the expression constructs (Diamos and Mason 2018). Importantly, whilst each of these DNA elements had been demonstrated to improve transgene expression, none of the DNA elements from different research had been combined, which could result in further improved protein expression.

3.3.1.2 Design, cloning and screening of Phase 1 constructs

To achieve this, two expression constructs were designed and created using Golden Gate cloning. The first and simplest construct, pNS-eGFP, consisted of the core CaMV 35S promoter with its upstream DNA elements, eGFP coding sequence, and nopaline synthase terminator, comparable to the initial P2X7 constructs. The second more complex construct, pNC-eGFP, contained the core

CaMV 35S promoter with its upstream DNA elements, synthetic 5’ UTR, eGFP coding sequence, CPMV 3’ UTR, an extensin terminator and Actin terminator from *N. tabacum* and *N. benthamiana*, respectively, and a RB7 matrix attachment region. These constructs are shown diagrammatically in Figure 3.3, and were assembled using Golden Gate cloning, using the restriction enzymes BsaI and BpiI which can be seen in Figure 3.1 above.

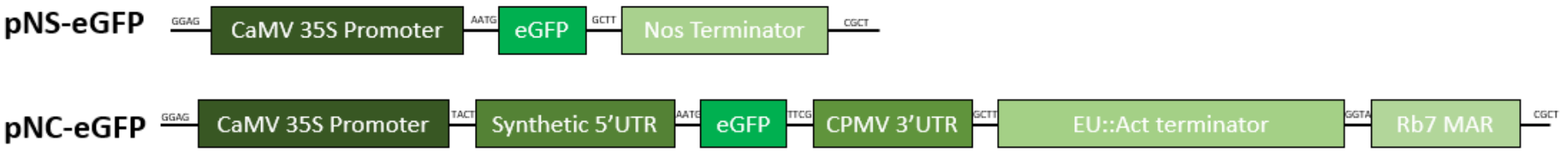


Figure 3.3: The Phase 1 constructs used in this research. pNS-eGFP = Simple construct containing (5' to 3') the CaMV 35S promoter, eGFP coding sequence, and nopaline synthase terminator in a non-replicating vector, pICH47732. pNC-eGFP = Complex construct containing (5' to 3') the CaMV 35S promoter, a synthetic 5' UTR (Peyret et al. 2019), eGFP coding sequence, the CPMV 3' UTR (Peyret et al. 2019), *N. tabacum* Extensin and *N. benthamiana* Actin double terminator (Diamos & Mason 2018), and RB7 matrix attachment region (Diamos & Mason 2018) in a non-replicating vector. All constructs shown are 5' to 3', left to right.

The initial Level 1 expression vector was a non-replicating (N) Golden Gate vector, pICH47742, shown in Figure 2.1, optimally designed for plant transformation, containing the left and right borders, which are DNA elements needed for *Agrobacterium* mediated transformation either side of the transcriptional unit.

Broadly, the cloning pipeline required the following steps:

1. Synthesis or amplification of insert with the appropriate overhangs containing the type IIS restriction enzyme recognition sites and sticky ends
2. Digestion and ligation of the insert into the target Golden Gate vector
3. Transformation of the reaction mix into *E. coli* and screening of white colonies by colony PCR
4. Plasmid DNA isolation and DNA sequencing
5. Repeat steps 2-4 until the target Level 1 construct has been created
6. Transform into *Agrobacterium* and test expression constructs in plants

In order to create these parts, each needed to be assembled into the Level 1 transcriptional units. Golden Gate cloning works on a modular basis, where custom sticky ends can be created to ligate DNA sequences into target vectors. The existing system utilises a series of Level 0 vectors, each designed with their own unique 4bp overhangs. This means that the promoter elements, UTRs, coding sequence, terminators, and matrix attachment modules could be assembled into their own Level 0 modules provided the inserts were designed or amplified to contain the relevant restriction enzyme recognition site and the appropriate sticky end. These level 0 modules can then be assembled into a fully functional Level 1 transcriptional unit. An example of this is shown diagrammatically in Figure 3.2 above, with the sticky ends for each part also shown. The detailed cloning procedure is documented in Section 2.3.7.

3.3.1.3 Qualitative in planta analyses of Phase 1 constructs

To assess the expression levels of each of the eGFP constructs, *N. benthamiana* leaves were transformed by hand-infiltration using a standardised bacterial OD600 of 1.0 and saturating the leaves to control the amount of *Agrobacterium* being infiltrated. After seven days post-infiltration (DPI), the leaves were analysed using photon imaging to check for eGFP expression, and a section of each leaf was taken to be studied by confocal microscopy. This showed a clear difference between expression constructs. Figure 3.4, Panel A shows photographs of whole leaves when eGFP is excited using photon imaging, and Panel B shows z-stacks taken on the confocal microscope. The same figure is shown in Figure S3.1 where confocal images have been digitally enhanced to increase visibility. Several clear differences between constructs are observed. Most notably, when eGFP expression is driven by the simple expression constructs, there is a low amount of eGFP fluorescence in each cell, with fluorescence mostly localising to the outer edges of the cell in the cytoplasm. Instead, when comparing eGFP expression driven by the complex expression construct, we see that the vast majority of each cell is fluorescent, including visible cytoplasmic strands within the central vacuolar space of each cell. Thus, qualitative analysis suggests that cell-to-cell expression is higher when eGFP is expressed by the complex constructs.

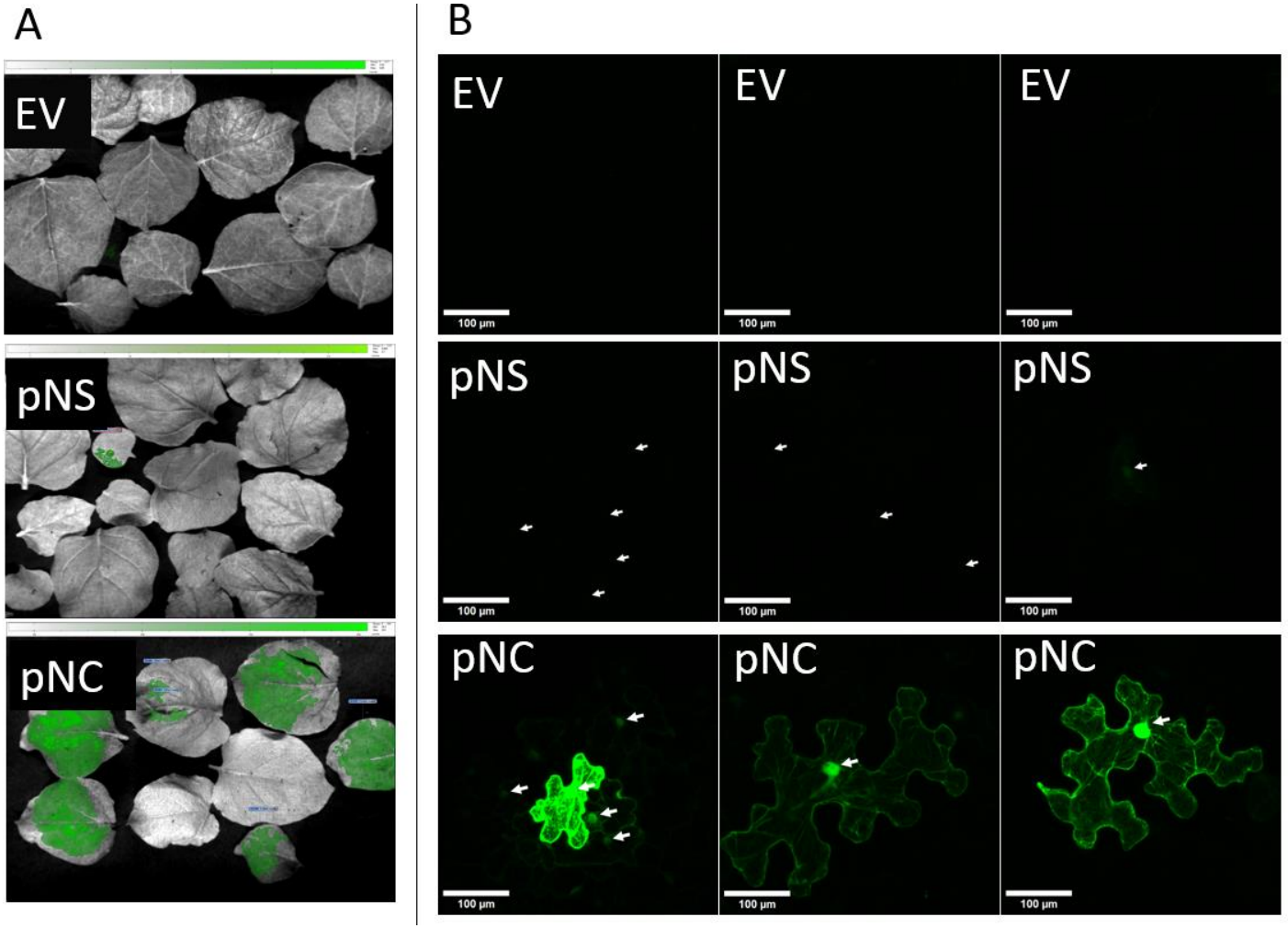


Figure 3.4 – Qualitative analyses of Phase 1 constructs. A - Photon Imager photographs of whole *N. benthamiana* leaves transformed with either empty vector, pNS-eGFP or pNC-eGFP constructs, viewed 7 DPI. Regions with expression are highlighted using the magic wand function on M3 Vision (shown by dark green highlighting). **B** – Confocal microscopy images showing 30 µm Z-stacks of transformed areas using a gain of 700 and an excitation wavelength of 488nm; scale bars represent 100 µm. White arrows indicate nuclei. Both pNS-eGFP and pNC-eGFP showed detectable levels of fluorescence in transformed cells, though expression levels are barely visible using these settings in pNS-eGFP. The same figure can be seen enhanced in Figure S3.1 where brightness and contrast have been equally enhanced using ImageJ.

3.3.1.4 Quantitative in planta analyses of Phase 1 constructs

To quantify the eGFP expression between each transformation method and expression construct, photon imaging was used as it is able to count the photons emitted from a scanned area. Three plants were transformed with either pNC-eGFP or pNS-eGFP, and transformed leaves had fluorescence quantified. Comparison of these initial constructs (Figure 3.5, Panel A) showed a mean fluorescence increase of 1.4-fold with leaves transformed with pNC-eGFP relative to pNS-eGFP, a statistically significant result when analysed using a one-way ANOVA. Interestingly, although the 35S CaMV promoter is widely used due to its high levels of expression, only low-level fluorescence was seen in this experiment when using pNS-eGFP. As a result, only pNC-eGFP produced significantly higher fluorescence than the empty vector control which was used to transform six plants. It is clear that most of the leaf remains untransformed using these constructs, so confocal images were used to quantify the expression of individually transformed cells (Figure 3.5, Panel B). Importantly, differences in expression using quantification of cellular fluorescence in this way cannot be wholly attributed to the construct, as transient transformants have variability in the number of T-DNA insertions between cells which is not accounted for. This limitation is explained in detail at the end of this chapter. Nonetheless, measuring the expression of cells at several infiltration sites for each construct can account for this variability by assuming that the number of T-DNA insertions will have a similar variability between constructs. For this reason, all eGFP-expressing cells around 10 infiltration sites had fluorescence measured. Quantification of transformed cells showed that pNC-eGFP produced significantly higher fluorescence than pNS-eGFP, at 20.1-fold. Both constructs produced significantly higher fluorescence than the empty vector. As the transformed area was relatively limited using this vector, it was decided that efforts needed to be made to improve the area of transformation, explored in Phase 2.

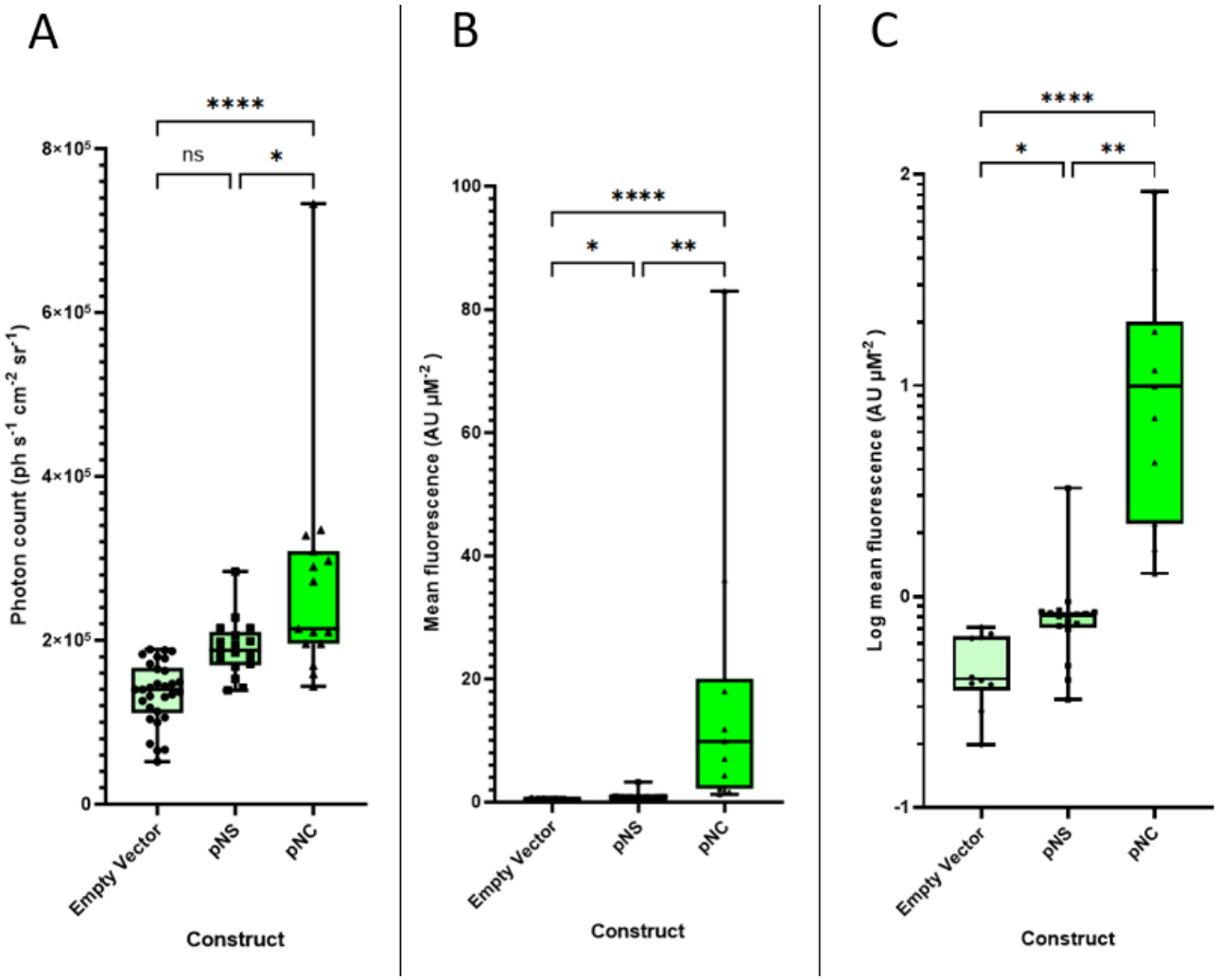


Figure 3.5 – Quantitative analyses of Phase 1 constructs. *N. benthamiana* plants were transformed with either empty vector (n=6), pNS-eGFP (n=3) or pNC-eGFP (n=3) and 5-6 transformed leaves from each plant had fluorescence measured 7 DPI. **A** – eGFP fluorescence quantified using a Biospace Optima photon imager. A one way ANOVA shows a significant difference between means ($F_{(2, 59)}$ = 15.67, p < 0.0001). A Tukey multiple comparisons test show that pNS-eGFP does not have significantly higher fluorescence than the negative control, but pNC-eGFP does (p = 0.0538 and p < 0.0001, respectively), and pNC-eGFP expresses significantly higher than pNS-eGFP (p = 0.0126). **B** – ImageJ quantification of per cell eGFP fluorescence within 10-17 cells transformed using each construct showing raw data on a linear scale. Panel **C** shows the same data on a log10 scale; A Kruskal-Wallis test shows a significant difference between means ($H_{(3,38)}$= 27.90, p < 0.0001). A Dunn's multiple comparisons test show that pNS-eGFP and pNC-eGFP have significantly higher fluorescence than the negative control (p = 0.0438 and p < 0.0001, respectively) and that pNC-eGFP has significantly higher fluorescence than pNS-eGFP (p = 0.0020). Significance values are: ns = not significant, * = p < 0.05, ** = p < 0.01, *** = p < 0.001, **** = p < 0.0001. In both panels, error bars show the minimum and maximum datapoints, showing high variance in the pNC-eGFP sample.

3.3.2 Design, cloning, screening, and experimental analyses of intermediate constructs (Phase 2)

3.3.2.1 Selection of expression vector

In order to improve expression replicating vectors can be used. There are many replicating vectors available which, as mentioned in Chapter 1, should be chosen on a case-by-case basis, with no single best vector for all applications. For this research, pJL-TRBO was used as the replicating (R) vector. pJL-TRBO is a variant of the tobacco mosaic virus, and it increases per cell expression through RNA-based overexpression by replicating RNA transcripts within the cell. It also increases transformation efficiency by condensing these RNAs into discrete packages, and transporting them between plant cells linked by plasmodesmata. This was chosen as TMV is well-studied, with a broad host range and is very effective at transforming *Nicotiana* species. Nonetheless, it is possible that alternative vectors may provide higher expression but have not been tested in this research, and are discussed at the end of this chapter.

3.3.2.2 Design, cloning and screening of Phase 2 constructs

pJL-TRBO requires conventional cloning to input a coding sequence, utilising either PacI, AvrII, or NotI as the appropriate restriction endonucleases. As a result, any inserts needed a PacI recognition site at the 5’ end, and an AvrII site at the 3’ end. Consequently, the simple expression construct, pNS-eGFP, from Phase 1 was PCR amplified with primers containing these overhangs, the product was separated using gel electrophoresis and the appropriate band extracted and digested with PacI and AvrII. The pNC-eGFP construct already contained these overhangs and was digested with PacI and AvrII. The vector was also digested with these enzymes, and the two inserts were each ligated into the pJL-TRBO plasmid, creating pRS-eGFP and pRC-eGFP. These constructs, along with the functional units of the vector, can be seen in Figure 3.6.

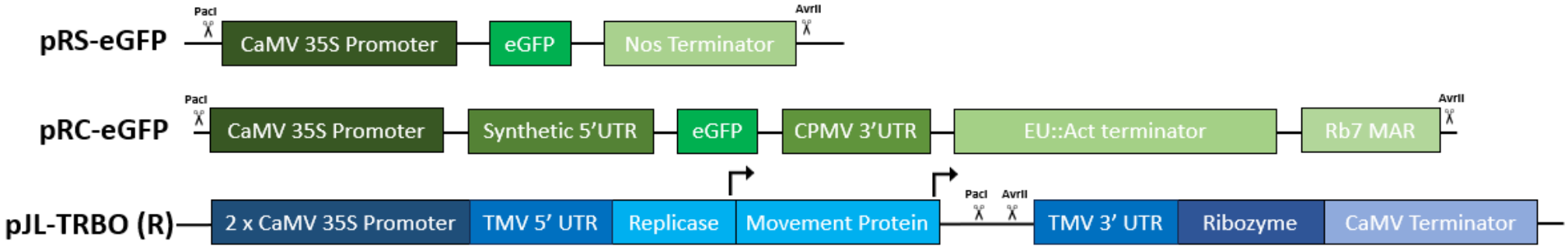


Figure 3.6: The Phase 2 constructs used in this research. pRS-eGFP = Simple construct containing (5' to 3') the CaMV 35S promoter, eGFP coding sequence, and nopaline synthase terminator in the replicating vector (R), pJL-TRBO. pRC-eGFP = Complex construct containing (5' to 3') the CaMV 35S promoter, a synthetic 5' UTR (Peyret et al. 2019), eGFP coding sequence, the CPMV 3' UTR (Peyret et al. 2019), *N. tabacum* Extensin and *N. benthamiana* Actin double terminator (Diamos & Mason 2018), and RB7 matrix attachment region (Diamos & Mason 2018) in the replicating vector (R), pJL-TRBO. pJL-TRBO, blue, contains a double CaMV 35S promoter to drive expression of a replicase protein enabling RNA-dependent RNA replication of any RNA between the two TMV UTRs. A subgenomic promoter drives expression of the movement protein that enables cell-to-cell spread of the amplified RNA between plasmodesmata. Each construct can be cloned into the vector using conventional cloning with the PacI and AvrII restriction enzymes. A downstream ribozyme enables efficient RNA cleavage, and the CaMV terminator enables transcript cleavage of any RNA that has not undergone ribozyme cleavage. All constructs shown are 5' to 3', left to right.

3.3.2.3 Optimising leaf infiltration

Before the Phase 2 constructs were compared, it was apparent from the patchy nature of the Phase 1 experiments that the leaf infiltration method required optimisation. In order to do this, vacuum infiltration was experimented with, using the new pRC-eGFP construct. Leaves were suspended in a solution of activated *Agrobacterium* containing the pRC-eGFP construct, and subject to a vacuum for varying lengths of time, ranging from 1 to 30 minutes, and the vacuum rapidly released. Theoretically, the displaced air in the spaces of the leaves should be replaced by the *Agrobacterium* solution, enabling transformation of the entire suspended area of the leaf. When qualitatively analysing leaves transformed using this method using blue-light excitation, it appeared that exposure to a vacuum for 2 minutes resulted in the greatest transformation efficiency (Figure 3.7).

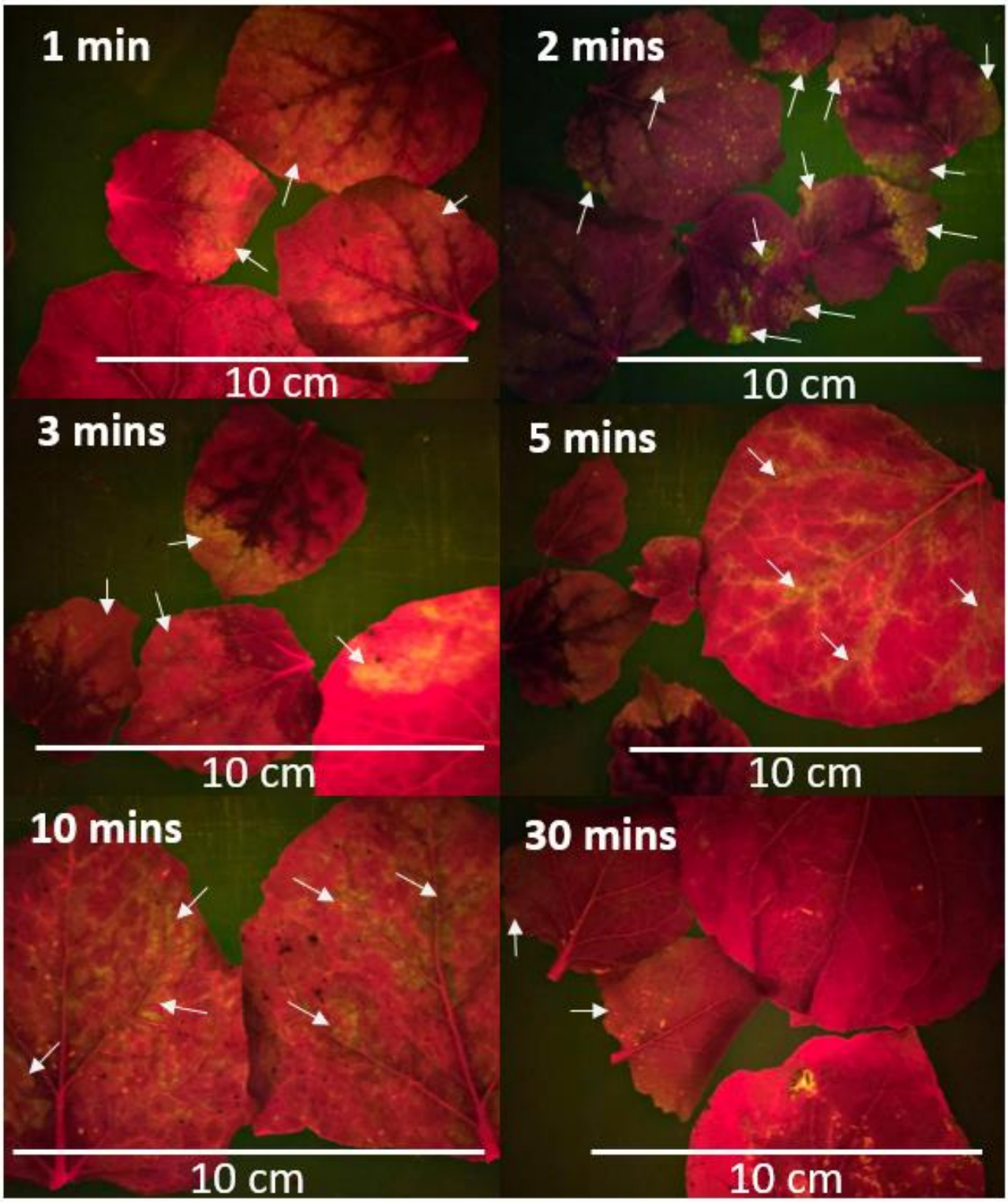


Figure 3.7: Photographs of *N. benthamiana* leaves transformed with pRC-eGFP using vacuum infiltration, viewed under blue-light. When a vacuum was applied for different lengths of time leaves showed minimal differences in transformation efficacy, but 2 minutes of vacuum infiltration produced marginally increased eGFP expression when analysed under blue-light with an orange filter. White arrows indicate some regions where eGFP is being expressed, visible in green of varying intensity. Red regions show chlorophyll autofluorescence of leaf tissue that is not expressing eGFP.

Importantly, it is clear that the vacuum infiltration is most effective in transforming the outer edges of exposed leaves, which are coincidentally the regions of the leaf that are usually left untransformed using the syringe infiltration method. As a result, it is feasible that a combination of transformation methods may result in the highest transformation efficiency and the highest overall recombinant protein expression. In order to test this, *N. benthamiana* leaves were again transformed using the pRC-eGFP construct, using either hand-infiltration, vacuum infiltration, or a mixed approach to transform three plants each, which were analysed by photon imaging. Transformed leaves were analysed using a photon imager 7 DPI. Whole leaves can be seen in Figure 3.8, Panels A-C, and the resulting eGFP fluorescence of the different infiltration methods can be seen in Figure 3.8, Panel D. Indeed, the mixed infiltration method produced the highest mean fluorescence, 1.23-fold and 5.68-fold higher than hand-infiltration or vacuum infiltration alone, though the difference between a mixed method and only hand-infiltration was not significant. As larger scale transformations are impractical when performed by hand, and industrial methods utilise vacuum infiltration on large scales, it is likely that the vacuum infiltration procedure requires further optimisation and more specialist equipment. Nonetheless, a mixed infiltration approach was consequently chosen for future experiments, as this produced the highest mean eGFP expression with the lowest variability both within and between leaves.

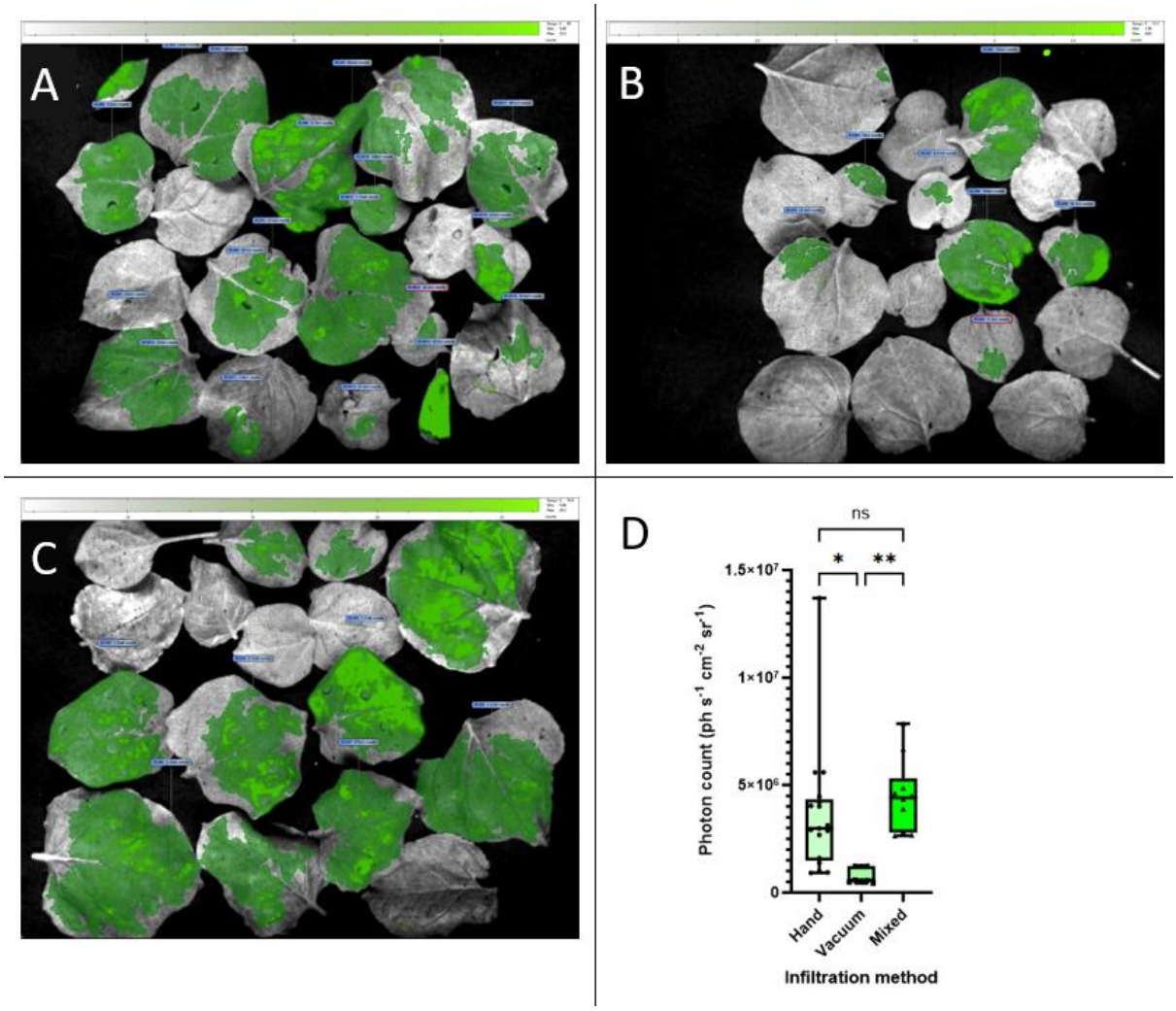


Figure 3.8: Analysis of *N. benthamiana* leaves transformed with pRC-eGFP using different agroinfiltration methods. *N. benthamiana* plants were transformed with pRC-eGFP using either hand-infiltration, vacuum infiltration, or both methods, with three plants transformed for each (n=3), and 3-6 transformed leaves measured per plant. **A** – Photon imager photograph of leaves transformed only using syringe infiltration, green areas show magic wand analysis where the BioSpaceLab software detects fluorescence. **B** – Photon imager photograph of leaves transformed using only vacuum infiltration, green areas show magic wand analysis where the BioSpaceLab software detects fluorescence. **C** - Photon imager photograph of leaves transformed using syringe infiltration followed by vacuum infiltration, green areas show magic wand analysis where the BioSpaceLab software detects fluorescence. **D** – Quantitative analysis of eGFP fluorescence from whole leaves using each infiltration method. A one way ANOVA test determined that the difference in eGFP fluorescence between these methods was statistically significant ($F_{(2, 32)} = 6.215$, $p = 0.0052$). Further analyses using a Tukey post hoc test revealed that the mixed infiltration method produced more eGFP fluorescence than both the hand-infiltrated leaves, though this was not significant, ($p= 0.6412$) and the vacuum-infiltrated leaves, which was statistically significant ($p= 0.0055$). Significance values on the graph are: ns = not significant, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$. Error bars show the minimum and maximum values, showing high variance in the hand-infiltrated sample.

3.3.2.4 Qualitative in planta analyses of Phase 2 constructs

When viewing transformed leaves using photon imaging (Figure 3.9, Panel A) and confocal microscopy (Figure 3.9, Panel B), the Phase 2 constructs were compared against the initial Phase 1 constructs, which can be seen above in Figure 3.4. When comparing expression by the different vectors there is a clear difference in homogeneity. When expressed using the Level 1 non-replicating Golden Gate vectors, in both the pNS-eGFP and pNC-eGFP constructs we saw large areas of darkness where individual leaf epidermal cells have not been transformed, meaning that there is a large area of leaf cells that are not expressing eGFP at all. Conversely, when comparing the images with eGFP expressed by pRS-eGFP and pRC-eGFP in the replicating vector pJL-TRBO, we can see that more cells are expressing eGFP and there is much more uniform expression across the analysed areas. Together, this suggests that the complex expression constructs drive increased eGFP expression within individual cells, relative to the simple expression constructs, and the replicating vector enables more cells to be transformed through virally induced RNA movement, particularly when paired with the mixed infiltration method to maximise infiltrated area. Each of these contribute to the overall protein expression, resulting in higher expression in the pRC-eGFP construct, which combines all features.

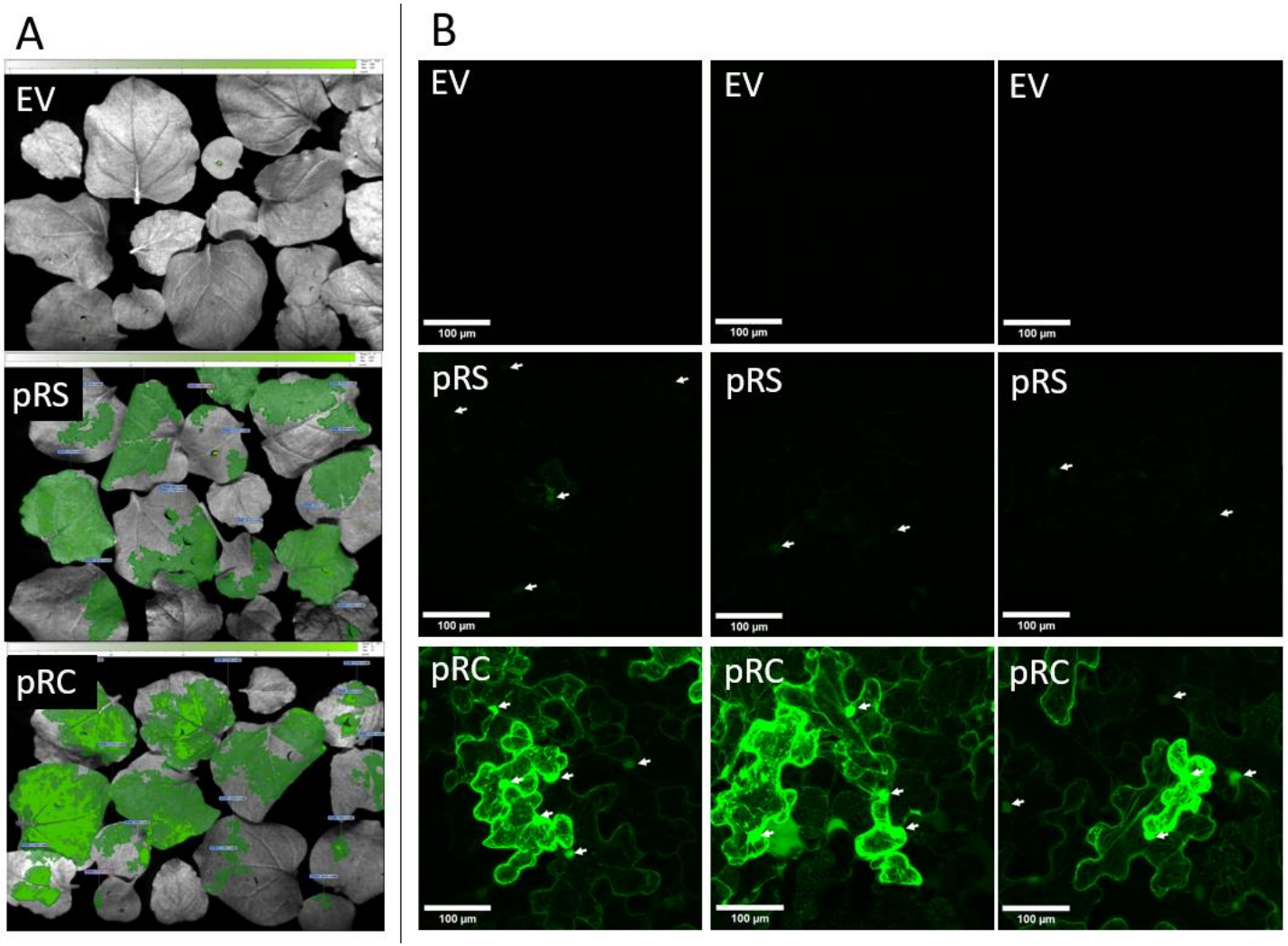


Figure 3.9: Qualitative analyses of *N. benthamiana* leaves transformed with Phase 2 constructs. A – Photon imager photographs of whole *N. benthamiana* leaves transformed with either empty vector, pRS-eGFP or pRC-eGFP constructs, 7 DPI. Regions with expression are highlighted using the magic wand function on M3 Vision, shown in dark green. **B** – Confocal microscopy images showing 30 µm Z-stacks of transformed areas using a gain of 700 and an excitation wavelength of 488nm; scale bars represent 100 µm. and white arrows indicate nuclei of epithelial cells.

3.3.2.5 Quantitative in planta analyses of Phase 2 constructs

Transformed *N. benthamiana* leaves expressing the constructs from Phase 2 were analysed by photon imaging alongside an untransformed control. These data, Figure 3.10, Panel A, showed that there is a 47-fold increase in fluorescence with pRC-eGFP relative to pRS-eGFP, which was statistically significant. Similarly, per cell quantification showed that pRC-eGFP expressed eGFP at 17.6-fold that of pRS-eGFP (Figure 3.10, Panel B). Thus, it was determined the complex expression cassette drives higher expression than the simple expression cassette, but that this effect is dramatically improved when combined with the replicating vector with a synergistic effect seen. The clear improvement in the pRC expression construct suggests that some of these regulatory elements are functional in the replicating vector, but it is difficult to determine which elements are responsible. It was hypothesised that the UTRs are the most likely candidates for this improvement, as UTRs function at the RNA level and TMV is a vector that replicates target RNA, so the two could work synergistically. As a result, the final expression constructs combining different combinations of UTRs were created and tested; these are the constructs of Phase 3.

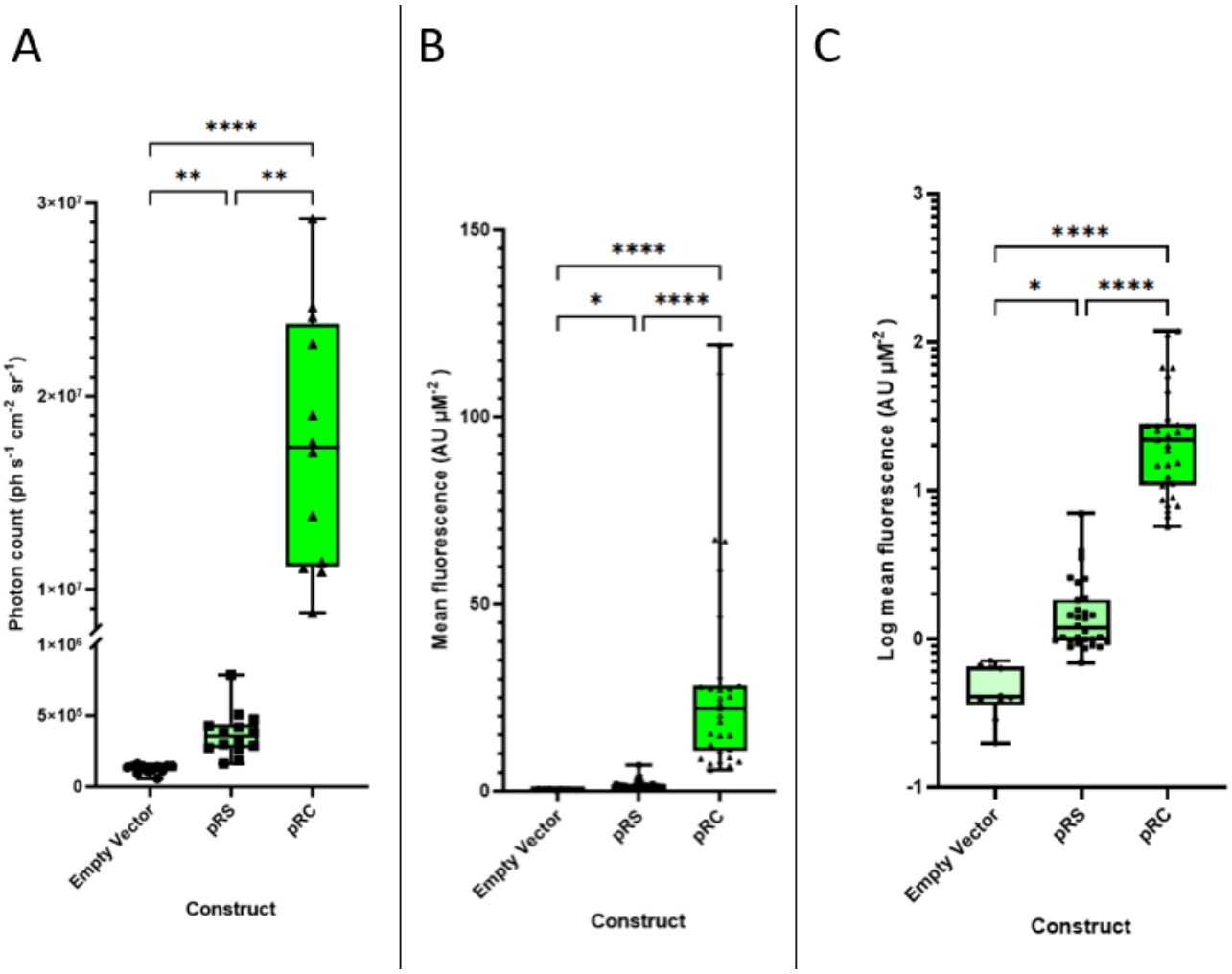


Figure 3.10: Quantitative analyses of *N. benthamiana* leaves transformed with Phase 2 constructs. *N. benthamiana* plants were transformed with either empty vector, pRS-eGFP, or pRC-eGFP (three plants each, n=3) and 4-5 transformed leaves were analysed for each plant 7 DPI. **A** – eGFP fluorescence quantified using a Biospace Optima photon imager. A Kruskal-Wallis test shows a significant difference between means ($H_{(2)}$= 32.85, $p < 0.0001$). A Dunn's multiple comparison test shows that pRC-eGFP produces significantly higher fluorescence than the negative control and pRS-eGFP ($p < 0.0001$ and $p = 0.0088$, respectively). **B** – ImageJ per cell quantification of eGFP fluorescence within 10-31 cells transformed with each construct. Panel **C** shows the same data as Panel B on a log10 scale; A Kruskal-Wallis test shows a significant difference between means ($H_{(3,71)}$= 58.33, $p < 0.0001$). A Dunn's multiple comparisons test show that pRS-eGFP and pRC-eGFP have significantly higher fluorescence than the negative control ($p = 0.$ 0.0245 and $p < 0.0001$, respectively) and that pRC-eGFP has significantly higher fluorescence than pRS-eGFP ($p < 0.0001$). Significance values are: ns = not significant, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$. Error bars show minimum and maximum values, with the highest variance seen in the pRC-eGFP sample. The empty vector

3.3.3 Design, cloning, screening, and experimental analyses of final constructs (Phase 3)

3.3.3.1 Design, cloning and screening of Phase 3 constructs

As the pJL-TRBO vector works on RNA amplification, it was hypothesized that the UTRs were the likely element driving the improvement in expression between pRS-eGFP and pRC-eGFP, as the UTRs are the only different elements that have a major impact at the RNA level. Consequently, the final eGFP constructs were created and tested. These were eGFP driven by either both the synthetic 5' and CPMV 3' UTRs (pRU-eGFP), just the 5' UTR (pR5-eGFP) or just the 3' UTR (pR3-eGFP), in case one improved expression whilst the other diminished it. A construct without any additional regulatory elements (pRI-eGFP) was also created, in case the additional UTRs could negatively affect expression; this is analogous to the pJL-TRBO-GFP construct previously reported (Lindbo 2007b), but with a slightly different CDS. These constructs were each created using the complex construct, pNC-eGFP, from Phase 1 as a template for PCR amplification to amplify eGFP with one or both UTRs, using primers that incorporated a 5' PacI and 3' AvrII recognition site. They are shown diagrammatically in Figure 3.11. The amplified fragments were separated by gel electrophoresis, extracted and digested alongside the pJL-TRBO vector. The inserts were then ligated into the vector, transformed into *E. coli* and grown overnight, followed by plasmid DNA purification and sequencing. *Agrobacterium* were transformed and the constructs were tested in plants, alongside pRC-eGFP and an empty vector control.

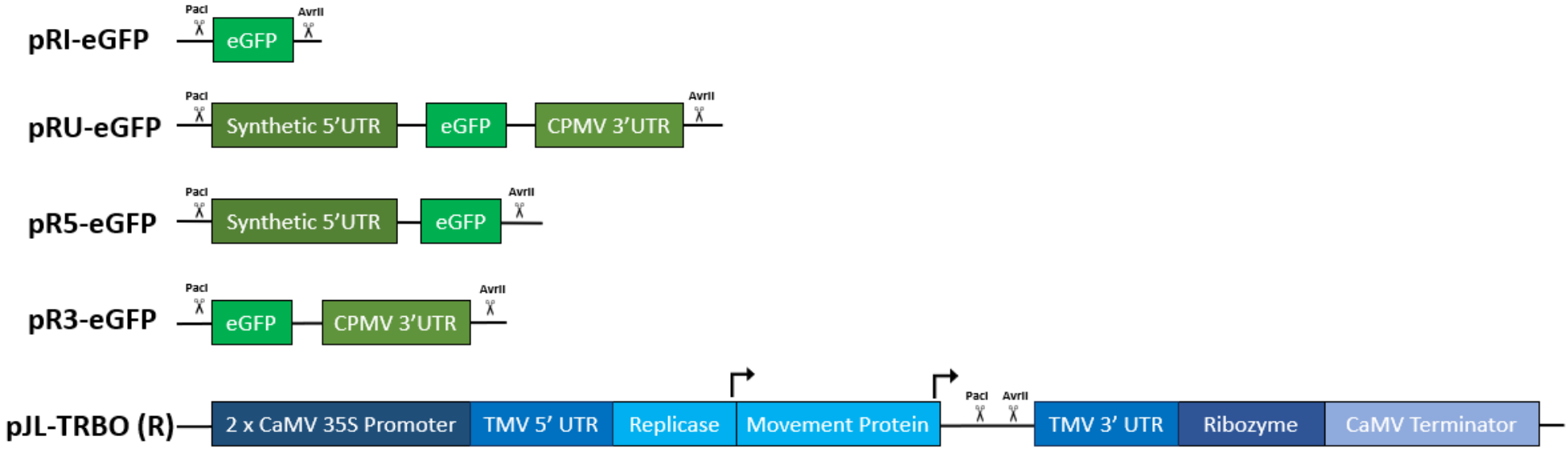


Figure 3.11: The Phase 3 constructs used in this research. pRI-eGFP = Isolated eGFP coding sequence in the replicating vector pJL-TRBO. pRU-eGFP = Double UTR construct containing (5' to 3') the synthetic 5' UTR, eGFP coding sequence, and CPMV 3' UTR in the replicating vector pJL-TRBO. pR5-eGFP = 5' UTR construct containing (5' to 3') the synthetic 5' UTR and eGFP coding sequence in the replicating vector pJL-TRBO. pR3-eGFP = 3' UTR construct containing (5' to 3') the eGFP coding sequence and CPMV 3' UTR in the replicating vector pJL-TRBO. pJL-TRBO, blue, contains a double CaMV 35S promoter to drive expression of a replicase protein enabling RNA-dependent RNA replication of any RNA between the two TMV UTRs. A subgenomic promoter drives expression of the movement protein that enables cell-to-cell spread of the amplified RNA between plasmodesmata. Each construct can be cloned into the vector using conventional cloning with the PacI and AvrII restriction enzymes. A downstream ribozyme enables efficient RNA cleavage, and the CaMV terminator enables transcript cleavage of any RNA that has not undergone ribozyme cleavage. All constructs shown are 5' to 3', left to right.

3.3.3.2 Qualitative in planta analyses of Phase 3 constructs

N. benthamiana leaves were transformed with the Phase 3 constructs and analysed alongside pRC-eGFP and an empty vector control. Figure 3.12 shows qualitative analysis of transformed leaves, using excitation of whole leaves via photon imaging and Figure 3.13 shows transformed cells when visualised using confocal microscopy. These qualitative data suggested that the new construct containing only the 5' UTR provided the highest *in planta* expression, especially when analysed by confocal microscopy. In contrast, the expression constructs containing both UTRs or the 3' UTR appear to have reduced expression relative to both the isolated CDS and the 5' UTR construct, suggesting that the 3' UTR diminishes expression. Of course, quantitative analyses are needed to provide conclusive evidence. Importantly, pRC-eGFP displayed by far the most uniform expression with the majority of leaf area being transformed. However, the Phase 3 constructs appear to primarily transform the vascular tissue, an area where pRC-eGFP appears to be unable to transform. This suggests that the additional DNA elements present in pRC that are absent in the other constructs result in more uniform eGFP expression across the leaf lamina. As a result, co-expression of both constructs may further improve expression by maximising the area of affected cells. For example, co-expressing a target protein under control of pR5 and pRC may enable high levels of expression in the vasculature and lamina, respectively, though this was not attempted in this research.

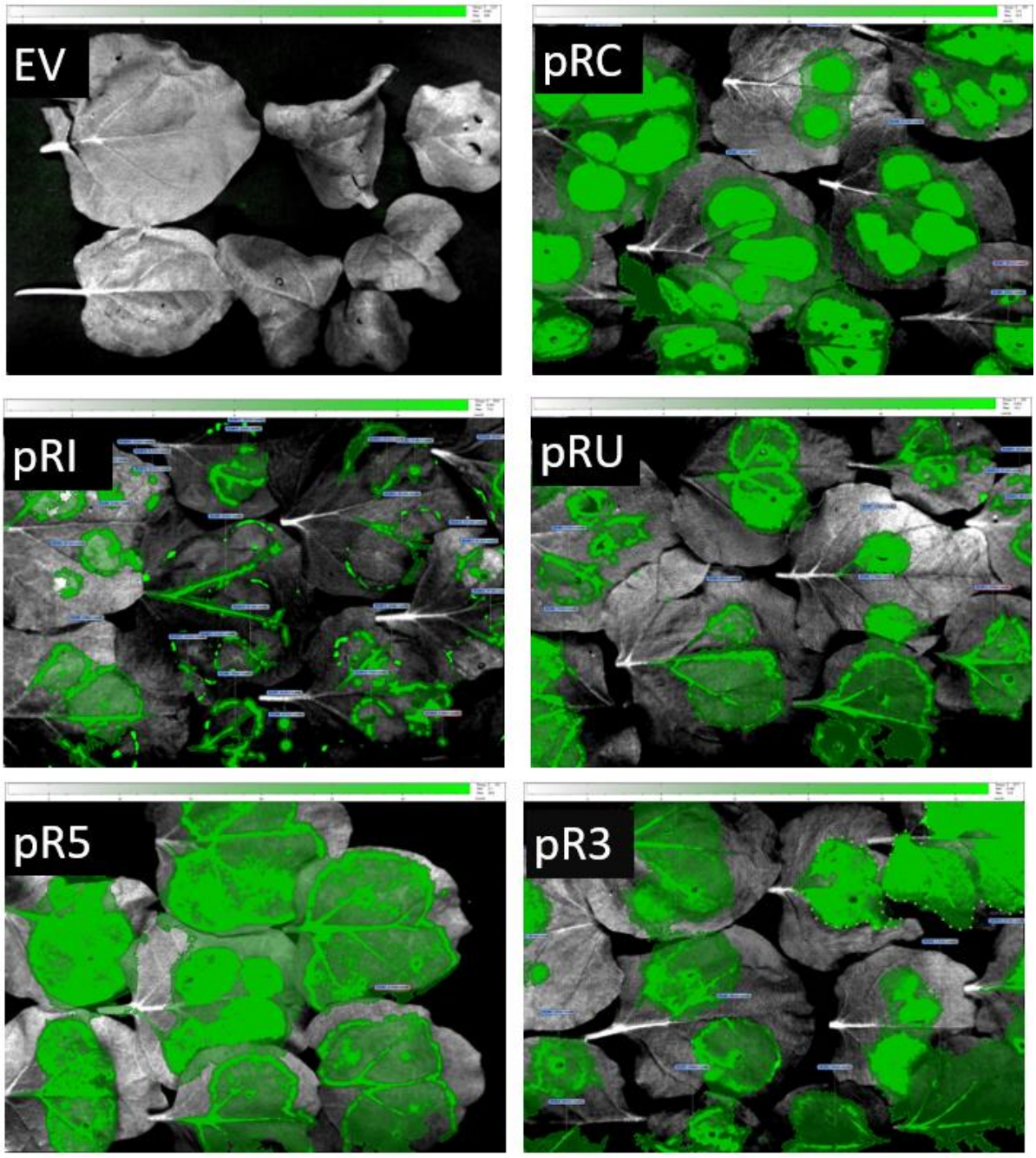


Figure 3.12: Photon Imager photographs of *N. benthamiana* leaves transformed with Phase 3 constructs and pRC-eGFP. Photon Imager photographs of whole *N. benthamiana* leaves transformed with either empty vector, pRC-eGFP, pRI-eGFP, pRU-eGFP, pR5-eGFP or pR3-eGFP, 7 DPI. Regions with expression are highlighted using the magic wand function on M3 Vision, shown as dark green areas. Only pRC shows high levels of eGFP expression throughout the leaf lamina, with low expression in the vasculature. In contrast, constructs pRI-eGFP, pRU-eGFP, pR5-eGFP and pR3-eGFP appear to show eGFP expression predominantly in the vascular tissue, suggesting that the additional DNA elements in pRC affect the area in which eGFP is expressed within leaves. In the empty vector control, the magic wand tool could not detect any fluorescence, so whole leaves are highlighted instead.

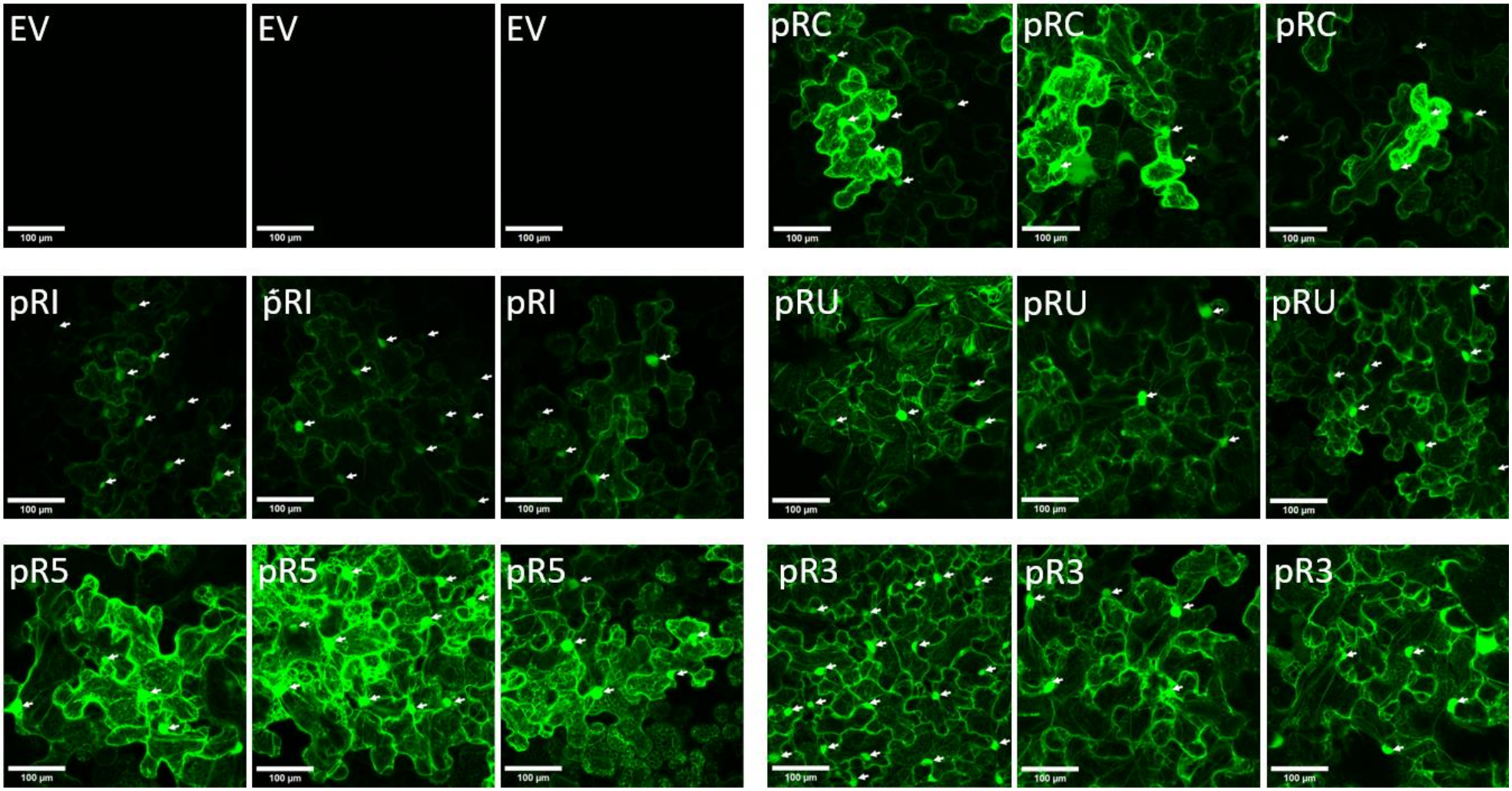


Figure 3.13: Confocal microscopy images of *N. benthamiana* leaf cells transformed with Phase 3 constructs and pRC, each expressing eGFP. Confocal microscopy images showing 30 µm Z-stacks of transformed areas using a gain of 700 and an excitation wavelength of 488nm; scale bars represent 100 µm. White arrows indicate nuclei of epidermal cells.

3.3.3.3 Co-expression of other proteins to enhance recombinant protein expression

Before these Phase 3 constructs were comparatively quantified, efforts were made to determine whether expression could be further enhanced through the co-expression of proteins. It is well documented in the literature that co-expression of other proteins, particularly gene silencing suppressors derived from plant viruses, can improve the expression of recombinant proteins. In fact, one research paper reported that co-expression of five proteins each improved the yields of plant recombinant protein (Norkunas et al. 2018). These proteins were Rep, RepA, CMV2b, BAG4, and p19. In order to further improve the yields of recombinant proteins in my expression system, a Level 2 Golden Gate construct termed 'Improving Infiltration' was created which contained multiple transcriptional units, enabling the expression of each of these coding sequences constitutively. The detailed cloning of this is documented in Section 2.3.9. The Rep and RepA proteins were ligated under control of a 5' Mas promoter from *A. tumefaciens* with a 3' nopaline synthase terminator. The BAG4 and CMV2b coding sequences were combined with a 5' RbcS2 promoter from *Solanum lycopersicum*, also with the 3' nopaline synthase UTR. A diagram of the construct can be seen in Figure 3.14. These different promoters were chosen for construct generation as they enable constitutive expression, and they are not utilised in any other constructs in this research. This reduces the chance of DNA-methylation of homologous promoters causing transcriptional gene silencing within the transformed plants. The p19 protein was expressed under control of the CaMV 35S promoter at the 5' end, and a nopaline synthase terminator at the 3' end. Once created, the construct was transformed into *Agrobacterium*, which was co-transformed into *N. benthamiana* leaves with pRC-eGFP. Unfortunately, when analysed using photon imaging, there was no significant difference in eGFP expression, suggesting that this Level 2 expression construct provides no advantages to recombinant protein expression (Figure 3.15, Panel A).

Figure 3.14: Level 2 expression cassette created to enhance recombinant protein expression when co-expressed. Five proteins were co-expressed, each shown to improve expression in the literature (Norkunas et al. 2018). Promoters are denoted with a p and are shown in dark green, coding sequences in mid-green and the nopaline synthase terminators (tNOS) in pale green. The silencing suppressor p19 had expression driven with the cauliflower mosaic virus CaMV 35S promoter; REP and REPA had expression driven by the *A. tumefaciens* mannopine synthase (AtuMaS); CMV2b and BAG4 were under control of the Ribulose bisphosphate carboxylase small subunit, chloroplastic 2 (RbcS2) promoter from *A. thaliana*. Each promoter drives consitutive expression and different promoters were chosen in an attempt to avoid transcriptional gene silencing of homologous promoters

As the viral silencing suppressors are the best characterised proteins for co-expression to improve recombinant protein yields, the co-expression of p19 was investigated further. One study tested several viral silencing suppressors and found p19 to be the most effective at improving transgene expression (Peyret et al. 2019). As a result, the p19 protein coding sequence was amplified with flanking restriction enzyme sites, with PacI and AvrII and the 5' and 3' ends, respectively. This was cloned into pJL-TRBO, creating pRI-p19. This was transformed into *Agrobacterium* which was used to co-transform *N. benthamiana* plants with either pRC-eGFP or pR5-eGFP.

When p19 was co-expressed with pRC-eGFP in a 1:1 v/v ratio, a 62% reduction in fluorescence was seen (Figure 3.15, Panel B), which was statistically significant. Similarly, when p19 was co-expressed in a 1:9 v/v ratio with pR5-eGFP, expression was reduced by 38%, though this was not statistically significant (Figure 3.15, Panel C). In addition, plant cell packs (PCPs) were also used to quantify the expression of these co-transformants, chosen because they have been reported to have low variability between samples and ease of use. These PCPs are condensed Tobacco Bright Yellow-2 cells which have had their culture media removed via low-speed centrifugation in 96-well filter plates and are mixed with activated *Agrobacterium* solution containing the expression construct. These can be visualised using the photon imager in the same way as the whole *N. benthamiana* leaves and are performed in triplicate. The same trend was also seen when co-expressing pR5-eGFP and p19 in plant cell packs in a 1:1 ratio, with a 71% reduction in eGFP fluorescence when co-expressed with p19 (Figure 3.15, Panel D), a statistically significant finding. In fact, plants expressing only pR5-eGFP appeared green and healthy (Figure 3.16, Panel A) but there was a lethal phenotype in the *N. benthamiana* plants which co-expressed p19, as they produced darkened flaccid leaves (Figure 3.16, Panel B). Further analysis of expression on blue light showed clear fluorescence in the pRC-eGFP transformed leaves (Figure 3.16, Panel C) with nearly no eGFP seen in leaves co-expressing p19 (Figure 3.16, Panel D). As a result, it was decided that co-expression of other proteins would not be utilised in this work but is a valuable avenue to explore for future expression improvements. Thus, the Phase 3 constructs could be quantified without co-expression of other proteins, as this was determined to produce the highest target protein expression in this research.

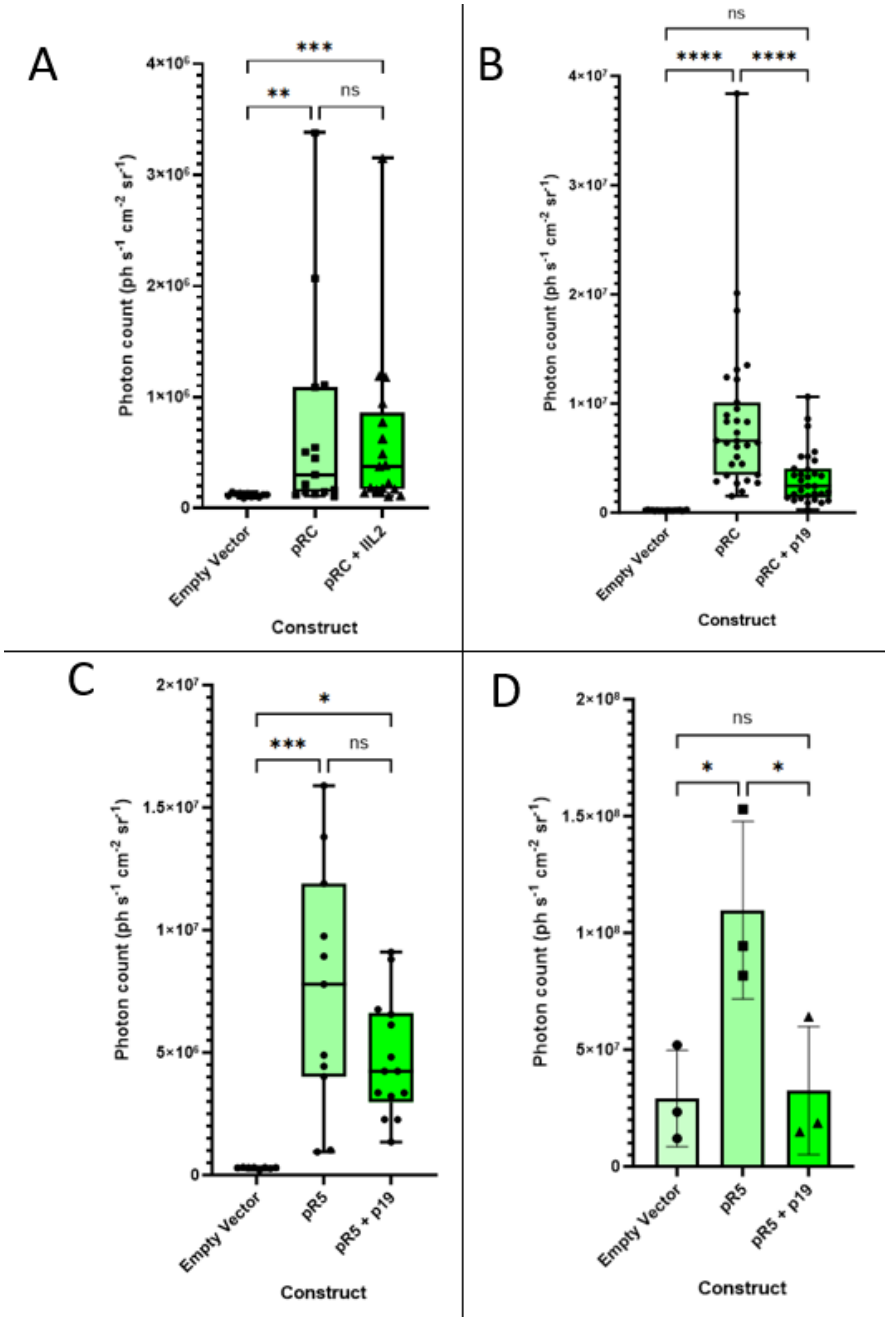


Figure 3.15: Expression data when eGFP is co-expressed with other proteins. A – pRC-eGFP was expressed in *N. benthamiana* plants (n=3, 4-5 leaves per plant) with or without the level 2 expression construct, IIL2, which had no impact on expression levels when analysed using a Dunn's multiple comparisons ($p > 0.9999$). B – pRC-eGFP was expressed with or without pRI-p19 in six plants (n=6, 3-6 leaves per plant), which significantly reduced eGFP fluorecence ($p < 0.0001$). C – pR5-eGFP was expressed with or without pRI-p19 in three plants (n=3, 3-5 leaves per plant), which reduced eGFP fluorecence but this was not significant ($p = 0.0998$). D – pR5-eGFP was co-expressed with pRI-p19 in plant cell packs where co-expression resulted in a significant reduction of eGFP expression ($p = 0.0427$). Overall, despite reports that these proteins can improve target protein expression, these data suggest that the opposite effect occurs, and co-expression can be deleterious to target protein expression. Significance values are: ns = not significant, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$. In all panels error bars show minimum and maximum values. Samples transformed with an eGFP expressing construct show the highest variance, showing a high degree of variability and expression levels, particularly between leaves.

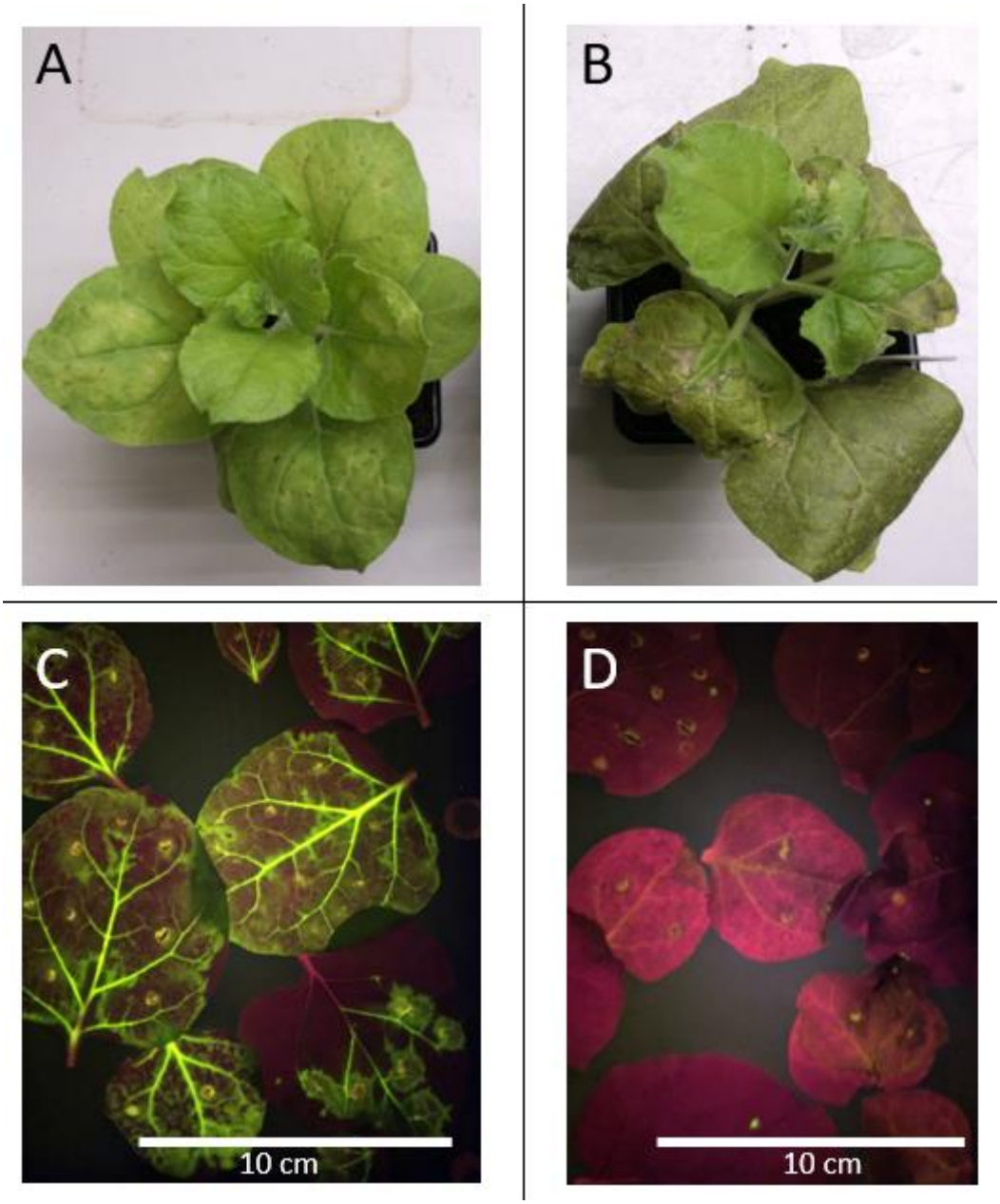


Figure 3.16: Photographs of stressed leaves when p19 is co-expressed with eGFP. *N. benthamiana* plants were transformed with only pR5-eGFP or co-transformed with pRI-p19 too. Photographs of whole plants showed that expressing only eGFP resulted in healthy plants (**A**) compared to stressed plants when p19 was co-expressed (**B**). When analysing leaf expression using blue light, leaves transformed with only pR5-eGFP (**C**) showed increased fluorescence (green) relative to leaves co-transformed with p19 (**D**) which show mostly untransformed tissue (red due to chlorophyll autofluorescence). Overall, these images suggest that co-expression of p19 results in dramatically reduced target protein expression through increased plant stress.

3.3.3.4 Quantitative in planta analyses of Phase 3 constructs

Quantitative analysis of Phase 3 constructs was performed using photon imaging to analyse eGFP expression across the whole leaf area. These data are show in Figure 3.17, Panel A. The most basic construct expressed the isolated eGFP coding sequence without any additional regulatory elements (pRI-eGFP) and was the lowest expressor in the group. As previously stated, this construct is analogous to the pJL-TRBO-G construct reported by Lindbo (2007b) but with a slightly different coding sequence. The next lowest expressor was the construct with both the 5' and 3' UTRs, pRU-eGFP, followed by pR3-eGFP which contained only the 3' UTR. The highest expressing construct was pR5-eGFP, which contained only the previously published synthetic 5' UTR (Peyret et al. 2019), which expressed eGFP at 2.2-fold, 1.1-fold, and 1.76-fold that of construct pRI-eGFP, pRU-eGFP, and pR3-eGFP, respectively. Interestingly, the largest construct, pRC-eGFP, was also expressed alongside the four UTR constructs, and displayed the highest level of eGFP expression. Photon imager quantification of the leaves shows that pRC-eGFP expressed eGFP at 5.5-fold, 2.7-fold, 2.5-fold, and 4.3-fold that of pRI-eGFP, pRU-eGFP, pR5-eGFP and pR3-eGFP, respectively, making it the highest expressor of the eight constructs created in this research.

However, confocal microscopy analysis of transformed cells (Figure 3.17, Panel B) showed that pRC-eGFP did not produce the highest level of eGFP fluorescence per cell. Quantification of per cell expression showed that pR5-eGFP had the highest eGFP production with 1.4-fold, 3.2-fold, 2-fold and 1.6-fold the expression of pRC-eGFP, pRI-eGFP, pRU-eGFP, and pR3-eGFP, respectively, and the difference in per cell expression between pR5-eGFP and pRC-eGFP was statistically significant.

Together these data suggest that pRC is the most robust construct in driving high levels of expression throughout the leaf tissue, but that pR5 drives the highest per cell expression. These constructs also appear to primarily transform different leaf tissues. As a result, it is possible that co-transformation using both pRC and pR5 constructs may enable further enhanced expression of a recombinant protein by maximising the area of expressing tissue, though this has not been tested.

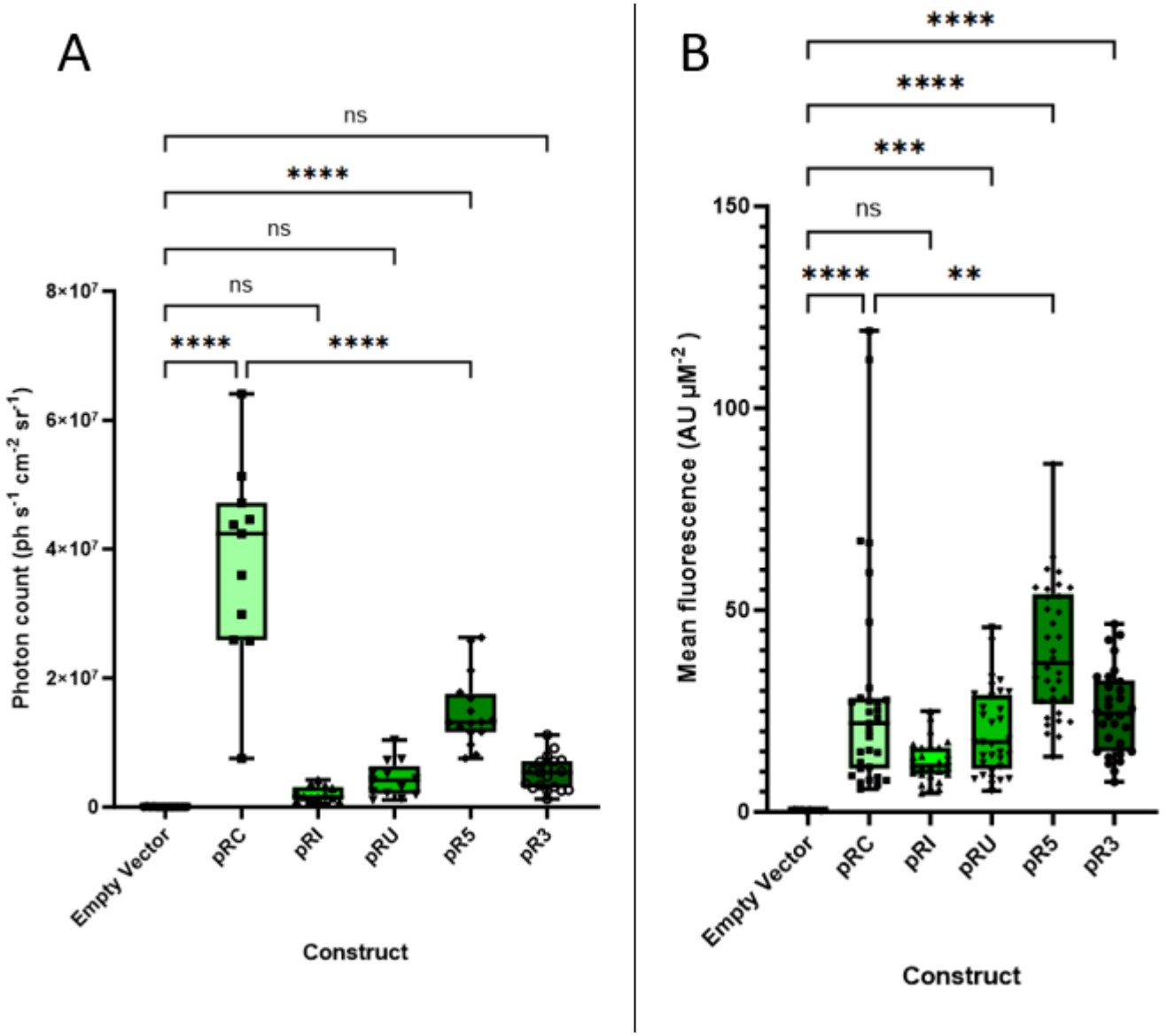


Figure 3.17: Quantitative analyses of *N. benthamiana* leaves transformed with Phase 3 constructs and pRC, each expressing eGFP. Three *N. benthamiana* plants were each transformed with phase 3 constructs; transformed leaves were analysed 7 DPI. **A** – eGFP fluorescence quantified using a Biospace Optima photon imager. A one way ANOVA showed a statistically significant difference between means ($F_{(5, 80)} = 63.73$, $p < 0.0001$). A Tukey test shows that pRC-eGFP and pR5-eGFP are significantly different from each other ($p < 0.0001$). **B** – ImageJ per cell quantification of eGFP fluorescence within 10-34 cells transformed with each construct; A Kruskal-Wallis test shows a significant difference between means ($H_{(6,165)} = 74.08$, $p < 0.0001$). A Dunn's multiple comparisons test show that pR5-eGFP produced significantly higher per cell fluorescence than pRC ($p = 0.0046$). Significance values are: ns = not significant, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$. Error bars show minimum and maximum values, with the highest variance seen in whole leaves in the pRC sample. The empty vector and pRC datapoints are the same as in Figure 3.10, as all constructs were transformed at the same time.

Finally, expression of the Phase 3 constructs was compared in plant cell packs, shown in Figure 3.18, which suggested that each of the expression constructs was similar with all constructs except pRC-eGFP showing significantly higher expression than the negative control, but no significant difference of the constructs between each other. However, this may be due to the increased green autofluorescence of the BY-2 cells, relative to whole *N. benthamiana* plants.

Overall, the data presented in this chapter shows that two of the developed constructs produce the highest eGFP expression. pRC-eGFP produced the highest expression across the whole leaf, with the largest area of transformed tissue, however pR5-eGFP resulted in the highest level of per cell expression.

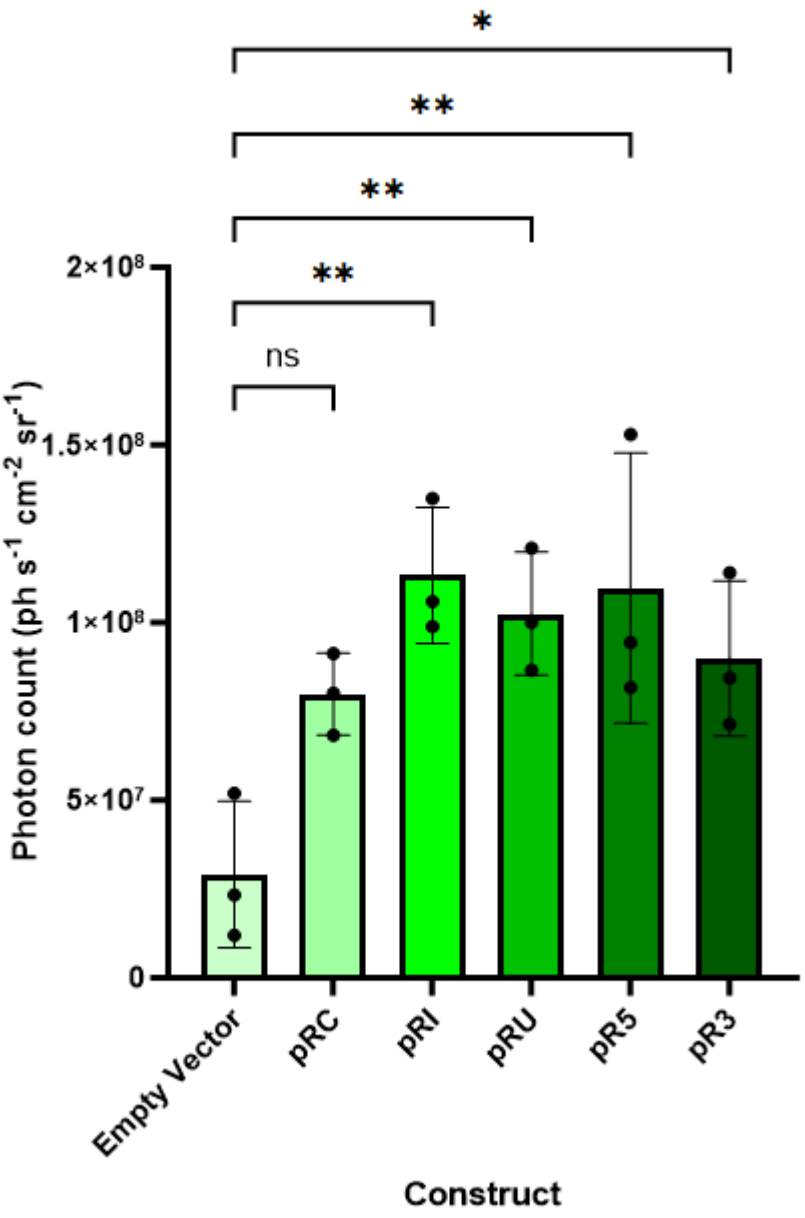


Figure 3.18: Photon imager fluorescence of plant cell packs transformed with Phase 3 constructs and pRC, each expressing eGFP. PCPs were prepared from BY-2 cells, transformed using agrobacterium and had eGFP fluorescence analysed by photon imaging 4 days post-transformation. The difference between means was statistically significant according to a one-way ANOVA ($F_{(5,12)} = 5.56$, $p = 0.007$). Further analyses using a TukeyHSD test revealed that neither the complex nor 3' UTR construct were statistically significantly different from the untransformed control. Though equally, none of the expression constructs were statistically significantly different from each other. Significance values are: ns = not significant, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$. Error bars show minimum and maximum values.

3.4 – Discussion

The main aim of this chapter was to develop an expression construct that produces enhanced recombinant protein expression to improve the overall yields that plant chassis can make. Initial simple expression constructs only utilised a promoter, CDS, and terminator and were considered the baseline for gene expression. This aim has been met, with several improved expression constructs made, and the best, pRC-eGFP, produced significantly higher *in planta* fluorescence than the simple initial construct, with an increase in fluorescence of up to 199-fold relative to the initial pNS-eGFP construct, and 2.2-fold the fluorescence of pRI-eGFP, analogous to the previously published pJL-TRBO vector. This section will discuss the achievements, comparing each construct and critically analysing why each construct showed the expression level that it did. Further, it will compare these constructs to the existing literature from which many of the DNA elements are derived, suggesting mechanisms of action, limitations and avenues for future work.

3.4.1 Comparison of Constructs

As mentioned at the start of the chapter, initial expression constructs were created that expressed C-terminally eGFP-tagged P2X7 isoforms under regulatory control of only a CaMV 35S promoter and a nopaline synthase terminator. Early quantification of these constructs (explored later in Chapter 5) suggested that the expression levels using such regulatory elements would be insufficient to produce high recombinant protein yields. As a result, a substantial genetic engineering effort was made to maximise the yield of recombinant proteins at the genetic level, through a combinatorial analysis of different regulatory elements and vector backbones. All the constructs discussed in this chapter utilised eGFP as a reporter, enabling the *in planta* measurement of fluorescence to estimate yield and localisation, which could be further quantified at the protein level when extracted.

The initial expression construct was a simple non-replicating construct, pNS-eGFP, with only the CaMV 35S promoter and nopaline synthase terminator to drive eGFP expression. This produced only low-level expression, despite the CaMV 35S promoter being widely used in the literature for its high expression levels. However, failure to detect high-levels of expression may be due to a poor transformation efficiency when using the pNC-eGFP construct. Through combining several published genetic elements, a complex composite construct was developed in a non-replicating vector, pNC, which combined, 5’ to 3’, the same CaMV 35S promoter, a synthetic CPMV-derived UTR, CPMV 3’ UTR, *N. tabacum* extensin terminator, *N. benthamiana* Actin terminator, and RB7 matrix attachment region. This construct resulted in 20.1-fold increased per cell *in planta* fluorescence. This increase must be attributed to the UTRs, terminators, and MAR. The UTRs are derived from a publication (Peyret et al. 2019) which attempted to synthetically improve the CPMV UTRs, already well

established in the literature and the basis for the commercially used HyperTrans system (Sainsbury and Lomonossoff 2008; Sainsbury et al. 2009). The 5' UTR is thought to work through promoting very efficient ribosome binding (Sainsbury et al. 2009), maximising translational efficacy, whilst the 3' UTR forms a complex secondary structure that is critical for improving RNA stability (Meshcheriakova et al. 2014). In contrast, the terminators and MAR are thought to act at the transcriptional level, through reducing transcript readthrough and ensuring the creation of a transcriptionally active domain by binding nuclear lamina, respectively (Diamos et al. 2016; Diamos and Mason 2018; Dolgova and Dolgov 2019). However, despite improved per cell expression, this construct resulted in only 1.4-fold increased fluorescence across the whole leaf, relative to pNS-eGFP. Spatial analyses suggested that transformed tissue was primarily localised to the infiltration sites, and failed to spread around the whole leaf, despite multiple infiltration sites per leaf.

This was the primary reason that a TMV replicating vector was chosen for later constructs, as these ensure RNA replication and movement through plasmodesmata, maximising the transformation area. As the simple and complex cassettes were already made, they were cloned into the TMV-derived replicating vector, pJL-TRBO, to make pRS-eGFP and pRC-eGFP, respectively. This vector does not require any additional DNA elements to function and is designed with a multiple cloning site (MCS) for the ligation of only a coding sequence; TMV is an RNA vector, and the replication is of the RNA between native TMV UTRs at the MCS. In theory, this means that any regulatory elements that affect transcription, such as promoters, terminators and matrix attachment regions should not be functional at the RNA level, so it was hypothesised that the simple and complex expression cassettes would not function at this level. In addition, these larger presumably non-functional DNA elements may impair RNA replication, although it has been suggested that there is no theoretical insert size limit (Lindbo 2007a). To our surprise, this was not the case, with both expression cassettes displaying improved expression in the replicating vector than in the non-replicating vector. pRS-eGFP showed an improvement in the number of expressing cells relative to pNS-eGFP, suggesting that the low-level expression seen was partly due to the number of transformed cells. However, expression levels are still low, which may be due to a lack of sensitivity with the photon imager, and due to the confocal microscopy settings used which were adjusted for visualisation and quantification of the higher expressing constructs using the complex expression cassette. This means that eGFP fluorescence that is below the range set for visualising the higher expressing constructs would not be detected. More sensitive detection methods such as a western blot, which are explored in Chapter 4, can detect protein expression more definitively. Furthermore, pRC-eGFP showed an improvement in gene expression relative to pRS-eGFP, suggesting that the additional regulatory elements in the complex cassette work synergistically to the replicating vector. As the transcriptional

DNA elements were unlikely to be functional in the replicating RNA, it was hypothesised that the UTRs are most likely responsible for this increase.

Consequently, the Phase three constructs were generated to test the relative effects of the UTRs. To my knowledge, a combination of viral UTRs had not been tested in the literature, but replacement of CPMV UTRs has shown to compromise CPMV replication function (Meshcheriakova et al. 2014), so it was plausible that combining the CPMV UTRs with the TMV UTRs could also compromise gene expression. As a result, each combination of UTRs was tested within the pJL-TRBO vector, including a variant without any additional regulatory elements with only the eGFP CDS ligated into the MCS. This analysis showed that utilising any combination of UTRs produced expression levels higher than no UTRs. Interestingly, utilising both UTRs produced a lower expression level than using only one UTR, and the best construct used only the 5' UTR. Together, these data suggest that both the synthetic 5' UTR and CPMV 3' UTR can improve expression levels when incorporated into TMV but are best utilised without the other UTR. As both additional UTRs used here are derived from CPMV, it is likely that the antagonistic effect of both UTRs is unique to combining them in another viral vector. In addition, as the function of the UTRs is related to their sequence and structure, it is likely that the order as well as the combination of them makes a substantial difference, though further work would be necessary to confirm this. Perhaps more interesting is that the highest expressing construct in whole leaves was the pRC construct, although pR5-eGFP showed significantly higher per cell expression when analysed by confocal microscopy. This means one of two things. Either the higher expression of pRC is real and the additional terminators and MAR unexpectedly provide a function when incorporated into the replicating vector, or that there is no difference between pRC and pR5, and that the 5' UTR is exclusively responsible for the improved gene expression within the pRC construct. This is plausible as the 5' UTR is an improved version of the commercially used HyperTrans UTR (Sainsbury and Lomonossoff 2008; Sainsbury et al. 2009; Peyret et al. 2019).

To my knowledge, no publication exists that has combined and tested UTRs within a TMV-derived viral vector. One study has assessed the effectiveness of replacing UTRs within CPMV and found that UTR truncations or full replacement diminishes expression levels (Meshcheriakova et al. 2014), and another has tested several UTRs within BeYDV (Diamos et al. 2016). However, combined synthetic UTRs have been tested in *E. coli* and *P. putida* (Le et al. 2020), as well as HEK293 cells (Cao et al. 2021), which found that multiple 5' UTRs in tandem can improve gene expression levels.

Furthermore, it is unlikely that the CaMV 35S promoter is providing functionality when transcribed within the TMV vector, as it is not thought to contain any internal ribosome entry sites (IRES; Ryabova and Hohn 2000). In addition, it has been reported that the TMV 3' UTR contains structures,

known as pseudoknot domains, that are essential for the translation, replication, and movement of TMV RNA (Zeenko et al. 2002). Thus, if the terminator elements in pRC and pRS are functional, the transcription termination should result in a polyadenylated sequence where the TMV 3' UTR is excluded from the transcript, suggesting that RNA replication and movement cannot occur in these constructs and that translation is occurring through a TMV 3' UTR independent mechanism. If this were true, then pRC and pRS would not be able to replicate, undergo cell-to-cell spread, or be translated any differently to their non-replicating counterparts, pNC and pNS. However, the data reported here shows a significant difference in transformation efficiency and gene expression when the simple and complex cassettes are utilised in the replicating vector, so at least one of the vector-enhanced functions must be maintained. Finally, less is known about the mechanism of MAR function, but they are known to bind the nuclear lamina and create a transcriptionally active domain (Dolgova and Dolgov 2019); this only occurs within the nucleus, so with our current understanding of MAR function it is highly unlikely that the MAR could provide a functional benefit at the RNA level in the cytoplasm. Overall, it was expected that pRC and pRS would have compromised expression levels because of these reasons, but the opposite was seen. This is why it was especially surprising to see that pRC-eGFP produced eGFP expression levels higher than all other construct variants, including the UTR variants which should not have such dramatic effects on TMV function.

3.4.2 Limitations and Future Work

There are several limitations associated with this research. The major limitation is variability in transformation efficiency which is variable between leaves and plants. Generally, older leaves have larger cells and larger gaseous spaces which *Agrobacterium* transformation solutions occupy during the infiltration process, enabling a better spread of transformation solutions within larger older leaves (Leuzinger et al. 2013). However, whilst younger leaves are substantially harder to transform using syringe-based infiltration methods, transformed cells express recombinant proteins at higher yields due to reduced protease activity (Robert et al. 2013). This means that there is a balance between leaf and plant size and age where the middle leaves of mature plants express recombinant eGFP better old and young leaves. Vacuum infiltration procedures can address this by ensuring an even infiltration procedure across the entire leaf surface rather than from syringe infiltration sites, and whilst these methods were tested and utilised, a lack of specialist equipment to measure the vacuum chamber pressure and ensure absolute consistency between samples hindered transformation efficiency, resulting in patchy expression even with our vacuum infiltration apparatus. For this reason, a dual infiltration approach was used to strike a balance between uniform transformation and expression levels, using standardised bacterial concentrations and vacuum infiltration times, however there was still significant variability in expression levels. Future

work should aim to utilise better vacuum infiltration equipment to maximise transformation efficiency and consistency between plants and leaves, maximising yields (Chen et al. 2013) and minimising the high standard deviation found here.

Another limitation is the lack of silencing suppressor effectiveness seen in this research. It is well established that p19 is an effective suppressor of silencing, found to be the best tested suppressor in one piece of research (Peyret et al. 2019). However, in this work, co-expression of p19 was found to induce a stress response within transformed plants, causing a dramatic reducing in eGFP expression. This finding was different to previous research reported by Shah et al. (2013) who found that co-expression of p19 with pJL-TRBO significantly improved GFP expression. Whilst the reason for this is unknown, optimisation of silencing suppressor co-expression could increase recombinant protein expression levels, particularly with replicating vectors that cause high levels of RNA accumulation within cells. As a result, future work should reduce the concentration of bacteria in transformation mixtures to assess whether this is causing the stress response, test other silencing suppressors such as p24 and p21 (Peyret et al. 2019; Mardanova et al. 2017) and attempt to fine-tune the co-transformation conditions by using different ratios of silencing suppressor and recombinant protein bacterial mixtures.

Additionally, as discussed, the use of terminators in the replicating vector, as seen in pRS and pRC, should create a cleaved transcript containing a polyadenylated tail, creating an mRNA that is isolated from the TMV 3’ UTR which is essential for TMV function (Zeenko et al. 2002). Despite this, it appears that the additional DNA elements enhance recombinant protein production, though the mechanism for this is unknown, and could have been elucidated by measuring the amount of target RNA within cells and comparing the relative differences of the *in planta* fluorescence and RNA levels between constructs. Alternatively, a novel expression construct could be developed using the same DNA elements in a different order that would not theoretically impair TMV function. Such a construct could include the use of the TMV 5’ UTR, synthetic CPMV-derived 5’ UTR, CDS, TMV 3’ UTR, double terminator and MAR, 5’ to 3’. This would ensure maximum recombinant protein production using only the additional 5’ UTR, as demonstrated by the Phase three constructs, whilst utilising the double terminator and MAR that improves expression levels as seen in pRC but without their incorporation into the replicated RNA. Assessing whether this construct improves expression levels could help elucidate the functions of each DNA element, and in theory would maximise transcription in the nucleus whilst maintaining TMV-based RNA replication and movement in the cytoplasm.

Furthermore, the transient expression constructs tested in this chapter showed variation when used in whole *N. benthamiana* plants compared to *N. tabacum* BY-2 cell cultures. In industrial settings it is preferential to utilise *N. tabacum* plants, which are larger and grow faster, producing higher yields of recombinant protein than the smaller *N. benthamiana* utilised here. However, these expression constructs have not been tested in whole *N. tabacum* plants, which could show different expression levels to those seen here. As a result, future work should assess expression levels in *N. tabacum* plants. This pairs with the vacuum infiltration method improvements discussed above, as vacuum agroinfiltration is utilised in industrial settings as it is not feasible to utilise syringe-mediated agroinfiltration with plants of this size or at large scales.

Alternatively, stably transformed plants would have been a useful tool to minimise the expression variability of constructs as the number of insertions can be validated during the creation of stable lines, and clonal plants can be obtained that express eGFP in every cell rather than just infiltrated areas seen in transient transformants. Whilst stable transformants are explored in Chapter 5, stable plant lines expressing eGFP with each construct in this chapter were not investigated as the design, build, test cycle would have been too slow. With transient transformants, data can be obtained in only one week, compared to many months for stable transformants because of the need to generate homozygous clonal lines. This strategy of testing constructs in transient transformants for use in stable transformations is employed in the literature (Shah et al. 2013). Nonetheless, producing stable plant transformants is a viable area for future work and further construct validation.

Similarly, quantification of transformed cells using confocal microscopy has some important limitations. Factors such as protein stability and folding, fluorophore maturation, and the efficiency of translation are all assumed to be equal between measured cells. Whilst these cannot be controlled for, differences in these factors would influence the values produced by quantification. More importantly, as mentioned in section 3.3.1.4, cells that are transiently transformed using agroinfiltration express transgenes from a variable number of T-DNA insertions, which was not controlled for in these experiments. Cells that carry more T-DNA insertions than others will express the transgene at higher levels. To control for this, constructs could be created using eGFP expression constructs with different regulatory elements, as seen here, which also contain a standardised transcriptional unit expressing a different fluorescent protein, such as Cyan Fluorescent Protein (CFP), using the same regulatory elements between constructs. An example of this can be seen in Figure 3.19 using the Phase 1 constructs presented in this chapter. Once differences in fluorophore maturation and bleaching have been accounted for, a ratiometric calculation can be performed which normalises the eGFP fluorescence against the CFP fluorescence, standardising the expression

and enabling more accurate quantification measurements with the number of T-DNA inserts controlled for.

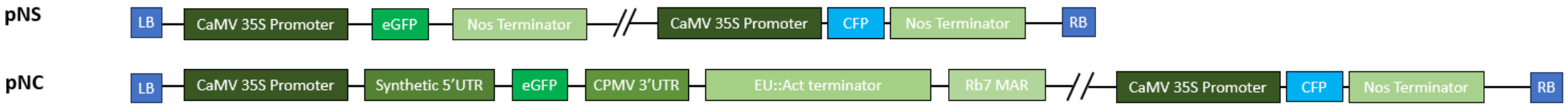


Figure 3.19: Suggested constructs that enable quantificaton of expression standardised to the number of T-DNA insertions. Utilising a second transcription unit expressing another quantifiable protein within the same T-DNA insertion enables standardisation of expression for each T-DNA molecule, removing differences in construct expression that are due to this. Once differences in the fluorescent protein properties are accounted for, these constructs would provide clearer information about the effects of the regulatory elements tested in Chapter 3 as differences in expression could be attributed to the regulatory elements rather than differences in transformation efficiency and T-DNA copy number.

Lastly, this chapter only analysed *in planta* eGFP expression. As the ultimate aim of this research is to develop a plant-based expression system for complex recombinant proteins, it is necessary to assess the eGFP expression levels at the protein level, and assess the expression levels of other proteins, as expression levels will vary depending on protein size, intracellular localisation, and post-translational modifications. As a result, the following chapter will assess protein expression levels between each of the discussed constructs, and the next chapter will assess the variation in protein expression using different proteins.

3.5 – Conclusions

The development of enhanced expression constructs is of paramount importance in this research, enabling eGFP expression at significantly higher levels than the initial expression construct. While this is an important step towards creating a viable plant protein production system, it is crucial to recognize that further investigation is required to determine the system's usefulness for more relevant recombinant proteins. Current assessments have been limited by only studying the expression of eGFP *in planta*; therefore, more in-depth experimental characterisation is necessary, especially at the protein level and with more complex proteins. The following two chapters of this research will explore these issues, focusing first on protein extraction, purification, and quantification. Next the versatility of the expression system will be assessed using proteins of varying complexities. These comprehensive analyses will provide valuable insights into the potential biological, medical, and industrial applications of this expression system.

Chapter 4 – Development of an Effective Protein Extraction and Purification Pipeline

4.1 – Chapter Summary

The research of the previous chapter demonstrated the successful development of an improved eGFP expression construct for use in transiently transformed *N. benthamiana*. However, methods used to measure expression were limited to *in planta* expression. The main aim of this research is to develop a plant recombinant protein expression system capable of producing recombinant eukaryotic proteins, meaning that the most important measurable factor is the overall yield of target protein. This chapter describes the developments made to improve target protein extraction and purification and shows that the expression constructs described in Chapter 3 result in high levels of extractable and quantifiable protein. This section details the development of optimised protein extraction procedures from *N. benthamiana* plants and investigates potential options for optimising the purification of extracted proteins. Following the testing of several C-terminal purification tags, an Octa-His tag was selected for the purification of eGFP using nickel beads, a process which also needed optimisation. Finally, once suitable protein extraction and purification methods were developed, plants transformed with the constructs described in Chapter 3 had eGFP extracted and quantified, which showed that pRC-eGFP produced the highest yield of extracted protein at 254-fold that of pNS-eGFP, and 16-fold that of pRI-eGFP, analogous to the previously published pJL-TRBO-GFP construct. The overall yield of eGFP produced by pRC-eGFP is estimated to be between 12-20% total soluble protein when quantified using RuBisCO as an internal control. According to the literature, this may result in a total concentration of target protein of between 0.4-4.0 mg of protein per gram of fresh leaf tissue (Mardanova et al. 2015; Lindbo 2007b).

4.2 – Introduction

Recombinant protein expression is influenced by many factors both at the expression level *in planta* and the extracted protein level. The previous chapter mostly investigated the considerations of the expression system and construct design, known as upstream processing. However, the downstream processing element of recombinant protein production represents the majority of costs (Singh & Herzer 2017; Schillberg et al. 2019). An effective expression system needs to be paired with effective downstream processing procedures, aimed to maximise recovered protein yield, ideally obtaining pure protein.

Here, plant expression systems have a major drawback relative to other expression systems, in that plant cells have a fibrous carbohydrate-rich cell wall, resulting in the need for additional steps when

extracting proteins. The disruption of the plant cell wall can be done mechanically or chemically, which each have their advantages and disadvantages. Mechanical disruption at small scales usually involves the use of cheap liquid nitrogen to freeze plant material, followed by the laborious grinding of leaf tissue into a fine powder, usually with a pestle and mortar. Conversely, chemical disruption is less laborious but usually much more expensive as chemical agents like cellulase are far more costly than liquid nitrogen.

Following disruption of the cell wall, proteins are liberated from the cell membrane, often using hypertonic solutions containing high molarity NaCl or sucrose. Extraction buffers usually have a buffering agent to maintain pH, such as Tris-HCl, salts to provide ionic strength and keep proteins in their native state, chelating agents such as EDTA to prevent metals from affecting proteins, reducing agents such as DTT minimise protein aggregation and denaturation, and often contain stabilising agents such as glycerol or sucrose which can also help reduce denaturation and aggregation (Leibly et al. 2012). The diversity of each of these agents means that there are many different extraction buffers in the literature, making it challenging to choose and optimise these buffers. Furthermore, different buffers have varying suitability for different recombinant proteins, as not all proteins require the same environmental conditions, meaning that extraction buffers should be carefully chosen on a case-by-case basis.

This complexity is increased further when downstream purification is also considered. Purification methods can often be very expensive depending on the required purity and protein complexity. Chromatographic techniques rely on separation of protein samples based on protein properties, such as size, ionic charge, or hydrophobicity (Coskun 2016). Alternatively, non-chromatographic techniques exist where proteins are purified through precipitation by salt concentration or pH (Wingfield 1998), filtration (Pires & Palmer 2021), or flocculation (Buyel and Fischer 2013). Purification tags can also be added to proteins to provide them with new properties that enable alternative purification methods and are often cleaved later once pure protein has been obtained (Pina et al. 2014).

This chapter documents the protein extraction and purification processes tested. Cell wall disruption was carried out mechanically using tissue grinding in liquid nitrogen, and protein extraction was tested using various extraction buffers from the literature. In addition, several purification methods were tested, and a histidine-tag was shown to be the most effective purification method. As a result, all constructs contained a C-terminal Octa-His tag for purification following total protein extraction. Plants transformed with each construct from Chapter 3 had total protein extracted and the relative amounts of eGFP quantified by western blot, which confirmed pRC-eGFP to be the highest

expressor. SDS-PAGE and InstantBlue™ staining was used to estimate the total yield of recombinant eGFP as a percentage of total soluble protein (% TSP) and as mg of protein per gram of fresh tissue weight (mg/g FW). Overall, this chapter shows the development of an effective protein extraction and purification pipeline which was used to quantify the amount of target eGFP produced by the constructs. This showed that pRC-eGFP produced the highest amount of target eGFP protein at 254-fold that of pNS-eGFP, and 16-fold that of pRI-eGFP, comparable to the previously published pJL-TRBO-GFP construct. The overall yield of eGFP produced by pRC-eGFP is estimated to be between 12-20% total soluble protein.

4.3 – Results

4.3.1 Development of an optimised protein extraction method to pair with purification

There are many different plant protein extraction methods in the literature, and as discussed above, a protein extraction buffer and method may vary on a case-by-case basis. As such, several extraction methods and buffers were tested in this research, each based on a successful published protein extraction method. In order to compare methods, quantification of the extracted protein is necessary. To do this, three methods were used; these were a Bradford assay to quantify the total amount of extracted protein in each method, SDS-PAGE separation of protein samples and either Coomassie blue staining of protein gels to reveal the identities and relative amounts of each extracted protein, or western blotting using an anti-eGFP primary antibody and a fluorescent secondary antibody, enabling highly specific, sensitive and accurate quantification of extracted eGFP. As a result, each extraction method was analysed by Bradford assay and Coomassie blue-stained SDS-PAGE gels, but only promising extraction methods were analysed by western blot. This section documents the developments in protein extraction. Importantly, RuBisCO is the most abundant plant protein, with eight copies of both the large (RbcL) and small (RbcS) subunit in the complex that produce a dominant protein band at around 50 kDa in size, corresponding to RbcL, when extracted effectively. As this protein is highly abundant in all plant tissue, amounting to between 30-50% of total soluble protein (Lin et al. 2021; Robert et al. 2015; Schmidt et al. 2021), it can be used as an indicator to measure the effectiveness of an extraction method, with a darker and cleaner band being indicative of a more successful extraction method. For most methods described here, leaf tissue was first finely ground in liquid nitrogen before suspension in protein extraction buffer. If promising, a histidine-pulldown assay was attempted in order to develop an effective extraction and purification pipeline. For method development, the tissue for each extraction method was the highest expressing tissue available at the time of experimentation. This is usually eGFP expressing tissue, but early experiments expressed rat ballast P2X7 fused to a C-terminal eGFP (see Chapter 5) in the simple expression construct, pNS.

The very initial protein extraction tested was utilising a minimal extraction buffer and benchtop fractionation method. Briefly, protein extraction buffer 1.0 (PEB 1.0) contained PBS and 1 x cOmplete protease inhibitor at pH 7.5 and was mixed with plant tissue expressing rat ballast P2X7 in a 2:1 v/w ratio. Debris was separated by centrifugation at 4,500 *x g*, detergent added, and the total protein separated via centrifugation at 20,800 *x g*. This method was chosen as an initial setup as the reagents and equipment necessary are readily available. Using *N. benthamiana* plants transiently transformed with pNS-rbP2X7, this extraction method was tested and whilst some protein was extracted as indicated by the Bradford assay, there were no visible bands when analysed by SDS-PAGE and stained using Coomassie blue, suggesting a poor extraction efficacy (Figure 4.1, Panel A). Nonetheless, a western blot was carried out on protein extracts from transiently transformed *N. benthamiana* leaves expressing human and rat P2X7 under control of the pNS construct (Figure 4.1, Panel E) using an anti-eGFP primary antibody. This failed to produce any distinct bands and the method was deemed unsuitable.

Following this, a second extraction buffer was created and tested utilising protein extraction buffer 2.0 (PEB 2.0), containing 50 mM HEPES-KOH, pH 7.6, 150 mM NaCl, 10% v/v glycerol, 0.25% FC-12, 1 mM PMSF, and 1 x cOmplete protease inhibitor, adapted from Qi and Katagiri (2009). This buffer was tested with rat ballast P2X7 expressing tissue in a 4:1 v/w ratio, initially using a benchtop fractionation method. Briefly, this involved a 4,000 *x g* centrifugation step to remove debris. The supernatant was filtered through double-layer Miracloth and centrifuged at 22,000 *x g* to separate the cytosolic proteins (supernatant) from the membrane fractions (pellet). The pellet was resuspended in PEB 2.0 containing 1% FC-12, incubated on ice, then centrifuged again to separate membrane proteins from debris. Quantification of extracted protein by Bradford assay suggested a successful extraction, and analysis by SDS-PAGE and Coomassie staining showed a single band at about 250 kDA, and a faint band at 75 kDa (Figure 4.1, Panel B). This method was also used to extract protein from tissue transformed with empty vector, and the two samples compared by western blot (Figure 4.1, Panel F), which showed no distinct band or difference between tissue samples.

As this produced only limited success a different fractionation method, adapted from Qi and Katagiri (2009), was tested using the same buffer. Instead, this utilised an initial debris separation spin of 18,000 *x g*, Miracloth filtration of the supernatant, and a 100,000 *x g* spin to separate the cytosolic fraction from the membrane fraction. The pellet was resuspended in PEB 2.0 containing 1% FC-12 and incubated on ice, followed by a final centrifugation at 100,000 *x g* to isolate the membrane proteins from debris. Again, quantification of extraction by Bradford assay suggested successful extraction of protein, but SDS-PAGE and Coomassie staining was unable to produce distinct bands

(Figure 4.1, Panel C), suggesting that this method was unsuitable. This method was also used to extract protein from full-length, cysteine-rich, and ballast domain rat P2X7 expressing tissue, and was analysed by western blot (Figure 4.1, Panel G), showing no difference between samples. The fraction containing debris produced an extremely dark band, suggesting that one of the antibodies used binds to a component of this debris. As a result, debris fractions were not analysed in future western blots. Of all the extraction buffers tested, this is the only buffer that utilised glycerol; whilst glycerol can be beneficial to maintain protein stability, it can interfere with SDS-PAGE, which may be the reason for the lack of distinct bands using this method. Nevertheless, as SDS-PAGE was the chosen method for comparing extraction conditions, this buffer was deemed unsuitable.

The next extraction method tested involved the use of ATPS, adapted from Dong, Ding and Wang, (2019) mentioned in Chapter 1. Once this buffer was tested, the pRC-eGFP construct had been created so 0.5 g of leaf tissue was tested with 4 volumes of protein extraction buffer 3.0 (PEB 3.0) containing 50 mM Tris-HCl, 150 mM NaCl, and 10 mM EDTA, pH 8.0. Debris was removed using a 12,000 *x g* centrifugation step and the supernatant subject to ATPS using a mixture of 5 M NaCl, saturated ammonium sulphate, and ethanol followed by centrifugation and separation of the upper phase using N-butanol, centrifugation and retaining of the lower phase, thought to contain extracted protein. This method was analysed by Bradford assay which suggested successful protein extraction, and SDS-PAGE separation followed by Coomassie blue staining showed faint bands, with the darkest bands at approximately 50 kDa, corresponding to the RuBisCO proteins of the large subunit (Figure 4.1, Panel D). This was promising, so the same sample was also analysed by Western blot, which produced some faint bands, including a band at approximately 27 kDa where eGFP is expected (Figure 4.1, Panel H). However, this protein separation was not clear enough to justify using as the main extraction method, considering the additional cost of reagents.

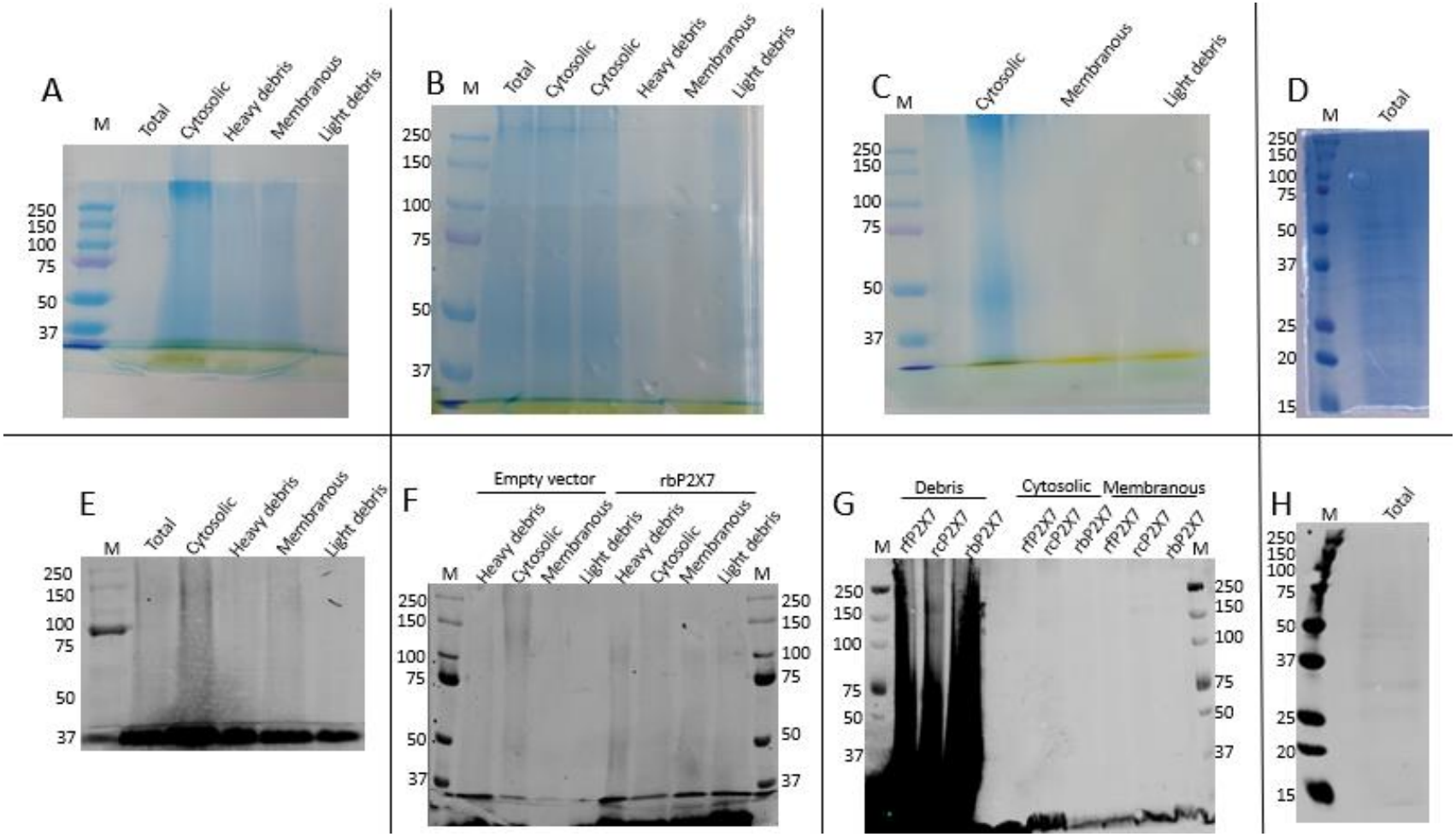


Figure 4.1: SDS-PAGE images of tested protein extraction methods with either InstantBlue™ staining (A-D) or western blotting (E-H) on tissue expressing recombinant proteins. A – PEB 1.0 method on tissue expressing pNS-rbP2X7. B – PEB 2.0 benchtop method on tissue expressing pNS-rbP2X7. C – PEB 2.0 ultracentrifuge method on tissue expressing pNS-rbP2X7. D – PEB 3.0 method on tissue expressing pRC-eGFP. E – western blot of A with the same lanes. F – western blot of protein extracts from *N. benthamiana* tissue transiently transformed with either empty pICH47742 vector or pNS-rbP2X7. G = western blot of protein extracts from *N. benthamiana* tissue transiently transformed with either pNS-rfP2X7, pNS-rcP2X7, or pNS-rbP2X7, showing debris, cytosolic, and membranous fractions for each. H = western blot of D. In all panels Lane M is the BioRad Precision Plus dual colour marker, with MW in kDa displayed alongside the corresponding band. For all western blots, a polyclonal rabbit anti-GFP primary antibody (1/5000 v/v) and a secondary 800nm IRDye goat anti-rabbit antibody (1/13500 v/v) were used.

The next method tested utilised buffer 4.1 (PEB 4.1), a different modified buffer from Qi and Katagiri (2009), containing 50 mM Tris-HCl, 330 mM sucrose, 5 mM EDTA, 5 mM DTT and 1 x sigma plant protease inhibitor cocktail. *N. benthamiana* tissue expressing pRC-eGFP was suspended in PEB 4.1 in a 1:4 w/v ratio, and subject to a debris removal step of 2,000 *x g* followed by the same 100,000 *x g* centrifugation method utilised with PEB 2.0. When analysed by Bradford assay, protein was extracted successfully. When separated by SDS-PAGE and stained using Coomassie blue, clear bands can be seen on the resulting gel, shown in Figure 4.2, Panel A, with the darkest band present at 50 kDa, corresponding to RuBisCO proteins. There are also darker smears at around 27 kDa, but the bands are not distinct so it is difficult to assess whether eGFP can be quantified using this method. A western blot was also performed and showed a marked improvement with distinct though non-specific bands showing that the antibody used was not specific to eGFP (Figure 4.2, Panel B).

As PEB 4.1 was promising, the buffer was modified slightly, replacing the 330 mM sucrose with 165 mM NaCl, creating PEB 4.2. Again using *N. benthamiana* leaf tissue expressing pRC-eGFP, the total protein was extracted utilising an almost identical fractionation method, with the only difference being a 4,000 *x g* spin instead of a 2,000 *x g* spin for debris removal. Analysis of protein extraction by Bradford assay suggested successful protein extraction, and separation of proteins by SDS-PAGE and Coomassie staining showed much clearer bands, suggesting a more effective protein extraction. Figure 4.2, Panel C shows the Coomassie stained gel, with very dark and clear bands at 50 kDa, representing RuBisCO. There is also a clear band at approximately 27 kDa which is thought to be eGFP, though it is more prominent in the membranous fraction. The results of this gel appeared promising, so another SDS-PAGE was performed followed by a western blot using the polyclonal rabbit anti-eGFP primary antibody, and an infrared goat anti-rabbit secondary antibody. Initially, four fractions were analysed which included the heavy debris, cytosolic, membranous, and light debris fractions. Similar to PEB 2.0, the debris fraction produced a very strong signal (Figure 4.2, Panel D), confirming the incompatibility of these antibodies with cellulose-rich fractions. As a result, the western blot was repeated with only the cytoplasmic and membranous fractions analysed; this produced a relatively clean blot, with an eGFP signal in the cytosolic fraction as well as several non-specific bands (Figure 4.2, Panel E). However, the main aim of using a western blot was to quantify eGFP specifically, and the non-specific bands suggested that the use of this primary antibody was not sufficient without purification.

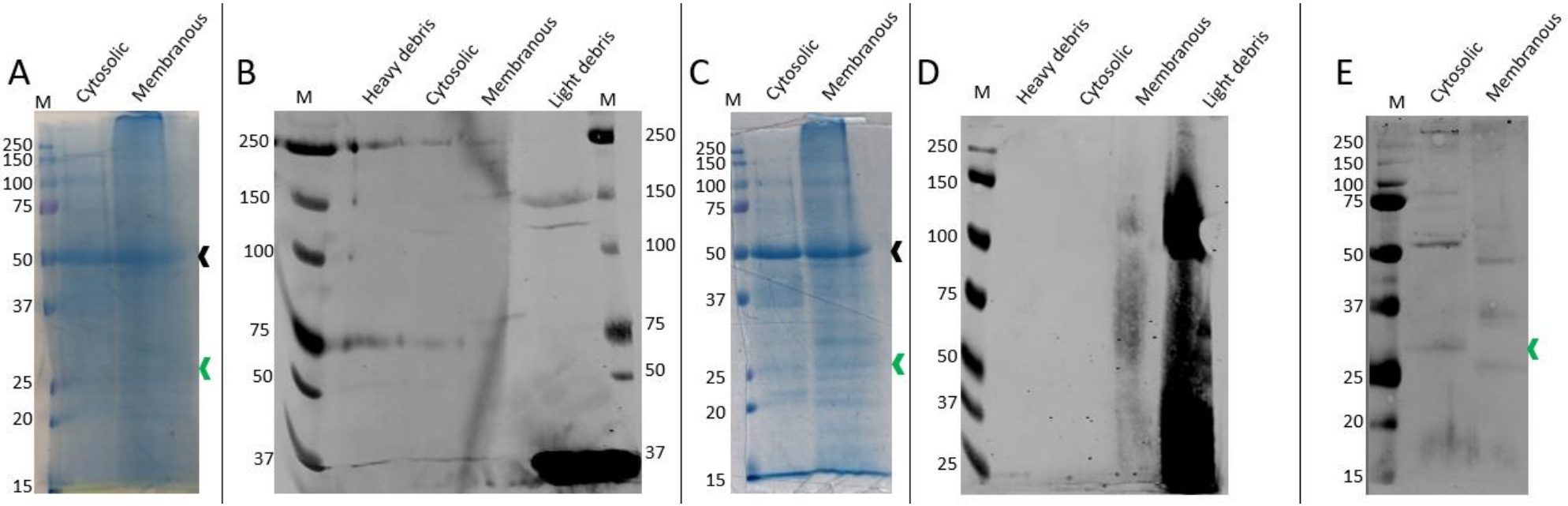


Figure 4.2: SDS-PAGE results using PEBs 4.1 and 4.2. A – pRC-eGFP expressing tissue extracted using PEB 4.1 method and stained with InstantBlue™ showing the cytosolic and membranous fractions. pRC-eGFP expressing tissue extracted using PEB 4.1 method and visualised using a western blot. Lanes show the heavy debris, cytosolic fraction, membranous fraction, and light debris. C – Same as A using PEB 4.2. D – same as C using PEB 4.2. E – Same as D but excluding debris fractions. In all panels Lane M is the BioRad Precision Plus dual colour marker, with MW in kDa displayed alongside the corresponding band. Green arrows indicate 27 kDa, the MW of eGFP, and black arrows indicate 50 kDa, the MW of RuBisCo proteins. For all western blots, a polyclonal rabbit anti-GFP primary antibody (1/5000 v/v) and a secondary 800nm IRDye goat anti-rabbit antibody (1/13500 v/v) were used.

As purification appeared necessary to improve the clarity of the western blots, an overnight histidine-pulldown was performed on both the cytosolic and membranous fractions using loose nickel beads, with the protein eluted the following day. This was possible as all eGFP sequences used in this research have a C-terminal Octa-His tag to enable purification, except for constructs described later in this chapter when alternative purification tags were tested instead. Both the supernatant and eluent fractions were analysed by western blot (Figure 4.3, Panel A), which showed a faint band at 80 kDa in the cytoplasmic eluent sample. Underexposure showed several more bands, including a band in each lane at approximately 27 kDa which could be eGFP (Figure 4.3, Panel B). This suggests that the pulldown method works but is non-specific and requires optimisation. It also suggests that the antibody used was non-specific and needed to be changed.

As a result, this western blot was repeated but this time using a different antibody, an anti-plant-GFP monoclonal rat antibody. An infrared goat-anti-rat secondary antibody was used for visualisation. This western blot, shown in Figure 4.3, Panel C, produced very distinct individual bands in the eluents corresponding to the proteins from the cytoplasm and membrane, respectively, that were eluted from the nickel beads. This was a breakthrough, showing that PEB 4.2 was effective for protein extractions and could successfully be paired with a histidine-pulldown on nickel beads. It also demonstrated that the new monoclonal anti-eGFP antibody was much more specific than the original and was a suitable primary antibody for quantification of western blots.

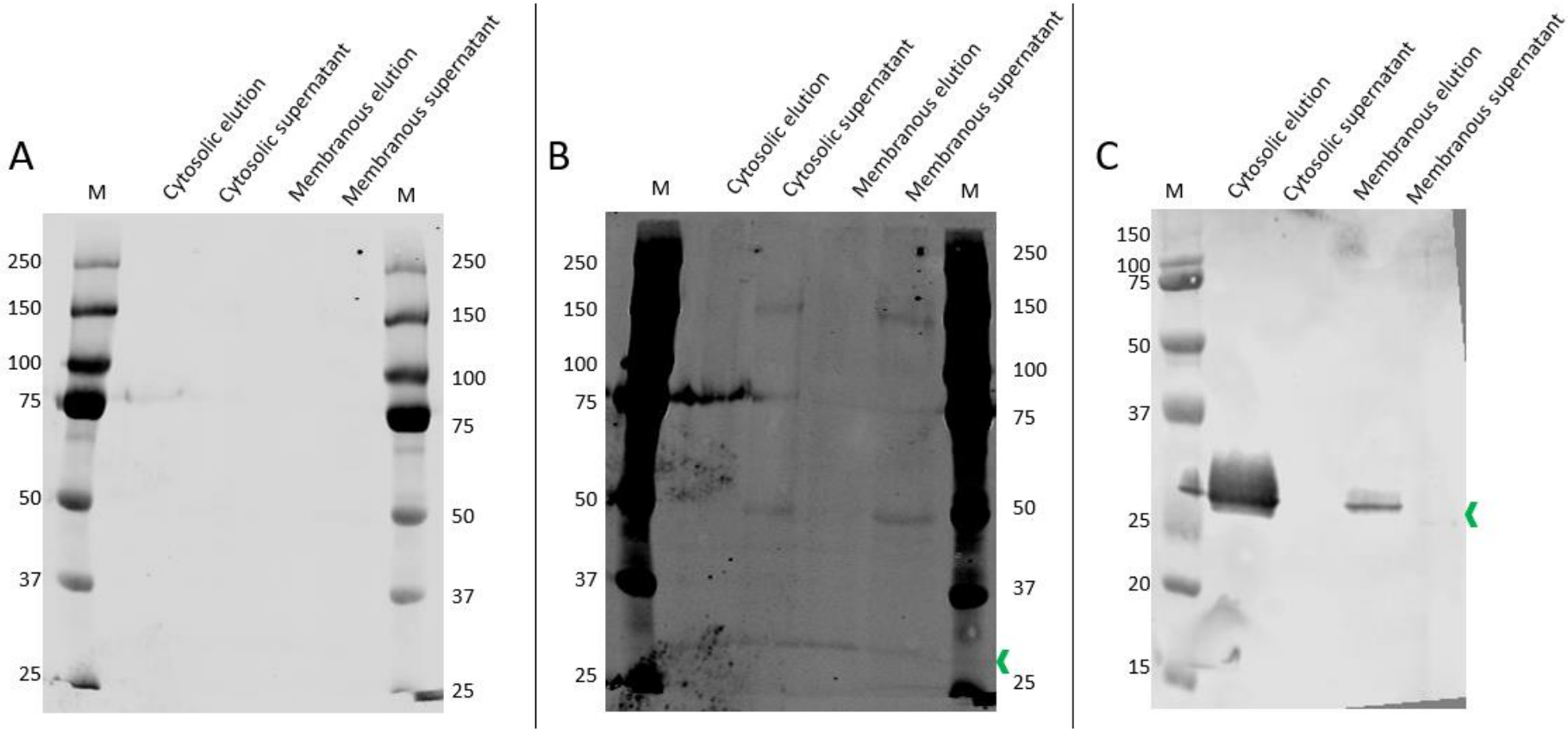


Figure 4.3: Western blots of protein extracted from pRC-eGFP expressing tissue using PEB 4.2 with different antibodies, following histidine pulldown. A – western blot using polyclonal rabbit anti-GFP primary antibody (1/5000 v/v) and a secondary 800nm IRDye goat anti-rabbit antibody (1/13500 v/v). Lanes show the eluent of cytoplasmic fraction, supernatant of cytoplasmic fraction, eluent of membranous fraction, and supernatant of membranous fraction. B – same as A but underexposed to show more bands. C – Same as A but using a monoclonal mouse plant-specific anti-GFP antibody (1/5000 v/v) followed by a secondary 800nm IRDye goat anti-mouse antibody (1/13500). In all panels Lane M is the BioRad Precision Plus dual colour marker, with MW in kDa displayed alongside the corresponding band. Green arrows indicate 27 kDa, the MW of eGFP.

As this method was promising, a scaled-up version of the experiment was performed using 8 g of the same tissue, and only 1 mM DTT in the extraction buffer. Extracted tissue was only subject to the initial centrifugation step without separation of cytosolic and membranous fractions. A his-pulldown was performed, and all fractions had proteins separated by SDS-PAGE and stained using Coomassie blue. The following fractions were analysed: the total protein extraction, the unbound supernatant during the overnight nickel-binding, a washed sample from the beads containing weakly bound proteins, the eluent containing the strongly bound proteins, and the remaining nickel beads following elution, containing any proteins that were not eluted (Figure 4.4). Analysis of these fractions showed two things; firstly, some eGFP bound the nickel beads and was eluted producing a single band on the SDS-PAGE, suggesting relatively high purity following elution. Secondly, much of the histidine-tagged eGFP remains in the unbound fraction following overnight nickel bead binding, suggesting that the binding process is inefficient and needs optimisation.

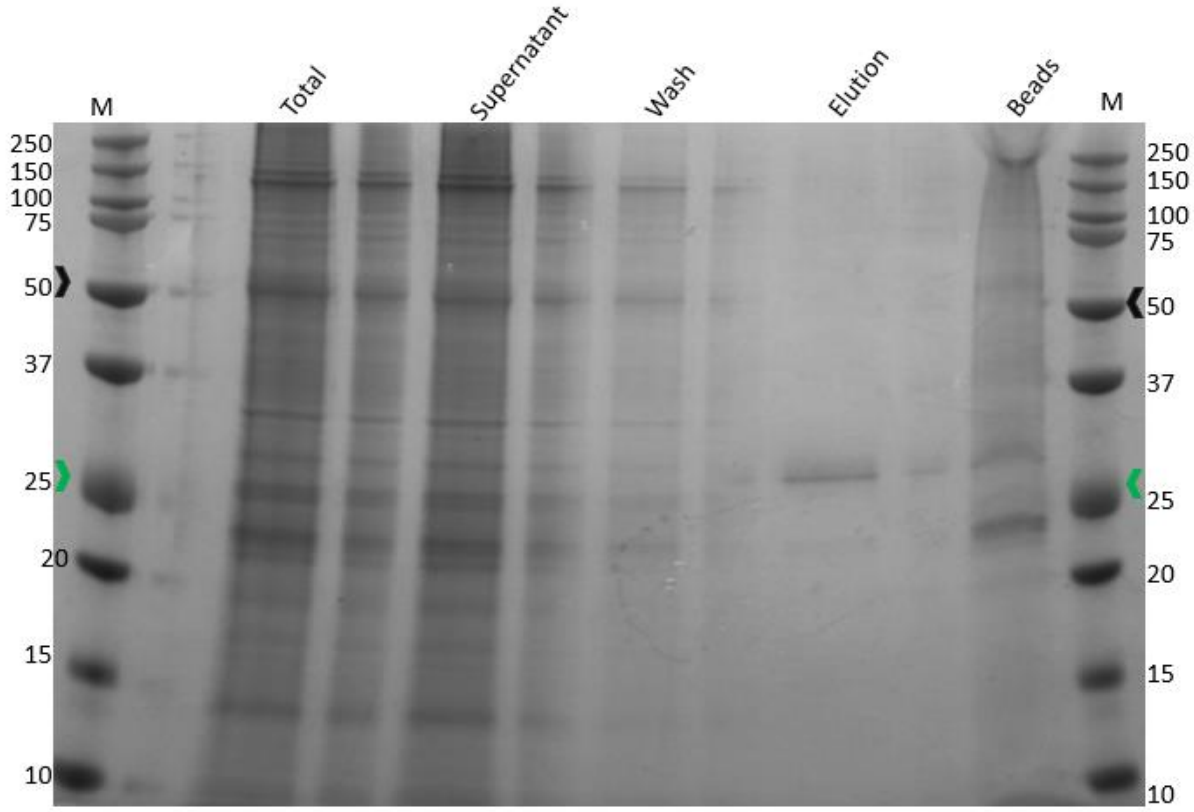


Figure 4.4: InstantBlue™ stained SDS-PAGE gel following a scaled-up histidine-pulldown of protein extracted from pRC-eGFP expressing *N. benthamiana* tissue using PEB 4.2. Lanes show the total protein before pulldown, the supernatant representing the unbound fraction, the wash removing loosely bound proteins, eluent of bound proteins, and the bead fraction, of proteins that were not eluted. Lane M is the BioRad Precision Plus dual colour marker, with MW in kDa displayed alongside the corresponding band. The green arrow indicates 27 kDa, the MW of eGFP, and the black arrow indicates 50 kDa, the MW of RuBisCo proteins.

The chemical compositions of PEB 4.2 and the his-pulldown buffers are relatively different, which could cause ineffective binding of extracted protein to nickel beads. In case a more optimal combined method could be developed, a third variant of the NaCl buffer was created, PEB 4.3, containing 50 mM phosphate, pH 8.0, 10 mM Tris, pH 8.0, 500 mM NaCl, 0.1% Tween-20, 0.1% FC-12, and 1 x sigma plant protease inhibitor, adapted from Lindbo (2007b), chosen as it was reported to function as both the extraction buffer and nickel-binding buffer. 5 g of plant tissue was extracted using this buffer in a 1:4 w/v, the sample filtered through Miracloth and centrifuged at 12,000 *x g* for 20 minutes, and a sample was taken. Overnight binding to nickel beads was performed using PEB 4.3 containing 20 mM imidazole. Following, the sample centrifuged to sediment the beads and the supernatant was retained, the beads washed once in extraction buffer, and the protein eluted from the beads using PEB 4.3 containing 500 mM imidazole. Analysis of the fractions by SDS-PAGE and Coomassie blue staining (Figure 4.5) suggested that this method could successfully extract protein, as several protein bands can be seen in the total fraction. However, there is no eGFP band in the eluent or the wash fractions, suggesting that the eGFP did not bind. Conversely, the supernatant from the overnight binding, representing the proteins that did not bind the beads, does contain a protein of around 27 kDa. Interestingly, many of the proteins seen in the total fraction are not present in any of the lanes, suggesting that they may remain bound to the beads or have been degraded as the few bands present in the eluent are faint and smeared. Overall, this extraction method was effective, but was worse than PEB 4.2 for the purification step so was not investigated further.

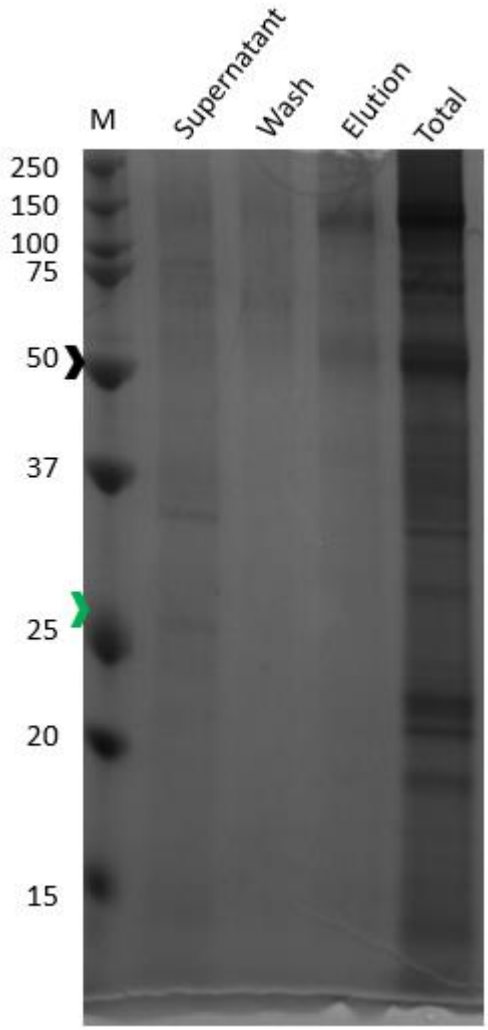


Figure 4.5: InstantBlue™ stained SDS-PAGE gel following a histidine-pulldown of protein extracted from pRC-eGFP expressing *N. benthamiana* tissue using PEB 4.3. Lanes show the supernatant representing the unbound fraction, the wash removing loosely bound proteins, the eluent of bound proteins, and the total protein fraction. Lane M is the BioRad Precision Plus dual colour marker, with MW in kDa displayed alongside the corresponding band. The green arrow indicates 27 kDa, the MW of eGFP, and the black arrow indicates 50 kDa, the MW of RuBisCo proteins.

Finally, another extraction buffer was made, PEB 5.0, containing 20 mM sodium phosphate, pH 7.5, 150 mM NaCl, 1% (v/v) Triton X-100, 5 mM DTT, 1 mM PMSF and 1 x sigma plant protease inhibitor cocktail, chosen as this more closely reflects the recommended binding conditions suggested for use with nickel beads. To test this, 1 g of pRC-eGFP expressing *N. benthamiana* tissue was ground and the protein extracted in PEB 5.0 in a 1:3 w/v ratio. The sample was centrifuged at 15,000 *x g* for 20 minutes at 4°C, the supernatant filtered through double layer Miracloth, and mixed in a 1:1 v/v ratio of sodium phosphate binding buffer. Nickel beads were added to this solution overnight at 4°C to allow binding. Following, the sample was centrifuged to sediment the beads and the supernatant removed. The beads were washed with sodium phosphate wash buffer, and the protein eluted using sodium phosphate elution buffer. Alongside this, 0.5 g of tissue had protein extracted using the same method, but instead purified using a HisTrapHP column, and the total, flow-through, wash, and elution fractions were taken. Each fraction was analysed by SDS-PAGE and Coomassie blue staining and can be seen in Figure 4.6. This extraction method was highly effective, with clear bands seen in the total protein lane. In both variations of the purification, the protein bound strongly to the nickel resin as no eGFP band can be seen in the supernatant or the wash fractions. In the eluents of both experiments there is a prominent eGFP band at 27 kDa, but other proteins are present suggesting the need for purification optimisation.

FPLC automation using the HisTrapHP column failed to work and is explored later in this chapter. As a result, the centrifugal purification method using loose beads was chosen for tissue extraction at scale. However, this failed to produce clean elutions due to contaminating plant material also being pelleted during centrifugation steps. For this reason, the final and optimal purification method involved the use of magnetic nickel beads, enabling washing to be performed without sedimentation of contaminating plant matter. This is discussed in more detail in the following chapter. In addition, a tabular summary of the protein extraction buffers employed in this research and their results can be seen in Table 4.1 below.

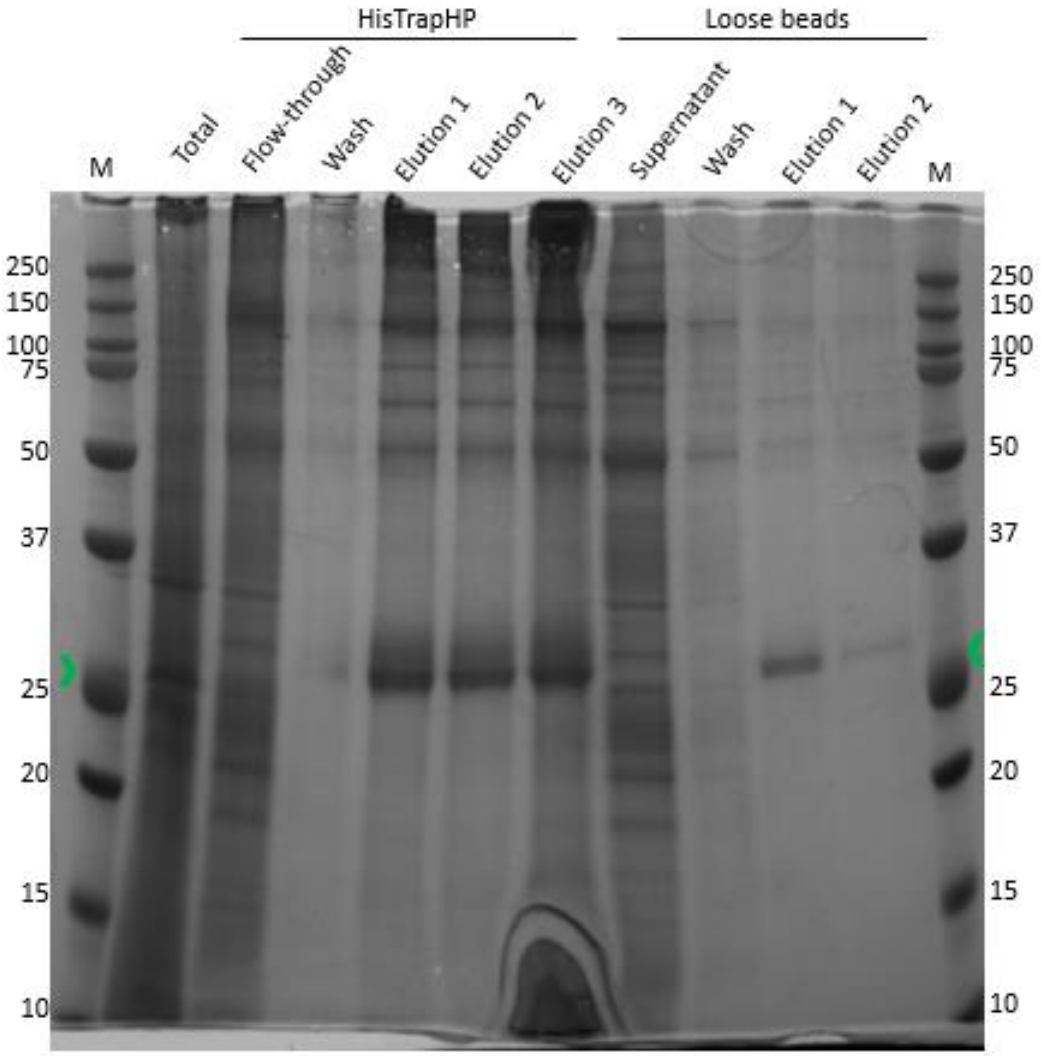


Figure 4.6: InstantBlue™ stained SDS-PAGE gel on protein extracted from pRC-eGFP expressing *N. benthamiana* tissue using PEB 5.0, followed by a manual histidine-pulldown either on a column or on loose beads. Two histidine pulldowns were performed either manually by hand using a HisTrapHP column using 0.5 g of tissue, in which three elutions were performed which are each shown, or overnight on loose Ni Sepharose® High Performance beads using 1 g of tissue, after which two elutions were performed. Lane M is the BioRad Precision Plus dual colour marker, with MW in kDa displayed alongside the corresponding band. The green arrow indicates 27 kDa, the MW of eGFP.

Table 4.1: Protein extraction buffers used in this research, showing reagents used and a summarised result.

Protein extraction buffer	Reagents	Result
1.0	Phosphate-buffered saline, pH 7.4, and 1 x Roche cOmplete protease inhibitor	Distinct protein bands could not be seen when analysed by SDS-PAGE and InstantBlue™ staining or western blot.
2.0	50 mM HEPES-KOH, pH 7.6, 150 mM NaCl, 10% v/v glycerol, 0.25% FC-12, 1 mM PMSF, 1 x cOmplete protease inhibitor	Distinct protein bands could not be seen when analysed by SDS-PAGE and InstantBlue™ staining or western blot.
3.0	50 mM Tris-HCl, pH 7.4, 150 mM NaCl, and 10 mM EDTA, pH 8.0	Some faint bands seen when analysed by SDS-PAGE and InstantBlue™ staining or western blot, but clarity was too poor to justify further use.
4.1	50 mM Tris-HCl, pH 7.4, 330 mM sucrose, 5 mM EDTA, 5 mM DTT and 1 x sigma plant protease inhibitor cocktail	Distinct but faint protein bands were visible when analysed by SDS-PAGE and InstantBlue™ staining or western blot.
4.2	50 mM Tris-HCl, pH 7.4, 165 mM NaCl, 5 mM EDTA, 5 mM DTT and 1 x sigma plant protease inhibitor cocktail	Distinct and clear protein bands were visible when analysed by SDS-PAGE and InstantBlue™ staining or western blot. Protein was able to bind nickel beads at small scales, but not large scales.
4.3	50 mM phosphate, pH 8.0, 10 mM Tris, pH 8.0, 500 mM NaCl, 0.1% Tween 20, 0.1% FC-12, and 1 x sigma plant protease inhibitor	Distinct and clear protein bands were visible when analysed by SDS-PAGE and InstantBlue™ staining or western blot. However, protein was unable to bind nickel beads.
5.0	20 mM sodium phosphate, pH 7.5, 150 mM NaCl, 1% (v/v) Triton X-100, 5 mM DTT, 1 mM PMSF and 1 x sigma plant protease inhibitor cocktail	Distinct and clear protein bands were visible when analysed by SDS-PAGE and InstantBlue™ staining or western blot, and protein was able to successfully bind nickel beads at all scales tested.

4.3.2 Optimisation of protein purification

4.3.2.1 Design, engineering and testing of purification tags

Mid-way through testing extraction conditions several other purification tags were investigated in case a histidine-tag was not the optimal purification tag to utilise. There many different purification tags that can be used but several have been utilised for protein purifications from plant extracts, and four were selecting for testing. These are a synthetic HPB tag (Qi and Katagiri 2009), a cellulose binding domain tag (CBD) with a self-cleaving intein (Islam et al. 2018), and two antibody-mediated purification methods for anti-FLAG and anti-GFP (DYKDDDDK Fab-Trap and GFP-Trap Magnetic Particles M-270, respectively). The HPB, CBD, and FLAG tags were synthesised using GeneWiz as a DNA fragment containing terminal BpiI overhangs enabling digestion and ligated into a Level -1 Golden Gate vector, to act as a 3' C-terminal eGFP tag. Importantly, eGFP was also amplified using high-fidelity PCR to create overhangs that also enables digestion and ligation into the Level -1 vector as the 5' fragment. When the 5' eGFP and the 3' tags, in their respective Level -1 vectors, were digested and ligated into the Level 0 coding sequence module, the three new C-terminally tagged eGFP variants were created. This method is identical in principle to the creation of the double terminator Level 0 module discussed in Chapter 3. Once the eGFP-tag variants were created, these were assembled into the complex Level 1 constructs, and then digested and ligated into pJL-TRBO, creating the final expression constructs. For testing the anti-GFP antibody purification method, the original pRC-eGFP construct was used, which contains the Octa-His tag.

Each of these tags were assessed together in transiently transformed *N. benthamiana* leaves. The *in planta* fluorescence of each eGFP-tag variant was analysed qualitatively using a photon imager, and the images of each construct can be seen in Figure 4.7, Panel A, showing a subset of the best expressing leaves for each construct. These initial analyses suggested that the HPB tag resulted in the highest expression, followed closely by the Octa-His tag and the FLAG tag, with the CBD tag resulting in low-to-no expression. In order to assess this further, the leaves were quantified using photon imaging to measure the expression more accurately, shown in Figure 4.7. In fact, this showed that the Octa-His tag resulted in the highest expression, followed by the HPB-tag, FLAG-tag, and CBD-tag with the lowest fluorescence. These data confirmed that the Octa-His, HPB and FLAG tag all result in similar levels of expression with no significant difference between each other, but a significant difference from the CBD tag and the empty vector control. It also confirms that the CBD-tag results in diminished fluorescence which is not significantly different from the negative control.

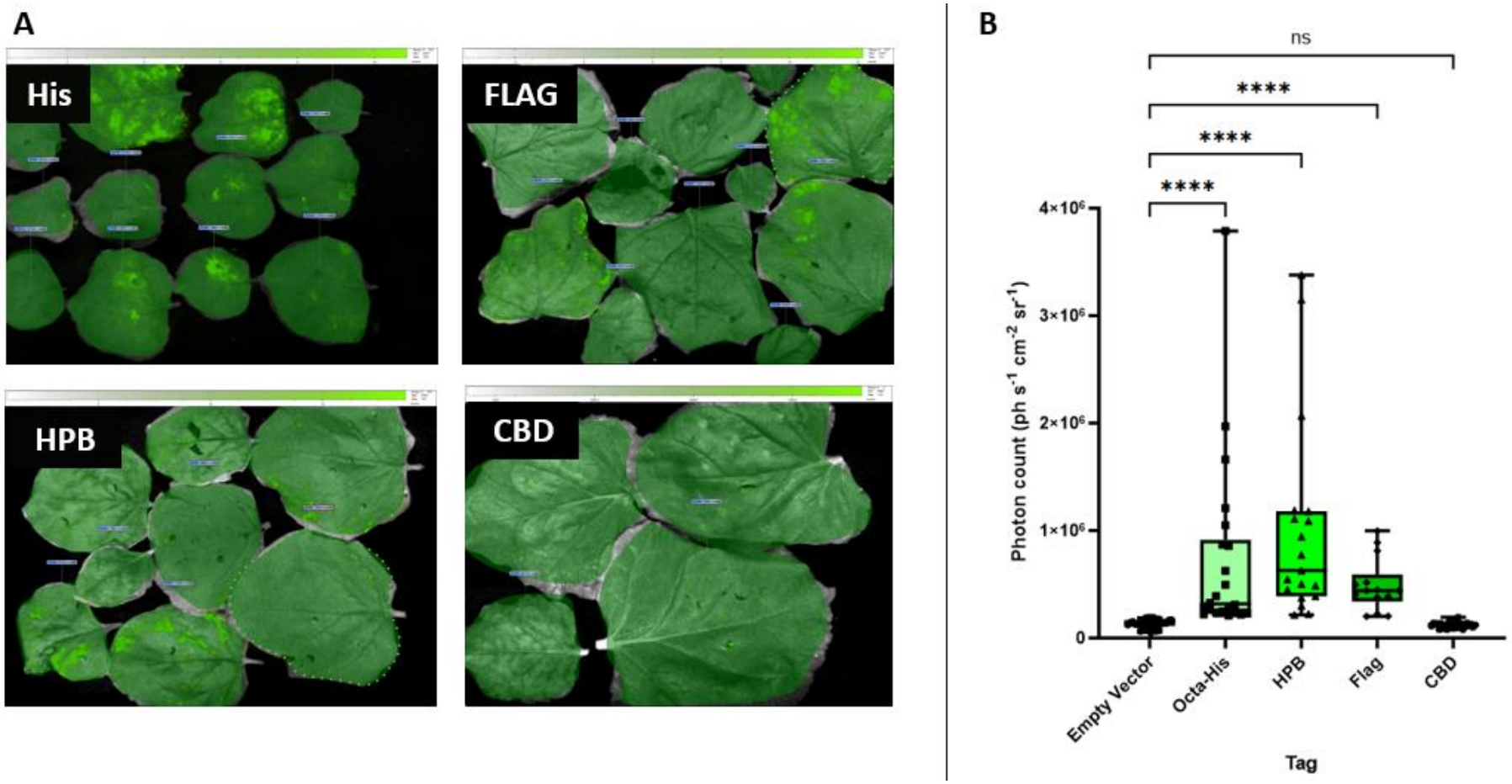


Figure 4.7: Fluorescence analysis of *N. benthamiana* leaves transformed with pRC-eGFP with different C-terminal purification tags. Four plants were transformed (n=4) using each coding sequence variant, but datapoints show technical replicates with 5-6 leaves taken per plant. A – Photon imager photographs of whole leaves where whole leaves are manually highlighted to measure expression, not using the magic wand tool. B – Photon imager quantification showing that the Octa-His, HPB, and FLAG tags do not impact protein fluorescence, with no statistically significant difference from each other ($p > 0.999$) when analysed using a Dunn's multiple comparisons test, but a highly statistically significant difference from the negative control leaves and the CBD tag ($p < 0.0001$). The fluorescence from leaves transformed with the CBD tag was not statistically significant from the negative control leaves ($p > 0.999$). A kruskal-wallis test showed an overall statistically significant difference between samples ($H_{(4)} = 86.59$, $p = 2.2 \times 10^{-16}$). Significance values on the graph are: ns = not significant, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$. Error bars show minimum and maximum values, with the highest variance seen in leaves transformed with eGFP tagged with the Octa-His and HPB tags.

Transformed leaves were finely ground in liquid nitrogen and had expression analysed at the protein level using PEB 4.2, as PEB 5.0 had not yet been developed. Analysis of extraction by Bradford assay suggested that protein was successfully extracted from each sample, and a purification assay was attempted using the method from each tag (described in Chapter 2). For the anti-GFP antibody, tissue expressing the His-tag was used. Importantly, the use of these different tags produces products of various molecular weights; the Octa-His-tagged, FLAG-tagged, HPB-tagged, and CBD-tagged eGFP variants have molecular weights of 27 kDa, 27 kDa, 39 kDa, and 63 kDa, respectively.

Initially, an 8% SDS-PAGE gel was used to separate samples but this failed to produce the appropriate resolution, with the appropriate molecular weights for all tags except CBD running off the gel. Nonetheless, a western blot was used to analyse purification efficiency and specificity using each method, which can be seen in Figure 4.8, Panel A. However, the lane containing the CBD-tagged protein was thought to contain residual cellulose as the western blot produced a very dark band resembling that of the early western blots of the debris extracts. This smear spanned most of the lane, with several other distinct non-specific bands, making it difficult to tell purified protein and non-specific signal apart. Whilst it is possible that further optimisation of the CBD-tag may make it a viable choice, its impact on *in planta* fluorescence indicates that it prevents proper protein folding and makes *in planta* fluorescence comparisons impossible. As a result, a second 12% SDS-PAGE was performed excluding the CBD-tagged sample and was analysed by western blot. This can be seen in Figure 4.8, Panel B, and shows that even the poorly optimised Octa-His tag produced by far the darkest band, in both the cytoplasmic and membranous fractions, suggesting that this was the better purification. The anti-GFP antibody pulldown also had the correct band but contained several others, suggesting that it was not specific, and neither the HPB-tagged variant or the FLAG-tagged variant could be seen in their respective lanes. Consequently, the Octa-His tag was chosen as the main method of purification, as it was shown to be a very effective and relatively inexpensive option, with well-established automation protocols available.

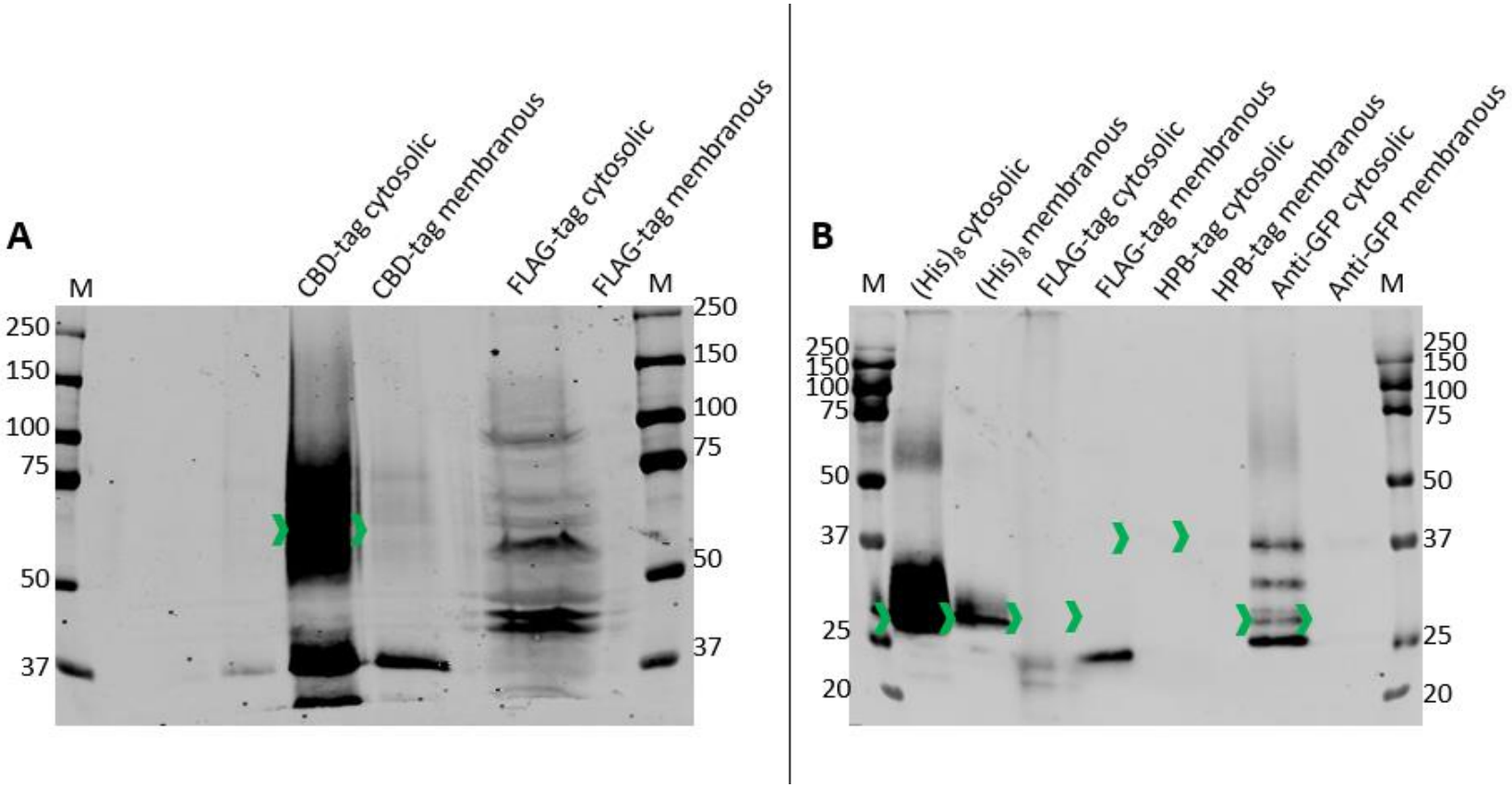


Figure 4.8: Western Blot Analysis of pRC-eGFP Extracted with Different Purification Tags using PEB 4.2. A – Western blot analysis to detect eGFP using a mouse plant-specific-anti-GFP primary antibody (1/5000) and a goat-anti-mouse 800nm secondary antibody (1/13500) on an 8% SDS-PAGE gel. Lanes show the CBD-tagged eGFP protein from the cytosolic and membranous fractions, and the FLAG-tagged protein from the cytosolic and membranouse fractions. B – Western blot analysis to detect eGFP using a mouse plant-specific-anti-GFP primary antibody (1/5000) and a goat-anti-mouse 800nm secondary antibody (1/13500) on a 12% SDS-PAGE gel. Lanes show the cytosolic and membranous fractions of eGFP tagged with the Octa-His tag, FLAG-tag, HPB-tag, and the His-tagged eGFP purified using the Anti-eGFP antibody pulldown. In both panels, Lane M is the BioRad Precision Plus dual colour marker, with MW in kDa displayed alongside the corresponding band and the green arrows indicate the expected MW of the purified product in each sample. These data show that the CBD tag may work but produces substantial degradation and several non-specific products. The histidine tag produces the cleanest purification with substantial yields, followed by the anti-GFP antibody pulldown method, though there are several smaller bands suggesting degradation. Neither the FLAG-tag nor HPB tag produced a band of the correct size.

4.3.2.2 Fast Protein Liquid Chromatography

As the original Octa-His was chosen as the best purification tag, efforts were made to optimise the histidine-pulldown process. Fast protein liquid chromatography using automated systems is the best way to minimise batch variability and ensure consistency in purifications, so this method was experimented with using AKTA purification systems. Initially, PEB 4.2 and PEB 4.3 method variants were utilised for extraction, and the pulldown was performed using the Tris-HCl method in Chapter 2. As previously seen in this chapter, this purification method worked with overnight binding in a falcon tube followed by manual washing and elution in microcentrifuge tubes. As a result, the same components were tested using a 1 mL HisTrapHP column and FPLC automation using an AKTA Prime. Several purification methods were attempted using PEB 4.3, but A280 absorbance by FPLC determined that no protein bound the column, and instead was lost during the loading phase (Figure 4.9, Panel A). These data were confirmed by SDS-PAGE analysis of collected fractions. Denaturing proteins can improve interactions with resins so the protein sample was denatured in 8 M urea and the purification re-tested. The denatured protein also failed to bind the column, shown in Figure 4.9, Panel B). As a result, it was determined that FC-12 may be preventing binding, due to its zwitterionic nature meaning it may be able to bind both the nickel ions and the histidine residues, preventing interaction and column binding.

When PEB 5.0 was used to extract plant tissue, manual purifications using a HisTrapHP column or overnight binding on nickel beads worked very effectively (see Figure 4.6). However, when automating the system using an AKTA prime, the protein still did not bind to the insert resin with a 1 mL/min flowrate (Figure 4.9, Panel C). The manual column purification utilises a higher flowrate and pressure which can improve protein-resin interactions, though this is variable due to human error. To assess whether the issue was with the AKTA Prime or the flowrate and pressure an AKTA Start was tested. Again, the protein failed to bind the resin when a 1 mL/min flowrate was used (Figure 4.9, Panel D), so a 4 mL/min flowrate was used but this also failed to improve binding (Figure 4.9, Panel E). To enable a further increase in flowrate, a 5 mL HisTrapHP column was used instead of the 1 mL, and a flow rate of 20 mL/min but still showed that the protein would not interact with the resin (Figure 4.9, Panel F). For this reason, loose nickel beads were used to scale the purification of future samples. However, due to contaminating plant tissue with the centrifugal wash method, magnetic nickel beads were used for the highest purity purifications, as seen in the following chapter.

Finally, hydrophobic interaction chromatography was tested using a HiTrap™ Capto™ HIC selection kit of 1 mL columns, including a HiTrap Capto -Phenyl, -Phenyl ImpRes, -Butyl, -Butyl ImpRes, and -Octyl column. Each of these, shown in Figure 4.10, Panels A-E, displayed no A280 peaks after the

loading phase, suggesting that no protein was binding any resin. Consequently, FPLC based purification methods were abandoned, and manual purification methods chosen instead. This concluded the optimisation of protein purification, with manual purification of Octa-His tagged eGFP determined to be the most effective method.

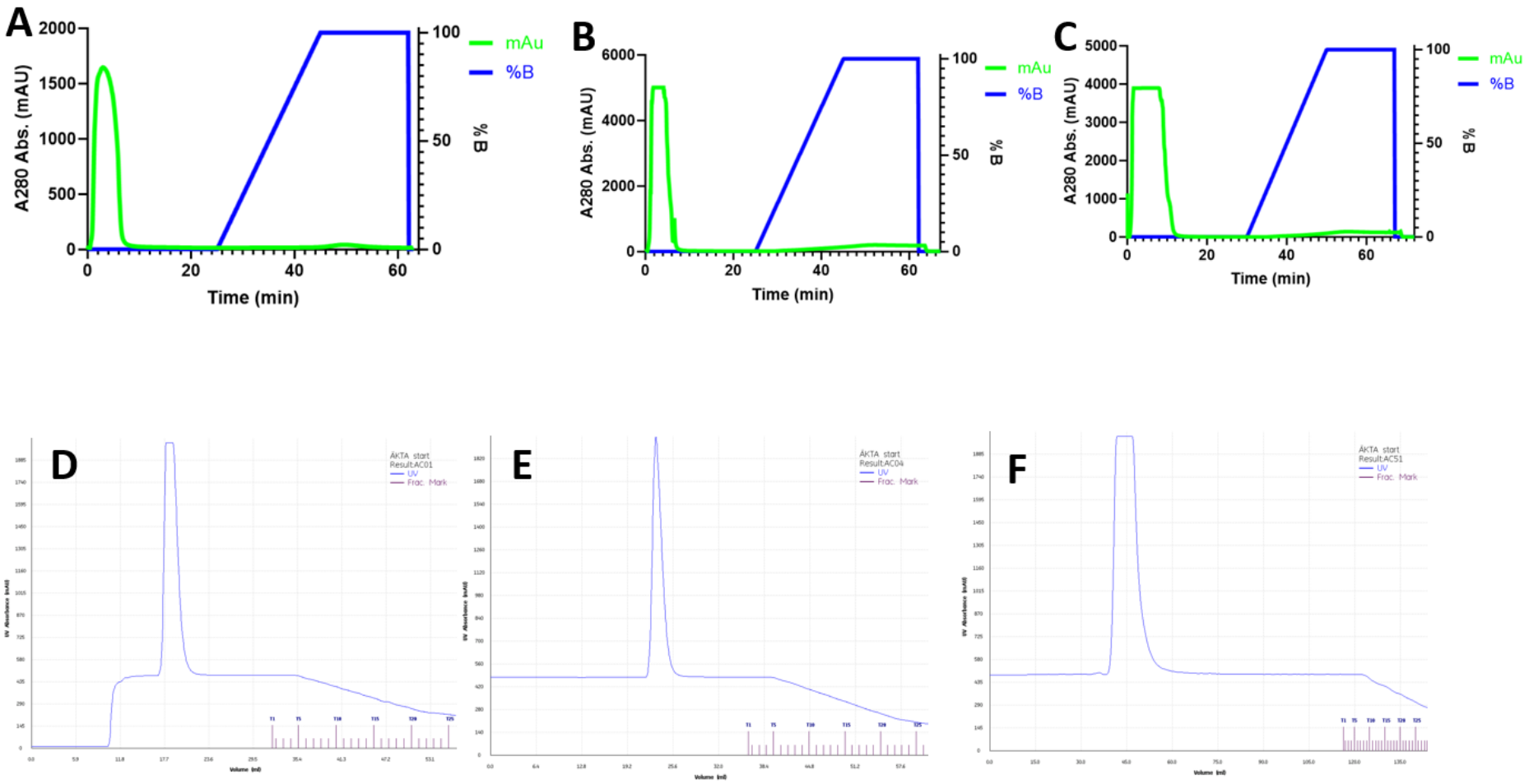


Figure 4.9: A280 traces of AKTA runs using HisTrapHP columns. In all panels tissue expressing pRC-eGFP was extracted and loaded onto each column. In all panels, the single high peak shows the protein eluting from the column during the loading phase, suggesting that protein does not bind the resin. A – 5 mL protein extract using an AKTA Prime in PEB 4.2. Flow rate used was 1 mL/min. B – 5 mL protein extract denatured in 8 M urea using AKTA Prime in PEB 4.2 and a flow rate of 1 mL/min. C – 10mL protein extract using an AKTA Prime in PEB 5.0 and a flow rate of 1 mL/min. D – 1 mL of extract used on the AKTA start with a 1 mL column and 1 mL/min flow rate. E – 1 mL of extract used on the AKTA start with a 1 mL column and 4 mL/min flowrate. F – 5 mL of extract used on the AKTA start with a 5 mL column and a flowrate of 20mL/min. In D to F, the X axis shows volume in mL, and the Y axis A280 absorbance in arbitrary units. In all panels tissue expressing pRC was extracted and loaded onto each column. In all panels, the single high peak shows the protein eluting from the column during the loading phase, suggesting that protein does not bind the resin.

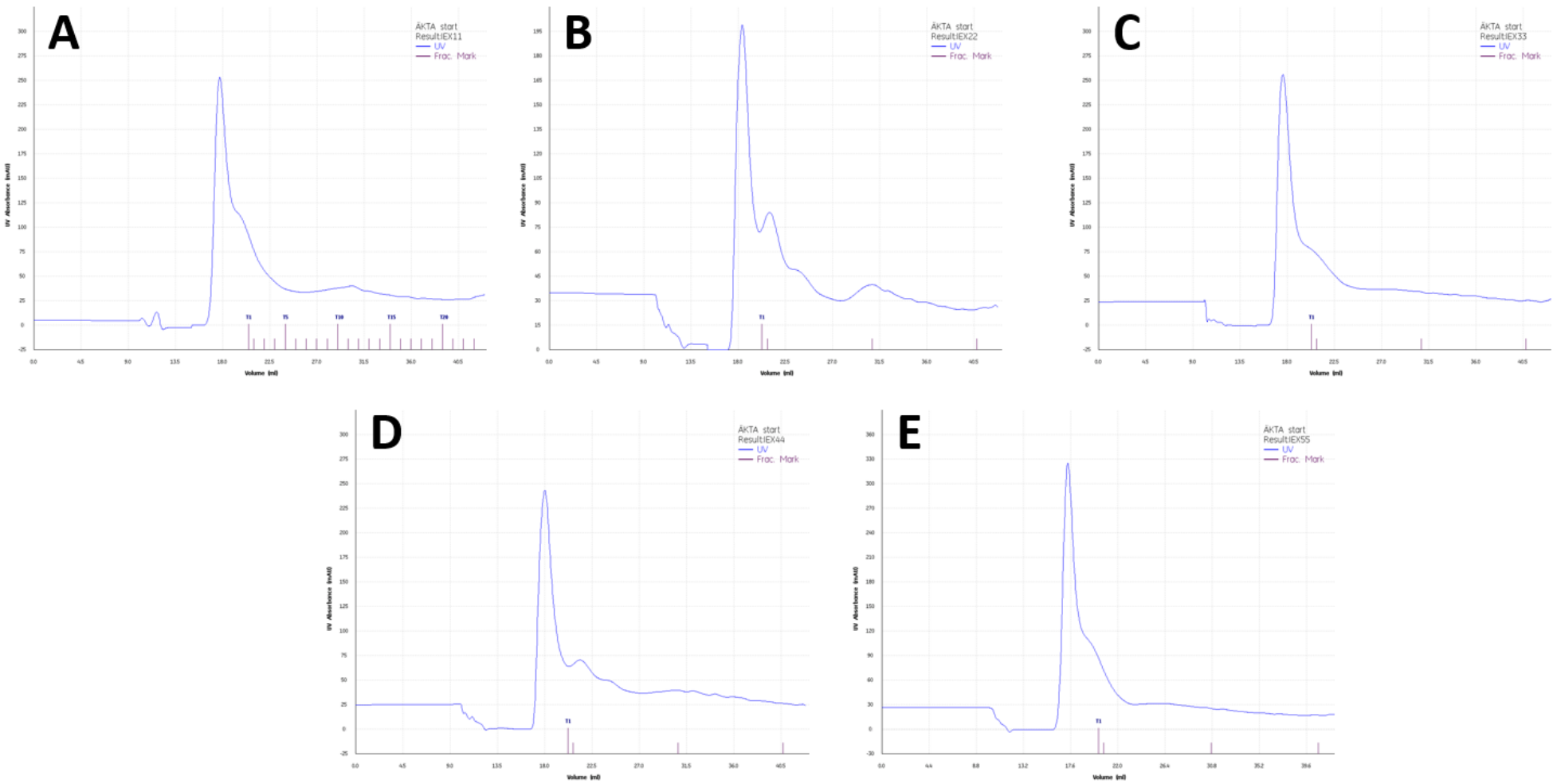


Figure 4.10: A280 traces of AKTA Start runs using the HiTrap™ Capto™ HIC selection kit. A – HiTrap Capto Phenyl column. B – HiTrap Capto Phenyl ImpRes column. C – HiTrap Capto Butyl column. D – HiTrap Capto Butyl ImpRes column. E - HiTrap Capto Octyl column. All columns used are 1 mL with a 1 mL/min flowrate. The X axis shows volume in mL, and the Y axis A280 absorbance in arbitrary units. 0.5 mL of pRC-eGFP extracted tissue solution was loaded onto each column. In all panels, the single high peak shows the protein eluting from the column during the loading phase, suggesting that protein does not bind the resin.

4.3.3 Quantification of extracted protein from expression constructs

With a protein extraction method optimised and an Octa-His tag determined to be optimal for purification, the constructs developed in Chapter 3 could be quantified. Three *N. benthamiana* plants transformed with each of the eGFP constructs had total protein extracted and separated by SDS-PAGE and stained with InstantBlue™. Figure 4.11, Panel A shows the total protein extracts of plants transformed with constructs pNS-eGFP, pNC-eGFP, pRS-eGFP, and pRC-eGFP, and Panel B shows an empty vector construct, pRC-eGFP, pRI-eGFP, pRU-eGFP, pR5-eGFP and pR3-eGFP, alongside a BSA loading control. According to the literature, RuBisCO comprises between 30-50% of total soluble protein in *N. benthamiana* (Lin et al. 2021; Robert et al. 2015; Schmidt et al. 2021), enabling relative quantification of the eGFP band in the pRC-eGFP sample. As a result, densitometric analyses were performed using the pRC-eGFP 27 kDa eGFP band and the 50 kDa RbcL band in Figure 4.11, Panel A. Each lane in this figure represents the pooled protein extracts from transformed leaves taken from three plants, a method chosen to reduce the variability seen from the *in planta* fluorescence data from each construct in Chapter 3, and reduce resource consumption across experiments. Pooling samples in this way also enables purification of target proteins resulting in increased signal, which may be necessary to detect lower abundance target proteins explored later in Chapter 5. As a result, each eGFP and RbcL band represents the average of three plants but is a sample size of one, as standard deviation and error cannot be calculated. This comparative analysis suggests that pRC-eGFP results in eGFP production at between 12-20% TSP. Comparing this to estimates in the literature, this corresponds to a lower estimate of 0.4-0.67 mg of protein per gram of fresh leaf weight (Mardanova et al. 2015), and a higher estimate of 1.44-4 mg of protein per gram (Lindbo 2007b). The protein extracts were also examined by western blot, enabling more accurate comparative quantification of the eGFP signals. Figure 4.11, Panel C shows the raw signal from the Phase 1 and 2 constructs, with Panel D showing quantification. Similarly, Figure 4.11, Panel E shows the signal from the Phase 3 constructs, alongside pRC-eGFP and an empty vector control, with Panel F showing quantification. As three plants were pooled for each sample to account for the variation seen in Chapter 3, statistical analyses are not possible. However, these analyses suggest that pRC-eGFP produces the highest eGFP fluorescence, with 254-fold, 16-fold, and 17-fold increased expression to pNS-eGFP, pRS-eGFP, and pNC-eGFP, respectively, and 16-fold, 17-fold, 1.76-fold, and 3.87-fold higher than pRI-eGFP, pRU-eGFP, pR5-eGFP and pR3-eGFP, respectively.

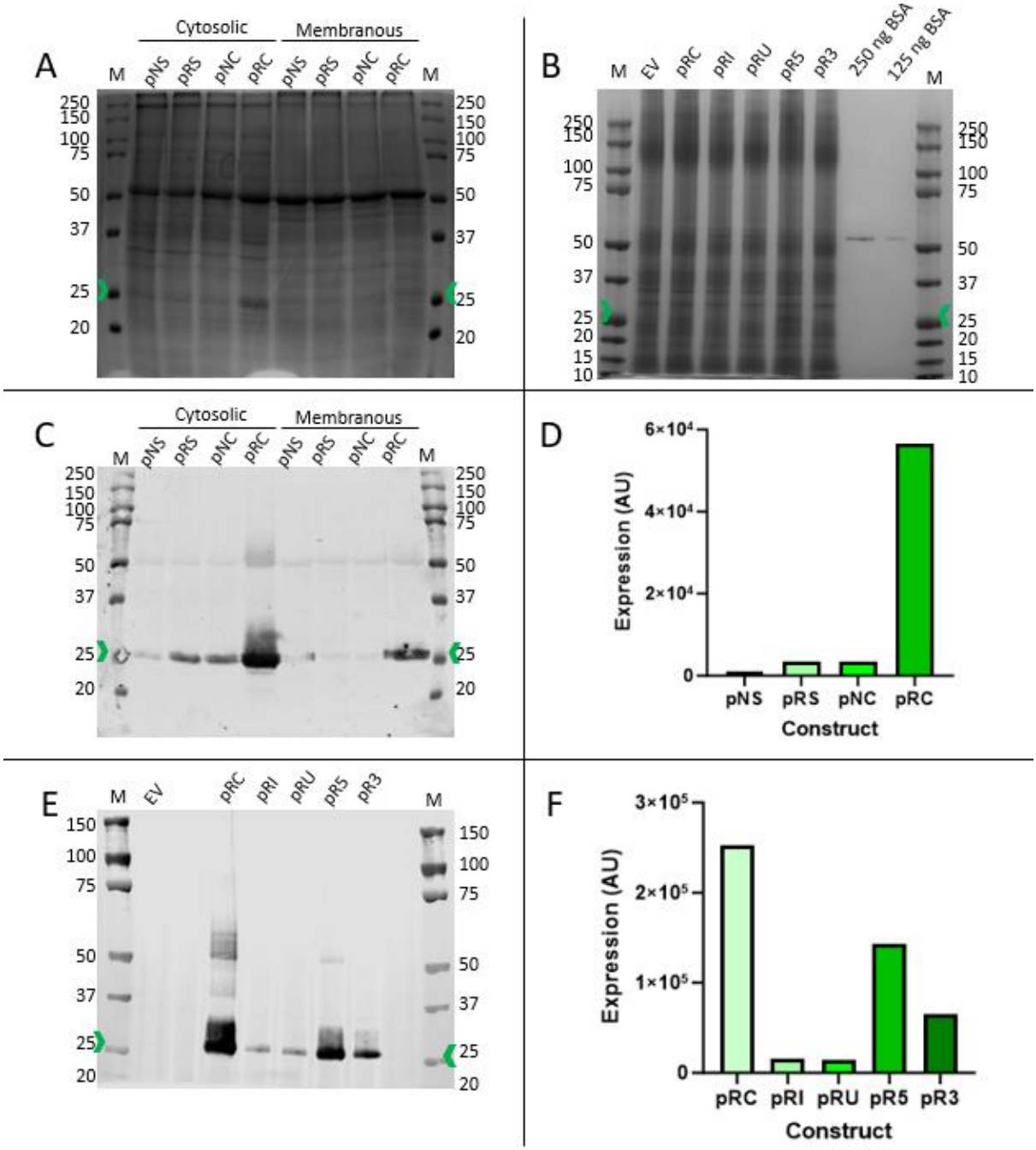


Figure 4.11: SDS-PAGE and Western Blot Analyses of Extracted Proteins from Different Samples. For each construct, three plants (n=3) were transformed with the constructs expressing eGFP, and protein extracts were pooled. **A** – an InstantBlue-stained SDS-PAGE gel showing the protein extracts from plants transformed with Phase 1 and Phase 2 constructs using PEB 4.2. **B** – an InstantBlue-stained SDS-PAGE gel showing the protein extracts from plants transformed with Phase 3 constructs and pRC using PEB 5.0. **C** - Western blot analysis to detect eGFP using a mouse plant-specific-anti-GFP primary antibody (1/5000) and a goat-anti-mouse 800 nm secondary antibody (1/13500). Lanes are the same as A. **D** – Quantification of the cytosolic bands in C. **E** – Western blot performed as C using samples shown in B, without BSA. **F** – Quantification of the bands in E. In all panels, Lane M is the BioRad Precision Plus dual colour marker, with MW in kDa displayed alongside the corresponding band and the green arrow indicates 27 kDa, the MW of eGFP. These data show that pRC-eGFP produces the highest eGFP expression over all other constructs. Raw densitometric analysis data can be seen in appendix Figure S4.1.

4.4 – Discussion

This chapter explored several different protein extraction and purification methods, aiming to optimise downstream processing whilst minimising costs. This section covers the effectiveness of different extraction and purification methods tested and delves into why some experimental conditions were more suitable than others. It discusses the challenges and limitations, compares the data obtained with existing literature, and discusses areas for future work. Finally, critically assesses whether downstream processing has been optimised and is suitable for a plant-based protein expression system.

4.4.1 Protein extraction

Whilst several different buffers were suitable for protein extraction, not all were compatible with purification. Commonly used plant protein extraction buffers utilise low osmolarity to lyse cell membranes and liberate cellular components, and each buffer utilised here contains salts or sucrose for this purpose. Other common reagents include detergents to solubilise membranes, reducing agents to reduce protein aggregation, protease inhibitors to reduce protein degradation, and a buffering agent to maintain a stable pH.

The unsuccessful buffers lacked classes of components that benefit protein extractions. PEB 1.0 contained phosphate-buffered saline with protease inhibitors, and lacked reducing agents and detergents, so may have been ineffective at extracting proteins from sub-cellular compartments and preventing aggregation. PEB 2.0 included glycerol and FC-12 as a stabiliser and detergent, respectively, but lacked a reducing agent like DTT, which could be crucial for breaking disulfide bonds to liberate proteins. This buffer was taken from Qi and Katagiri (2009), which was effective at purifying low abundance *A. thaliana* membrane complexes in their research, but failed to be effective when employed here. PEB 3.0 was paired with an ATPS purification method, but lacked reducing agents, detergents, and protease inhibitors, with a higher pH that may be unsuitable for many proteins. This buffer was taken from Dong, Ding and Wang (2019), which effectively purified eGFP when paired with the ATPS method, and was compatible with the HIC purification columns tested in this chapter, however these data could not be reproduced in this research. These first three extraction buffers failed to produce distinct bands when analysed by SDS-PAGE, whereas later buffers did.

In contrast, the effective extraction buffers now contained more essential reagents. PEB 4.1 utilised sucrose instead of salt to reduce osmolarity and stabilise proteins, but also contained EDTA, DTT, FC-12, and a plant protease inhibitor cocktail. The presence of DTT likely assisted in breaking disulfide bonds and reducing protein aggregation. EDTA may chelate divalent cations and reduce protease

activity but was later removed as it can have a significant effect on immobilised metal affinity chromatography, the chosen method of purification. As a result, PEB 4.2 tested the effectiveness of sodium chloride instead of sucrose and removed the EDTA. The substitution of sucrose for salts improved band quality when analysed using SDS-PAGE and InstantBlue™ staining and could be due to promoting a more favourable ionic environment improving protein stability. The FC-12 in both PEB 4.1 and 4.2 likely assisted in liberating compartmentalised proteins, but likely interfered with purification due to its zwitterionic nature. Consequently, PEB 5.0 included Triton X-100, a non-ionic detergent, which can effectively solubilise membrane proteins. It also includes a phosphate buffer, instead of Tris-HCl, encouraged for use with His-tag purification. This fine tuning resulted in marked improvements in protein extraction, but further experimentation could provide a more effective extraction buffer. Increasing the DTT concentration and adding more general protease inhibitors such as PMSF could further improve the clarity of extracts, but there is a balance between DTT concentration and His-tag purification efficiency, as nickel beads have a reported maximum DTT concentration of 5 mM before purification efficacy is affected.

4.4.2 Protein purification

Several purification methods were tested, each taken from successful examples in the literature, chosen for the range of purification mechanisms and costs. However, these showed limited success when tested, with the original Octa-His tag outperforming other purification tags and methods, both in terms of yield and purification. The HPB tag was chosen as it has been successfully employed in plant production systems to purify low abundance membrane proteins and appeared extremely suitable for this research (Qi and Katagiri 2009). However, there was no evidence of successful purification when tested which may be due to the different extraction buffer used in this experiment compared to their published data. In contrast, the CBD tag was taken from Islam et al. (2018) and may have been effective at purifying protein as the western blot using this purification method produced a large signal at the appropriate molecular weight, though it is difficult to definitively say whether that this was true signal or an artefact of the use of cellulose as seen in earlier western blots. In addition, several non-specific bands were also present, suggesting that there was substantial degradation, and *in planta* fluorescence was compromised with the use of this large purification tag, making upstream expression quantification inaccurate. The CBD purification utilised very cheap reagents, so this method is worth optimising in the future to minimise downstream processing costs. The antibody-based methods to purify the FLAG-tag and eGFP directly each produced non-specific bands with several products. Furthermore, the high cost of antibody-based purification methods makes these unfavourable for large scale purifications. As a result, a histidine pulldown was chosen as the purification method for future experiments. However, FPLC automation

of the his-purification proved difficult, with binding only occurring using manual methods, despite several experimental conditions being tested. Manual purification using the column was effective, though cannot be done with multiple samples at the same time, so purification using loose beads was chosen, followed by a series of washes after bead pelleting by centrifugation. Whilst effective, the centrifugation step often pelleted plant material despite the vast majority of this being removed during the protein extraction steps, which utilised high-speed centrifugations and MiraCloth filters. Consequently, magnetic nickel beads were used for the final purification method, as these can be separated from the protein sample without centrifugation, resulting in fewer cleaner and more effective washes, and the ability to elute purified protein in a smaller elution volume, maximising purity and yield. This could then be paired with spin-column protein concentration to maximise protein concentration.

There are many other potential avenues that could have been explored for protein purification, including both chromatographic and non-chromatographic techniques. The histidine-pulldown demonstrated here was performed with a single elution of 500 mM imidazole, and produced moderately pure eGFP, but unwanted products were also present. Improving the purity could require fine tuning of the his-pulldown elution conditions, performing multiple stepwise elutions to remove unwanted products at lower imidazole concentrations, and perform a final elution to obtain pure recombinant protein. However, this would require strict control over the conditions used, ideally with constant flow using an FPLC system, which failed to work in this research. Alternatively, secondary purification methods such as filtration or flocculation, or other chromatography techniques like size exclusion chromatography or ion exchange chromatography could provide enhanced purity, though this would typically be at the cost of yield.

4.4.3 Limitations and future work

There are several limitations in the methods tested. The first and major limitation is inconsistency between samples, as it is very difficult to control for the mechanical grinding efficiency when breaking the cell walls using liquid nitrogen. Incomplete tissue grinding leads to larger fragments and less effective protein liberation. Fortunately, this can be controlled for at a later stage, when the amount of protein loaded onto SDS-PAGE gels is normalised. A second limitation is the use of detergent in the initial extraction buffer, which starts to liberate the cellular membranes even at a low concentration, the efficiency of which is determined largely by particle size after grinding. This has been accounted for as best as possible by limiting the time in detergent and using low concentrations of detergent in initial extraction buffers, increasing the detergent concentration when complete membrane solubilisation is desired. The final detergent used, Triton X-100, is a mild detergent commonly used in plant extraction buffers, but alternative detergents such as DDM are

commonly used and effective for this. FC-12 is a harsher detergent that may improve membrane solubilisation, but was found to be unsuitable when used in conjunction with IMAC purification, presumably as it is able to interact with both the resin and protein due to its zwitterionic nature. The purity of product was the next limitation, with several off-target proteins also purified. As mentioned, FPLC fine tuning of the histidine-pulldown could improve this, or the use of secondary purification methods.

As a result, future work should aim to standardise the mechanical disruption of the tissue, potentially through the use of food processors and tissue homogenisers, as is often done at large scales in industrial settings (Buyel, Hubbuch and Fischer 2016). For cytosolic proteins specifically, screw presses have shown success, able to crush tissue to isolate intracellular mediums whilst retaining matter attached to the cell walls (Hansen et al. 2022). This was not investigated in this research as the ultimate aim was to utilise plant expression systems for the production of membrane proteins but may prove promising for cytosolic protein targets. Following cell wall disruption, the extraction buffers should be fine tuned and different concentrations of reducing agents tested to identify the perfect balance between reducing protein aggregation and ensuring effective purification. One alternative solution would be to optimise the CBD tag, which may have shown some success but was not worth investigating further for this research due to the upstream effects. Redesigning this tag may improve the *in planta* protein folding and make it suitable for this research, but the purification methods used do not rely on IMAC, and so would likely be compatible with higher concentrations of DTT. Finally, the his-pulldown process itself should be optimised so that pure recombinant protein is obtained, potentially using gradient or stepwise elutions. Ideally, this would involve FPLC automation to maximise consistency and reproducibility.

4.4.4 General suitability

Overall, the extraction buffer and method utilised here should be suitable for all cytosolic proteins but may not be suitable for all membrane proteins. As mentioned, different detergents and solubilisation times may be necessary for different membrane proteins. It may also be worth experimenting with the concentration of detergents in the initial extraction buffer to enable full separation of the cytosolic and membranous fractions, as this was likely not optimised in this research. By far the most expensive reagent used in downstream processing are the protease inhibitors, representing 99.25% of the protein extraction buffer costs in this research, with the extraction buffer costing £2073 per kg of tissue. If a cheaper but more generic protease inhibitor was used, such as PMSF, these costs could be reduced to £34 per kg of tissue, so cheaper alternatives would be favourable. These financial and temporal costs could likely be further reduced when

extractions and purifications are performed at scale with automation, making the downstream processing of plant materials cost effective.

4.5 – Conclusions

To conclude, the downstream processing of plant made recombinant proteins has seen a dramatic improvement during the course of this work, starting with ineffective protein extraction methods and resulting in the development of an effective protein extraction buffer for total protein, with an adaptable method that can isolate cytosolic and membranous compartments using centrifugal fractionation. This buffer can be effectively paired with a histidine-pulldown purification assay, that provides moderately pure recombinant protein, though further optimisation would be necessary. A cost analysis shows that cheaper options could be found by replacing only the protease inhibitors used in this research, which would be further reduced at larger scales in better controlled environments with automation. Future fine tuning would enable plant-based expression systems to produce recombinant proteins with similar downstream costs and purities to other expression systems. Combining this with optimised protein production constructs to maximise *in planta* yields could also enable similar yields to conventional expression systems.

Chapter 5 – Comparative Expression Analysis of Proteins with Varying Complexities in Plants

5.1 – Chapter Summary

The previous two chapters focussed on the development of upstream and downstream processing of plant-made recombinant proteins, but successful methods only used eGFP as the expressed protein. Whilst this is effective for process optimisation, an expression system is only useful if it can produce commercially, industrially, or medically relevant proteins. In particular, the main aim of this research was to develop a production system for functional, correctly folded, post-translationally modified eukaryotic integral membrane proteins. Thus, to assess the versatility of the expression system, comparing the expression levels of proteins varying in complexity, with different sizes, intracellular localisations, and endogenous PTMs, can provide insights into the relationship between recombinant protein complexity and yield. This main aim of this chapter was to produce a range of target proteins and measure the impact of complexity on yield, both *in planta* and at the extracted level. In addition, alternative plant expression hosts were explored to produce full-length P2X7 with varying effectiveness. This section first describes the protein targets chosen for expression, along with the rationale and design of the constructs. Second, the *in planta* expression data is analysed and compared in an identical manor to that described in Chapter 3. Thirdly, the protein expression data is analysed and compared analogous to Chapter 4. Together, these data suggest that protein complexity does impact protein yields, with protein size and presence of transmembrane domains both having an impact. Proteins that are reported to contain post-translational modifications are also produced in lower yields than proteins reported to be unmodified, but the presence of these potential modifications was not directly measured in this research. The data also suggest that certain plant expression systems may be better suited to producing different proteins than others, an important finding for future research.

5.2 – Introduction

As previously mentioned, there is a strong need for a protein production system that can produce low-abundance post-translationally modified proteins cheaply. Conventional expression systems cannot do this, as whilst prokaryotic expression systems are cheaply scalable, they cannot produce many PTMs, and whilst eukaryotic expression systems can, these are very costly to scale. This creates an obvious problem for low-abundance proteins with complex PTMs, which require a cheaply scalable production system to enable affordable production. In fact, approximately 60% of human proteins are modified in some tissues (Bagwan et al. 2021), meaning prokaryotic systems

cannot produce them, and 25-30% of proteins are membranous (Dobson et al. 2015; Almén et al. 2009), meaning a cost-effective production system should be cheaply scalable to obtain the yields necessary for downstream analyses.

In addition, to my knowledge, there is no single study which has investigated the impact of recombinant protein complexity on yields from plant species, so the relative impacts of protein size, localisation, and post-translational modifications have not been assessed. Protein size can significantly affect yields, as larger proteins are more challenging to express and purify due to increased energetic demand, chance of misfolding, and chance of aggregation relative to smaller proteins (Goldenzweig and Fleishman 2018). PTMs can affect the yield of recombinantly produced proteins, and may be essential for function, folding, and stability. These are carried out in specific organelles which can incur a production bottleneck on the yield of modified proteins. Lastly, the sub-cellular localisation can significantly affect recombinant protein yields. Proteins are naturally located in specific compartments, such as the plasma membrane, nucleus, mitochondria, chloroplast, or endoplasmic reticulum which may be essential for correct protein folding and function. Together, these mean that the yields of recombinant protein will vary based on the nature of the protein itself. Consequently, several protein targets varying in complexity were expressed in this research and are each discussed in the following subsections. Importantly, each CDS was expressed with a C-terminal eGFP tag enabling direct comparisons of expression *in planta* and at the extracted protein level. Together, this selection and design enabled the relative yields of each protein to be compared to eGFP, a small cytoplasmic unmodified protein and the simplest in this research. The main hypothesis was that protein size, PTMs, and localisation each affect the resulting recombinant protein yield.

5.2.1 P2X7

P2X7 is a post-translationally modified mammalian integral membrane protein found on the cell surface that is activated by extracellular ATP. It plays a crucial role in the immune and inflammatory responses by mediating the release of cytokines and has been implicated in various pathological conditions such as chronic pain, arthritis, and neurological disorders (Tewari and Seth 2015; Lara et al. 2020). P2X7 has an intracellular C-terminal signalling domain, but the molecular mechanisms of the protein and its contribution to these disorders is unknown. Structural studies could provide functional insights into these mechanisms, and whilst a nearly complete structure of rat P2X7 has been published (Figure 5.1, Panel B; McCarthy et al. 2019), the structure of full-length human P2X7 has not been. P2X7 is industrially and medically relevant as obtaining structural information could be vital for developing drugs that modulate its activity and treat associated disorders. The protein, shown diagrammatically in Figure 5.1, Panel A, contains an integral membrane P2X domain and a cytoplasmic C-terminal signalling domain, and is heavily post-translationally modified with

glycosylation and palmitoylation (Kopp et al. 2019), meaning that prokaryotic expression systems are unable to produce P2X7 recombinantly. To date, this protein has not been expressed recombinantly in plants, which could enable human P2X7 to be produced in sufficient yields for further research. Additionally, as both intracellular localisation and PTMs affect recombinant protein yields, truncated versions of P2X7 were also expressed which exclude either the transmembrane domain, or both the transmembrane domain and a palmitoylated region of P2X7, seen in Figure 5.1. These truncated versions are valid targets as studying the C-terminal signalling domain could provide functional information on P2X7 signalling events associated with inflammation and associated disorders.

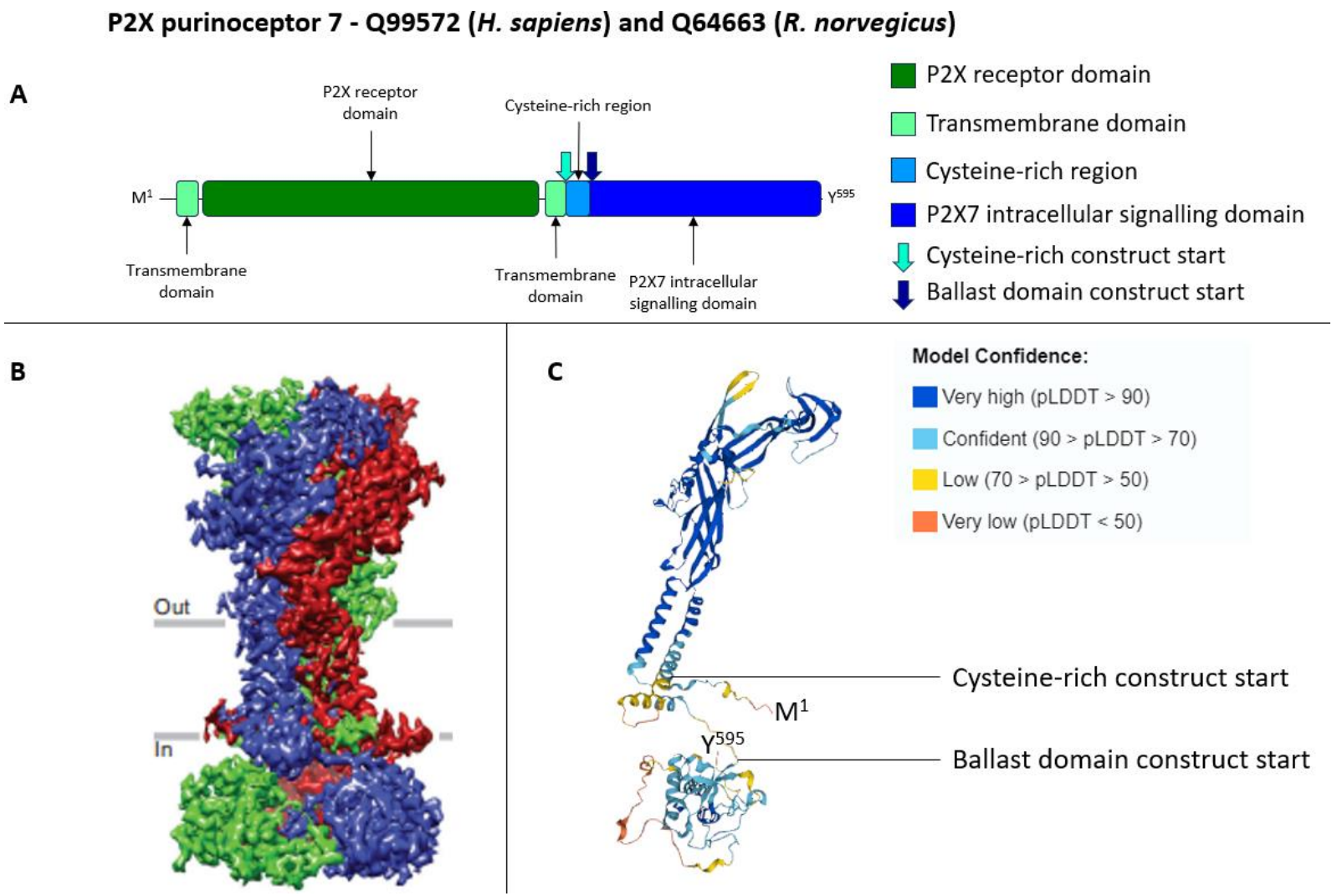


Figure 5.1: The structure of P2X7. A – line diagram of domains of P2X7 and sites of importance, where different colours show different domains. **B** – Published full-length rat P2X7 structure by McCarthy et al. (2019) where different colours show different monomers of the P2X7 trimer. **C** – AlphaFold predicted model of P2X7 with confidence values in pLDDT, where a high score (dark blue) represents a higher confidence of that region relative to a very low score (orange).

5.2.2 Does Not Respond to Nucleotides 1 (DORN1)

DORN1 is a plant post-translationally modified plasma membrane purinoreceptor that binds extracellular ATP and is a member of the L-type lectin-domain containing receptor kinase family of proteins. The structure of DORN1 can be seen in Figure 5.2. It is an essential protein involved in calcium signalling cascades that affect processes such as growth, stress responses, and cell death (Chen et al. 2017; Kumar et al. 2020). DORN1 is relevant from an industrial and agricultural perspective, as understanding its function may enable the development of crops that have increased growth and yields or improved stress tolerance. DORN1 was primarily chosen as a target in this research as it has a similar level of complexity to P2X7, is also heavily glycosylated, but is endogenous to plants meaning it can be correct folded and trafficked within transformed cells. Therefore, whilst it is a relevant target, it acts primarily as a proof-of-concept protein here for endogenous transmembrane plant proteins that are post-translationally modified.

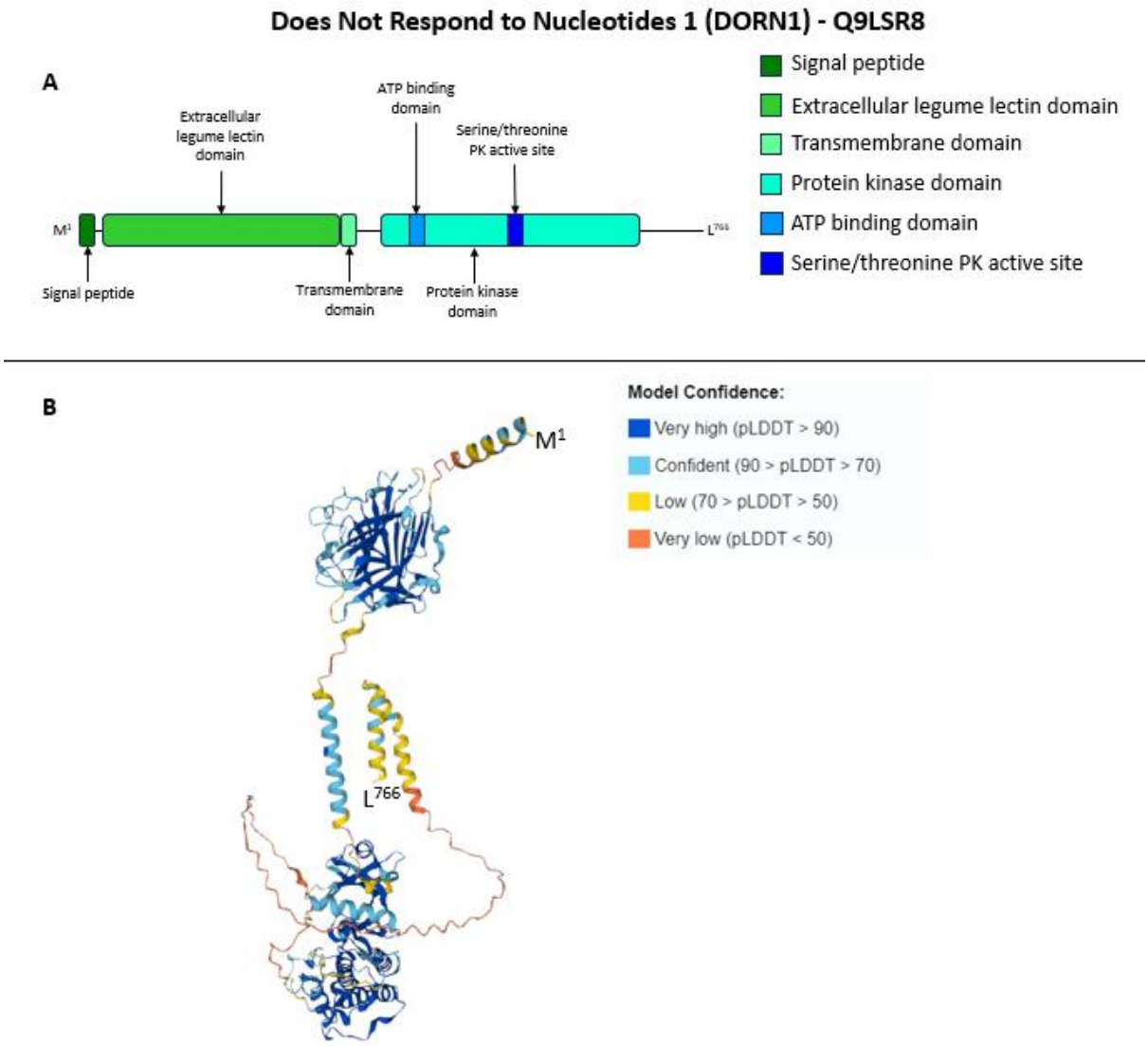


Figure 5.2: The structure of DORN1. A – line diagram of domains of DORN1 and sites of importance where different colours show different domains. **B** – AlphaFold predicted model of DORN1 with confidence values in pLDDT, where a high score (dark blue) represents a higher confidence of that region relative to a very low score (orange).

5.2.3 Putative cadmium/zinc-transporting ATPase HMA4 (HMA4)

HMA4 is an unmodified plasma membrane protein of the P-type ATPase family. It is involved in the transport of heavy metals, particularly zinc and cadmium, across cell membranes. In *A. thaliana*, it plays a crucial role in heavy metal homeostasis and detoxification and is a relevant production target as a tool for bioremediation, where several publications have identified it as a useful tool for removing heavy metals from soil (An et al. 2020; Ibuot et al. 2020). Expressing HMA4 can allow studies into the mechanism of heavy metal uptake by plants, potentially enabling the design of improved engineered or synthetic receptors useful in industrial settings for bioremediation. HMA4 has an extracellular metal binding domain, eight transmembrane domains that constitute transport, and two cytosolic catalytic domains (Laurent et al. 2016); a diagram can be seen in Figure 5.3, Panel A. An NMR structure of the metal-binding domain has been solved (Figure 5.3, Panel B; Zimmermann et al. 2009), but the full structure has not been. HMA4 was chosen as a target in this research as it is endogenous to plants so can be produced correctly in transformed cells and acts primarily as a proof-of-concept here for endogenous transmembrane plant proteins that are not post-translationally modified..

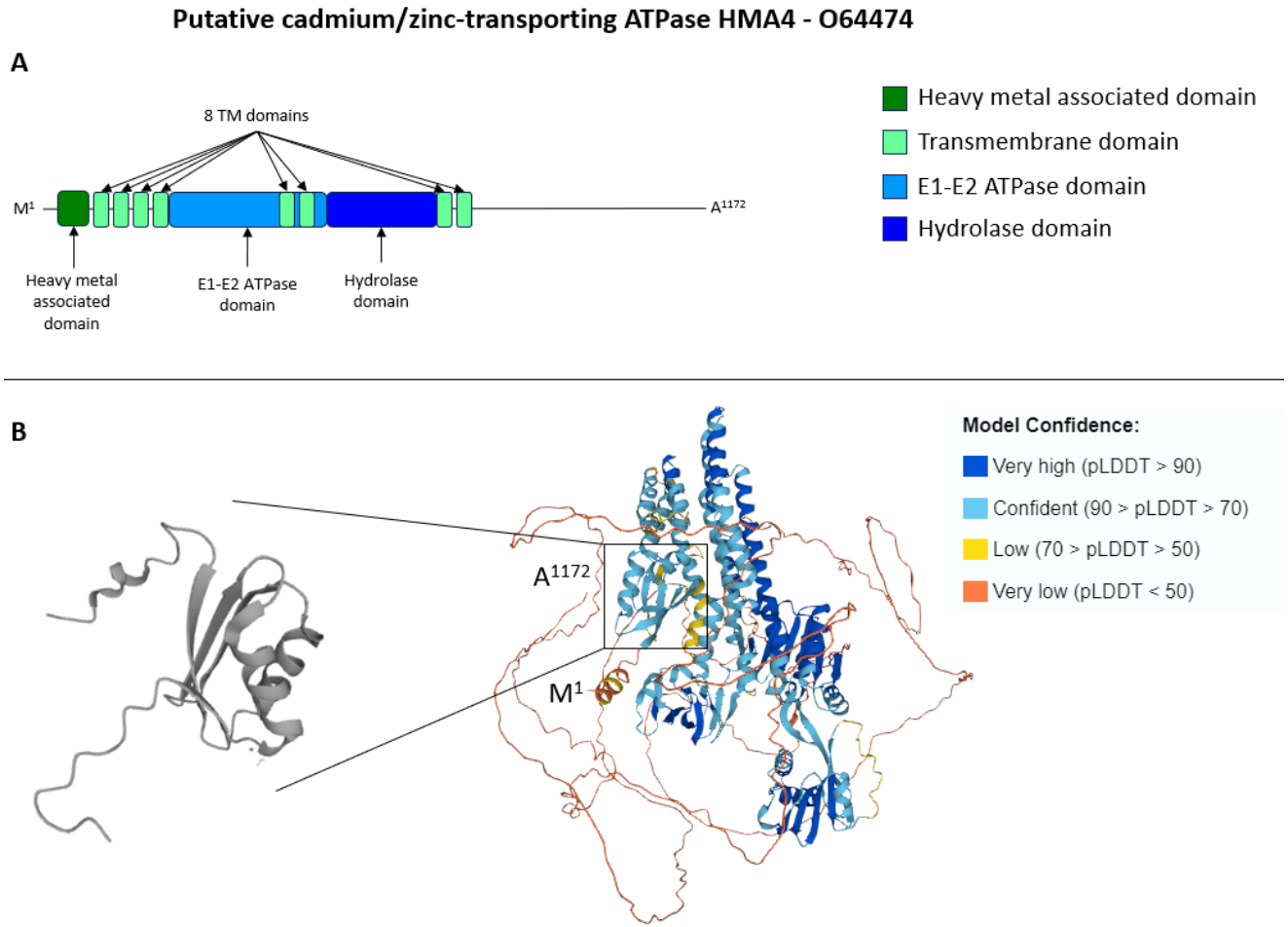


Figure 5.3: The structure of HMA4. A – line diagram of domains of HMA4 and sites of importance where different colours show different domains. **B** – AlphaFold predicted model of HMA4 with confidence values in pLDDT, where a high score (dark blue) represents a higher confidence of that region relative to a very low score (orange). An NMR study of the metal binding domain of HMA4 has been published by Zimmermann et al. (2009) and is shown in grey, with its localisation in the AlphaFold model indicated.

5.2.4 Factor C

Factor C is a clotting protein present in horseshoe crab blood, which is extremely sensitive to bacterial endotoxins. It is large, cytosolic and heavily glycosylated with several core protein domains, shown in Figure 5.4. The protein is medically, ecologically, and industrially relevant, as it is conventionally harvested from live horseshoe crabs where it is used in the Limulus Amoebocyte Lysate (LAL) assay for detection of endotoxins ensuring sterility in medical devices and vaccines. Many species, some of which are endangered, feed on horseshoe crabs, their eggs, or their larvae (Dellinger and Kepley 2018). LAL harvesting from crabs has a high mortality rate, up to 30%, and females that are released following harvesting are lethargic and have reduced reproductive activity (Dellinger and Kepley 2018), causing a decline in horseshoe crab populations. Therefore, recombinant production of Factor C is significant due to ethical concerns over the harvesting of horseshoe crabs and the sustainability of natural populations, and recombinant production may reduce dependence on wild-caught horseshoe crabs while ensuring a consistent and contaminant-free supply for medical testing. LAL and its components have been produced synthetically and recombinantly (Dubczak et al. 2021), though plants could offer a cheaper and potentially more effective alternative to recombinant Factor C expressed by current systems.

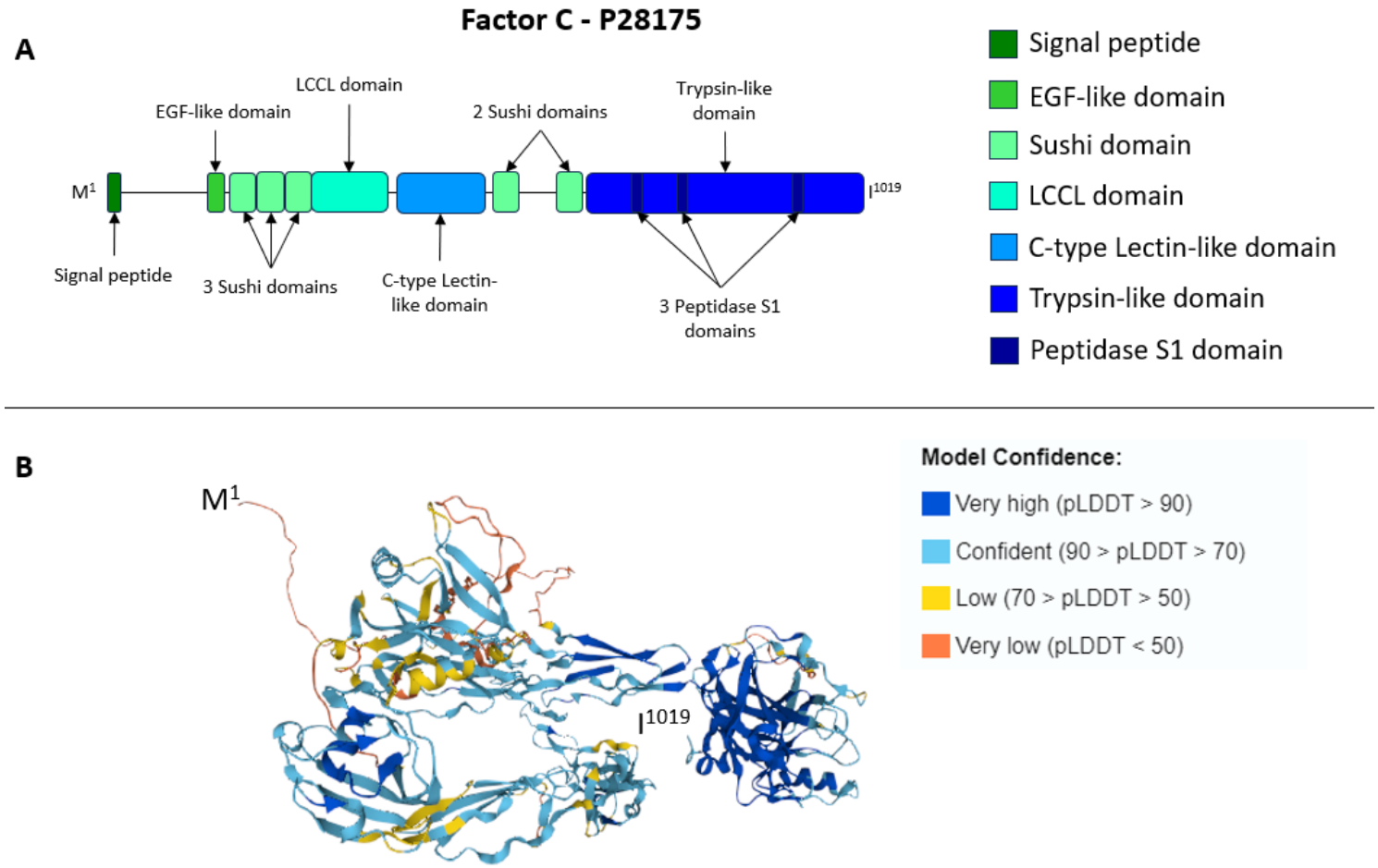


Figure 5.4: The structure of Factor C. A – line diagram of domains of Factor C and sites of importance where different colours show different domains. **B** – AlphaFold predicted model of Factor C with confidence values in pLDDT, where a high score (dark blue) represents a higher confidence of that region relative to a very low score (orange).

5.2.5 RAN GTPase-activating protein 1 (RANGAP1)

RANGAP1 is a modified cytosolic cell cycle regulator in eukaryotic cells that stimulates the GTPase activity of Ran, a small GTP-binding protein involved in cell-cycle control, nucleocytoplasmic transport, mitotic spindle assembly, and nuclear envelope formation (Xu et al. 2008). In *A. thaliana*, RANGAP1 is crucial for proper cell division and plant development. Shown in Figure 5.5, it is a simple protein with only two core domains, a WPP domain that traffics the protein to the nuclear envelope (Xu et al. 2008), and leucine-rich repeat domain needed for Ran binding and stimulation of GTPase activity (Haberland and Gerke 1999). Conserved charged residues in the leucine-rich repeat domain of the Ran GTPase activating protein are required for Ran binding and GTPase activation. Recombinant production of RANGAP1 is valuable for research into cell cycle regulation and plant development, with potential applications in improving plant growth for use in crop biotechnology. However, like DORN1 and HMA4, it was primarily chosen in this research as a proof-of-concept protein target for endogenous cytosolic plant proteins that are post-translationally modified.

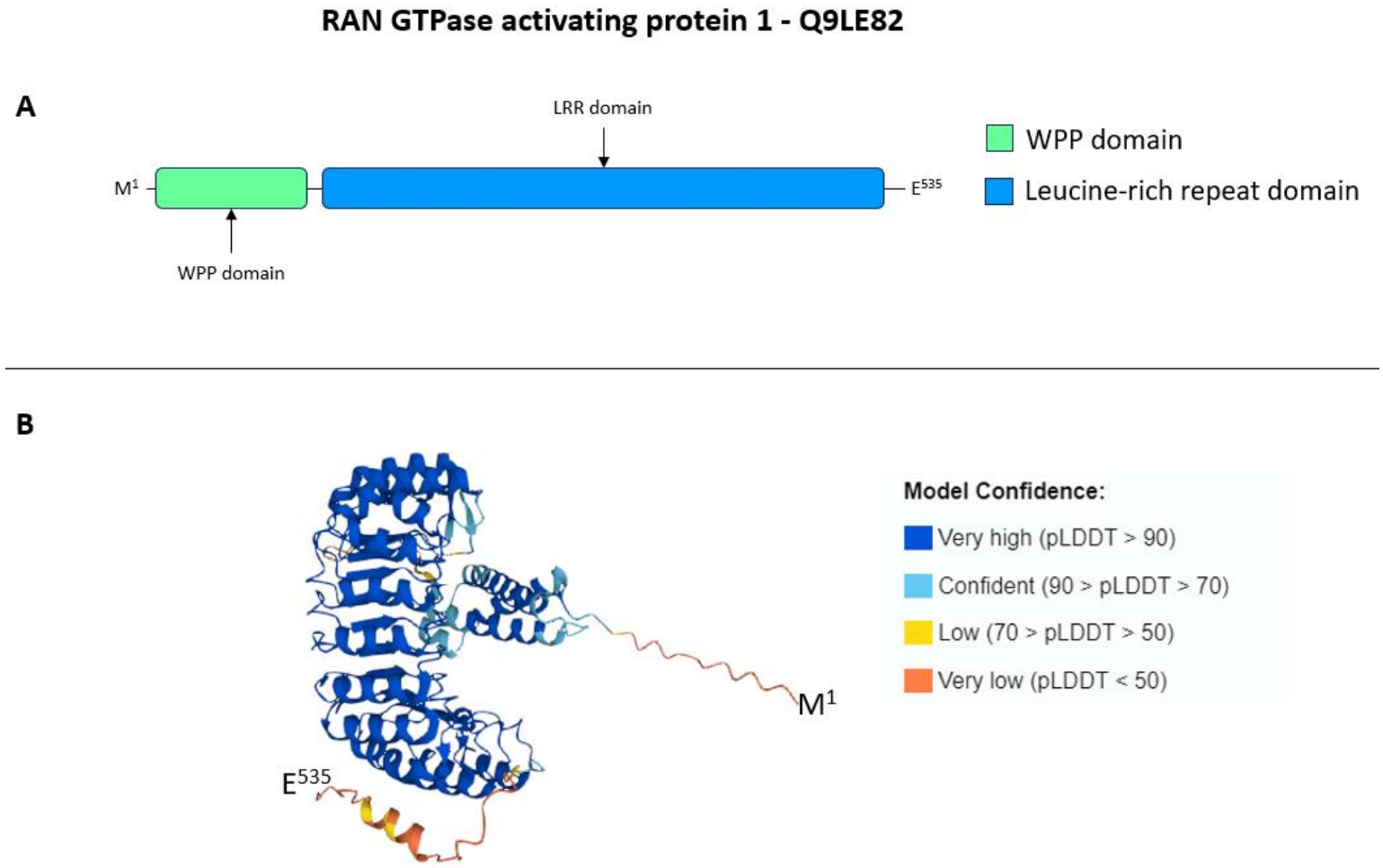


Figure 5.5: The structure of RANGAP1. A – line diagram of domains of RANGAP1and sites of importance where different colours show different domains. **B** – AlphaFold predicted model of RANGAP1 with confidence values in pLDDT, where a high score (dark blue) represents a higher confidence of that region relative to a very low score (orange).

Consequently, each protein was produced using the best expression constructs from each phase of Chapter 3, pNC, pRC, and pR5, and the protein extracted and quantified using the best methods from Chapter 4. That is, the expression levels of these proteins were quantified *in planta* using an eGFP tag and fluorescence detection methods, and at the protein level using an anti-GFP antibody with western blotting. The aim was to provide a comprehensive view of how the expression constructs of Chapter 1 perform in producing proteins of varying complexities. As anticipated, the constructs and methods optimised for eGFP expression showed a marked difference when proteins of different complexities were expressed. The general trend seen was that smaller unmodified and cytoplasmic proteins expressed the highest, followed by modified cytoplasmic proteins, with all membranous proteins having the lowest expression levels. This chapter explores these data in comprehensive detail and draws insights into the versatility of plant-based expression systems to produce industrially and medically relevant proteins, paving the way for future studies and potential industrial applications.

This chapter will first discuss the early expression constructs utilised to produce the three different isoforms of rat and human P2X7 in various plant-based expression systems, which highlighted the need to enhance expression levels through genetic engineering efforts. Following this, it will discuss the expression levels of each protein when expressed using the constructs developed in Chapter 3. Finally, it will describe experiments that were performed to determine whether full-length P2X7 is functional *in planta* and draw all the data together to assess whether the plant protein expression system developed in this research is suitable for complex recombinant proteins.

5.3 – Results

5.3.1 Design, cloning, and screening of expression constructs with different coding sequences

5.3.1.1 Design

To be used in this research, each of the different protein coding sequences needed to be designed with compatibility for use with both the Golden Gate Type IIS restriction endonucleases, and the conventional endonucleases required to insert them into the replicating vector, pJL-TRBO. Consequently, each part was carefully designed so that they contained no internal endonuclease recognition sites that would interfere with cloning. Two native P2X7 sequences, rat and human full-length P2X7, were previously used in the Young lab but contained internal Type IIS restriction endonuclease recognition sites. In addition, the proteins native to *A. thaliana*, RanGAP1, HMA4, and DORN1, also contained internal recognition sites and were amplified from genomic DNA previously obtained by researchers in the Scofield lab. Consequently, these recognition sites were removed by

a process called domestication, which uses partial fragment amplification with primers that create single nucleotide changes to remove these recognition sites. The fragments can then be cloned into Level -1 Golden Gate vectors as partial coding sequences, designed so that C-terminal eGFP can be incorporated, which can then be combined back into a full-length C-terminally tagged coding sequence when each fragment is cloned into a Level 0 vector. Rat P2X7 had three internal sites, human P2X7 had two, RanGAP1 had two, DORN1 had two, and HMA4 had five. However, after difficulty amplifying DORN1 from gDNA, either in fragments or as a full-length coding sequence, it was synthesised as a domesticated fragment using GeneWiz to be cloned into a Level -1 vector. As there was no existing template of Factor C, this was also synthesized using GeneWiz to be cloned into a Level -1 vector. The truncated versions of P2X7 were amplified from their domesticated parent sequence and cloned into their appropriate Level 0 vectors. This enabled the creation of the C-terminally eGFP-tagged protein coding sequences used in this chapter.

5.3.1.2 Cloning and screening

To test each of the different proteins in plants, several expression constructs were created in case different proteins benefitted from different expression constructs. Firstly, each CDS was assembled into the complex Level 1 expression construct, in an identical process to pNC-eGFP in Chapter 3, containing the CaMV 35S promoter, a synthetic 5' UTR, variable coding sequence, CPMV 3' UTR, extensin and Actin terminator, and a RB7 matrix attachment region. These are denoted as pNC-CDS, where CDS is the name of the protein used. Following this, each construct was digested with PacI and AvrII and ligated into the replicating vector pJL-TRBO, with an identical process to the creation of pRC-eGFP in Chapter 3. Only the rat P2X7 sequences were not created in this way due to difficulties in cloning, despite domestication of the sequences to remove all internal restriction endonuclease sites. Finally, each CDS was also expressed in the pR5 expression construct from Chapter 3; these were created in an identical manner to pR5-eGFP discussed in Chapter 3. It has been reported that RNA replicating vectors, such as TMV from which pJL-TRBO is derived, are unable to express large proteins as the RNA virus size is a limiting factor (Hefferon 2012), though other publications suggest that there is no theoretical size limit (Lindbo 2007b). Chapter 3 of this research showed that the large complex construct had markedly improved expression when expressed in the pJL-TRBO vector, suggesting that this size limit may not be strictly true as the pNC construct is over 5 kB in length. For this reason, the best expression constructs from each phase in Chapter 3, pNC, pRC, and pR5, were chosen to express other recombinant proteins.

5.3.2 Testing and analyses of different protein coding sequences

As mentioned in previous chapters, the very first expression constructs made and tested in this research expressed the six rat and human P2X7 in the pNS expression system seen in Chapter 3. This

is because pNS utilises well characterised DNA elements including the CaMV 35S promoter which is widely considered to drive high constitutive gene expression levels (Seternes et al. 2016) and may have been suitable to produce P2X7 at high yields. When transformed into *N. benthamiana* plants, expression could be seen using the confocal microscope in some cells, though the transformation efficiency was low, with most cells remaining untransformed reflecting the low transformation rate when using the non-replicating vector. In addition, early protein extractions using plants transformed with this tissue failed to produce a band corresponding to the target protein or tag when visualised using SDS-PAGE followed by InstantBlue™ staining or a western blot. These data are discussed in detail later in this chapter; however, it was decided that the expression levels of recombinant protein using the pNS construct were too low, which sparked the development of the enhanced expression constructs seen in Chapter 3.

5.3.2.1 In planta expression analyses of different proteins

The different coding sequences were first tested in the complex non-replicating construct, pNC. As expected, when eGFP fluorescence was measured *in planta*, plants expressing different proteins produced variable expression levels, shown in Figure 5.6. Surprisingly, despite all proteins being C-terminally tagged with eGFP, the plant expressing eGFP as a reporter instead of as a fusion did not produce the highest expression, with low fluorescence that was not significantly higher than the empty vector controls. Instead, only the truncated P2X7 variants produced high levels of expression when expressed in pNC. In fact, rat cysteine-rich, rat ballast, and human ballast P2X7 showed 33.6, 18.3, and 21-fold increased fluorescence than the eGFP protein expressed by itself, each of which was statistically significantly higher than the eGFP and empty vector controls. This was surprising as typically fluorescent protein fusions express at lower levels that the fluorescent protein when not fused, as fused partners have increased protein complexity and require further folding. None of the other proteins produced fluorescence that was significantly different from the empty vector control, suggesting that expression of these proteins is either absent, very low, or, if the proteins are being produced, that the eGFP fusion tag isn't folding properly to produce detectable fluorescence.

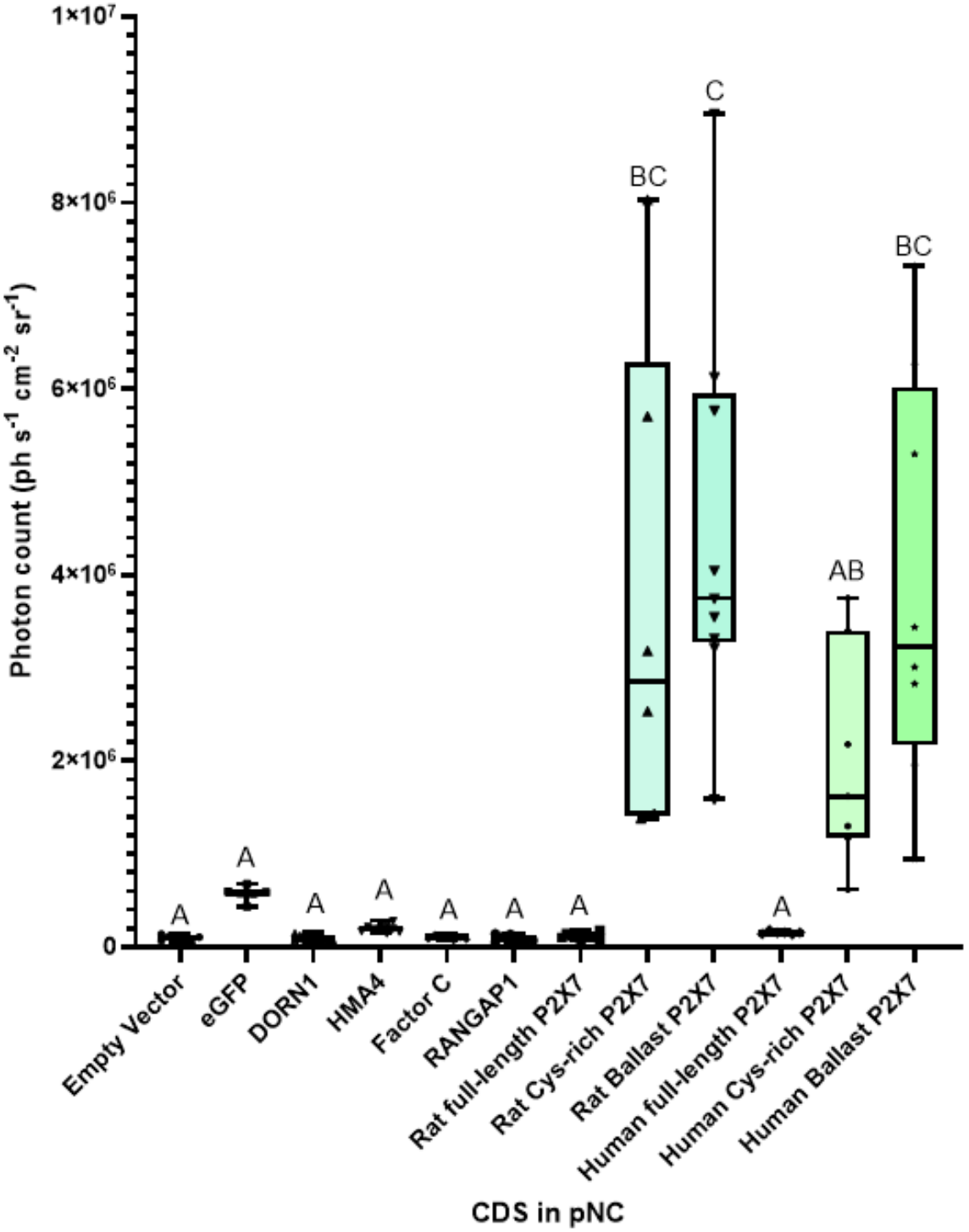


Figure 5.6: eGFP fluorescence of *N. benthamiana* leaves transformed with pNC-CDS, quantified using a Biospace Optima photon imager. Three plants were transformed with each construct (n=3) and 2-3 transformed leaves were taken and analysed from each plant. A Kruskal-Wallis test shows a statistically significant difference between samples overall ($\chi2(11) = 81.722$, $p = 6.85x10^{-13}$). Interestingly, the expression of pNC-eGFP was very low with this construct despite several repeats. Only the truncated P2X7 variants, rat cysteine-rich and ballast P2X7, and human ballast P2X7 had significantly higher fluorescence than the empty vector control. Letters denote compact letter display where the alpha value is 0.05. Samples that do not share a common letter have significantly different expression, and samples that share a common letter show expression that is not significantly different. Error bars show minimum and maximum values.

Next, the different proteins were tested in the complex construct within the replicating vector, pRC. Unfortunately, due to difficulties in obtaining positive clones, the rat P2X7 constructs were not tested. When expressing the different proteins using pRC (Figure 5.7), we see that the eGFP expression is enhanced and is significantly different to the empty vector control. Further, the same pattern is seen with the truncated P2X7 variants, which express at a similar level to eGFP; in fact, human cysteine-rich and ballast P2X7 express at 0.81-fold and 1.15-fold that of eGFP, respectively, but neither difference is statistically significant. Interestingly, we also see enhanced expression for RANGAP1, the plant endogenous modified cytosolic protein, though this is not significantly different from the empty vector levels. Overall, assuming the tagged protein is not affecting eGFP folding, these data suggest that the plants are unable to produce substantial amounts of recombinant DORN1, HMA4, Factor C, RANGAP1, or full-length P2X7 proteins using the complex expression cassette in either a non-replicating or replicating vector.

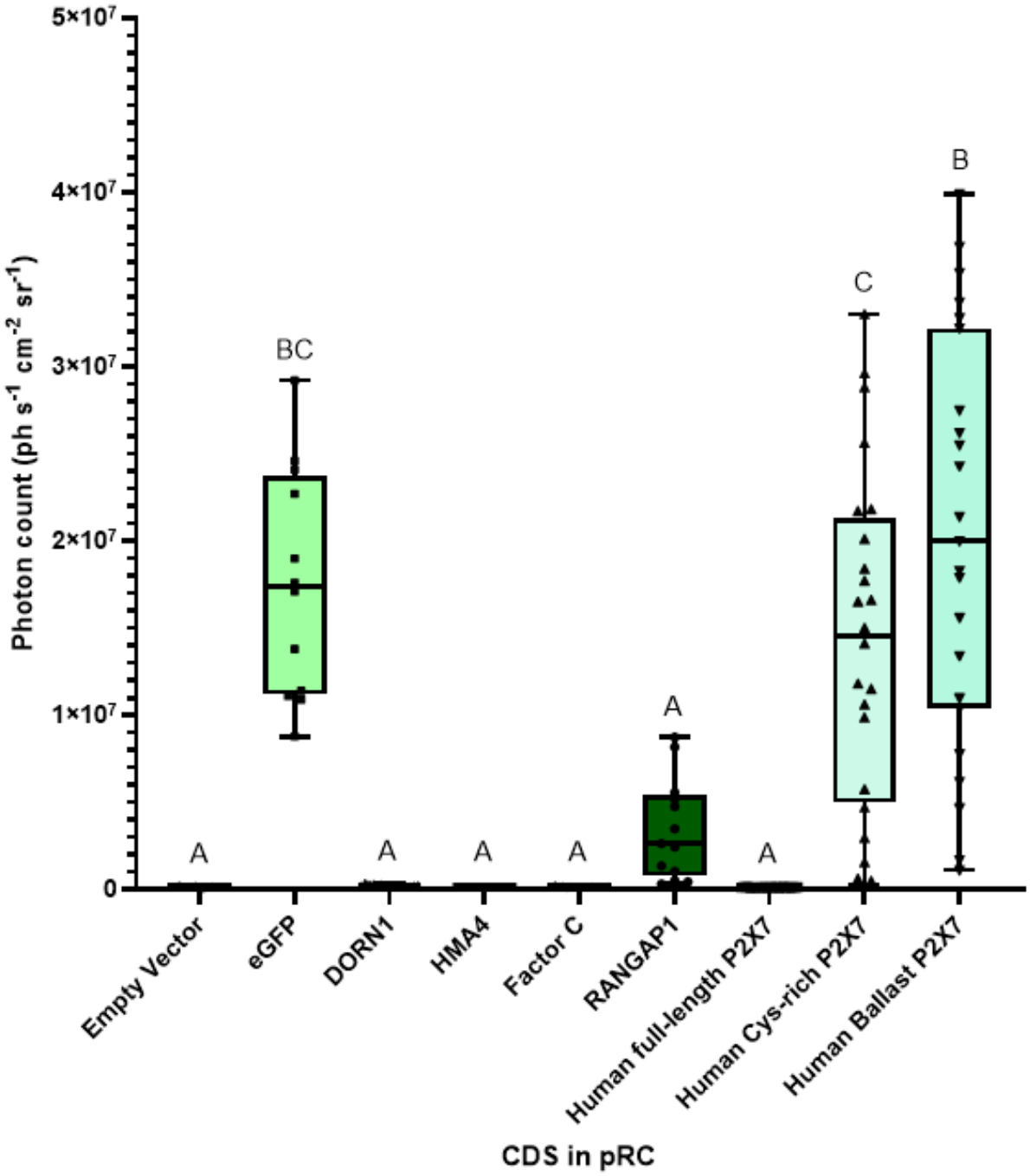


Figure 5.7: eGFP fluorescence of *N. benthamiana* leaves transformed with pRC-CDS, quantified using a Biospace Optima photon imager. Four plants were transformed with each construct (n=4) and 3-4 transformed leaves were taken and analysed from each plant. A Kruskal-Wallis test shows a statistically significant difference between samples overall ($\chi2(8) = 122.92$, $p = 2.2x10^{-16}$). A more in-depth analysis shows that this time, the eGFP fluorescent signal is dramatically increased and is statistically significantly different from the untransformed control ($p = 1.1x10^{-6}$). The truncated P2X7 constructs, human cysteine-rich and human ballast, produce fluorescence that is not statistically significantly different from the eGFP expressing construct ($p = 0.86$, and $p = 0.97$, respectively), but are statistically significantly different from the untransformed control ($p = 1.68x10^{-5}$, and $p = 0$, respectively). The only major difference in these experiments is the expression of the cytoplasmic modified *A. thaliana* protein, RANGAP1, which expresses detectable fluorescence that is not significantly different from the negative control. Letters denote compact letter display where the alpha value is 0.05. Samples that do not share a common letter have significantly different expression, and samples that share a common letter show expression that is not significantly different. Error bars show minimum and maximum values.

Finally, the highest expressing eGFP construct from Phase 3 of Chapter 3, pR5, was used to express the other coding sequences. In this experiment a similar trend is seen to the CDS expressed in pRC, with only the cytosolic proteins showing *in planta* fluorescence, but in this circumstance only eGFP was statistically significantly different from the empty vector control, with the other cytosolic proteins only showing slightly elevated fluorescence relative to the empty vector (Figure 5.8). Together, the variable expression data of different proteins between constructs suggests that some constructs may be better suited to producing some proteins than others. In pNC, the truncated P2X7 variants expressed at the highest level, significantly higher than eGFP, whereas in pRC, eGFP was expressed higher and was not significantly different to the truncated P2X7 variants. More peculiar is that the opposite trend is seen when expressing proteins in pR5, where eGFP is the only protein that produces significantly higher fluorescence than the negative control, with the truncated P2X7 variants showing very little fluorescence. As different proteins have different complexities, it was anticipated that expression levels would vary between proteins. However, it was not expected that the expression construct would change the relative production of each protein, as was found here; further research could reveal whether these differences have biological significance or are simply caused by variability between experiments. However, as mentioned previously, a lack of fluorescence does not necessarily mean that the protein is not being expressed as the eGFP tag could be misfolded due to the unique attachment to the protein. As a result, it is necessary to perform analyses at the protein level.

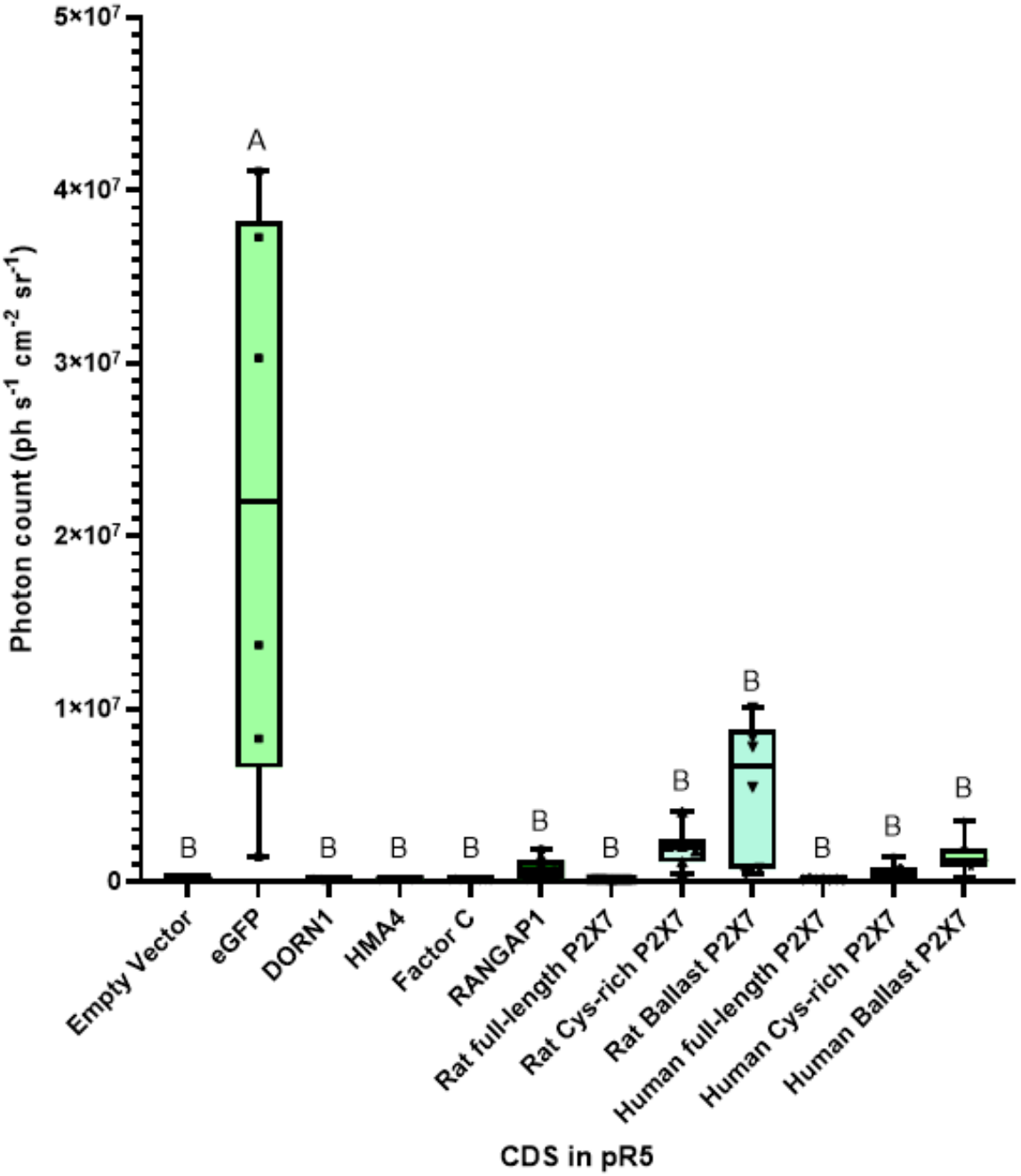


Figure 5.8: eGFP fluorescence of *N. benthamiana* leaves transformed with pR5-CDS, quantified using a Biospace Optima photon imager. Three plants were transformed with each construct (n=3) and 2-4 transformed leaves were taken and analysed from each plant. A Kruskal-Wallis test shows a statistically significant difference between samples overall ($\chi2(11) = 73.492$, $p = 2.636x10^{-11}$). A more in-depth analysis shows only the pR5-eGFP fluorescent signal is significantly higher than the empty vector control ($p = 0$). The expression of RANGAP1 and the truncated P2X7 variants is detectable but not significantly different from the empty vector control; this could be due to the smaller number of replicates used in this experiment resulting in reduced statistical power. Letters denote compact letter display where the alpha value is 0.05. Samples that do not share a common letter have significantly different expression, and samples that share a common letter show expression that is not significantly different. Error bars show minimum and maximum values.

5.3.2.2 Quantifying the expression of different proteins by InstantBlue™ staining and western blot

As the ultimate aim of this research is to generate plant-based expression constructs to recombinantly produce complex eukaryotic proteins, the relative amounts of each protein need to be compared. To do this, *N. benthamiana* tissue expressing the highest yields of each protein, from constructs pRC and pR5, was finely ground and the total protein extracted using the best extraction buffer and method, PEB 5.0, from Chapter 4. The mass of plant tissue expressing different proteins was variable between samples, ranging from 2.2 g to 7.5 g for pR5, and 5.4 g to 11 g for pRC, with the exact masses shown in Table S2.

Initially, total protein extracts of pRC-CDS and pR5-CDS were separated using SDS-PAGE and visualised using InstantBlue™ staining, shown in Figure 5.9. When analysing the protein expression from pRC, only the eGFP band can be seen, with no other bands corresponding to target proteins identifiable (Figure 5.9, Panel A). When expressing the different proteins under pR5 control, more protein aggregation was seen with less clear separation by SDS-PAGE. However, the same pattern is seen with no target proteins visible (Figure 5.9, Panel B), suggesting that a more sensitive method of analysis is needed.

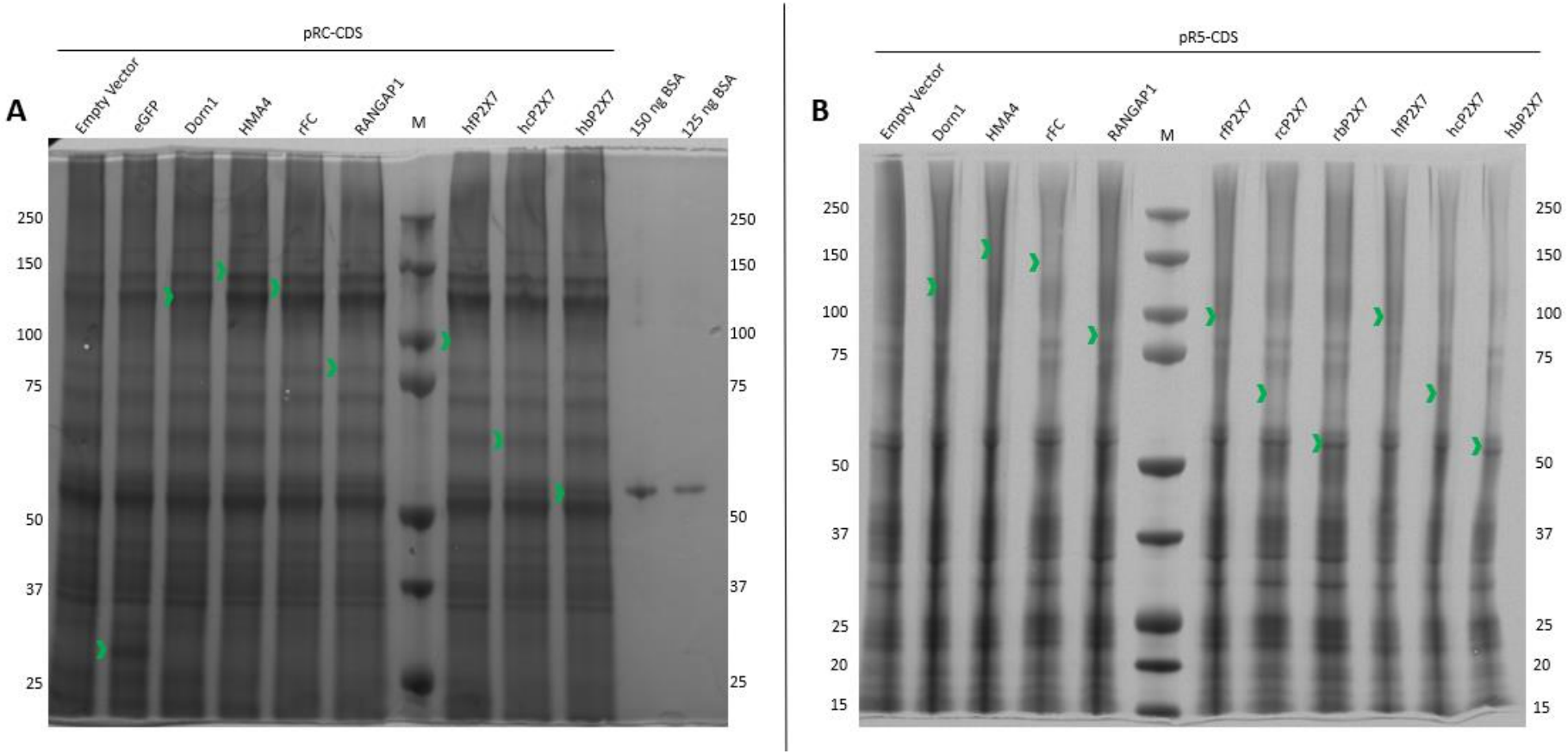


Figure 5.9: InstantBlue™ stained SDS-PAGE gel of total protein extracts from three (pR5) or four (pRC) *N. benthamiana* plants transformed with pRC-CDS (A) or pR5-CDS (B). In both panels, Lane M is the BioRad precision plus dual colour molecular weight marker, with the molecular weights in kDa shown either side of each gel. The green arrows indicate the expected location of the target protein in each lane, with only pRC-eGFP showing the visible product.

Consequently, using a standardised total protein mass of 20 µg, total protein extracts from pRC-CDS and pR5-CDS were separated by SDS-PAGE analysed by western blot. Initially this was performed on the pR5-CDS extracted protein using the plant-specific monoclonal antibody, which failed to produce a signal for any of the proteins except eGFP (Figure 5.10, Panels A and B) suggesting that either there is not enough target recombinant protein for the western blot to detect, or the monoclonal antibody used is unsuitable for eGFP when used as a C-terminal tag. To test which was true, the western blot was repeated using the original polyclonal anti-GFP antibody discussed in Chapter 4. This produced bands at approximately 27 kDa in size in the tissue extracts containing RANGAP1, rat cysteine rich P2X7, and rat ballast P2X7, corresponding to cleaved eGFP (Figure 5.10, Panel C). No full-length products could be seen in any lane, except for a faint smeared band at approximately 53 kDa in the rat ballast P2X7 lane, which could be the full-length product but suggests that it is unstable and is degraded. The presence of new 27 kDa bands suggests that the monoclonal antibody was unsuitable for eGFP when used as a fusion protein, and the original polyclonal antibody is necessary for detection of these bands. However, the absence of full-length products suggests that the C-terminally fused target proteins are not produced in sufficient yields for detection or are degraded. Importantly, the 27 kDa eGFP bands in each lane have variable intensities, suggesting that the different target protein fusions are having an impact on the intensity of these bands. As a result, further analysis is limited to using these 27 kDa eGFP bands as a proxy for full-length target expression, though it is important to note that no actual full-length target products can be confirmed. Densitometric quantification of these bands, seen in Figure 5.10, Panel D, suggests that the expression of RANGAP1, rat cysteine rich P2X7, and rat ballast P2X7 are 171-fold, 84.5-fold, and 10.8-fold lower than eGFP expression, respectively. Interestingly, the human P2X7 variants express at much lower levels than the rat P2X7 variants, despite being orthologous proteins with very similar amino acid sequences. Densitometric analysis of the 53 kDa C-terminally tagged rat ballast P2X7 product suggest that it is produced at a yield 36.7-fold lower than eGFP.

Similarly, the same western blot was performed on pRC-CDS extracts using the polyclonal anti-GFP antibody, seen Figure 5.10, Panel E. Here, there is a more obvious 53 kDa band in the human ballast P2X7 lanes, suggesting the C-terminally tagged target is produced. Only RANGAP1 and the truncated P2X7 variants produce a signal at 27 kDa, suggesting that the expression of these proteins is much higher than the others when using the eGFP band as an indicator. Further analyses, shown in Figure 5.10, Panel F, show that RANGAP1, human cysteine-rich P2X7 and human ballast P2X7 are expressed at 16.2-fold, 12.3-fold, and 4.3-fold lower than eGFP, respectively. Quantification of the smeared 53 kDa human ballast P2X7 band suggest that it is produced at a yield 3.2-fold lower than eGFP.

Together, these data suggest that only the ballast P2X7 proteins can be produced in this expression system, with pRC producing higher target protein yields than pR5. Using the 27 kDa eGFP bands as a proxy, this suggests that RANGAP1 and cysteine-rich P2X7 may be produced in low yields. In addition, C-terminally tagged ballast P2X7 products were substantially degraded, suggesting that further optimisation of extraction methods is necessary to preserve protein state. Importantly, the protein samples that had an eGFP signal were each cytosolic proteins, suggesting that they can be produced at higher yields than membranous proteins. In addition, the densitometric analysis of the bands showed that rat ballast P2X7 was produced at nearly 8-fold that of rat cysteine-rich P2X7 in pR5, suggesting that the additional palmitoylated domain also substantially affects yield. Similarly, the expression of human ballast P2X7 was nearly 3-fold that of human cysteine-rich P2X7 when expressed in pRC. However, as full products cannot be clearly seen it is likely not accurate to compare protein samples because it assumes that eGFP cleavage and protein degradation is proportionate between samples. Ideally, whole protein products would be compared, but corresponding bands show degradation and aren't clear enough to draw definitive conclusions. There are several potential reasons for this. The first and most likely reason is that there may not be enough full-length tagged product to be detected on the western blot, suggesting that most of the eGFP is cleaved off within the cell and target proteins are degraded. Second, an alternative translational start site could cause only the eGFP protein to be synthesised independent of the target protein. This is unlikely as variable bands would be expected in each lane, as there are multiple potential translation start sites within each coding sequence. Finally, it may be that the primary antibody used is unsuitable for detecting eGFP when it is used as a C-terminal tag. This is also unlikely, as recent research has successfully employed the primary antibody to detect eGFP fused to full-length rat P2X7 produced by HEK-293 cells (Figure 10, Panel G; Jaradat 2021), suggesting that it is suitable, at least in mammalian cell extracts.

However, between 2.2 g and 11 g of tissue had protein extracted from these samples, resulting in a final volume of approximately 6.6 mL to 33 mL of total protein, depending on the sample. As the SDS-PAGE wells can only hold approximately 25 µL of sample, this results in a maximum load of between 0.37-0.076% of the total protein extracts. This means that increasing the purity and protein concentration may allow more target protein to be analysed, increasing the likelihood of seeing full C-terminally eGFP-tagged protein products.

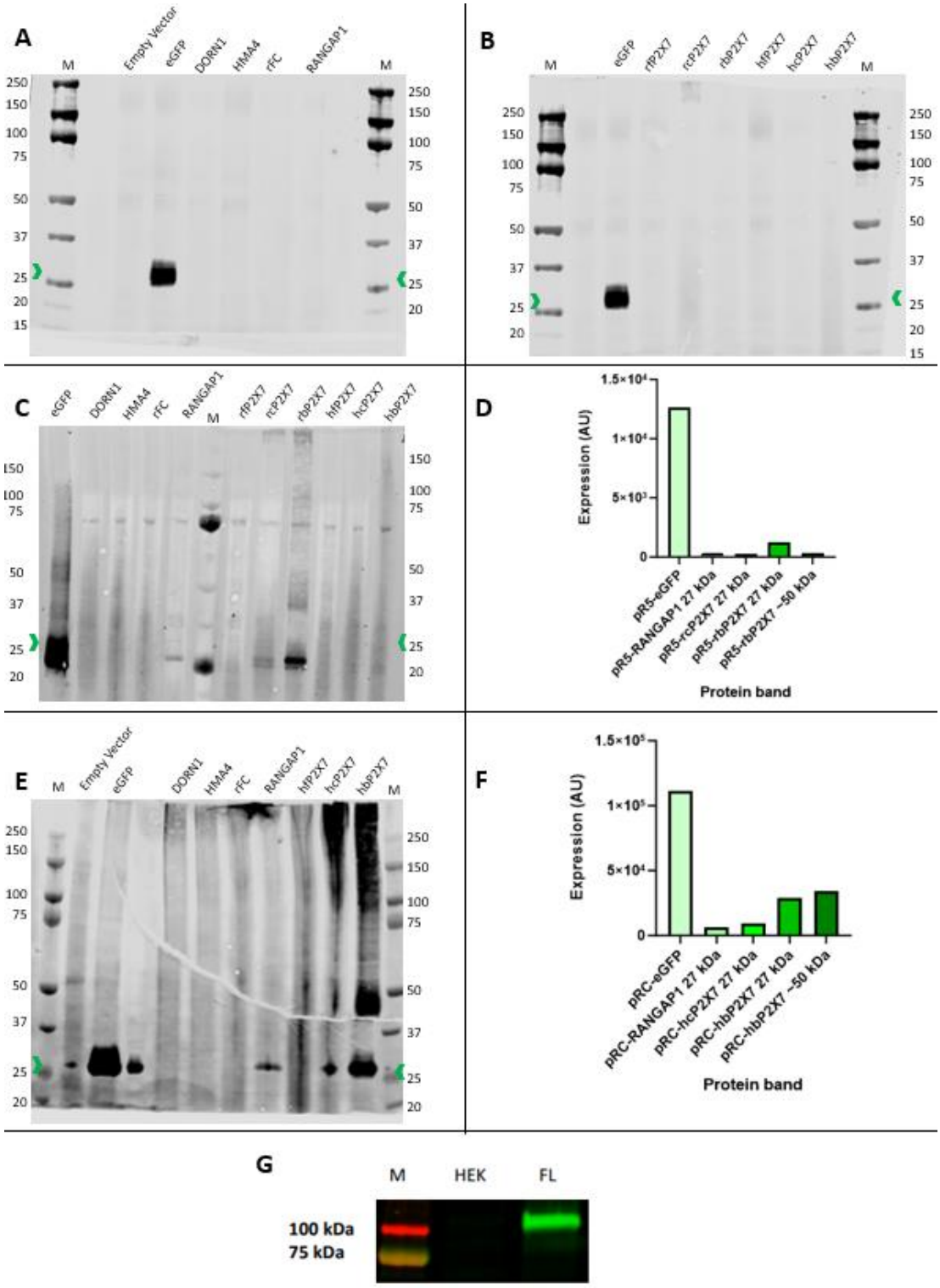


Figure 5.10: Western blot analyses of extracted proteins from pRC-CDS and pR5-CDS. Four *N. benthamiana* plants were transformed each pRC construct, and 3-4 leaves were taken from each plant. Three *N. benthamiana* plants were transformed each pR5 construct, and 2-4 leaves were taken from each plant. **A and B** – Western blot analysis to detect eGFP using a monoclonal mouse plant-specific-anti-GFP primary antibody (1/5000) and a goat-anti-mouse 800nm secondary antibody (1/13500). **C** - Western blot analysis to detect eGFP using a polyclonal rabbit anti-GFP primary antibody (1/10000) and a goat-anti-rabbit 800nm secondary antibody (1/13500). **D** = Densitometric quantification of bands in C. **E** - Western blot analysis to detect eGFP using a polyclonal rabbit anti-GFP primary antibody (1/10000) and a goat-anti-rabbit 800nm secondary antibody (1/13500). **F** = Densitometric quantification of bands in E. **G** = Western blot from Jaradat (2021) showing that the antibodies used can detect full-length P2X7 where the red band shows the 100 kDa marker, orange band shows the 75 kDa marker, and the green band shows full-length P2X7 expressed in HEK293 cells transfected with P2X7 but not untransfected HEK293 cells. In all panels, Lane M is the BioRad precision plus dual colour molecular weight marker, with the molecular weights in kDa shown either side of each gel. The green arrows indicate the molecular weight of eGFP at 27 kDa. The only target protein that produces a band corresponding to full-length protein is rat ballast P2X7 at 53 kDa, though this is degraded. This suggests that the presented expression system cannot produce most of these full-length targets. The data showing raw densitometric analyses can be seen in appendix Figure S5.1.

Therefore, to increase the amount of protein loaded onto the gel, a his-pulldown was performed on total protein extracts, firstly using pR5-CDS samples. Proteins were eluted in 500 µL of elution buffer, enabling 5% of the total protein to be analysed by SDS-PAGE. When analysed by western blot, seen in Figure 5.11, Panel A, a faint 27 kDa signal could now be seen in the truncated human P2X7 samples, confirming that human P2X7 variants can be produced but with lower yields than rat P2X7 variants, when using the 27 kDa eGFP band as a proxy. In addition, there is a faint band corresponding to C-terminally tagged rat ballast P2X7, though this is faint and smeared suggesting degradation. Densitometric analyses of the 27 kDa bands show that human cysteine-rich P2X7 and rat ballast P2X7 were expressed 12.8-fold and 42.6-fold lower than their rat counterparts (Figure 5.11, Panel B). When normalised against the total starting tissue masses, RANGAP1 expressed at 158.2 AU g^{-1}, rat cysteine rich P2X7 at 186.7 AU g^{-1}, rat ballast P2X7 at 935.4 AU g^{-1}, human cysteine rich P2X7 at 20.5 AU g^{-1-}, and human ballast P2X7 at 18.5 AU g^{-1}; this means that human cysteine rich and ballast P2X7 express 9.1-fold and 50.3-fold lower than their rat counterparts, per gram of transformed tissue. Densitometric analysis of the only C-terminal fusion protein, rat ballast P2X7, suggests that it is produced at 7.8-fold lower than its corresponding 27 kDa band, suggesting that 89% of the protein is degraded. No other C-terminally tagged protein target bands could be seen, but this may still be due to the amount of loaded protein; increasing protein concentrations may reveal more bands corresponding to C-terminally eGFP-tagged target proteins. Finally, only 1 µL of purified pR5-eGFP was analysed by SDS-PAGE and western blot, which resulted in a very dense band that is likely saturated and cannot be accurately measured.

The pR5-CDS pulldown method utilised centrifugal washes to separate the nickel beads from the unbound proteins, which also resulted in the pelleting of plant material that hindered washing and elution efficiency, requiring many washes which likely resulted in the loss of yield and less-pure product; ten washes were performed yet contaminating plant matter remained. Therefore, magnetic nickel beads were used instead, offering a favourable alternative to maximising pulldown effectiveness using magnetic bead separation instead of centrifugation. Indeed, this purification method produced much cleaner elutions with only three washes needed. When these purified samples were analysed by western blot the same trend is seen, with clear 27 kDa bands in the samples of purified eGFP, RANGAP1, human cysteine-rich P2X7, and human ballast P2X7 (Figure 5.11, Panel C), all expressed at similar levels according to densitometric analyses (Figure 5.11, Panel D). When normalised to the starting tissue mass, the expression of RANGAP1 is higher than the P2X7 variants, at 4072.7 AU g^{-1}, compared to 2017.8 AU g^{-1} for human cysteine-rich P2X7 and 2304.3 AU g^{-1} for human ballast P2X7, corresponding to RANGAP1 expression at 1.93-fold and 1.77-fold higher than human cysteine-rich and ballast P2X7, respectively. Here, 25 µL of pRC-eGFP was loaded,

resulting in a band that is likely saturated and spilled into neighbouring lanes. Interestingly, additional products can be seen in the pRC-RANGAP1 sample, including a band of the expected molecular weight at 87 kDa, as well as several smaller products suggesting degradation. A similar trend is seen for the pRC-hbP2X7 sample expressing human ballast P2X7, with a full-length degraded product detectable around 53 kDa and larger bands above 100 kDa and 150 kDa, which may be dimers or trimers of the target protein, respectively. Finally, there are additional bands in the human cysteine-rich P2X7 lane, though these are not the full-length C-terminally tagged product which would be 62 kDa and may instead be degraded product. Densitometric analyses of these bands relative to their 27 kDa band suggest that at least 72.5% of RANGAP1 protein and 51.6% of human ballast P2X7 is completely degraded, with further protein being partially degraded, creating the smearing.

Together, these western blots suggest that only the smaller cytosolic proteins can be made using the pRC and pR5 expression constructs in transiently transformed *N. benthamiana* plants. The pR5-CDS data also suggest that human P2X7 variants are more difficult to produce than rat P2X7 variants. Data from both constructs suggests that cysteine-rich P2X7 is more complex than ballast P2X7, resulting in lower yields of the former. This suggests that the additional size or presence of post-translational modification in cysteine-rich P2X7 does have an impact on yield.

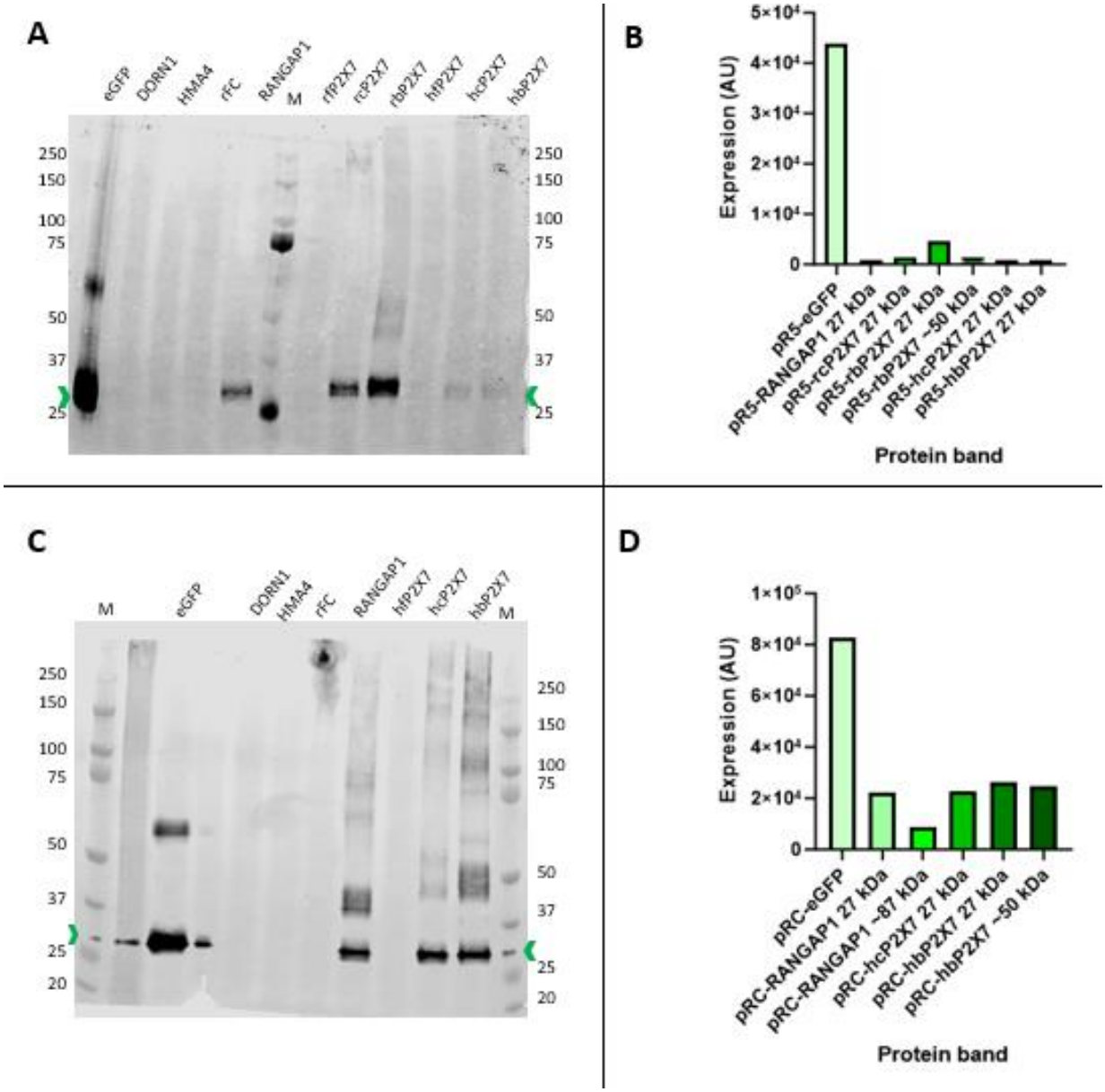


Figure 5.11: Western blot analyses of extracted proteins from p5-CDS (A and B) and pRC-CDS (C and D) following a his-pulldown. A – Western blot analysis to detect target proteins when produced by pR5, following his-pulldown using a polyclonal rabbit anti-GFP primary antibody (1/10000) and a goat-anti-rabbit 800 nm secondary antibody (1/13500). **B** = A quantified using densitometric analysis. **C** - Western blot analysis to detect target proteins produced by pRC following his-pulldown using a polyclonal rabbit anti-GFP primary antibody (1/10000) and a goat-anti-rabbit 800 nm secondary antibody (1/13500). **D** = C quantified using densitometric analysis. In all panels, Lane M is the BioRad precision plus dual colour molecular weight marker, with the molecular weights in kDa shown either side of each gel. The green arrows indicate the molecular weight of eGFP at 27 kDa. The signal of eGFP in several lanes was increased suggesting that the purification method was successful and resulted in more concentrated protein. Following concentration, the 53 kDa band corresponding to C-terminally tagged rat ballast P2X7 is fainter, suggesting a loss of target protein. However, new bands are present that correspond to C-terminally tagged targets, including human ballast P2X7 at 53 kDa, and RANGAP1 at 87 kDa, when produced by the pRC expression construct. There are also additional bands that are smaller than the full-length targets that may represent degraded target protein in the lanes corresponding to RANGAP1 and human cysteine rich P2X7 when produced by pRC. Finally, a laddering effect can be seen in the pRC-hbP2X7 sample, suggesting target protein dimers and trimers at 106 and 159 kDa, respectively. These data suggest that the expression system presented here can produce cytosolic post-translationally modified proteins, particularly when expressed using the pRC expression construct, albeit in relatively low yields compared to eGFP. The data showing raw densitometric analyses can be seen in appendix Figure S5.2.

Finally, as approximately 5% of the protein was loaded following purification, purified protein samples were concentrated using 3000 MWCO concentrator columns. These were chosen as the concentrating method as it has been reported that proteins can be concentrated with 96% yield retention, enabling up to 90% of the total protein to be analysed by western blot. When pR5-CDS proteins were concentrated and analysed (Figure 5.12, Panel A), there was a clear increase in the 27 kDa eGFP signal of RANGAP1 and the truncated P2X7 variants, suggesting that the concentrating method was effective. Densitometric analyses can be seen in Figure 5.12, Panel B, but it is impossible to estimate the recovered yield following column concentration as the signal is saturated. Conversely, when analysing the concentrated protein samples from pRC-CDS (Figure 5.12, Panel C), only RANGAP1 produces a strong signal, with a very faint signal present in human cysteine-rich P2X7. The signal in the DORN1 lane is due to a mistake in sample loading, where 1 µL of eGFP was loaded into the first two lanes. The reduction in signal in the truncated P2X7 isoform samples relative to the Figure 5.11 above suggests that the protein concentration procedure resulted in a substantial reduction in the yield. This is peculiar, as the purified P2X7 proteins would be expected to behave similarly when produced by either pRC or pR5, as the protein product should remain the same. However, as pRC produced more of the human P2X7 isoforms initially, it is possible that this increase in protein concentration caused the protein to become insoluble and aggregated, reaching a critical threshold resulting in the loss of yield during concentration that pR5-produced human P2X7 isoforms did not reach. Importantly, in all protein extracts from both pRC and pR5, no new bands are present, and the only full-length product detectable was degraded rat ballast P2X7 when produced by pR5. In other words, the larger molecular weight products that were present following purification of the proteins were lost after concentration, suggesting that the concentrating method was unsuitable for the full-length tagged targets.

Overall, these data suggest that the transient plant expression system presented here can produce only cytosolic proteins in detectable yields, with post-translational modifications potentially having an impact although further experiments are needed to detect these and confirm this. The only full-length C-terminally tagged target proteins that had the correct molecular weight bands present on western blots were RANGAP1 and ballast P2X7. These had reduced signals relative to the eGFP signal in their lanes, suggesting that the target proteins were degraded. If using the 27 kDa eGFP band as a proxy, this suggests that all the small cytosolic proteins can be produced but are degraded. There was no detectable eGFP signal *in planta* or at the extracted protein level for any large or membranous proteins, suggesting that the expression system may not be suitable for proteins of this nature. However, there could be other reasons for the apparently absent yields that should be investigated before definitively making this conclusion; briefly, the detection methods and

downstream processing utilised may not be suitable for all proteins, and membranous proteins may require higher masses of transformed plant material than utilised here. These are valid avenues for future work which are discussed in detail in the following Chapter.

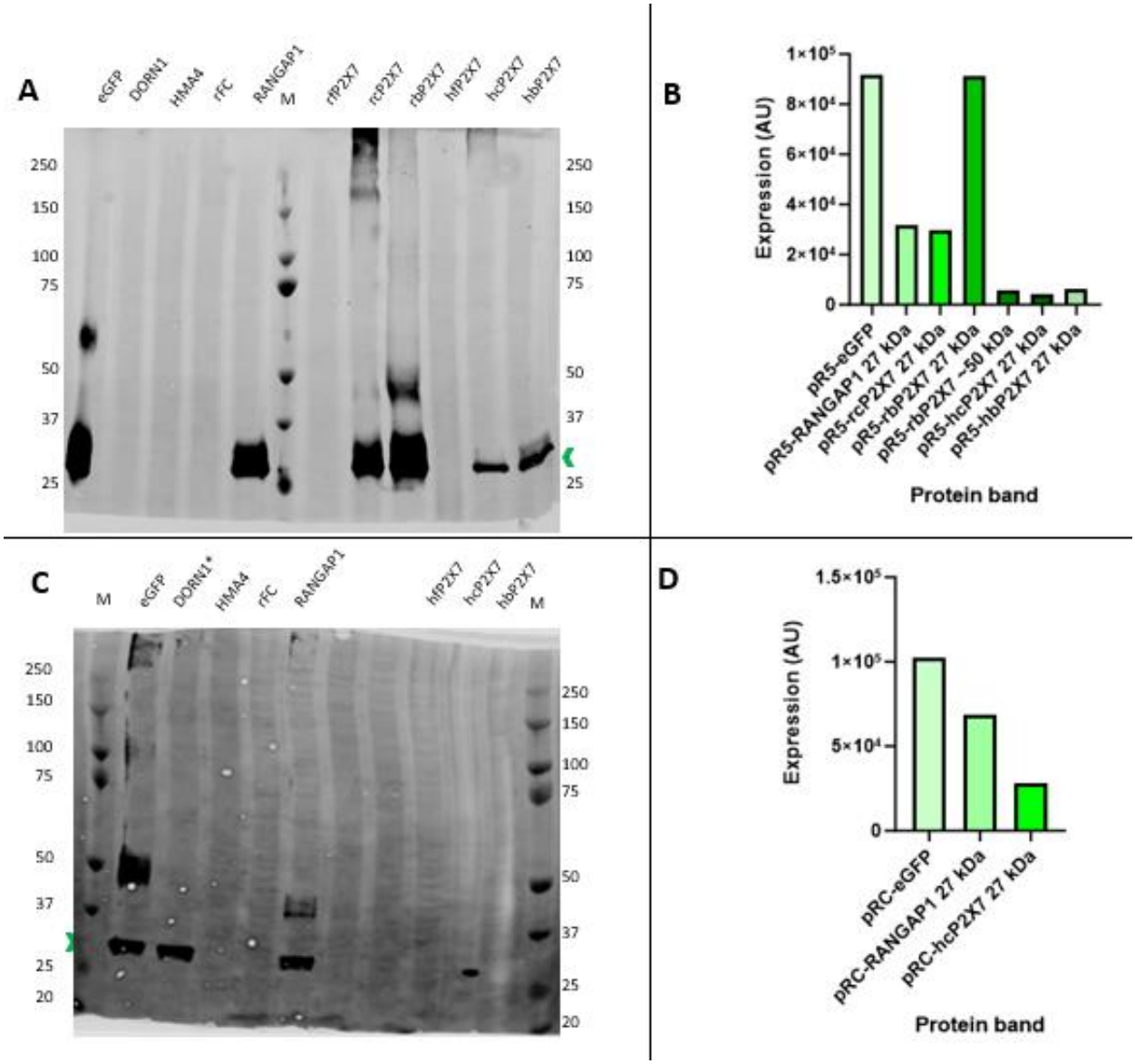


Figure 5.12: Western blot analyses of extracted proteins from pRC-CDS and pR5-CDS following a His-pulldown and protein concentration. A – Western blot analysis to detect eGFP following his-pulldown and protein concentration using VivaSpin columns using a polyclonal rabbit anti-GFP primary antibody (1/10000) and a goat-anti-rabbit 800nm secondary antibody (1/13500). **B** = Densitometric quantification of bands in A. **C** - Western blot analysis to detect eGFP following his-pulldown and protein concentration using VivaSpin columns using a polyclonal rabbit anti-GFP primary antibody (1/10000) and a goat-anti-rabbit 800nm secondary antibody (1/13500). *The pRC-DORN1 lane was accidentally loaded with eGFP too resulting in false signal. **D** = Densitometric quantification of bands in C. In all panels, Lane M is the BioRad precision plus dual colour molecular weight marker, with the molecular weights in kDa shown either side of each gel. The green arrows indicate the molecular weight of eGFP at 27 kDa. The signal of eGFP in several lanes was increased suggesting that the purification method was successful and resulted in more concentrated protein. However, all full-length C-terminally tagged target proteins, except pR5-rbP2X7, were lost during the concentrating process, suggesting that this method was unsuitable for preserving the yields of the C-terminally tagged targets. The data showing raw densitometric analyses can be seen in appendix Figure S5.3.

5.3.2.3 RT-PCR detection of target protein transcripts

As discussed above, most of the full-length target proteins were not detected by western blot, even following protein purification and concentration. There could be several reasons for this, which are discussed at the end of this chapter, but one possibility is that the full-length transcripts were not produced, or transcripts were not produced at all. Reverse transcription polymerase chain reaction (RT-PCR) enables detection of transcripts and could rule out this scenario. However, to do this an RNA extraction method needed to be optimised for *N. benthamiana* plants, a suitable reference gene identified for use as a control, and a suitable eGFP primer pair found that produces a single specific amplicon for transcript detection. To optimise RNA extraction, several methods were tried, documented in detail in Chapter 2. Briefly, a pre-established method for RNA extraction from *A. thaliana* was performed, involving a phenol, chloroform and isopropanol extraction, but this failed to provide substantial yields of RNA when used on *N. benthamiana* leaves, with large amounts of DNA present. Following, a plant RNA extraction kit, RNeasy, was tested, using the buffers optimised for plants with high levels of secondary metabolites as *Nicotiana* plants are typically high in compounds such as phenolics and alkaloids. This was an improvement as there was very little contaminating DNA, but the RNA yields were very low. Instead, a third RNA extraction was attempted, with similar principles to the *A. thaliana* phenol, chloroform and isopropanol method, but with different ratios of each reagent. This used TRI Reagent from Zymo Research but the protocol for TRIzol™ Reagent from Invitrogen instead. Using this method, RNA of sufficient quality and quantity was obtained, with minimal DNA contamination as DNase treatment did not significantly change the concentration of nucleic acid by nanodrop, which showed an A260/280 ratio of ~2.0. The RNA using this method was of high quality when analysed by agarose gel electrophoresis (Figure 5.13). As a result, this was the chosen method for subsequent RNA extractions. With a suitable RNA extraction protocol developed, cDNA was synthesised. The pR5 constructs utilise replicating viral RNA to create transcripts that lack poly-adenylated tails meaning that random primers, instead of Oligo dT primers, had to be used to produce cDNA.

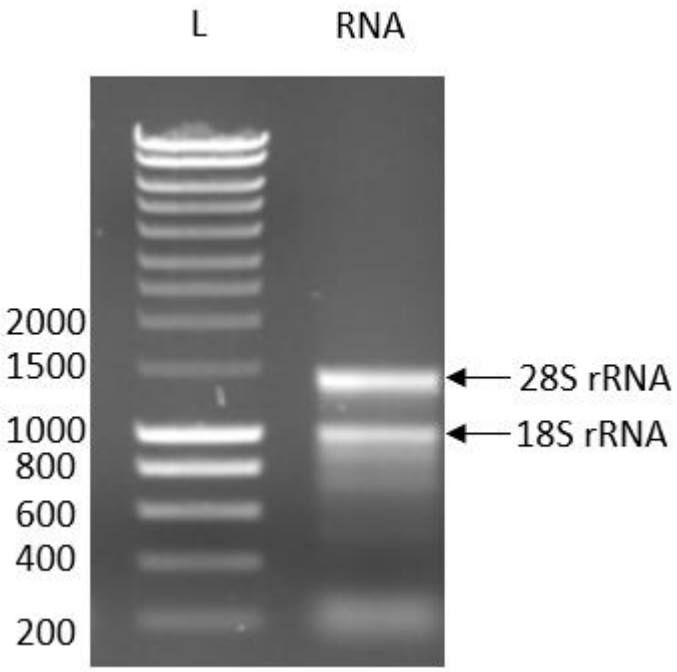


Figure 5.13: RNA quality when extracted using TriReagent with the TriZol method shown to work for *N. benthamiana* leaves. Of the tested RNA extraction methods in this research, this method resulted in the cleanest RNA with the fewest impurities so was chosen as the main RNA extraction method. Lane L is the Eurogentec SmartLadder, the lane containing the RNA shows the 28S and 18S rRNA (labelled) as wel as smaller RNA around 200 bp.

In addition, a series of reference genes were assessed using cDNA made from RNA from untransformed *N. benthamiana* tissue to find a suitable reference gene. After reviewing the literature, five reference genes were selected and tested using a an RT-PCR; these were Actin-1, F-box, GAPDH, L23, and PP2A, reported to be suitable reference genes in the literature (Liu et al. 2012). When the products were analysed, it was decided that Actin-1 and GAPDH were the most suitable reference genes to use as these produced distinct single products. However, as GAPDH is involved in metabolism, there is a chance that expressing high levels of target protein may influence cellular metabolism which could affect the transcription of GAPDH, making it unsuitable for comparison. As a result, Actin-1, involved in cellular structure, was the preferred reference gene, as these RNA levels were less likely to be influenced by target protein expression.

With an RNA extraction and cDNA synthesis pipeline established and a suitable reference gene identified, a primer pair to quantify the eGFP transcript needed to be designed; this should produce a single clear and specific product. A total of 13 primer pairs to amplify eGFP were created using the Primer3 design tool, either built into Benchling or using the Primer3 website, and their sequences can be found in Table S1. These primers were designed to amplify eGFP as each transcript should have the eGFP sequence at the 3' end, so all transcripts should be able to be detected.

Using Actin-1 as a reference (Genbank: AY594294.1), a series of RT-PCRs were performed on plant tissue transformed with either empty pJL-TRBO vector or pR5-eGFP, and the products were analysed by gel electrophoresis. Most of these produced either no product, the wrong product, multiple products, or also amplified in the wild-type sample, and are not shown. Only two primer pairs, both designed using the Primer3 tool, were carried forward for testing. The results of the RT-PCR reactions to test these primers are shown in Figure 5.14. In the initial experiment, each produced a single product of the correct size in the eGFP-expressing sample with a faint band in the wild-type sample but not the no-template control (NTC) sample, suggesting contamination had occurred before the RT-PCR (Figure 5.14, Panel A). In addition, the Actin-1 control only showed amplification in the empty vector sample.

As a result, a fresh RNA extraction was performed from *N. benthamiana* tissue transformed with all pR5-CDS constructs, including the pR5-eGFP positive control and an empty pJL-TRBO vector negative control. pR5 was chosen as this expression construct was successfully made to express all the target proteins, including the rat P2X7 variants which were not expressed using pRC, and showed high levels of expression in Chapter 3. The extracted RNA was DNase treated and cDNA was synthesised using random primers. Shown in Figure 5.14, Panel B, a quality-control RT-PCR was first performed on the controls to detect the presence of suitable reference genes in both samples, which showed

amplification of Actin-1, GAPDH, and L23, showing successful RNA extraction and cDNA synthesis in both. In addition, a test for contamination used Primer3 pair 3 to detect eGFP transcripts, which produced the expected band in only the eGFP sample, with both the negative controls showing no amplification. This showed that the reference gene primers and eGFP detection primers used were suitable for the detection of other pR5-CDS transcripts.

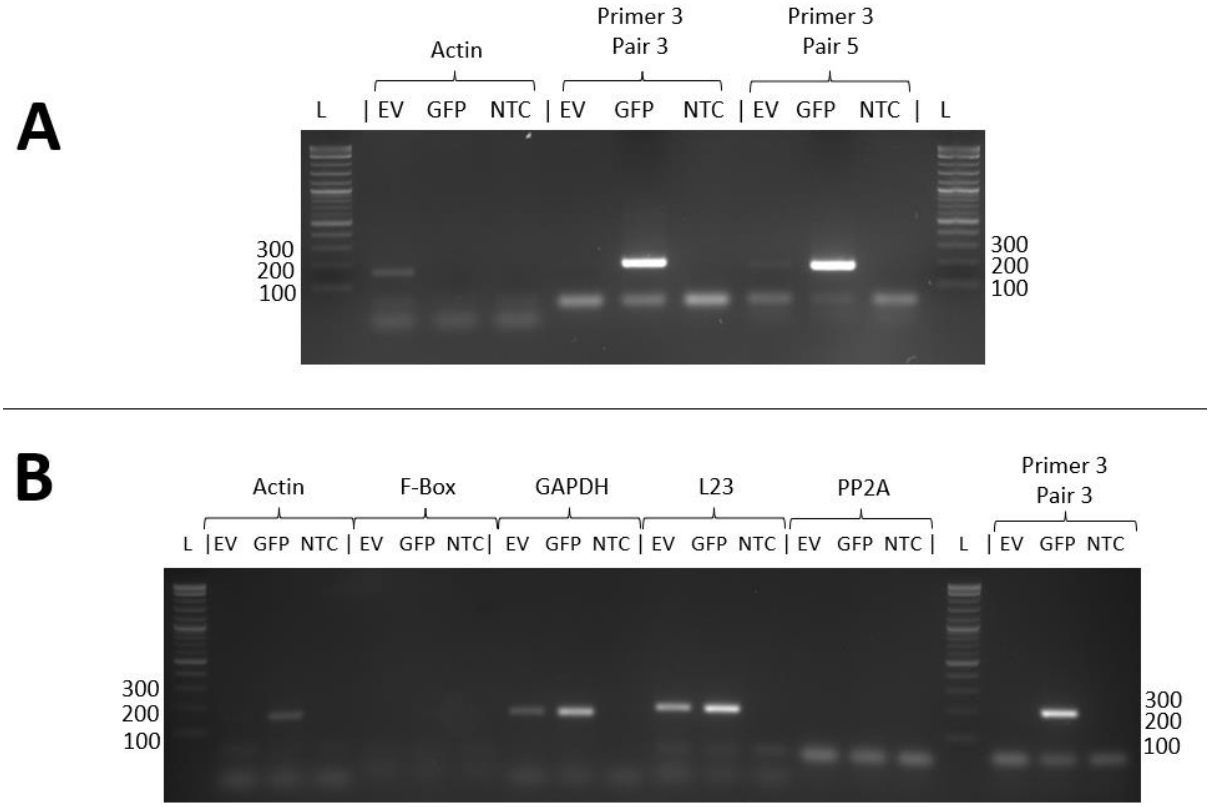


Figure 5.14: RT-PCR products when testing eGFP primer pairs P3.3 and P3.5, alongside reference gene controls. RNA was extracted from plants transformed with either empty pJL-TRBO vector or pR5-eGFP, and cDNA was synthesised using random primers. Panel **A** shows the results of an RT-PCR to test the eGFP-amplifying primers Primer3 pair 3 (P3.3) and Primer3 pair 5 (P3.5). Both produced the correct product in the GFP-expressing sample but also produced fainter products in the empty vector sample. The Actin-1 control only produced a band in the empty vector sample, so all references genes were tested using an RT-PCR in panel **B**. This shows that Actin-1, GAPDH, and L23 can all amplify the correct transcript in tissue transformed with either empty vector or pR5-eGFP. These data showed that the use of Actin-1 as a reference gene and eGFP detection primers P3.3 and P3.5 were suitable.

With suitable primers for transcript detection developed an RT-PCR was performed to each transcript. Interestingly, Actin-1 was not detected in every sample (Figure 5.15, Panel A), but each sample showed detection of eGFP with one or both primer pairs, except the empty vector and no-template controls (Figure 5.15, Panels B and C). Importantly, the Actin-1 amplicon was not detected in the empty vector control sample in this RT-PCR, but several reference genes were detected using the same RNA sample in Figure 5.14, showing that this sample does contain RNA. These data suggest that the transcripts were successfully being produced with each construct but does not determine whether they were full-length transcripts.

As a result, an additional RT-PCR reaction was performed using primers that amplify between the 5' UTR and the 3' end of the C-terminal eGFP tag, enabling full-length amplification of target RNA molecules. When products were analysed by agarose gel electrophoresis, shown in Figure 5.15, Panel D, only the pR5-eGFP sample showed production of the full-length amplicon, with no other full-length products shown. These primers are suitable for the amplification of the full-length targets, as they produced bands of the correct size when amplifying the full coding sequence using pR5-CDS vector as a template (Figure 5.15, Panel E). Seeing amplification of only eGFP was expected as the full-length RNA products range from 1.4 kb for the ballast P2X7 transcripts to 4.3 kb for the HMA4 transcript; longer RNA molecules are harder to detect than shorter molecules as it relies on a longer RNA sequence being intact to amplify from when making cDNA, particularly when cDNA was synthesised using random primers which can bind anywhere in the sequence. Overall, whilst this experiment does not definitively show that the full-length transcripts were produced for each target protein, it does suggest that the transcripts were present as the short eGFP amplicon was detected in each sample, and the full-length target transcripts were different to the eGFP positive control because the eGFP band at 800 bp was not detected in other samples.

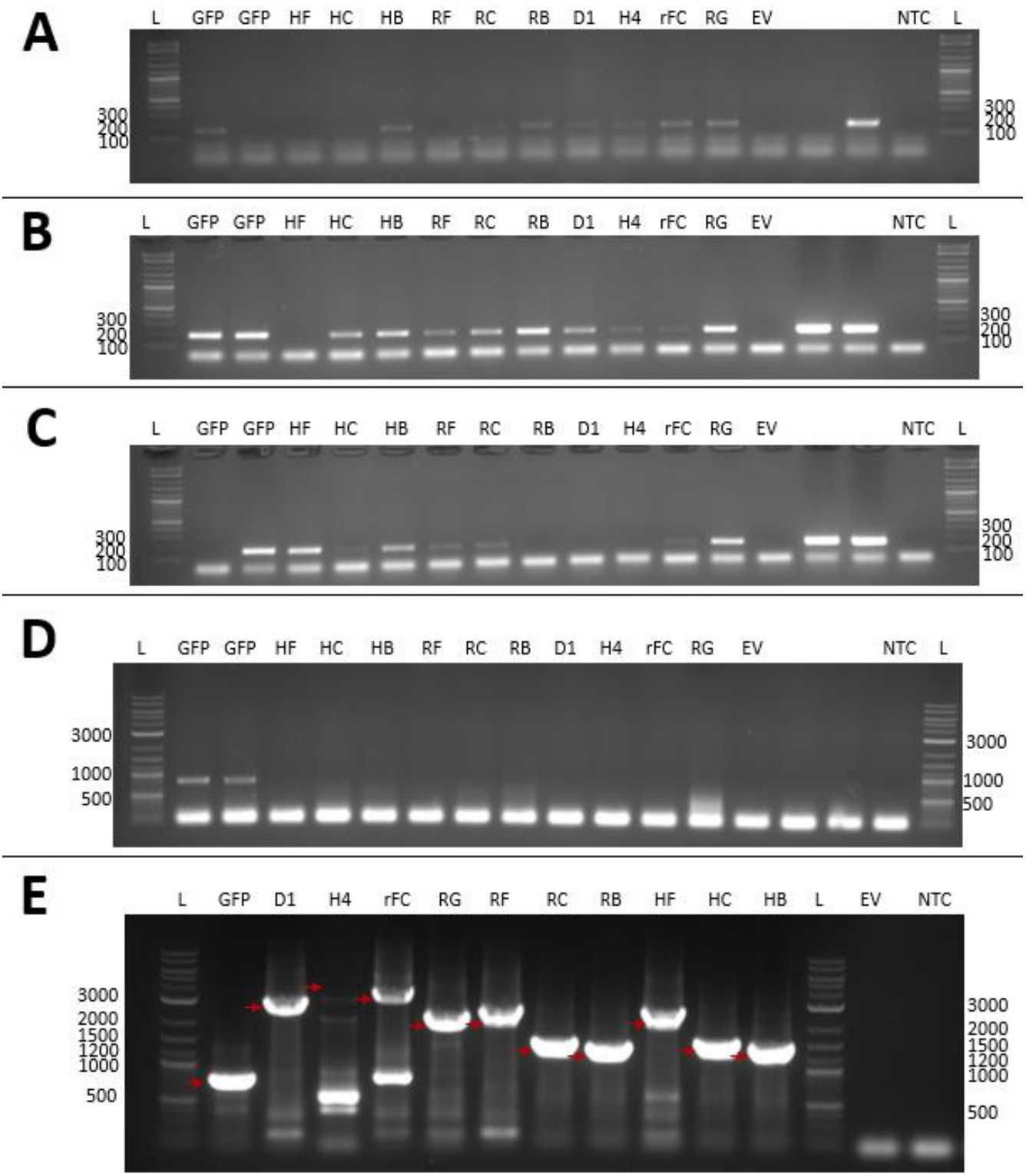


Figure 5.15: RT-PCR products when assessing whether plants transformed with the pR5-CDS constructs produce target eGFP transcripts. A – amplification of Actin-1 in samples creating an amplicon of 163 bp. **B** – amplification of eGFP using primer pair P3.3 creating an amplicon of 202 bp. **C** – amplification of eGFP using primer pair P3.5 creating an amplicon of 184 bp. Although the empty vector control sample did not produce any amplification of Actin-1 in this experiment, it did in previous experiments using cDNA created from the same RNA sample, shown in Figure 5.14. **D** – amplification of the full-length trancripts using a forward primer that binds to the 5' UTR, and reverse primer that binds to the C-terminal eGFP tag. Only eGFP produced the correct amplicon, with no amplification seen in other samples suggesting that the pR5-CDS amplicons seen in panels B and C are not the eGFP product. **E** – PCR amplification of the pR5-CDS vectors using the same primers used in Panel D, showing that these primers can produce the full-length products, indicated by red arrows. Only HMA4 did not produce the correct product, instead amplifying smaller several products. In all panels, NEB 1Kb Plus DNA Ladder (L) was used.

5.3.3 Expressing different target proteins using different expression systems

Throughout this research, other plant-based protein expression systems were also tested, primarily to assess whether they could produce the P2X7 isoforms. These included *A. thaliana* plants, both through transient transformation of protoplasts and stable transformation of whole plants, transiently transformed plant cell packs derived from *N. tabacum* BY-2 cells, and an almost living cell free expression (ALiCE) system. Each showed varying levels of effectiveness at producing different recombinant proteins and will be discussed in this subsection.

Firstly, each of the different coding sequences were expressed in PCPs, under control of the pR5 construct. When analysing the *in planta* fluorescence 4 days post transformation, some constructs showed significantly higher fluorescence than the empty vector controls (Figure 5.16). Specifically, HMA4, rat ballast P2X7, and human cysteine rich P2X7 each displayed significantly higher fluorescence than the negative control ($p < 0.05$). However, only the human cysteine rich construct displayed significantly higher fluorescence than all other samples. These expression data are different from the transiently transformed *N. benthamiana* experiment where the eGFP and ballast P2X7 constructs expressed at the highest level, which may suggest that different plant expression hosts have varying capacities to produce different recombinant proteins. There are many hosts available, discussed in Chapter 1, which may represent important avenues for future work.

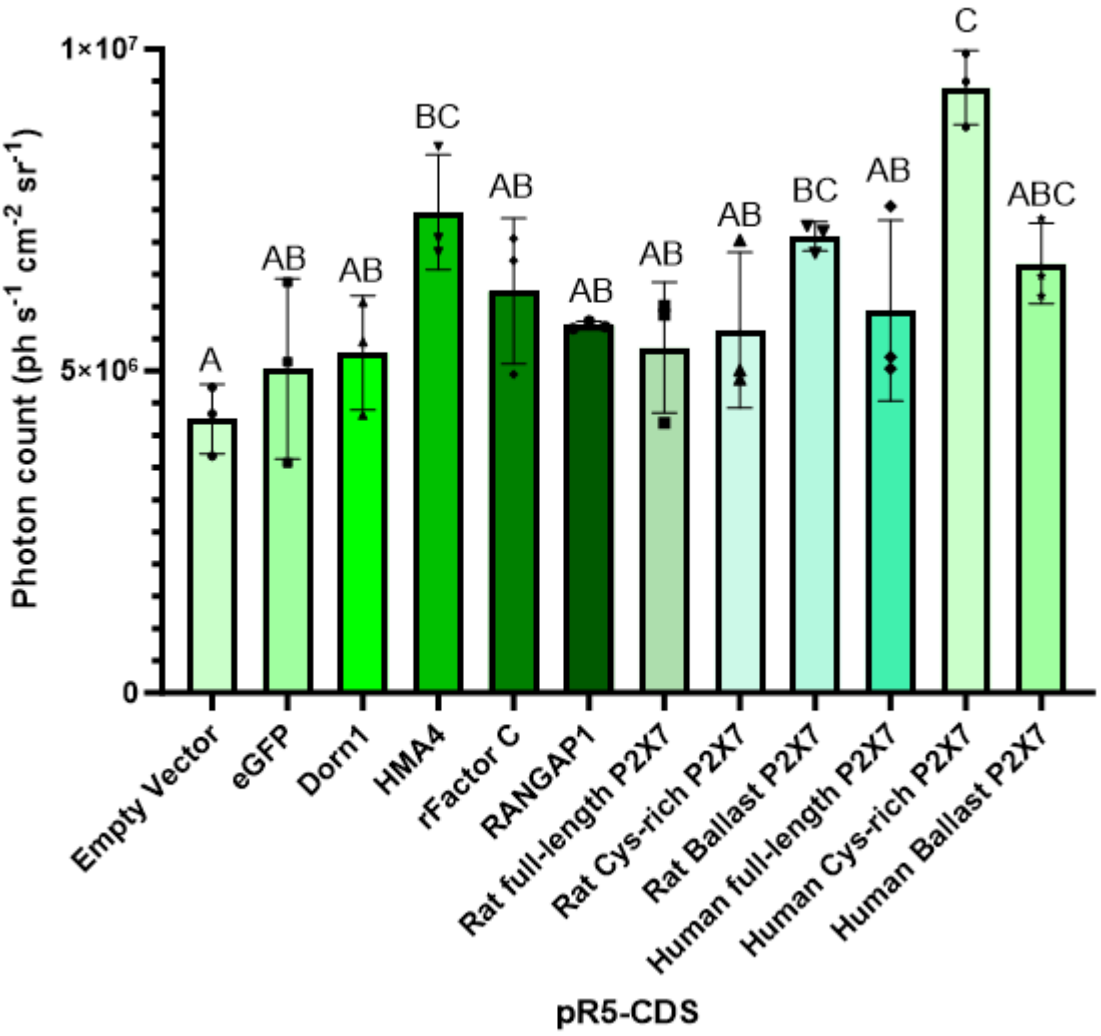


Figure 5.16: eGFP fluorescence of PCPs transformed with pR5-CDS, quantified using a Biospace Optima photon imager. A one-way ANOVA test shows a statistically significant difference between samples overall ($F(11, 24) = 6.381$, $p = 7.69 \times 10^{-5}$). However, these data show that only HMA4, rbP2X7, and hcP2X7 showed significantly higher fluorescence than the empty vector control. This may be due to the high autofluorescence of this expression system causing a reduction in sensitivity in this experiment. However, these data do suggest that HMA4 can be expressed in PCPs. Letters show compact letter display with an alpha value of 0.05. Samples that do not share a common letter have significantly different expression, and samples that share a common letter show expression that is not significantly different. Error bars show minimum and maximum values. The high fluorescence in the empty vector negative control sample exhibits the high level of green autofluorescence produced by PCPs. This reduces the sensitivity of detecting target proteins tagged with eGFP.

As mentioned at the start of this chapter, early P2X7 expression was carried out in the pNS construct and transiently transformed into *N. benthamiana* leaves with very low expression levels seen for all, only detectable using a confocal microscope (Figure 5.17) and not visible using a photon imager, leading to the creation of the constructs discussed in Chapter 3. The expression of human full-length P2X7 was not definitively seen, with only a single cell showing fluorescence nearly indistinguishable from background levels. In addition, *A. thaliana* protoplasts were also extracted and transformed with the same constructs. Again, expression of five of the six P2X7 isoforms was seen in *A. thaliana* protoplasts, with only full-length human P2X7 failing to show any transformants when screened (Figure 5.18, Panel A). Interestingly, corrected total cell fluorescence (CTCF) analyses of these protoplasts (Figure 5.18, Panel B) suggest that human cysteine-rich P2X7 was expressed at the highest level, a similar finding to the PCP experiment mentioned above, but different from the transient *N. benthamiana* transformations discussed earlier in this chapter. Importantly, due to the low transformation efficiency of protoplasts, this analysis was only able to be performed on one protoplast from each sample, so the variability in the expression is unknown. This is in part because the protoplast transformation method used relies on several washes using centrifugation, often resulting in the lysis and subsequent death of protoplasts. These data may also indicate the importance of different expression hosts for different recombinant proteins, but an increased sample size and statistical power is necessary to make this conclusion.

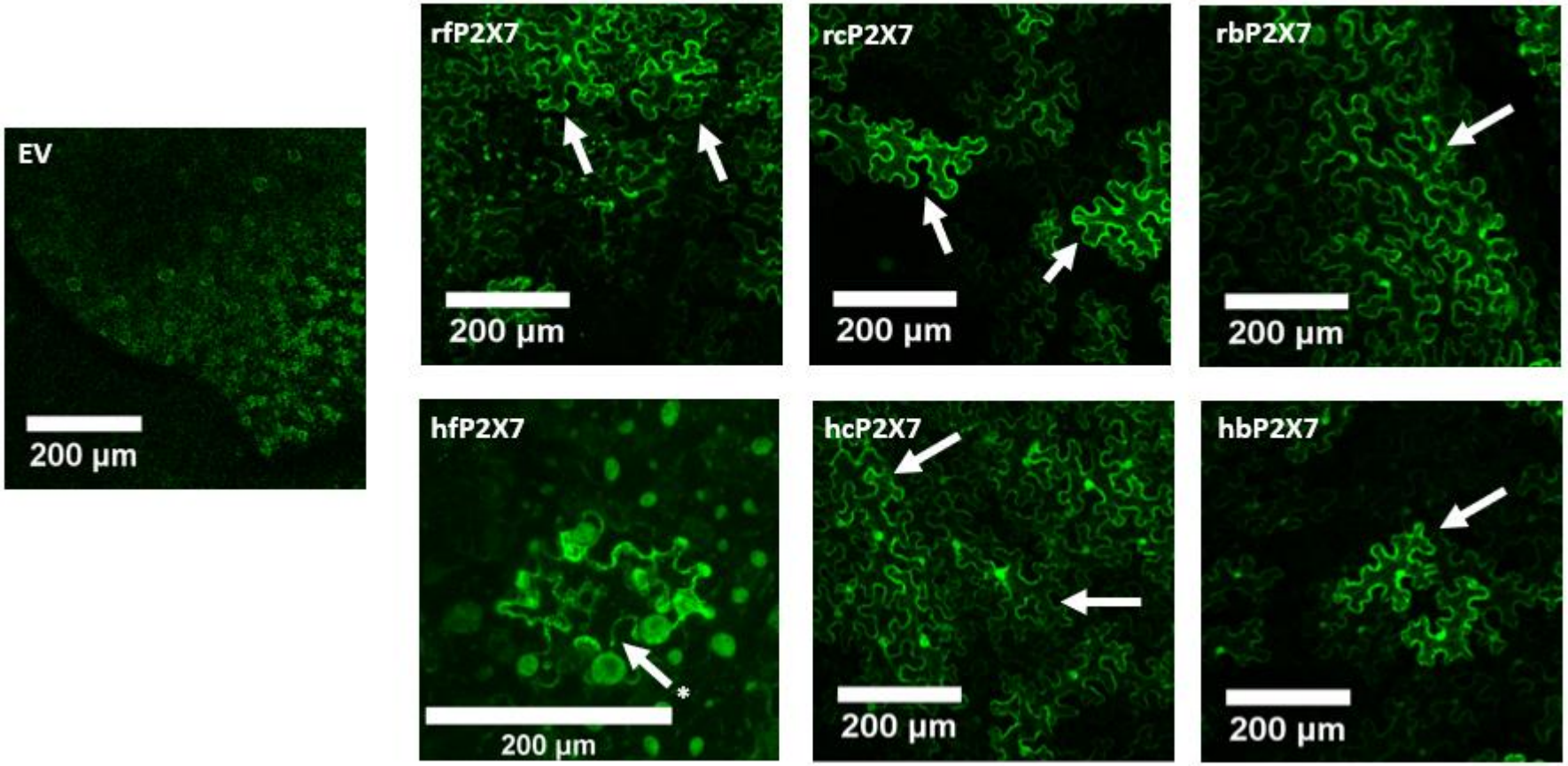


Figure 5.17: Confocal images of pNC-P2X7 variants expressed transiently in *N. benthamiana* plants. Images show 50 µm Z-stacks of transformed areas using a gain of 731 and an excitation wavelength of 488nm; scale bars represent 200 µm. These data suggest that all P2X7 variants can be expressed under control of pNC, though expression of human full length is disputable. Examples of cells that show eGFP fluorescence, suggesting successful expression of the tagged P2X7 protein, are indicated with an arrow. When transforming plants with human full-length P2X7, only a single cell showed very low-level fluorescence which was too close to background fluorescence levels to definitively tell that it is expressing the protein.

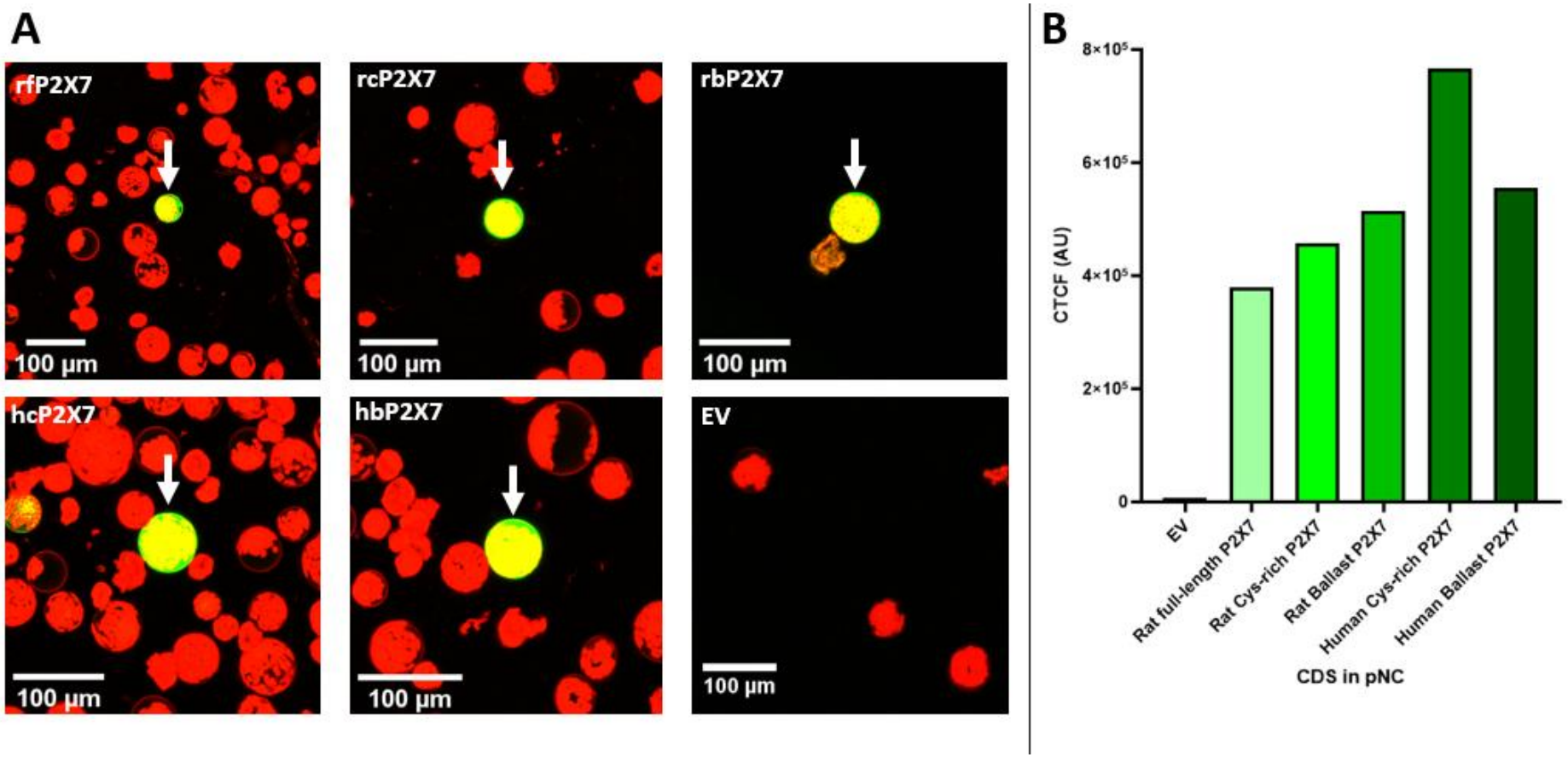


Figure 5.18: Confocal images of pNC-P2X7 variants expressed transiently in *A. thaliana* protplasts. A – confocal images of protoplasts, stained with FM4-64, showing Z-stacks of transformed areas using a gain of 731 and an excitation wavelength of 488nm; scale bars represent 100 µm. White arrows indicate protoplasts expressing the target protein. B – ImageJ CTCF analysis of only one protoplast from each construct, indicated by white arrows. These data suggest that all P2X7 variants except human full length P2X7 can be expressed under control of pNC.

Throughout this research, stable transformations were also carried out, though with limited success largely due to poor transformation efficiencies, the loss of early stable lines during the COVID-19 pandemic, and time constraints to make transformants with later constructs. Stably transformed plants are rarely used to produce recombinant proteins due lower variable yields and lengthy generation times needed to produce homozygous lines, relative to transient transformants (Tyurin et al. 2020; Garabagi et al. 2012). However, valuable insights can be gained from studying stable transformants, as there is reduced variability and noise between stable transformed lines than between transiently transformed plants. Due to the lengthy times required to make stably transformed plant lines only a few transgenic lines were prioritised. These were plants expressing the P2X7 constructs using pNS and pNC. Unfortunately, only some of these were successful and are discussed below.

The six early pNS-P2X7 constructs were stably transformed into *N. benthamiana* plants through leaf disc transformation and calli regeneration, and *A. thaliana* plants were transformed using floral dipping. *N. benthamiana* transformants were recovered from only the rat cysteine rich P2X7 construct, with no antibiotic resistant plants emerging from the other five P2X7 constructs. Similarly, the transgenic *A. thaliana* plants expressed rat cysteine rich P2X7 and rat ballast P2X7. Attempts to analyse these plants using the photon imager were unsuccessful, showing no detectable fluorescence (not shown). However, confocal microscopy did reveal low levels of eGFP expression. Importantly, no T3 transformants were obtained due to time constraints, but T2 transformants were analysed which were not assessed for homozygosity at the transgene locus, an important avenue for future work. The *N. benthamiana* T2 transformants expressing rat cysteine-rich P2X7 in the construct pNS-rcP2X7 showed variable patchy expression, seen in Figure 5.19. In contrast, the same construct transformed into *A. thaliana* plants showed more uniform expression, seen in Figure 5.20. The T2 *A. thaliana* plants expressing rat ballast P2X7 (Figure 5.21) under control of pNS-rbP2X7 showed much lower fluorescence relative to rat cysteine-rich P2X7, which could be due to genome integration effects.

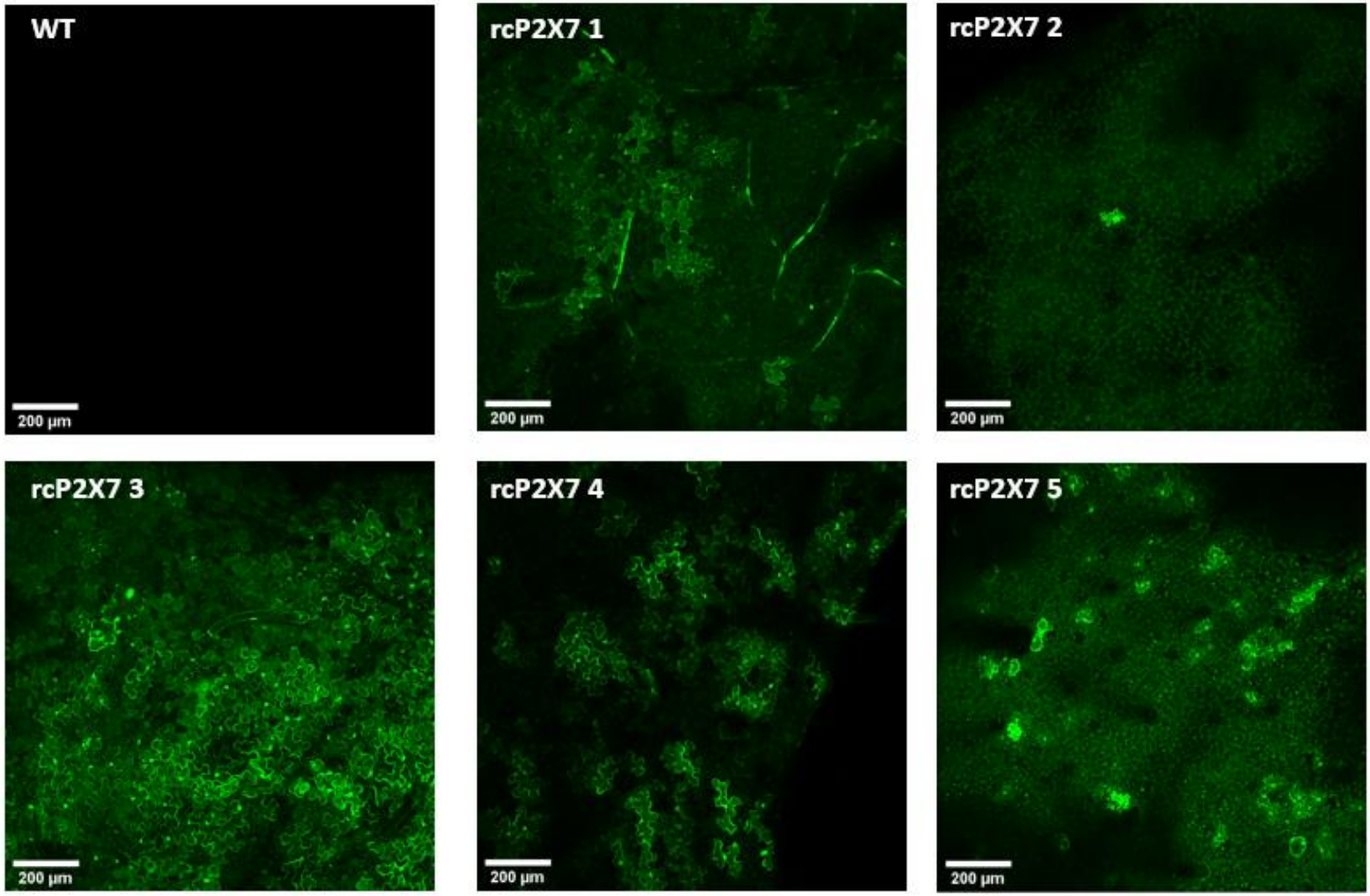


Figure 5.19: Confocal images stably transformed T2 *N. benthamiana* plants expressing rat cysteine-rich P2X7. Images show 50 µm Z-stacks of transformed areas using a gain of 455 and an excitation wavelength of 488nm; scale bars represent 200 µm. Brightness and contrast were equally increased across images to aid viewing.

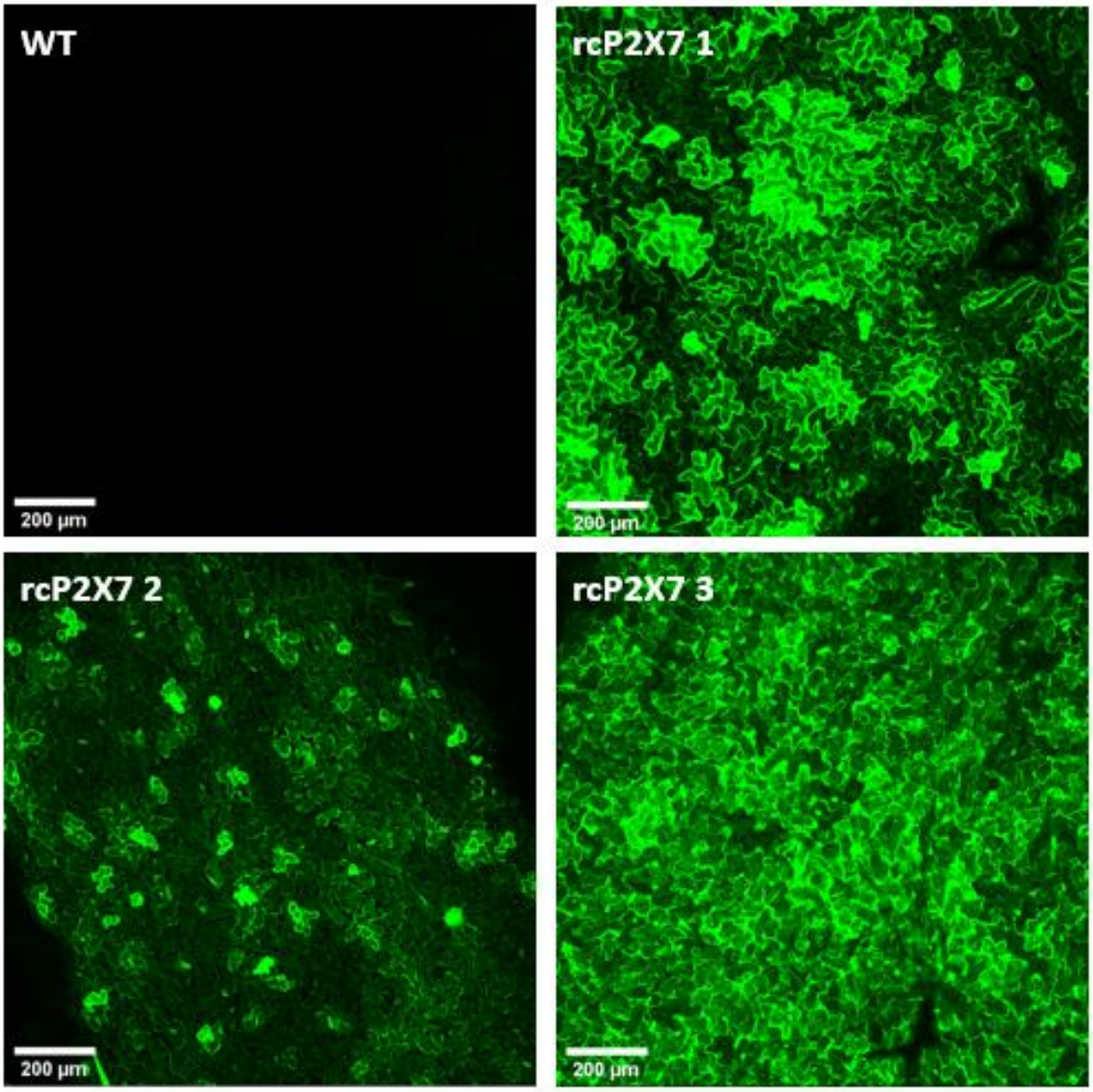


Figure 5.20: Confocal images stably transformed T2 *A. thaliana* plants expressing rat cysteine-rich P2X7. Images show 50 µm Z-stacks of transformed areas using a gain of 455 and an excitation wavelength of 488nm; scale bars represent 200 µm. Brightness and contrast were equally increased across images to aid viewing.

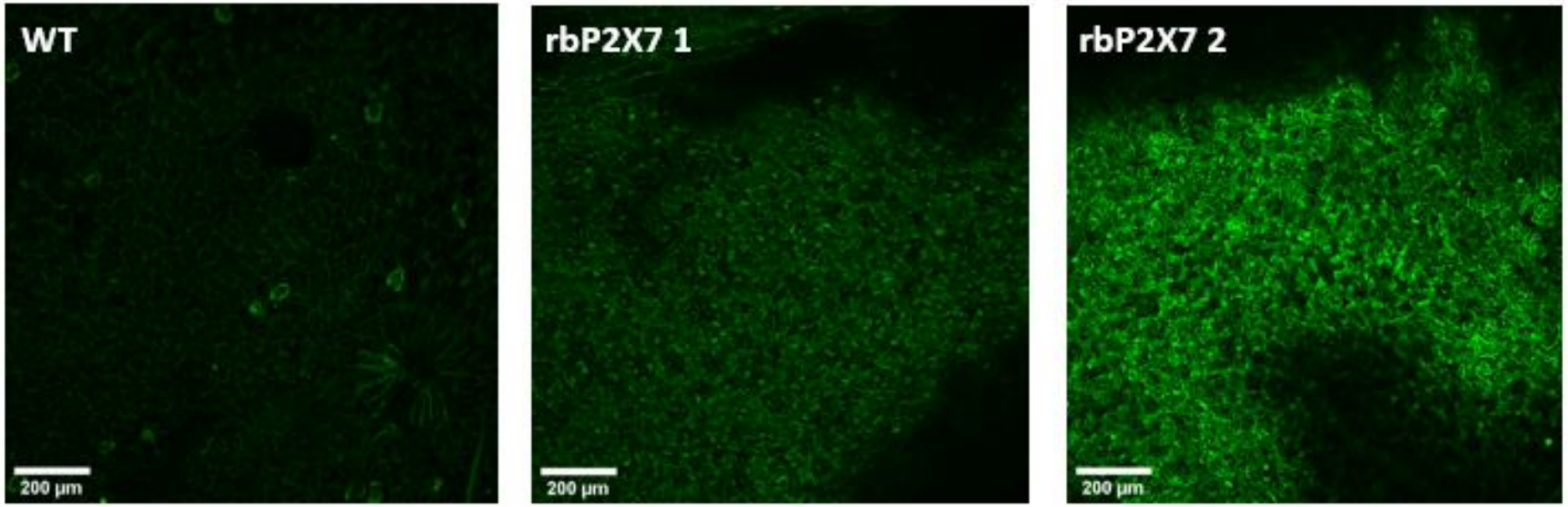


Figure 5.21: Confocal images stably transformed T2 *A. thaliana* plants expressing rat ballast P2X7. Images show 50 µm Z-stacks of transformed areas using a gain of 455 and an excitation wavelength of 488nm; scale bars represent 200 µm. The expression of these constructs is barely distinguishable from the negative control, suggesting very low expression. Brightness and contrast were equally increased across images to aid viewing.

Integration effects are another potential downside of stably transformed plants, as the expression level of a protein is largely dependent on the chromatin environment that the transgene integrates into; if integrating into heterochromatin, the expression is likely to be decreased compared to integration into a euchromatic environment. To some extent, the matrix attachment region found in the complex expression cassette, pNC, can help alleviate this by ensuring an euchromatic environment. As a result, several attempts were made to transform *A. thaliana* plants using with the full length P2X7 isoforms under control of the complex expression cassette, pNC. This had limited success, with only T1 and T2 transformants being produced in time. These plants were analysed by confocal microscopy; only one plant transformed with human full-length P2X7 was resistant to selection media and showed faint but unconvincing levels of expression when analysed by confocal microscopy, despite the antibiotic resistance. For rat full-length P2X7, only two plants were resistant to the selection media, but both plants showed detectable eGFP fluorescence with one expressing higher than the other. These images are shown in Figure 5.22 and suggest that the full-length rat P2X7 isoform is being expressed in these T1 transgenic plants. The three transgenic plants were grown to maturity, allowed to self-fertilise and produce seed to generate T2 plants. These were grown on selective media and resistant seedlings had cotyledons analysed by confocal microscopy using a Zeiss LSM 880 confocal microscope. Shown in Figure 5.23, fluorescence could be seen in each of the T2 transformants, suggesting that stable transgenic plants had been generated that express both full-length human and rat P2X7 variants. In addition, RNA was extracted from several seedlings, DNase treated, and the quality analysed (Figure 5.24, Panel A). cDNA was then created using oligo(dT) primers to amplify mRNA transcripts. When analysed by RT-PCR, Actin-2 and eGFP were only detected in the tissue sample expressing human full-length P2X7 and were absent from the wild-type control and samples expressing rat P2X7 (Figure 5.24, Panel B). Furthermore, a band corresponding to the full-length human P2X7 transcript was detected when cDNA was amplified using a forward primer in the 5' UTR and a reverse primer in the C-terminal eGFP sequence (Figure 5.24, Panel C). This confirms that the mRNA transcript in the human full-length P2X7 T2 transformant encodes the full-length protein. However, the expression of the rat P2X7 variant remains unknown, due to the failure to detect either Actin-2, eGFP, or the full-length P2X7 transcripts. Further research could involve producing clonal T3 transgenic plants, where expression analyses could be performed at the protein level. Alternatively, functional assays could be attempted using these stable transgenics lines in both whole plants and protoplasts to assess P2X7 function in plant cells with and without a cell wall.

Importantly, there were no stable transformant pairs suitable for comparison, as construct comparisons would require the expression of the same protein using the different constructs, pNS

and pNC, which were not successfully produced. In addition, protein expression comparisons would require different proteins produced using the same construct. To some extent this was successful, but there were not enough plant lines produced to account for the random insertion into different chromatin environments and clonal T3 plant lines were not made due to time constraints. Therefore, robust comparisons of these stable plant lines are not possible.

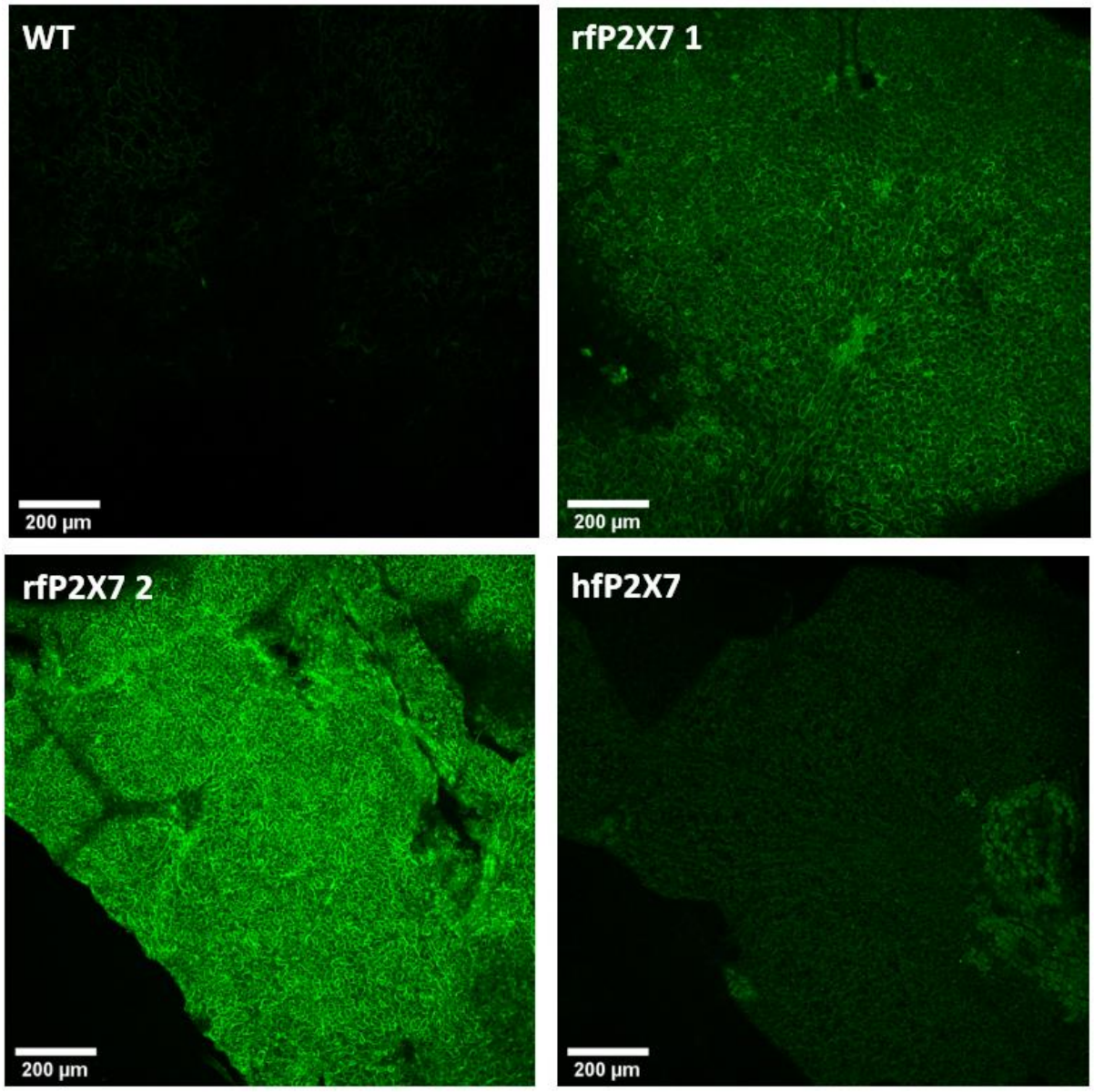


Figure 5.22: Confocal images stably transformed T1 *A. thaliana* plants expressing rat and human full-length P2X7. Images show 50 µm Z-stacks of transformed areas using a gain of 455 and an excitation wavelength of 488nm; scale bars represent 200 µm. This suggest that both full-length rat P2X7 and human P2X7 are being expressed in these stable T1 transgenics, however the generation of T3 lines are necessary to confirm this. Brightness and contrast were equally increased across images to aid viewing.

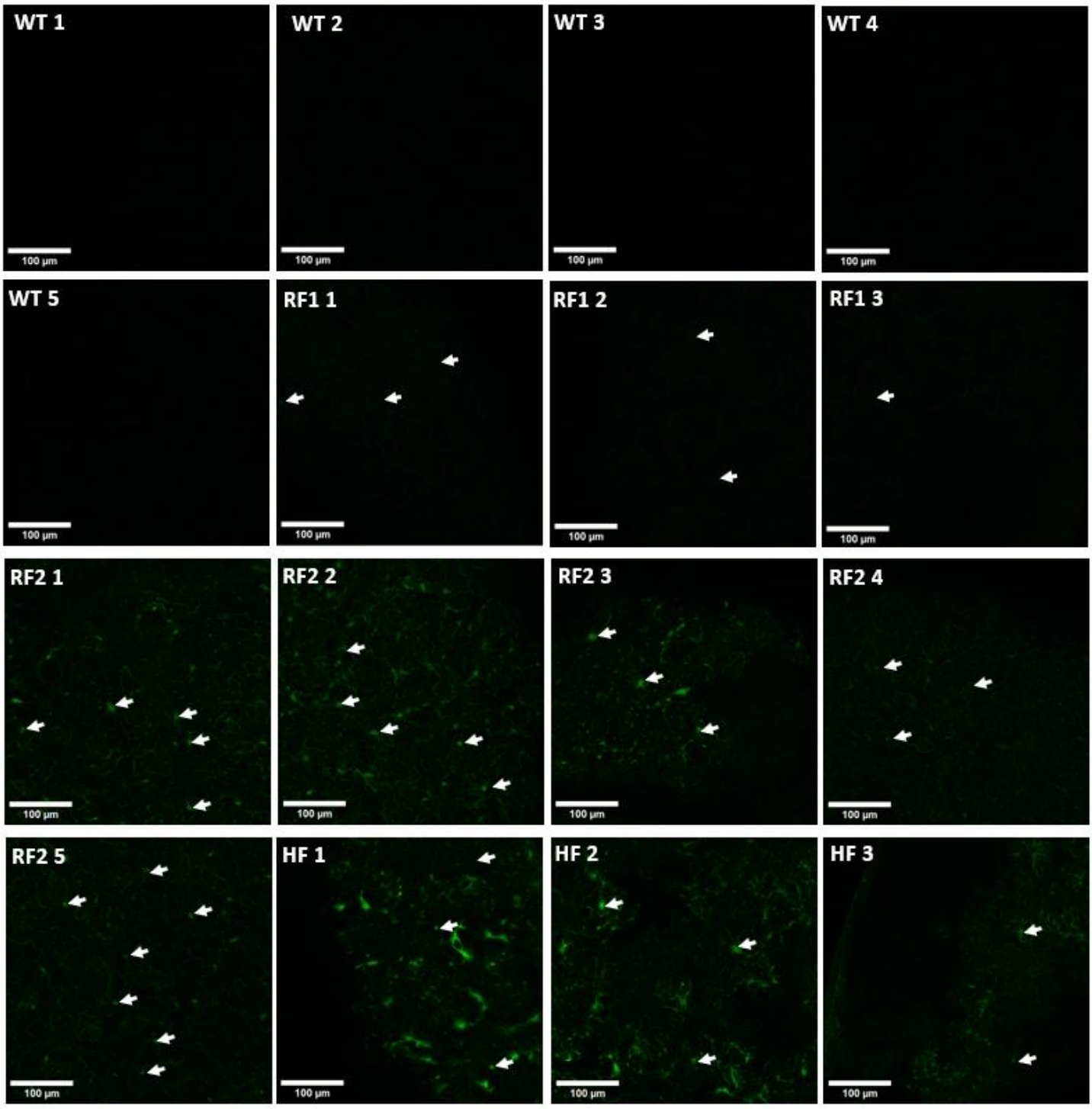


Figure 5.23: Zeiss LSM 880 confocal microscope images stably transformed T2 *A. thaliana* plants expressing rat and human full-length P2X7. Images show 30 µm Z-stacks of transformed areas using a gain of 750 and an excitation wavelength of 488nm; scale bars represent 100 µm. This suggest that both full-length rat P2X7 and human P2X7 are being expressed in these stable T2 transgenics. White arrows indicate the nuclei of fluorescent cells.

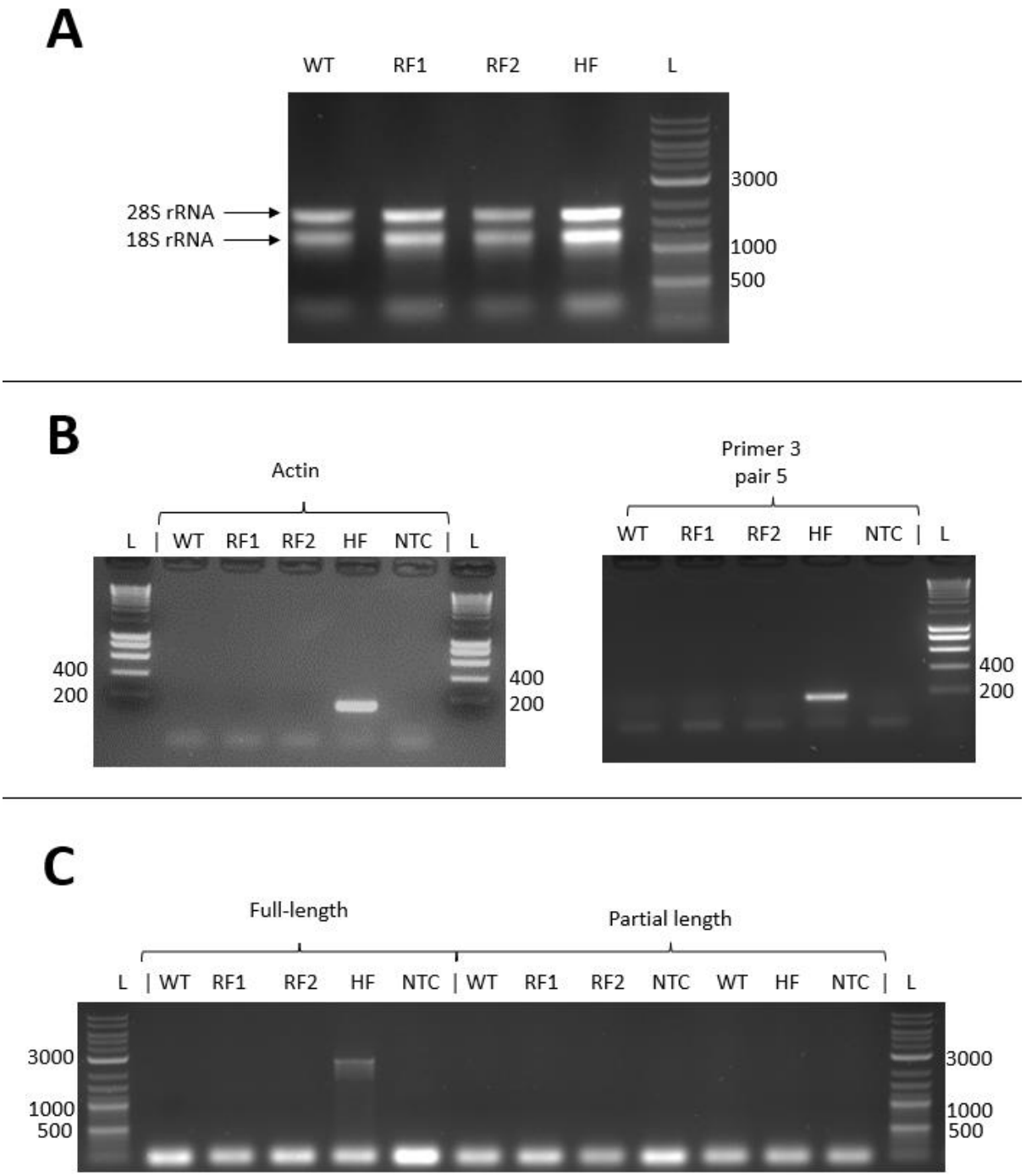


Figure 5.24: RT-PCR products from stably transformed T2 *A. thaliana* plants expressing rat and human full-length P2X7. A – RNA quality from DNase treated RNA extracts showed suitable quality RNA. **B** – Detection of Actin-2 and the C-terminal eGFP tag. The sample expressing human full-length P2X7 showed clear amplification of eGFP and Actin-2 though no other bands were seen suggesting that the RNA samples for the wild-type and rat P2X7 expressing tissue were poor. **C** – Detection of full-length transcripts and partial length transcripts at the P2X7 and eGFP junction. Only full-length human P2X7 could be detected, and no partial products were amplified suggesting that primers used were unsuitable. These data suggest that transcripts encoding human full-length P2X7 are produced. For Panels A and C the Eurogentec SmartLadder was used, and for the gels in panel B, NEB 1kB plus ladder was used.

Finally, rat full length P2X7 was expressed in the almost living cell free expression (ALiCE) system. Following transformation, faint eGFP fluorescence was detectable when excited using blue-light. However, when analysing the protein expression using a western blot, several non-specific bands were present but the band corresponding to either rat full-length P2X7 or cleaved eGFP was not present. Whilst expression using this system was attempted it was not optimised, and the data can be seen in the supplementary material (Figure S5.4). The ALiCE system costs approximately £100 for a 50 µL reaction, and whilst larger reaction volumes are available, these costs are vastly higher than expression in whole plants and lack the high level of upstream scalability that makes plant expression systems a favourable option, which is why it was not investigated further.

Overall, the transiently transformed *N. benthamiana* plants represented the best expression system for the different cytosolic proteins, able to produce high yields of target protein whilst being cheaply scalable, and less time consuming than using stable transgenic lines. Alternatively, the preliminary research with stably transformed *A. thaliana* plants suggests that these may be suitable to produce full-length P2X7 and other transmembrane proteins. Protein extraction and P2X7 purification from T3 *A. thaliana* plants would confirm this. In summary, the different expression systems explored here appeared to be able to produce different levels of each recombinant protein, so even within plant expression systems it is unlikely that there is a single chassis that is optimal for every protein. As a result, future work could assess more expression systems and identify the niche each occupies for different proteins; *N. tabacum* plants are particularly worth exploring as these have a much higher biomass than *N. benthamiana* plants, enabling unparalleled upstream scalability and increased transformed biomass, particularly when paired with large volume vacuum infiltration of plants, so would be the next obvious candidate for large scale protein expression.

5.3.3.1 Assessing full-length P2X7 function in plant cell packs

Throughout this chapter attempts have been made to express full-length P2X7 in plants. However, expression has been measured using the C-terminal eGFP tag and there is no guarantee that the P2X7 protein is correctly folded or trafficked within the plant cells. P2X7 is an ATP-gated transmembrane cation channel, and usually functions in an environment with a different pH that lacks a carbohydrate rich cell wall. A functional assay would be able to determine whether the P2X7 protein is correctly processed within plant cells, but it is difficult to design an assay to assess function within whole plants. Fortunately, P2X7 functional assays in mammalian cells are well established in the literature and can easily be recreated in our plant cell pack expression system. When activated by ATP, P2X7 opens a pore that enables influx of cations into the cell. This can be paired with a cationic molecule such as ethidium bromide or Yo-Pro™-1 iodide which each bind nucleic acids and fluoresce. This signal can be measured over time, and the uptake of the molecule calculated. If P2X7 is functional, one would expect cells transfected with P2X7 to produce a larger fluorescent signal than cells not transfected with P2X7. Consequently, PCPs expressing the full length P2X7 isoforms were activated with ATP and incubated with either ethidium bromide or Yo-Pro™-1 for 10 minutes, either with or without the presence of a reported inhibitor, JNJ-47965567, and the uptake of each molecule measured using a photon imager.

Initially ethidium bromide was used to measure activity, but this failed to produce stable reaction kinetics curves for each sample with highly variable data points, including negative datapoints when normalised suggesting that the experimental setup is unsuitable (Figure 5.25, Panel A). This could be due to differences in the number of cells within the PCPs or a lack of sensitivity as the photon imager was not able excite ethidium bromide at its main peak of 300 nm, and instead excitation was achieved at a secondary excitation wavelength of 500 nm, which may result in a less clear signal. Consequently, ethidium bromide was replaced with Yo-Pro™-1 in case there could be more specific uptake, and the molecule could be excited with the optimal wavelengths. This experimental setup more closely resembles the functional assays performed in mammalian expression systems. However, whilst the uptake of Yo-Pro™-1 was much clearer than ethidium bromide, a non-specific response was seen, with increased fluorescence over time in all samples, suggesting that the dye was being incorporated regardless of P2X7 transfection (Figure 5.25, Panel B). In addition, the presence of the P2X7 antagonist, JNJ-47965567, had no effect on the rate of uptake. Together these data suggest that P2X7 is non-functional within the PCPs or is not expressed at all.

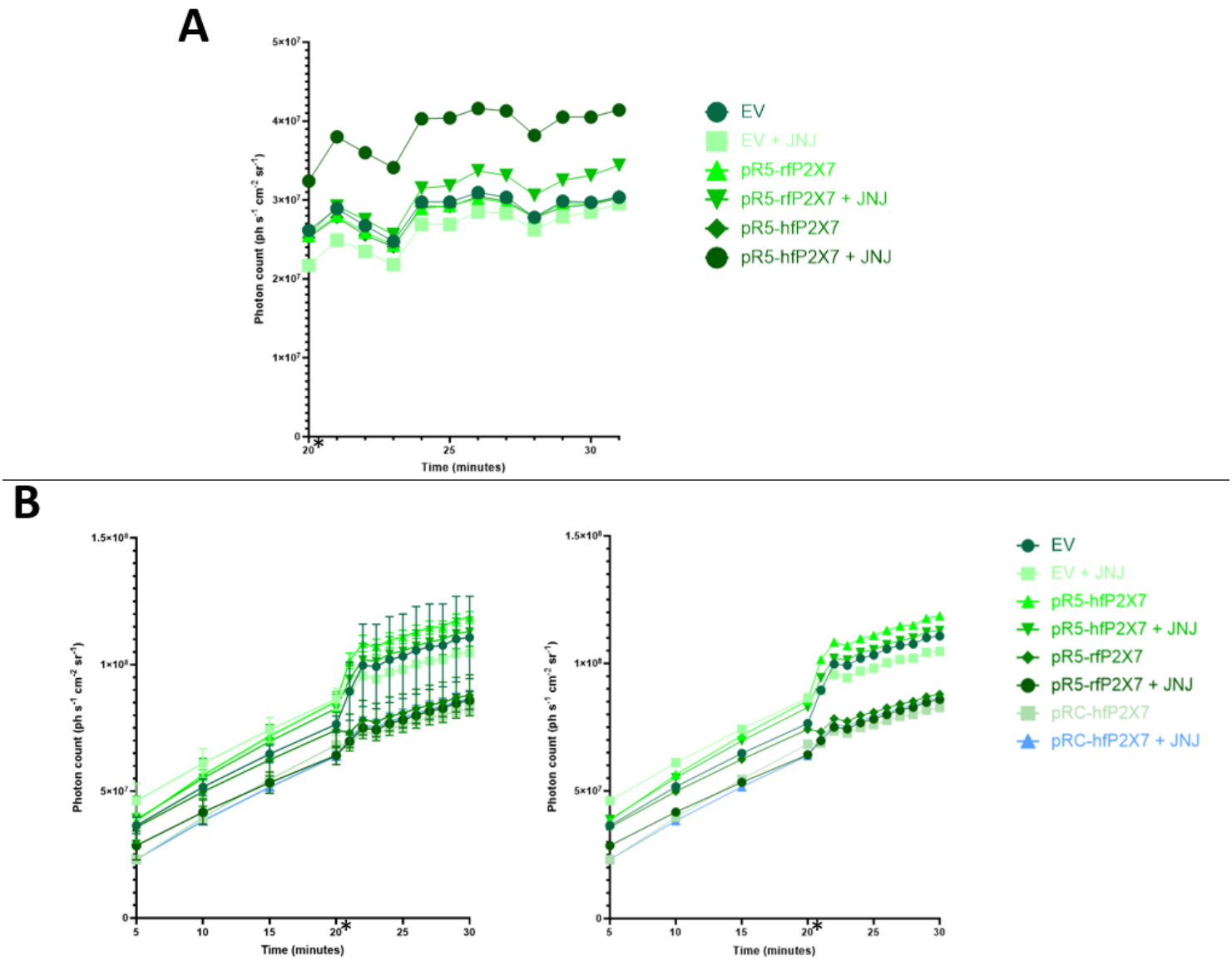


Figure 5.25: Functional assays performed on PCPs expressing P2X7 variants. A – pR5-P2X7 variants were transformed into PCPs which were incubated with ethidium bromide (A) or Yo-Pro-1 (B) with and without a P2X7 antagonist, JNJ-47965567. After a 20-minute preincubation with these compounds, ATP was added, denoted by * and fluorescence was measured every minute. As the cationic dye from each experiment migrates to the nucleus an increase in fluorescence should be seen. With ethidium bromide, an excitation wavelength of 520-540 nm was used, and emission detected at 590-610 nm, but the data was noisy and showed only a slight non-P2X7-specific increase. With Yo-Pro-1 the excitation wavelength was 450-470 nm and the emission wavelength was 500-520 nm. This showed a non-specific increase in fluorescence in all samples suggesting that P2X7 is either non-functional or not expressed in transformed cells. It also suggests that uptake is increased by ATP in a mechanism independent of P2X7. Both graphs in panel B show the same data, with and without error bars showing standard error of the mean. They key shows which constructs were transformed into PCPs, with and without the JNJ-47965567 P2X7 antagonist.

5.4 – Discussion

This chapter aimed to determine the limits of recombinant protein complexity when produced in plant expression systems. This section will discuss the overall findings and assess why some proteins could be produced in transient expression systems and not others. It will also discuss the limitations of the experiments performed and suggest avenues for future work to address these and provide more conclusive answers.

5.4.1 Protein complexity and yield

When proteins were expressed transiently in *N. benthamiana* it is clear that only cytosolic proteins can be expressed in detectable yields. Expectedly, these yields decrease as protein size increases and potentially with the addition of post-translational modifications, though further experiments to detect specific modifications are necessary to confirm this. The smallest unmodified protein, eGFP, was produced at the highest yields, followed by ballast P2X7 which lacks post-translational modifications, then RANGAP1 which is larger, cytosolic and post-translationally modified. These were the only full-length protein targets that could be detected using these methods. Even when paired with purification and protein concentration no other proteins were detectable using SDS-PAGE and western blotting.

Instead, several of the lanes show a 27 kDa band that corresponds to eGFP, rather than the full-length target. If eGFP is cleaved from the fused target proteins only the cleaved eGFP would be visible as the detection method utilised used anti-GFP primary antibodies. There are several potential reasons for seeing only the 27 kDa eGFP band. Firstly, the detection method used may be unsuitable for detecting some full-length target products. However, previous research in our laboratory has shown that the primary antibody used is suitable for detection of full-length rat P2X7 with a C-terminal eGFP fusion when produced in HEK-293 cells (Figure 5.10, Panel G; Jaradat 2021), so it is unlikely that this is the reason. Secondly, each the transcripts contain several additional start codons which could be used as translational start sites. However, this cause is also unlikely as there are several start codons upstream of the eGFP which could also serve as alternate translation start sites which would produce multiple visible bands larger than the 27 kDa band seen. Furthermore, the P2X7 fusion proteins do not contain a methionine at the start of the eGFP sequence, as it is mutated to a valine instead. This means that there is not a start codon at the beginning of the eGFP sequence within these samples, so the production of the 27 kDa band seen cannot be due to this reason. As the 27 kDa band is of identical size in each lane, it is unlikely that alternative translation start sites are the cause of this singular band in multiple samples. Instead, the most likely cause is the presence of a thrombin cleavage site immediately 5’ of the fused eGFP sequence. This was

incorporated into the constructs in case cleavage of the eGFP protein was necessary post-purification to obtain pure target protein suitable for further downstream analyses. However, several publications have shown that some plant species express proteases that exhibit thrombin-like protease activity (Urs et al. 2019; Urs et al. 2021). Whilst this has not been detected in *N. benthamiana* plants that are used here, protease activity at or near this site would produce the banding pattern seen in these experiments.

Interestingly, human P2X7 appears to have dramatically reduced yields relative to rat P2X7, despite very similar amino acid sequence identity. In fact, neither sequence was codon optimised for plant expression, with rat P2X7 containing two more rare codons than human P2X7 so it is unlikely the expression difference is due to codon usage. Instead, human P2X7 could have reduced stability or contain more post-translational modifications in the C-terminal domain; the UniProt database suggests that the latter is true as both sequences contain extensive glycosylation and disulfide bridging, but human P2X7 has an additional four residues that are phosphorylated, which may affect production yields though this would only be expected if the modification occurs during protein folding and trafficking. However, only the full-length structure for rat P2X7 has been published (McCarthy et al. 2019), and human P2X7 has substantially less structural information available, so whether these differences are true is uncertain.

As for the other target proteins, DORN1, full-length P2X7 and HMA4 are each larger plasma membrane proteins, which would dramatically reduce the yield due to the lower intracellular volume of plasma membranes relative to the cytoplasm, and the hydrophobic nature of these proteins (Puthenveetil et al. 2020). The biochemical properties of eukaryotic membrane proteins also present challenges when produced in prokaryotic expression systems resulting in protein misfolding and low yields (Puthenveetil et al. 2020); as a result, very few mammalian membrane proteins have been produced for structural studies in *E. coli* (Hattab et al. 2015) and similar problems may result in the low yields seen here. In addition, the detergent used can determine membrane protein extraction effectiveness, with DDM being reported as suitable for most proteins, with detergent concentrations at 1-2% v/v (Puthenveetil et al. 2020). Here, detergent concentrations in this range were used, but FC-12 was utilised in early extraction buffers, and triton X-100 in later ones. Nonetheless, DORN1 and HMA4 were chosen as targets because they are endogenous plant proteins, meaning plants can produce them correctly; this suggests that the absence of detectable expression is mostly likely due to low yields resulting in expression below western blot detection limit. Existing research has successfully extracted *A. thaliana* membrane protein complexes using three to six times the starting mass of tissue utilised here (Qi and Katagiri 2009); scaling up the experiments using membrane proteins would provide clarity on whether low yields are simply the

issue. Alternatively, the experimental setup may be unsuitable; *in planta* fluorescence from C-terminally eGFP tagged DORN1 and HMA4 may not detected due to protein misfolding, degradation, premature eGFP cleavage or improper fluorophore maturation resulting in the inability to detect the protein. However, transcripts were detected in plant tissue transformed with each of the pR5-CDS constructs that were not simply the eGFP transcript, suggesting that the target RNA molecules were made. This indicates that the lack of expression seen at the protein level is likely due to issues at the protein level, including protein misfolding, instability and subsequent degradation, or premature eGFP cleavage.

The final target protein was factor C, a modified cytosolic protein derived from the horseshoe crab, *Tachypleus tridentatus*. This protein is very large at 1019 amino acids long before the addition of the eGFP tag which may cause difficulties in protein production. In addition, the protein is naturally secreted from the cells, heavily glycosylated, and contains lipopolysaccharide-induced protein cleavage sites. These additional factors may contribute to the difficulty in producing the protein, as the secretory peptide signal may behave differently when expressed in plant systems, plant protein glycosylation is often different to animal protein glycosylation (Strasser 2016), and premature protein cleavage could result in either misfolded protein or the inability to detect extracted protein.

Finally, when these proteins were produced in BY-2 cell derived plant cell packs, the only difference seen was that pR5-HMA4 eGFP-fluorescence was significantly higher than the negative control. This is suggestive that PCPs are better suited for the production of membrane proteins, but their limited scalability hinders the usefulness as an expression system, providing no real benefits over mammalian cell cultures. A major disadvantage seen here was the high level of background fluorescence seen in the PCPs, resulting in pR5-eGFP expression that was not significantly higher than the negative control; as a result, utilising a different fusion protein such as far-red fluorescent proteins could dramatically improve experimental sensitivity. Nonetheless, PCPs may be worth investigating further and are a useful expression system for rapid screening of constructs.

5.4.2 P2X7 in other expression systems

Whilst it was clear that full-length P2X7 was not produced in high yields in transiently transformed *Nicotiana*, limited success was seen in other expression systems. When analysed using confocal microscopy, full-length rat P2X7 expression could be seen in *A. thaliana* protoplasts and later in early stage transgenics; interestingly, full-length human P2X7 was not seen despite several transformation attempts, further suggesting that the human variant of P2X7 is more complex and challenging to produce. Interestingly, when expressed in *A. thaliana* protoplasts, the truncated human P2X7 variants produced more fluorescence than the truncated rat P2X7 variants; this would need further

validation through experimental repeats and measuring more protoplasts. Additionally, the same trend was seen with rat ballast P2X7 producing the highest fluorescence, followed by cysteine-rich then full-length P2X7 as anticipated. Importantly, more experimental repeats are necessary to make valid conclusions as the low sample size used here means it is impossible to perform statistical analyses on the data. Additionally, full-length rat P2X7 expression was attempted in the ALiCE production system with limited success. No extracted P2X7 protein was detectable by western blot, with several non-specific products seen. This may be due to the extraction and detection methods, so alternative methods may be worth investigating. However, like PCPs, the lack of scalability and high cost of the ALiCE expression system results in few advantages over mammalian cell cultures.

Finally, functional assays were performed on PCPs transformed with full-length P2X7. There was no detectable fluorescence in the transformed PCPs, suggesting that full-length P2X7 may not be produced at all. However, in case this was not true, P2X7 receptor activation was achieved through the addition of either ethidium bromide or Yo-Pro-1 with or without a P2X7 antagonist, JNJ-47965567, to transformed cells, followed by the addition of BzATP. However, a non-specific increase in fluorescence was seen in all samples, with no difference between empty vector controls, or transformed cells with or without the P2X7 antagonist. These experiments suggest that either P2X7 is not expressed at all in PCPs, or that it is non-functional. Stable *A. thaliana* transgenic plants appear to express full-length rat and human P2X7, and may represent an important avenue for future functional assays. The full-length human P2X7 transcript was detected in stably transformed T2 *A. thaliana* plants, though RT-PCRs and RNA work performed in this research was inconsistent, with some references genes able to be detected in some experiments and not others.

5.4.3 Limitations and future work

Overall, the research presented in this chapter has revealed several avenues for future work. Firstly, the different expression levels seen between chassis suggests that there may not be a single optimal plant expression system for every protein target. As a result, more protein targets should be investigated to identify the niche that each expression system occupies. Secondly, construct expression should be screened in PCPs using a different fluorescent protein, such as mCherry or mPlum, as these would be more suitable with reduced background autofluorescence, that would enable heightened sensitivity to provide more conclusive data of the relationship between protein complexity and yield. Thirdly, scaled up experiments should be performed on transiently transformed *Nicotiana* plants expressing the membrane proteins to determine if low yields are the issue. This could either be in the form of more *N. benthamiana* plants, or utilising larger *N. tabacum* plants to provide more material. Ideally, this would be paired with an optimised vacuum infiltration method to minimise transformation variability and would require a different extraction method such

as the use of a food processor for tissue extraction as mechanical grinding would be unfeasible at very large scales. Finally, the stable transgenic plants expressing full-length human and rat P2X7 should be self-fertilised and propagated into T3 lines. Using these T3 plants, functional experiments could be performed using similar reagents to the PCP assays discussed above. This could involve incubating sections of leaves with BzATP, Yo-Pro-1 and an antagonist to determine whether the P2X7 protein is functional. Alternatively, in case the cell wall hinders function, protoplasts could be made from leaves expressing P2X7 and the functional assays performed on these. Both experiments would provide useful insights into the structure and folding of P2X7 produced in plants.

5.5 – Conclusions

Overall, the data presented here suggests that there is an inverse relationship between recombinant protein complexity and yield, with larger proteins containing transmembrane domains unable to be detected, and only cytosolic proteins detected in transient plant expression systems at high yields. Further work is necessary to produce transmembrane proteins in this system, and to directly measure the presence and impact of post-translational modifications in successfully produced targets. These data also suggest that the production of membrane proteins may be more suitable in stably transformed plants, in which low levels of *in planta* expression was detected. It is likely that further developments in construct design, transformation procedures, and protein extraction methods could improve the versatility of the plant recombinant protein expression systems discussed here. Whilst the data in this chapter suggests that capacity of plants to produce modified membranous recombinant proteins is limited, the areas of future work discussed above would provide much more conclusive evidence on what these limits are, and how they differ between plant expression systems.

Chapter 6 – Discussion, Conclusions and Future Work

6.1 – Overview

The main aim of this research was to generate a plant expression system for recombinant proteins, ideally capable of producing post-translationally modified membrane proteins. This would be achieved through a series of objectives. Firstly, methods to enhance expression of target proteins would be explored to maximise *in planta* protein expression, measured using an eGFP reporter. Secondly, following the development of an effective downstream processing pipeline, the protein would be extracted, purified, and measured. Thirdly, with improved upstream and downstream methods established, the expression of target proteins would be measured. Originally, the main target protein was full-length P2X7 protein, but experiments suggested that it could only be produced at low levels. As a result, this was expanded to a range of proteins with a secondary aim of determining the maximum protein complexity; other protein targets were chosen for their varying sizes, intracellular localisation, and endogenous post-translational modifications.

The main strategy employed during this research was to maximise protein expression by addressing low yields and difficult downstream processing, two major limitations of plant expression systems that are widely reported (Schillberg et al. 2019). To improve protein yields, several factors were experimented with to varying degrees. Primarily, construct design was experimented with by combining previously published DNA regulatory elements that have each been shown to improve protein expression in *Nicotiana* hosts (Peyret et al. 2019; Diamos and Mason 2018). The selection of these DNA elements was by no means exhaustive, with many other possible regulatory elements available for selection with potential synthetic modifications possible that remain unexplored. Other factors that were tested to some degree include co-expression of silencing suppressors, improving transformation efficiency, and exploring various host species, each successfully employed in the literature (Peyret et al. 2019; Norkunas et al. 2018; Gengenbach et al. 2020). However, several other factors were not rigorously experimented with and include codon optimisation, subcellular targeting, suppression of *in planta* protease activity, co-expression of folding chaperones, and optimising harvest time. To improve the downstream processing process, several previously published protein extraction protocols were tested, including different protein extraction buffers, and different fractionation techniques. The final method developed combined aspects of several methods, with the creation of an extraction buffer that was suitable for downstream purification. Protein purification methods were also experimented with to a lesser degree, with the testing of several previously published purification tags (Qi and Katagiri 2011; Islam et al. 2018) and methods that showed varying degrees of success. However, there are many different approaches available for

downstream processing, particularly with purification methods, that were not explored. These include other chromatographic techniques to separate proteins by size, charge, or hydrophobicity. Alternatively, untested non-chromatographic methods include flocculation, tangential flow filtration, and precipitation of proteins.

Using this strategy, this research produced several key findings. Chapter 3 described genetic engineering efforts to enhance *in planta* protein expression. This involved the creation of several genetic constructs that contained various regulatory DNA elements that enhance expression by increasing transcriptional and translational efficiency. Chapter 4 explored several downstream processing methods to extract and purify target proteins. Following experimentation with several extraction and purification procedures, a method was developed that enabled successful extraction of target proteins and purification using an Octa-His tag. In Chapter 5, the expression of several other target proteins was investigated, with only cytosolic proteins being detected in substantial yields both *in planta* and at the extracted protein level. Furthermore, several other plant expression systems were explored that were separate from the main pipeline that revealed unexpected differences between expression chassis and opened avenues for future work. Overall, this research failed to produce detectable levels of membrane proteins, including full-length P2X7. However, the research determined the limit of protein complexity possible for detection at these scales. It is possible that increasing the scale of experiments may enable detection of these proteins, but this needs to be determined.

6.2 – Implications and Applications

The best expressing construct created here, pRC, combined several published DNA elements that have each been shown to improve gene expression, including a promoter, UTRs, terminators, and a matrix attach region. However, this expression construct was combined with the replicating vector, pJL-TRBO, derived from the TMV, an RNA virus. Normally, promoters, terminators and MARs do not function at the RNA-level so several other constructs were created that expressed eGFP with only the UTRs, with one construct, pR5, producing the highest expression. Across whole transformed leaves, pRC-eGFP produced significantly higher fluorescence than pR5-eGFP, but the opposite trend was seen when per cell expression was analysed. In addition, the expression pattern of the two was different, with pR5 primarily transforming the leaf vasculature, and pRC transforming the rest of the leaf. This was unexpected, as the genetic elements employed are not predicted to affect transformation efficiency. Together, these data strongly suggest that the additional promoter, terminators, and MAR within pRC have additional effects even when combined with the RNA virus; this is surprising as these DNA elements are each thought to work at the DNA-level by influencing

transcription rather than at the RNA level. As a result, more research is necessary to determine if and how these parts function in this context. Importantly, the pRC-eGFP and pR5-eGFP expression constructs developed here outperformed pRI-eGFP, equivalent to pJL-TRBO but with a different GFP sequence. pJL-TRBO has been reported to produce GFP expression at 25% TSP (Lindbo 2007b), a finding that could not be replicated in this work. Nonetheless, these data suggest that pRC and pR5 could be useful tools for enhancing protein expression, particularly if the regulatory elements can be re-ordered to synergise better with the replicating vector. Alternatively, these findings may be sufficient evidence to further investigate the use of exogenous DNA elements within other replicating vectors, potentially further enhancing transgene expression when using other deconstructed viruses.

The major finding of this research is that the transient expression system utilised here was not capable of producing high yields of membrane proteins. To definitively confirm this, scaled up experiments using higher masses of transformed plant tissue would be necessary. However, to my knowledge there is only a single example of a recombinant membrane protein produced in plants, which utilised stably transformed *A. thaliana* plants and extracted protein from 30 g of leaf tissue (Qi and Katagiri 2009). Generally, transient protein expression systems produce higher yields than stable transformants, but this research suggests that transient expression systems may only be viable for low complexity proteins, such as modified cytosolic proteins. This may mean that plants can still occupy their own niche, as prokaryotes are unable to produce the modifications and eukaryotic cell cultures are still costly to scale for these proteins. For membrane proteins, it is possible that stably transformed plants are required to obtain high yields. This is not because per cell yields are higher, but because more uniform expression was seen in the stably transformed T2 *A. thaliana* plants expressing full-length rat and human P2X7 that was not seen when using similar expression constructs in transiently transformed *N. benthamiana* plants. Alternatively, these results may be dependent on the selected proteins used in this research as none have been expressed in the literature to my knowledge. However, various proteins have been expressed in stable transgenic plants in the wider literature, usually to provide advantageous traits to plants. These include the expression of pathogenesis-related proteins (Santos and Franco 2023), insecticidal and herbicidal proteins (Gupta et al. 2021; Dong et al. 2021), nutritional proteins (Le et al. 2016), stress resistance proteins (Parmar et al. 2017), and proteins that improve shelf life (Aziz et al. 2022). These expressed proteins vary in size and complexity, but no study has directly compared the production capabilities of stable and transgenic plants expressing the same proteins in each. A major reason for this is that stable transgenics are often produced to provide beneficial traits to plants that are the product, without the need for protein extraction and purification. This contrasts with transient transgenics,

which are more often used as a production host to make a product, with plant tissue harvested and target proteins extracted. Consequently, further work with scaled up transient expression systems and comparative research expressing different proteins in both stable and transient transgenic plants is necessary to confirm these findings.

6.3 – Limitations and Future Work

As mentioned in previous sections, there are several limitations with this research that should be addressed with future work. These include developments to *in planta* gene expression and downstream protein purification, as well as expanding the diversity of the research.

In terms of the upstream pipeline, whilst the additional regulatory elements in pRC improved gene expression, it is likely that changing the order of these regulatory elements may improve expression further. Specifically, TMV is thought to amplify RNA between the TMV 5' and 3' UTRs, and the pRC construct contains several regulatory elements between these UTRs in the pJL-TRBO vector. As a result, reordering these DNA elements may further improve function by putting the promoter elements 5' of the 5' TMV UTR, and the terminator and MAR elements 3' of the 3' TMV UTR. This would minimise the replicon length which would only contain UTRs and the CDS within the replicated RNA, ensuring that the other DNA elements are only present in the DNA to influence transcription as they are thought to do. A proposed expression construct is shown in Figure 6.1, however other regulatory sequences could be investigated, including the use of different promoters, terminators, UTRs, MARs, replicating vectors, and unexplored elements such as introns. Furthermore, it is widely reported that co-expression of other proteins can improve gene expression, but the opposite finding was seen here; co-expression of silencing suppressors has been shown to substantially improve gene expression (Peyret et al. 2019; Mardanova et al. 2017; Daròs 2017), so it is likely that further expression enhancement could be obtained if the co-expression was optimised. This could include testing other silencing suppressors such as p24, and lowering the bacterial concentrations used to infiltrate plants in case this is the detrimental factor that causes plant stress. Finally, vacuum infiltration is usually favourable as it ensures even agroinfiltration across leaves, but the method utilised here only showed patchy effectiveness which produced lower expression levels than syringe-mediated agroinfiltration. At industrial scales, it is not feasible to hand infiltrate hundreds of large plants, so vacuum infiltration is instead used to transiently transform plants where it has the additional benefit of transforming shoot tissue, which cannot be infiltrated (Chen et al. 2013). As a result, optimising the vacuum infiltration method should make it more favourable than hand-infiltration, enabling large scale transformations with reduced variability and increased areas of transformation.

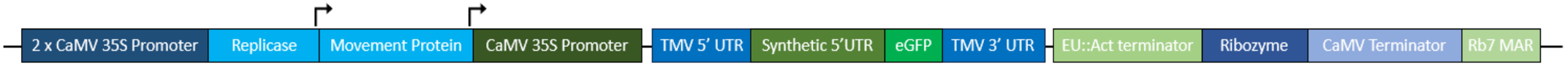


Figure 6.1: Proposed expression construct to further increase recombinant protein expression. This expression construct only includes UTRs and a coding sequence (here eGFP) between the two TMV UTRs, resulting in a replicating RNA that contains only DNA elements with functions at the RNA level. Instead, the additional regulatory elements from pRC are excluded from the replicating RNA and only function at the DNA level to improve transcriptional efficiency. This is an example construct utilising DNA elements from the novel expression cassettes developed here (green) and elements from the replicating pJL-TRBO vector (blue), though other DNA elements could also be explored.

The downstream processing also has substantial room for improvement, particularly through the testing of different methods. Firstly, whilst protein could be successfully extracted and purified using the final methods developed here, it is likely that further fine tuning could improve this. This may include increasing the concentration of DTT in the extraction buffer to reduce smearing and experimenting with different detergents and concentrations to maximise membrane solubility and protein stability. In addition, scaled up experiments using larger volumes of tissue would require automation. This could involve the use of food processors and screw presses to grind tissue whilst suspended in extraction buffer (Colas et al. 2013; Hansen et al. 2022; Yang et al. 2022). Furthermore, the histidine tag purification method utilised here could not be automated using by FPLC, and whilst the manual protein purification method can be performed at scale it would likely result in higher variability than FPLC automation due to human error. Lastly, the cost of downstream processing could be dramatically reduced. Protease inhibitors represent the major cost in extraction buffers, and cheaper alternatives should be sought to minimise costs. In addition, the cheapest purification method would utilise the cellulose binding domain tag discussed in Chapter 4, as this uses the cheapest reagents. Further investigation of this tag could reduce purification costs dramatically, particularly as the microcrystalline cellulose resin can be loaded onto a column for automated FPLC purifications at scale. Developing a purification tag that effectively enables target proteins to bind to microcrystalline cellulose without compromising the *in planta* protein expression and/or folding seen here could minimise costs.

Finally, broader limitations include the use of the selected target proteins and plant species utilised here. These proteins were chosen to encompass a wide range of complexities to identify the maximum yields obtainable for each class of protein. However, whilst this was somewhat successful, more protein candidates would be necessary to draw robust conclusions. This would include expressing as many proteins as possible with differences in size and intracellular localisation, and performing additional experiments to directly measure the presence of specific post-translational modifications. This research also showed differences in expression levels of different proteins depending on the plant host and transformation method used. Some other plant species have shown success at producing recombinant proteins, including mosses and aquatic plants (Reski et al. 2015; Yang et al. 2021), both of which are worth investigating further. To pair with this, a wide range of protein reporters would be beneficial, as the experiments within PCPs showed high levels of background green autofluorescence, reducing the sensitivity of experiments when eGFP was used as a reporter. Consequently, having a range of reporters including blue, yellow or far-red fluorescent proteins may be useful. Finally, due to time constraints developing the stable *A. thaliana* transgenic lines, only T2 transformants could be obtained expressing full length rat and human P2X7.

Generating T3 lines would enable functional assays to be performed on whole plants or protoplasts that would assess whether the mammalian protein is functional with and without the cell wall, respectively. Extracting the proteins from these plants would also enable comparisons between transient and stable transformants to determine whether each occupies their own niche for producing recombinant protein targets.

6.4 – Concluding Remarks

Overall, this research aimed to develop a plant-based expression system to produce post-translationally modified membrane proteins. Whilst there was some evidence for *in planta* expression of proteins of this nature, extracted protein could not be detected within this timeframe, and the presence of post-translational modifications was not assessed. However, this research successfully developed several promising expression constructs for use in stably and transiently transformed plants, shed light on the discrepancies in protein production between different plant systems, and determined the maximum level of protein complexity possible for detection at this scale. Future work to improve these expression constructs and downstream processing could further enhance expression levels, and scaled up experiments using a wider array of target proteins with additional experiments to detect specific post-translational modifications would strengthen these findings. This research has also suggested avenues for streamlining processes used, including automation of transformation, protein extraction and purification procedures to improve effectiveness and minimise costs. These suggested improvements may result in the detection of the lower abundance membrane proteins that were not detected here; otherwise, plant recombinant protein expression systems may still occupy their own niche of producing post-translationally modified cytosolic proteins that prokaryotic systems cannot produce, whilst being cheaper and more scalable than conventional eukaryotic expression systems.

References

Abd-Aziz, N., Tan, B.C., Rejab, N.A., Othman, R.Y. and Khalid, N. 2020. A new plant expression system for producing pharmaceutical proteins. *Molecular Biotechnology* 62(4), pp. 240–251. Available at: https://doi.org/10.1007/s12033-020-00242-2.

Abrahamian, P., Hammond, R.W. and Hammond, J. 2020. Plant Virus–Derived Vectors: Applications in agricultural and medical biotechnology. *Annual Review of Virology* 7(1), pp. 513–535. Available at: https://doi.org/10.1146/annurev-virology-010720-054958.

Ahmad, N., Mehmood, M.A. and Malik, S. 2020. Recombinant protein production in microalgae: emerging trends. *Protein and Peptide Letters* 27(2), pp. 105–110. Available at: https://doi.org/10.2174/0929866526666191014124855.

Ali, S. and Kim, W. 2019. A fruitful decade using synthetic promoters in the improvement of transgenic plants. *Frontiers in Plant Science* 10. Available at: https://doi.org/10.3389/fpls.2019.01433.

Allen, G.C. 2008. The role of nuclear matrix attachment regions in plants. In: *Plant cell monographs*. Available at: https://doi.org/10.1007/7089_2008_21.

Almén, M.S., Nordström, K., Fredriksson, R. and Schiöth, H.B. 2009. Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biology* 7(1). Available at: https://doi.org/10.1186/1741-7007-7-50.

An, Y., Wang, Y., Tan, S.N., Yusof, M.L.M., Ghosh, S. and Chen, Z. 2020. Phytoremediation: a promising approach for revegetation of Heavy Metal-Polluted land. Frontiers in Plant Science 11. Available at: https://doi.org/10.3389/fpls.2020.00359.

Avesani, L. et al. 2007. Stability of Potato virus X expression vectors is related to insert size: implications for replication models and risk assessment. *Transgenic Research* 16(5), pp. 587–597. Available at: https://doi.org/10.1007/s11248-006-9051-1.

Avesani, L., Merlin, M., Gecchele, E., Capaldi, S., Brozzetti, A., Falorni, A. and Pezzotti, M. 2013. Comparative analysis of different biofactories for the production of a major diabetes autoantigen. *Transgenic Research* 23(2), pp. 281–291. Available at: https://doi.org/10.1007/s11248-013-9749-9.

Aziz, M.A., Brini, F., Rouached, H. and Masmoudi, K. 2022. Genetically engineered crops for sustainably enhanced food production systems. *Frontiers in Plant Science* 13. Available at: https://doi.org/10.3389/fpls.2022.1027828.

Bagwan, N., Ali, H.H.E. and Lundby, A. 2021. Proteome-wide profiling and mapping of post translational modifications in human hearts. *Scientific Reports* 11(1). Available at: https://doi.org/10.1038/s41598-021-81986-y.

Baur, A., Reski, R. and Gorr, G. 2005. Enhanced recovery of a secreted recombinant human growth factor using stabilizing additives and by co-expression of human serum albumin in the moss Physcomitrella patens. *Plant Biotechnology Journal* 3(3), pp. 331–340. Available at: https://doi.org/10.1111/j.1467-7652.2005.00127.x.

Beringer, J. et al. 2017. Comparison of the impact of viral and plant-derived promoters regulating selectable marker gene on maize transformation and transgene expression. *Plant Cell Reports* 36(4), pp. 519–528. Available at: https://doi.org/10.1007/s00299-017-2099-y.

Bhattacharjee, B. and Hallan, V. 2022. Geminivirus-Derived Vectors as tools for functional genomics. *Frontiers in Microbiology* 13. Available at: https://doi.org/10.3389/fmicb.2022.799345.

Bidula, S., Dhuna, K., Helliwell, R. and Stokes, L. 2019. Positive allosteric modulation of P2X7 promotes apoptotic cell death over lytic cell death responses in macrophages. *Cell Death and Disease* 10(12). Available at: https://doi.org/10.1038/s41419-019-2110-3.

Bourdon, V., Harvey, A. and Lonsdale, D. 2001. Introns and their positions affect the translational activity of mRNA in plant cells. *EMBO Reports* 2(5), pp. 394–398. Available at: https://doi.org/10.1093/embo-reports/kve090.

Burnett, M. and Burnett, A.C. 2019. Therapeutic recombinant protein production in plants: Challenges and opportunities. *Plants, People, Planet* 2(2), pp. 121–132. Available at: https://doi.org/10.1002/ppp3.10073.

Burnstock, G., Nistri, A., Khakh, B.S. and Giniatullin, R. 2014. ATP-gated P2X receptors in health and disease. *Frontiers in Cellular Neuroscience* 8. Available at: https://doi.org/10.3389/fncel.2014.00204.

Buyel, J.F. and Fischer, R. 2013. Flocculation increases the efficacy of depth filtration during the downstream processing of recombinant pharmaceutical proteins produced in tobacco. *Plant Biotechnology Journal* 12(2), pp. 240–252. Available at: https://doi.org/10.1111/pbi.12132.

Buyel, J.F. and Fischer, R. 2014. Downstream processing of biopharmaceutical proteins produced in plants. *Bioengineered* 5(2), pp. 138–142. Available at: https://doi.org/10.4161/bioe.28061.

Buyel, J.F., Hubbuch, J. and Fischer, R. 2016. Comparison of tobacco host cell protein removal methods by blanching intact plants or by heat treatment of extracts. *Journal of Visualized Experiments* (114). Available at: https://doi.org/10.3791/54343.

Buyel, J.F., Stöger, E. and Bortesi, L. 2021. Targeted genome editing of plants and plant cells for biomanufacturing. *Transgenic Research* 30(4), pp. 401–426. Available at: https://doi.org/10.1007/s11248-021-00236-z.

Cao, J. et al. 2021. High-throughput 5’ UTR engineering for enhanced protein production in non-viral gene therapies. *Nature Communications* 12(1). Available at: https://doi.org/10.1038/s41467-021-24436-7.

Carpenter, E.P., Beis, K., Cameron, A.D. and Iwata, S. 2008. Overcoming the challenges of membrane protein crystallography. *Current Opinion in Structural Biology* 18(5), pp. 581–586. Available at: https://doi.org/10.1016/j.sbi.2008.07.001.

Castilho, A. and Steinkellner, H. 2012. Glyco-engineering in plants to produce human-like N-glycan structures. *Biotechnology Journal* 7(9), pp. 1088–1098. Available at: https://doi.org/10.1002/biot.201200032.

Chanroj, S., Jaiprasert, A. and Issaro, N. 2021. A novel technique for recombinant protein expression in duckweed (*Spirodela polyrhiza*) turions. *Journal of Plant Biotechnology* 48(3), pp. 156–164. Available at: https://doi.org/10.5010/jpb.2021.48.3.156.

Chen, D., Cao, Y., Li, H., Kim, D.W., Ahsan, N., Thelen, J.J. and Stacey, G. 2017. Extracellular ATP elicits DORN1-mediated RBOHD phosphorylation to regulate stomatal aperture. Nature Communications 8(1). Available at: https://doi.org/10.1038/s41467-017-02340-3.

Chen, Q. and Davis, K. 2016. The potential of plants as a system for the development and production of human biologics. *F1000Research* 5, p. 912. Available at: https://doi.org/10.12688/f1000research.8010.1.

Chen, Q. and Lai, H. 2015. Gene Delivery into Plant Cells for Recombinant Protein Production. *BioMed Research International* 2015, pp. 1–10. Available at: https://doi.org/10.1155/2015/932161.

Chuang, L., Enders, A., Offermann, S., Bahnemann, J. and Franke, J. 2022. 3D-printed autoclavable plant holders to facilitate large-scale protein production in plants. Engineering in Life Sciences 22(12), pp. 803–810. Available at: https://doi.org/10.1002/elsc.202200001.

Clemente, M., Corigliano, M.G., Pariani, S., Sánchez-López, E. and Sander, V.A. 2019. Plant serine protease inhibitors: Biotechnology application in agriculture and molecular farming. *International Journal of Molecular Sciences* 20(6), p. 1345. Available at: https://doi.org/10.3390/ijms20061345.

Coates, R.J., Young, M.T. and Scofield, S. 2022. Optimising expression and extraction of recombinant proteins in plants. *Frontiers in Plant Science* 13. Available at: https://doi.org/10.3389/fpls.2022.1074531.

Cohen, S.N., Chang, A.C.Y., Boyer, H.W. and Helling, R.B. 1973. Construction of Biologically Functional Bacterial Plasmids *In Vitro*. *Proceedings of the National Academy of Sciences of the United States of America* 70(11), pp. 3240–3244. Available at: https://doi.org/10.1073/pnas.70.11.3240.

Colas, D., Doumeng, C., Pontalier, P.-Y. and Rigal, L. 2013. Twin-screw extrusion technology, an original solution for the extraction of proteins from alfalfa (Medicago sativa). *Food and Bioproducts Processing* 91(2), pp. 175–182. Available at: https://doi.org/10.1016/j.fbp.2013.01.002.

Daniell, H., Ruiz, G., Dénes, B., Sandberg, L. and Langridge, W.H.R. 2009. Optimization of codon composition and regulatory elements for expression of human insulin like growth factor-1 in transgenic chloroplasts and evaluation of structural identity and function. *BMC Biotechnology* 9(1), p. 33. Available at: https://doi.org/10.1186/1472-6750-9-33.

Daròs, J. 2017. Viral suppressors: Combatting RNA silencing. *Nature Plants* 3(7). Available at: https://doi.org/10.1038/nplants.2017.98.

Day, C., Lee, E., Kobayashi, J., Holappa, L.D., Albert, H.H. and Ow, D.W. 2000. Transgene integration into the same chromosome location can produce alleles that express at a predictable level, or alleles that are differentially silenced. *Genes & Development* 14(22), pp. 2869–2880. Available at: https://doi.org/10.1101/gad.849600.

De Martinis, D., Hitzeroth, I.I., Matsuda, R., Pérez, N.I. and Benvenuto, E. 2022. Editorial: Engineering the Plant Biofactory for the production of Biologics and Small-Molecule Medicines—Volume 2. *Frontiers in Plant Science* 13. Available at: https://doi.org/10.3389/fpls.2022.942746.

Debler, J.W., Henares, B.M. and Lee, R.C. 2021. Agroinfiltration for transient gene expression and characterisation of fungal pathogen effectors in cool-season grain legume hosts. *Plant Cell Reports* 40(5), pp. 805–818. Available at: https://doi.org/10.1007/s00299-021-02671-y.

Decker, E.L. and Reski, R. 2007. Moss bioreactors producing improved biopharmaceuticals. *Current Opinion in Biotechnology* 18(5), pp. 393–398. Available at: https://doi.org/10.1016/j.copbio.2007.07.012.

Dellinger, A. and Kepley, C.L. 2018. The role of horseshoe crabs in the biomedical industry and recent trends impacting species sustainability. Frontiers in Marine Science 5. Available at: https://doi.org/10.3389/fmars.2018.00185.

Diamos, A.G. and Mason, H.S. 2018. Chimeric 3’ flanking regions strongly enhance gene expression in plants. *Plant Biotechnology Journal* 16(12), pp. 1971–1982. Available at: https://doi.org/10.1111/pbi.12931.

Diamos, A.G., Rosenthal, S.H. and Mason, H.S. 2016. 5’ and 3’ Untranslated Regions Strongly Enhance Performance of Geminiviral Replicons in Nicotiana benthamiana Leaves. *Frontiers in Plant Science* 7. Available at: https://doi.org/10.3389/fpls.2016.00200.

Ding, X. et al. 2022. Different Fruit-Specific promoters drive ATMYB12 expression to improve phenylpropanoid accumulation in tomato. *Molecules* 27(1), p. 317. Available at: https://doi.org/10.3390/molecules27010317.

Dobson, L., Reményi, I. and Tusnády, G. 2015. The human transmembrane proteome. *Biology Direct* 10(1). Available at: https://doi.org/10.1186/s13062-015-0061-x.

Dolgova, A. and Dolgov, S. 2019. Matrix attachment regions as a tool to influence plant transgene expression. *3 Biotech* 9(5). Available at: https://doi.org/10.1007/s13205-019-1709-5.

Dong, H., Huang, Y. and Wang, K. 2021. The development of herbicide resistance crop plants using CRISPR/CAS9-Mediated Gene Editing. Genes 12(6), p. 912. Available at: https://doi.org/10.3390/genes12060912.

Dong, J., Ding, X. and Wang, S. 2019. Purification of the recombinant green fluorescent protein from tobacco plants using alcohol/salt aqueous two-phase system and hydrophobic interaction chromatography. *BMC Biotechnology* 19(1). Available at: https://doi.org/10.1186/s12896-019-0590-y.

Donnelly-Roberts, D.L., Namovic, M.T., Han, P. and Jarvis, M.F. 2009. Mammalian P2X7 receptor pharmacology: comparison of recombinant mouse, rat and human P2X7 receptors. British Journal of Pharmacology 157(7), pp. 1203–1214. Available at: https://doi.org/10.1111/j.1476-5381.2009.00233.x.

Dotsenko, G. and Lange, L. 2016. Enzyme Enhanced Protein Recovery from Green Biomass Pulp. Waste and Biomass Valorization 8(4), pp. 1257–1264. Available at: https://doi.org/10.1007/s12649-016-9718-7.

Dubczak, J., Reid, N. and Tsuchiya, M. 2021. Evaluation of limulus amebocyte lysate and recombinant endotoxin alternative assays for an assessment of endotoxin detection specificity. European Journal of Pharmaceutical Sciences 159, p. 105716. Available at: https://doi.org/10.1016/j.ejps.2021.105716.

Egelkrout, E., Rajan, V. and Howard, J.A. 2012. Overproduction of recombinant proteins in plants. *Plant Science* 184, pp. 83–101. Available at: https://doi.org/10.1016/j.plantsci.2011.12.005.

Elmer, J., Harris, D.R., Sun, G. and Palmer, A.F. 2009. Purification of hemoglobin by tangential flow filtration with diafiltration. *Biotechnology Progress* 25(5), pp. 1402–1410. Available at: https://doi.org/10.1002/btpr.217.

Farías-Rico, J.A., Selin, F.R., Myronidi, I., Frühauf, M. and Von Heijne, G. 2018. Effects of protein size, thermodynamic stability, and net charge on cotranslational folding on the ribosome. *Proceedings of the National Academy of Sciences of the United States of America* 115(40). Available at: https://doi.org/10.1073/pnas.1812756115.

Fischer, R. and Emans, N. 2000. Molecular farming of pharmaceutical proteins. *Transgenic Research* 9(4/5), pp. 279–299. Available at: https://doi.org/10.1023/a:1008975123362.

Forestier, E., Czechowski, T., Cording, A., Gilday, A.D., King, A.J., Brown, G.D. and Graham, I.A. 2021. Developing a Nicotiana benthamiana transgenic platform for high-value diterpene production and candidate gene evaluation. Plant Biotechnology Journal (Print) 19(8), pp. 1614–1623. Available at: https://doi.org/10.1111/pbi.13574.

Gallego-Bartolomé, J. 2020. DNA methylation in plants: mechanisms and tools for targeted manipulation. *New Phytologist* 227(1), pp. 38–44. Available at: https://doi.org/10.1111/nph.16529.

Garabagi, F., McLean, M.D. and Hall, J.C. 2012. Transient and stable expression of antibodies in Nicotiana species. In: *Methods in molecular biology*. pp. 389–408. Available at: https://doi.org/10.1007/978-1-61779-974-7_23.

Gecchele, E., Merlin, M., Brozzetti, A., Falorni, A., Pezzotti, M. and Avesani, L. 2015. A comparative analysis of recombinant protein expression in different biofactories: bacteria, insect cells and plant systems. *Journal of Visualized Experiments* (97). Available at: https://doi.org/10.3791/52459.

Gengenbach, B.B., Opdensteinen, P. and Buyel, J.F. 2020. Robot Cookies – plant cell packs as an automated High-ThroughPut screening platform based on transient expression. *Frontiers in Bioengineering and Biotechnology* 8. Available at: https://doi.org/10.3389/fbioe.2020.00393.

Gerszberg, A. and Hnatuszko-Konka, K. 2022. Compendium on Food Crop Plants as a platform for pharmaceutical protein production. *International Journal of Molecular Sciences* 23(6), p. 3236. Available at: https://doi.org/10.3390/ijms23063236.

Gitzinger, M., Parsons, J., Reski, R. and Fussenegger, M. 2008. Functional cross-kingdom conservation of mammalian and moss (*Physcomitrella patens*) transcription, translation and secretion machineries. *Plant Biotechnology Journal* 7(1), pp. 73–86. Available at: https://doi.org/10.1111/j.1467-7652.2008.00376.x.

Goldenzweig, A. and Fleishman, S.J. 2018. Principles of protein stability and their application in computational design. *Annual Review of Biochemistry* 87(1), pp. 105–129. Available at: https://doi.org/10.1146/annurev-biochem-062917-012102.

Gomes, A.M.V., Carmo, T.S., Carvalho, L.S., Bahia, F.M. and Parachin, N.S. 2018. Comparison of yeasts as hosts for recombinant protein production. *Microorganisms* 6(2), p. 38. Available at: https://doi.org/10.3390/microorganisms6020038.

Gomord, V. and Faye, L. 2004. Posttranslational modification of therapeutic proteins in plants. *Current Opinion in Plant Biology* 7(2), pp. 171–181. Available at: https://doi.org/10.1016/j.pbi.2004.01.015.

Gupta, M., Kumar, H. and Kaur, S. 2021. Vegetative Insecticidal Protein (Vip): A Potential Contender From Bacillus thuringiensis for Efficient Management of Various Detrimental Agricultural Pests. Frontiers in Microbiology 12. Available at: https://doi.org/10.3389/fmicb.2021.659736.

Haberland, J. and Gerke, V. 1999. Conserved charged residues in the leucine-rich repeat domain of the Ran GTPase activating protein are required for Ran binding and GTPase activation. Biochemical Journal 343(3), pp. 653–662. Available at: https://doi.org/10.1042/bj3430653.

Hamada, A., Yamaguchi, K., Ohnishi, N., Harada, M., Nikumaru, S. and Honda, H. 2004. High-level production of yeast (Schwanniomyces occidentalis) phytase in transgenic rice plants by a combination of signal sequence and codon modification of the phytase gene. *Plant Biotechnology Journal* 3(1), pp. 43–55. Available at: https://doi.org/10.1111/j.1467-7652.2004.00098.x.

Hansen, M., Andersen, C.A., Jensen, P.R. and Hobley, T.J. 2022. Scale-Up of Alfalfa (Medicago sativa) Protein Recovery Using Screw Presses. *Foods* 11(20), p. 3229. Available at: https://doi.org/10.3390/foods11203229.

Hattab, G., Warschawski, D.E., Moncoq, K. and Miroux, B. 2015. Escherichia coli as host for membrane protein structure determination: a global analysis. *Scientific Reports* 5(1). Available at: https://doi.org/10.1038/srep12097.

He, Y., Wang, K. and Yan, N. 2014. The recombinant expression systems for structure determination of eukaryotic membrane proteins. *Protein & Cell* 5(9), pp. 658–672. Available at: https://doi.org/10.1007/s13238-014-0086-4.

Hefferon, K. 2017. Plant Virus Expression vectors: a powerhouse for global health. *Biomedicines* 5(4), p. 44. Available at: https://doi.org/10.3390/biomedicines5030044.

Hiatt, A., Cafferkey, R. and Bowdish, K.S. 1989. Production of antibodies in transgenic plants. *Nature* 342(6245), pp. 76–78. Available at: https://doi.org/10.1038/342076a0.

Hiwasa-Tanase, K., Nyarubona, M., Hirai, T., Kato, K., Ichikawa, T. and Ezura, H. 2010. High-level accumulation of recombinant miraculin protein in transgenic tomatoes expressing a synthetic miraculin gene with optimized codon usage terminated by the native miraculin terminator. *Plant Cell Reports* 30(1), pp. 113–124. Available at: https://doi.org/10.1007/s00299-010-0949-y.

Huang, J. et al. 2002. Expression of natural antimicrobial human lysozyme in rice grains. *Molecular Breeding* 10(1/2), pp. 83–94. Available at: https://doi.org/10.1023/a:1020355511981.

Huang, Z., Chen, Q., Hjelm, B.E., Arntzen, C.J. and Mason, H.S. 2009. A DNA replicon system for rapid high-level production of virus-like particles in plants. *Biotechnology and Bioengineering* 103(4), pp. 706–714. Available at: https://doi.org/10.1002/bit.22299.

Huether, C.M., Lienhart, O., Baur, A., Stemmer, C., Gorr, G., Reski, R. and Decker, E.L. 2005. Glyco-Engineering of Moss Lacking Plant-Specific sugar Residues. *Plant Biology* 7(3), pp. 292–299. Available at: https://doi.org/10.1055/s-2005-837653.

Ibuot, A., Webster, R.E., Williams, L.E. and Pittman, J.K. 2020. Increased metal tolerance and bioaccumulation of zinc and cadmium in Chlamydomonas reinhardtii expressing a AtHMA4 C-terminal domain protein. bioRxiv (Cold Spring Harbor Laboratory). Available at: https://doi.org/10.1101/2020.02.13.948307.

Iqbal, M. et al. 2016. Aqueous two-phase system (ATPS): an overview and advances in its applications. *Biological Procedures Online* 18(1). Available at: https://doi.org/10.1186/s12575-016-0048-8.

Islam, R. et al. 2018. Cost-effective production of tag-less recombinant protein in *Nicotiana benthamiana*. *Plant Biotechnology Journal* 17(6), pp. 1094–1105. Available at: https://doi.org/10.1111/pbi.13040.

Jackson, D.A., Symons, R.H. and Berg, P. 1972. Biochemical Method for Inserting New Genetic Information into DNA of Simian Virus 40: Circular SV40 DNA Molecules Containing Lambda Phage Genes and the Galactose Operon of *Escherichia coli*. *Proceedings of the National*

Academy of Sciences of the United States of America 69(10), pp. 2904–2909. Available at: https://doi.org/10.1073/pnas.69.10.2904.

Jaradat, A.Q, 2021, 'Investigating the role of the P2X7 intracellular domains in coupling to downstream signalling', PhD thesis, Cardiff University, Cardiff

Jeong, I.S. et al. 2018. Purification and characterization of *Arabidopsis thaliana* oligosaccharyltransferase complexes from the native host: a protein super-expression system for structural studies. *Plant Journal* 94(1), pp. 131–145. Available at: https://doi.org/10.1111/tpj.13847.

Joseph, M., Ludevid, M.D., Torrent, M., Rofidal, V., Tauzin, M., Rossignol, M. and Peltier, J.-B. 2012. Proteomic characterisation of endoplasmic reticulum-derived protein bodies in tobacco leaves. *BMC Plant Biology* 12(1). Available at: https://doi.org/10.1186/1471-2229-12-36.

Karasawa, A. and Kawate, T. 2016. Structural basis for subtype-specific inhibition of the P2X7 receptor. eLife 5. Available at: https://doi.org/10.7554/elife.22153.

Karasawa, A., Michalski, K., Mikhelzon, P. and Kawate, T. 2017. The P2X7 receptor forms a dye-permeable pore independent of its intracellular domain but dependent on membrane lipid composition. *eLife* 6. Available at: https://doi.org/10.7554/elife.31186.

Karki, U., Fang, H., Guo, W., Unnold-Cofre, C. and Xu, J. 2021. Cellular engineering of plant cells for improved therapeutic protein production. *Plant Cell Reports* 40(7), pp. 1087–1099. Available at: https://doi.org/10.1007/s00299-021-02693-6.

Khvatkov, P. et al. 2018. Development of Wolffia arrhiza as a Producer for Recombinant Human Granulocyte Colony-Stimulating Factor. *Frontiers in Chemistry* 6. Available at: https://doi.org/10.3389/fchem.2018.00304.

Kimple, M.E., Brill, A.L. and Pasker, R.L. 2013. Overview of affinity tags for protein purification. *Current Protocols in Protein Science* 73(1). Available at: https://doi.org/10.1002/0471140864.ps0909s73.

Kopp, R., Krautloher, A., Ramírez-Fernández, A. and Nicke, A. 2019. P2X7 interactions and signaling – making head or tail of it. Frontiers in Molecular Neuroscience 12. Available at: https://doi.org/10.3389/fnmol.2019.00183.

Kumar, D., Patro, S., Ranjan, R., Sahoo, D.K., Maiti, I.B. and Dey, N. 2011. Development of useful recombinant promoter and its expression analysis in different plant cells using confocal laser scanning microscopy. *PLOS ONE* 6(9), p. e24627. Available at: https://doi.org/10.1371/journal.pone.0024627.

Kumar, M. et al. 2021. Advances in the plant protein extraction: Mechanism and recommendations. *Food Hydrocolloids* 115, p. 106595. Available at: https://doi.org/10.1016/j.foodhyd.2021.106595.

Kumar, S., Tripathi, D., Okubara, P.A. and Tanaka, K. 2020. Purinoceptor P2K1/DORN1 Enhances Plant Resistance Against a Soilborne Fungal Pathogen, Rhizoctonia solani. Frontiers in Plant Science 11. Available at: https://doi.org/10.3389/fpls.2020.572920.

Kuo, Y.C., Tan, C.C., Ku, J.T., Hsu, W.C., Su, S.C., Lu, C.A. and Huang, L. 2013. Improving Pharmaceutical Protein Production in Oryza sativa. *International Journal of Molecular Sciences* 14(5), pp. 8719–8739. Available at: https://doi.org/10.3390/ijms14058719.

Kurup, V.M. and Thomas, J.M. 2019. Edible vaccines: promises and challenges. *Molecular Biotechnology* 62(2), pp. 79–90. Available at: https://doi.org/10.1007/s12033-019-00222-1.

Lara, R., Adinolfi, E., Harwood, C.A., Philpott, M.P., Barden, J.A., Di Virgilio, F. and McNulty, S. 2020. P2X7 in cancer: From molecular mechanisms to therapeutics. *Frontiers in Pharmacology* 11. Available at: https://doi.org/10.3389/fphar.2020.00793.

Laurent, C. et al. 2016. Metal binding to the N-terminal cytoplasmic domain of the PIB ATPase HMA4 is required for metal transport in Arabidopsis. Plant Molecular Biology 90(4–5), pp. 453–466. Available at: https://doi.org/10.1007/s11103-016-0429-z.

Le, D.T., Chu, H.D. and Le, N.Q. 2016. Improving nutritional quality of plant proteins through genetic engineering. Current Genomics 17(3), pp. 220–229. Available at: https://doi.org/10.2174/1389202917666160202215934.

Le, S.B., Onsager, I., Lorentzen, J.A. and Lale, R. 2020. Dual UTR-A novel 5' untranslated region design for synthetic biology applications. *Synthetic Biology* 5(1). Available at: https://doi.org/10.1093/synbio/ysaa006.

Lehtimäki, N., Koskela, M.M. and Mulo, P. 2015. Posttranslational modifications of chloroplast proteins: an emerging field. Plant Physiology 168(3), pp. 768–775. Available at: https://doi.org/10.1104/pp.15.00117.

Lehman. 1974. DNA ligase: structure, mechanism, and function. *Science* 186(4166), pp. 790–797. Available at: https://doi.org/10.1126/science.186.4166.790.

Leuzinger, K., Dent, M., Hurtado, J., Stahnke, J., Lai, H., Zhou, X. and Chen, Q. 2013. Efficient agroinfiltration of plants for high-level transient expression of recombinant proteins. *Journal of Visualized Experiments* (77). Available at: https://doi.org/10.3791/50521.

Lico, C., Chen, Q. and Santi, L. 2008. Viral vectors for production of recombinant proteins in plants. Journal of Cellular Physiology (Print) 216(2), pp. 366–377. Available at: https://doi.org/10.1002/jcp.21423.

Lim, C.J. et al. 2012. Screening of tissue-specific genes and promoters in tomato by comparing genome wide expression profiles of Arabidopsis orthologues. *Molecules and Cells* 34(1), pp. 53–59. Available at: https://doi.org/10.1007/s10059-012-0068-4.

Limkul, J., Misaki, R., Kato, K. and Fujiyama, K. 2015. The combination of plant translational enhancers and terminator increase the expression of human glucocerebrosidase in Nicotiana benthamiana plants. *Plant Science* 240, pp. 41–49. Available at: https://doi.org/10.1016/j.plantsci.2015.08.018.

Lin, M.T., Orr, D.J., Worrall, D., Parry, M. a. J., Carmo-Silva, E. and Hanson, M.R. 2021. A procedure to introduce point mutations into the Rubisco large subunit gene in wild-type plants. The Plant Journal 106(3), pp. 876–887. Available at: https://doi.org/10.1111/tpj.15196.

Lindbo, J.A. 2007a. High-efficiency protein expression in plants from agroinfection-compatible Tobacco mosaic virus expression vectors. *BMC Biotechnology* 7(1), p. 52. Available at: https://doi.org/10.1186/1472-6750-7-52.

Lindbo, J.A. 2007b. TRBO: a High-Efficiency Tobacco Mosaic Virus RNA-Based Overexpression Vector. *Plant Physiology* 145(4), pp. 1232–1240. Available at: https://doi.org/10.1104/pp.107.106377.

Liu, D., Shi, L., Han, C., Yu, J., Li, D. and Zhang, Y.L. 2012. Validation of Reference Genes for Gene Expression Studies in Virus-Infected Nicotiana benthamiana Using Quantitative Real-Time PCR. *PLOS ONE* 7(9), p. e46451. Available at: https://doi.org/10.1371/journal.pone.0046451.

Liu, H. and Timko, M.P. 2022. Improving Protein Quantity and Quality—The next level of plant molecular farming. International Journal of Molecular Sciences 23(3), p. 1326. Available at: https://doi.org/10.3390/ijms23031326.

Ma, J., Drake, P. and Christou, P. 2003. The production of recombinant pharmaceutical proteins in plants. *Nature Reviews Genetics* 4(10), pp. 794–805. Available at: https://doi.org/10.1038/nrg1177.

Maclean, J. et al. 2007. Optimization of human papillomavirus type 16 (HPV-16) L1 expression in plants: comparison of the suitability of different HPV-16 L1 gene variants and different cell-compartment localization. Journal of General Virology 88(5), pp. 1460–1469. Available at: https://doi.org/10.1099/vir.0.82718-0.

Mahmood, M.A., Naqvi, R.Z., Rahman, S.U., Amin, I. and Mansoor, S. 2023. Plant Virus-Derived vectors for plant genome engineering. *Viruses* 15(2), p. 531. Available at: https://doi.org/10.3390/v15020531.

Makatsa, M. et al. 2021. SARS-COV-2 antigens expressed in plants detect antibody responses in COVID-19 patients. *Frontiers in Plant Science* 12. Available at: https://doi.org/10.3389/fpls.2021.589940.

Malaquias, Â.D.M. et al. 2021. A review of plant-based expression systems as a platform for single-domain recombinant antibody production. *International Journal of Biological Macromolecules* 193, pp. 1130–1137. Available at: https://doi.org/10.1016/j.ijbiomac.2021.10.126.

Malla, A., Rosales-Mendoza, S., Phoolcharoen, W. and Vimolmangkang, S. 2021. Efficient transient expression of recombinant proteins using DNA viral vectors in freshwater microalgal species. *Frontiers in Plant Science* 12. Available at: https://doi.org/10.3389/fpls.2021.650820.

Mamat, U. et al. 2015. Detoxifying Escherichia coli for endotoxin-free production of recombinant proteins. *Microbial Cell Factories* 14(1). Available at: https://doi.org/10.1186/s12934-015-0241-5.

Mardanova, E.S., Blokhina, E., Цыбалова, Л.М., Peyret, H., Lomonossoff, G.P. and Ravin, N.V. 2017. Efficient transient expression of recombinant proteins in plants by the novel PEFF Vector based on the genome of potato virus X. *Frontiers in Plant Science* 8. Available at: https://doi.org/10.3389/fpls.2017.00247.

Mardanova, E.S., Kotlyarov, R.Y. and Ravin, N.V. 2009. The optimization of viral vector translation improves the production of recombinant proteins in plants. *Molecular Biology* 43(3), pp. 524–527. Available at: https://doi.org/10.1134/s0026893309030212.

Mardanova, E.S., Kotlyarov, R.Y., Kuprianov, V.V., Степанова, Л.А., Цыбалова, Л.М., Lomonosoff, G.P. and Ravin, N.V. 2015. Rapid high-yield expression of a candidate influenza vaccine based on the ectodomain of M2 protein linked to flagellin in plants using viral vectors. *BMC Biotechnology* 15(1). Available at: https://doi.org/10.1186/s12896-015-0164-6.

Marillonnet, S., Giritch, A., Gils, M., Kandzia, R., Klimyuk, V. and Gleba, Y. 2004. *In planta* engineering of viral RNA replicons: Efficient assembly by recombination of DNA modules delivered by

Agrobacterium. *Proceedings of the National Academy of Sciences of the United States of America* 101(18), pp. 6852–6857. Available at: https://doi.org/10.1073/pnas.0400149101.

Marillonnet, S., Thoeringer, C., Kandzia, R., Klimyuk, V. and Gleba, Y. 2005. Systemic Agrobacterium tumefaciens–mediated transfection of viral replicons for efficient transient expression in plants. *Nature Biotechnology* 23(6), pp. 718–723. Available at: https://doi.org/10.1038/nbt1094.

McCarthy, A.E., Yoshioka, C. and Mansoor, S. 2019. Full-Length P2X7 structures reveal how palmitoylation prevents channel desensitization. *Cell* 179(3), pp. 659-670.e13. Available at: https://doi.org/10.1016/j.cell.2019.09.017.

Menzel, S., Holland, T., Boes, A., Spiegel, H., Bolzenius, J., Fischer, R. and Buyel, J.F. 2016. Optimized blanching reduces the host cell protein content and substantially enhances the recovery and stability of two Plant-Derived Malaria vaccine candidates. *Frontiers in Plant Science* 7. Available at: https://doi.org/10.3389/fpls.2016.00159.

Merlin, M., Gecchele, E., Capaldi, S., Pezzotti, M. and Avesani, L. 2014. Comparative evaluation of recombinant protein production in different biofactories: The Green Perspective. *BioMed Research International* 2014, pp. 1–14. Available at: https://doi.org/10.1155/2014/136419.

Meshcheriakova, Y., Saxena, P. and Lomonossoff, G.P. 2014. Fine-tuning levels of heterologous gene expression in plants by orthogonal variation of the untranslated regions of a nonreplicating transient expression system. *Plant Biotechnology Journal* 12(6), pp. 718–727. Available at: https://doi.org/10.1111/pbi.12175.

Millar, A.H., Heazlewood, J.L., Giglione, C., Holdsworth, M.J., Bachmair, A. and Schulze, W.X. 2019. The scope, functions, and dynamics of posttranslational protein modifications. *Annual Review of Plant Biology* 70(1), pp. 119–151. Available at: https://doi.org/10.1146/annurev-arplant-050718-100211.

Mirzaee, M., Osmani, Z., Frébortová, J. and Frébort, I. 2022. Recent advances in molecular farming using monocot plants. *Biotechnology Advances* 58, p. 107913. Available at: https://doi.org/10.1016/j.biotechadv.2022.107913.

Moon, K.-B. et al. 2019. Development of systems for the production of Plant-Derived biopharmaceuticals. *Plants* 9(1), p. 30. Available at: https://doi.org/10.3390/plants9010030.

Nagaya, S., Kawamura, K., Shinmyo, A. and Kato, K. 2009. The HSP Terminator of Arabidopsis thaliana Increases Gene Expression in Plant Cells. *Plant and Cell Physiology* 51(2), pp. 328–332. Available at: https://doi.org/10.1093/pcp/pcp188.

Niederkrüger, H., Dabrowska-Schlepp, P. and Schaaf, A. 2014. Suspension culture of plant cells under phototrophic conditions. In: *Wiley-VCH Verlag GmbH & Co. KGaA eBooks*. pp. 259–292. Available at: https://doi.org/10.1002/9783527683321.ch08.

Nirenberg, M.W., Leder, P., Bernfield, M., Brimacombe, R., Trupin, J.S., Rottman, F.M. and O'Neal, C. 1965. RNA codewords and protein synthesis, VII. On the general nature of the RNA code. *Proceedings of the National Academy of Sciences of the United States of America* 53(5), pp. 1161–1168. Available at: https://doi.org/10.1073/pnas.53.5.1161.

Norkunas, K., Harding, R.M., Dale, J.L. and Dugdale, B. 2018. Improving agroinfiltration-based transient gene expression in Nicotiana benthamiana. *Plant Methods* 14(1). Available at: https://doi.org/10.1186/s13007-018-0343-2.

O'Flaherty, R. et al. 2020. Mammalian cell culture for production of recombinant proteins: A review of the critical steps in their biomanufacturing. *Biotechnology Advances* 43, p. 107552. Available at: https://doi.org/10.1016/j.biotechadv.2020.107552.

Oey, M., Lohse, M., Kreikemeyer, B. and Bock, R. 2009. Exhaustion of the chloroplast protein synthesis capacity by massive expression of a highly stable protein antibiotic. *Plant Journal* 57(3), pp. 436–445. Available at: https://doi.org/10.1111/j.1365-313x.2008.03702.x.

Oliveira-Giacomelli, Á., Petiz, L.L., Andrejew, R., Turrini, N., Silva, J.B., Sack, U. and Ulrich, H. 2021. Role of P2X7 receptors in immune responses during neurodegeneration. Frontiers in Cellular Neuroscience 15. Available at: https://doi.org/10.3389/fncel.2021.662935.

Opdensteinen, P., Clodt, J.I., Müschen, C., Filiz, V. and Buyel, J.F. 2019. A combined Ultrafiltration/Diafiltration step facilitates the purification of Cyanovirin-N from transgenic tobacco extracts. *Frontiers in Bioengineering and Biotechnology* 6. Available at: https://doi.org/10.3389/fbioe.2018.00206.

Opdensteinen, P., Lobanov, A. and Buyel, J.F. 2021. A combined pH and temperature precipitation step facilitates the purification of tobacco-derived recombinant proteins that are sensitive to extremes of either parameter. *Biotechnology Journal* 16(4). Available at: https://doi.org/10.1002/biot.202000340.

Owczarek, B., Gerszberg, A. and Hnatuszko-Konka, K. 2019. A brief reminder of systems of Production and Chromatography-Based Recovery of Recombinant Protein biopharmaceuticals. *BioMed Research International* 2019, pp. 1–13. Available at: https://doi.org/10.1155/2019/4216060.

Parmar, N. et al. 2017. Genetic engineering strategies for biotic and abiotic stress tolerance and quality enhancement in horticultural crops: a comprehensive review. 3 Biotech 7(4). Available at: https://doi.org/10.1007/s13205-017-0870-y.

Parmenter, D.L., Boothe, J.G., Van Rooijen, G.J.H., Yeung, E.C. and Moloney, M.M. 1995. Production of biologically active hirudin in plant seeds using oleosin partitioning. *Plant Molecular Biology* 29(6), pp. 1167–1180. Available at: https://doi.org/10.1007/bf00020460.

Pelegrín, P. 2011. Many ways to dilate the P2X7 receptor pore. *British Journal of Pharmacology* 163(5), pp. 908–911. Available at: https://doi.org/10.1111/j.1476-5381.2011.01325.x.

Perlak, F.J., Fuchs, R.L., Dean, D.A., McPherson, S.A. and Fischhoff, D.A. 1991. Modification of the coding sequence enhances plant expression of insect control protein genes. *Proceedings of the National Academy of Sciences of the United States of America* 88(8), pp. 3324–3328. Available at: https://doi.org/10.1073/pnas.88.8.3324.

Peyret, H., Brown, J.K.M. and Lomonossoff, G.P. 2019. Improving plant transient expression through the rational design of synthetic 5′ and 3′ untranslated regions. *Plant Methods* 15(1). Available at: https://doi.org/10.1186/s13007-019-0494-9.

Pina, A.S., Lowe, C.R. and Roque, A.C.A. 2014. Challenges and opportunities in the purification of recombinant tagged proteins. *Biotechnology Advances* 32(2), pp. 366–381. Available at: https://doi.org/10.1016/j.biotechadv.2013.12.001.

Pizarro, L. and Norambuena, L. 2014. Regulation of protein trafficking: Posttranslational mechanisms and the unexplored transcriptional control. *Plant Science* 225, pp. 24–33. Available at: https://doi.org/10.1016/j.plantsci.2014.05.004.

Plotkin, J.B. and Kudla, G. 2010. Synonymous but not the same: the causes and consequences of codon bias. *Nature Reviews Genetics* 12(1), pp. 32–42. Available at: https://doi.org/10.1038/nrg2899.

Puetz, J. and Wurm, F.M. 2019. Recombinant Proteins for Industrial versus Pharmaceutical Purposes: A Review of Process and Pricing. *Processes* 7(8), p. 476. Available at: https://doi.org/10.3390/pr7080476.

Puthenveetil, R., Lee, C. and Banerjee, A. 2020. Production of Recombinant Transmembrane Proteins from Mammalian Cells for Biochemical and Structural Analyses. *Current Protocols in Cell Biology* 87(1). Available at: https://doi.org/10.1002/cpcb.106.

Qi, Y. and Katagiri, F. 2009. Purification of low-abundance Arabidopsis plasma-membrane protein complexes and identification of candidate components. *Plant Journal* 57(5), pp. 932–944. Available at: https://doi.org/10.1111/j.1365-313x.2008.03736.x.

Quianzon, C.C.L. and Cheikh, I.E. 2012. History of insulin. *Journal of Community Hospital Internal Medicine Perspectives* 2(2), p. 18701. Available at: https://doi.org/10.3402/jchimp.v2i2.18701.

Rabert, C., Weinacker, D., Pessoa, A. and Farías, J.G. 2013. Recombinants proteins for industrial uses: utilization of Pichia pastoris expression system. *Brazilian Journal of Microbiology* 44(2), pp. 351–356. Available at: https://doi.org/10.1590/s1517-83822013005000041.

Rademacher, T.W., Sack, M., Blessing, D., Fischer, R., Holland, T. and Buyel, J.F. 2019. Plant cell packs: a scalable platform for recombinant protein production and metabolic engineering. *Plant Biotechnology Journal* 17(8), pp. 1560–1566. Available at: https://doi.org/10.1111/pbi.13081.

Rattanapisit, K. et al. 2020. Rapid production of SARS-CoV-2 receptor binding domain (RBD) and spike specific monoclonal antibody CR3022 in Nicotiana benthamiana. *Scientific Reports* 10(1). Available at: https://doi.org/10.1038/s41598-020-74904-1.

Reski, R., Bae, H.C. and Simonsen, H.T. 2018. Physcomitrella patens, a versatile synthetic biology chassis. *Plant Cell Reports* 37(10), pp. 1409–1417. Available at: https://doi.org/10.1007/s00299-018-2293-6.

Reski, R., Parsons, J. and Decker, E.L. 2015. Moss-made pharmaceuticals: from bench to bedside. *Plant Biotechnology Journal* 13(8), pp. 1191–1198. Available at: https://doi.org/10.1111/pbi.12401.

Rigano, M.M., De Guzman, G., Walmsley, A.M., Frusciante, L. and Barone, A. 2013. Production of pharmaceutical proteins in Solanaceae food crops. *International Journal of Molecular Sciences* 14(2), pp. 2753–2773. Available at: https://doi.org/10.3390/ijms14022753.

Robert, S., Goulet, M., D'Aoust, M., Sainsbury, F. and Michaud, D. 2015. Leaf proteome rebalancing inNicotiana benthamianafor upstream enrichment of a transiently expressed recombinant protein. Plant Biotechnology Journal 13(8), pp. 1169–1179. Available at: https://doi.org/10.1111/pbi.12452.

Robert, S., Khalf, M., Goulet, M.-C., D'Aoust, M.-A., Sainsbury, F. and Michaud, D. 2013. Protection of Recombinant Mammalian Antibodies from Development-Dependent Proteolysis in Leaves of Nicotiana benthamiana. *PLOS ONE* 8(7), p. e70203. Available at: https://doi.org/10.1371/journal.pone.0070203.

Rose, A.B. 2008. Intron-Mediated regulation of gene Expression. In: *Current Topics in Microbiology and Immunology*. pp. 277–290. Available at: https://doi.org/10.1007/978-3-540-76776-3_15.

Ruhlman, T.A., Verma, D., Samson, N. and Daniell, H. 2010. The role of heterologous chloroplast sequence elements in transgene integration and expression. *Plant Physiology* 152(4), pp. 2088–2104. Available at: https://doi.org/10.1104/pp.109.152017.

Ryabova, L.A. and Hohn, T. 2000. Ribosome shunting in the cauliflower mosaic virus 35S RNA leader is a special case of reinitiation of translation functioning in plant and animal systems. *Genes & Development* 14(7), pp. 817–829. Available at: https://doi.org/10.1101/gad.14.7.817.

Sainsbury, F. and Lomonossoff, G.P. 2008. Extremely High-Level and Rapid Transient Protein Production in Plants without the Use of Viral Replication. *Plant Physiology* 148(3), pp. 1212–1218. Available at: https://doi.org/10.1104/pp.108.126284.

Sainsbury, F., Thuenemann, E.C. and Lomonossoff, G.P. 2009. pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnology Journal* 7(7), pp. 682–693. Available at: https://doi.org/10.1111/j.1467-7652.2009.00434.x.

Santos, C.D. and Franco, O.L. 2023. Pathogenesis-Related Proteins (PRs) with Enzyme Activity Activating Plant Defense Responses. Plants 12(11), p. 2226. Available at: https://doi.org/10.3390/plants12112226.

Schaaf, A., Tintelnot, S., Baur, A., Reski, R., Gorr, G. and Decker, E.L. 2005. Use of endogenous signal sequences for transient production and efficient secretion by moss (Physcomitrella patens) cells. *BMC Biotechnology* 5(1). Available at: https://doi.org/10.1186/1472-6750-5-30.

Schillberg, S. and Finnern, R. 2021. Plant molecular farming for the production of valuable proteins – Critical evaluation of achievements and future challenges. *Journal of Plant Physiology* 258–259, p. 153359. Available at: https://doi.org/10.1016/j.jplph.2020.153359.

Schillberg, S., Raven, N., Spiegel, H., Rasche, S. and Buntru, M. 2019. Critical analysis of the commercial potential of plants for the production of recombinant proteins. *Frontiers in Plant Science* 10. Available at: https://doi.org/10.3389/fpls.2019.00720.

Schmidt, J.A., Richter, L.V., Condoluci, L.A. and Ahner, B.A. 2021. Mitigation of deleterious phenotypes in chloroplast-engineered plants accumulating high levels of foreign proteins. Biotechnology for Biofuels 14(1). Available at: https://doi.org/10.1186/s13068-021-01893-2.

Seternes, T., Tonheim, T.C., Myhr, A.I. and Dalmo, R.A. 2016. A plant 35S CaMV promoter induces long-term expression of luciferase in Atlantic salmon. *Scientific Reports* 6(1). Available at: https://doi.org/10.1038/srep25096.

Shah, K.H., Almaghrabi, B. and Bohlmann, H. 2013. Comparison of expression vectors for transient expression of recombinant proteins in plants. Plant Molecular Biology Reporter 31(6), pp. 1529–1538. Available at: https://doi.org/10.1007/s11105-013-0614-z.

Shanmugaraj, B., Bulaon, C.J.I. and Phoolcharoen, W. 2020. Plant Molecular Farming: a viable platform for recombinant biopharmaceutical production. *Plants* 9(7), p. 842. Available at: https://doi.org/10.3390/plants9070842.

Shapiro, J., MacHattie, L.A., Eron, L., Ihler, G.M., Ippen, K. and Beckwith, J. 1969. Isolation of Pure lac Operon DNA*. *Nature* 224(5221), pp. 768–774. Available at: https://doi.org/10.1038/224768a0.

Shen, J. et al. 2015. Mannose receptor-mediated delivery of moss-made α-galactosidase A efficiently corrects enzyme deficiency in Fabry mice. *Journal of Inherited Metabolic Disease* 39(2), pp. 293–303. Available at: https://doi.org/10.1007/s10545-015-9886-9.

Strasser, R. 2016. Plant protein glycosylation. *Glycobiology* 26(9), pp. 926–939. Available at: https://doi.org/10.1093/glycob/cww023.

Tewari, M. and Seth, P. 2015. Emerging role of P2X7 receptors in CNS health and disease. *Ageing Research Reviews* 24, pp. 328–342. Available at: https://doi.org/10.1016/j.arr.2015.10.001.

Tippens, N.D., Vihervaara, A. and Lis, J.T. 2018. Enhancer transcription: what, where, when, and why? *Genes & Development* 32(1), pp. 1–3. Available at: https://doi.org/10.1101/gad.311605.118.

Torrent, M., Llop-Tous, I. and Ludevid, M.D. 2009. Protein body induction: a new tool to produce and recover recombinant proteins in plants. In: *Methods in molecular biology*. pp. 193–208. Available at: https://doi.org/10.1007/978-1-59745-407-0_11.

Tripathi, N.K. and Shrivastava, A. 2019. Recent developments in bioprocessing of recombinant proteins: expression hosts and process development. *Frontiers in Bioengineering and Biotechnology* 7. Available at: https://doi.org/10.3389/fbioe.2019.00420.

Tyurin, A.A., Suhorukova, A.V., Кабардаева, К.В. and Голденкова-павлова, И.В. 2020. Transient Gene Expression is an Effective Experimental Tool for the Research into the Fine Mechanisms of Plant Gene Function: Advantages, Limitations, and Solutions. *Plants* 9(9), p. 1187. Available at: https://doi.org/10.3390/plants9091187.

Urs, A.P. et al. 2019. Plant latex thrombin-like cysteine proteases alleviates bleeding by bypassing factor VIII in murine model. Journal of Cellular Biochemistry 120(8), pp. 12843–12858. Available at: https://doi.org/10.1002/jcb.28555.

Urs, A.P., Manjuprasanna, V.N., Rudresha, G.V., Hiremath, V., Sharanappa, P., Rajesh, R. and Vishwanath, B.S. 2021. Thrombin-like serine protease, antiquorin from Euphorbia antiquorum latex induces platelet aggregation via PAR1-Akt/p38 signaling axis. Biochimica Et Biophysica Acta (BBA) - Molecular Cell Research 1868(3), p. 118925. Available at: https://doi.org/10.1016/j.bbamcr.2020.118925.

Usmani, S.S. et al. 2017. THPdb: Database of FDA-approved peptide and protein therapeutics. *PLOS ONE* 12(7), p. e0181748. Available at: https://doi.org/10.1371/journal.pone.0181748.

Vain, P., Worland, B., Kohli, A., Snape, J.W., Christou, P., Allen, G.C. and Thompson, W.F. 1999. Matrix attachment regions increase transgene expression levels and stability in transgenic rice plants and their progeny. *Plant Journal* 18(3), pp. 233–242. Available at: https://doi.org/10.1046/j.1365-313x.1999.00446.x.

Von Schaewen, A. et al. 2018. Improved recombinant protein production in Arabidopsis thaliana. *Plant Signaling & Behavior* 13(6), p. e1486149. Available at: https://doi.org/10.1080/15592324.2018.1486149.

Webster, G., Teh, A.Y. and K-c, J. 2016. Synthetic gene design-The rationale for codon optimization and implications for molecular pharming in plants. *Biotechnology and Bioengineering* 114(3), pp. 492–502. Available at: https://doi.org/10.1002/bit.26183.

Wilken, L.R. and Nikolov, Z.L. 2012. Recovery and purification of plant-made recombinant proteins. *Biotechnology Advances* 30(2), pp. 419–433. Available at: https://doi.org/10.1016/j.biotechadv.2011.07.020.

Wu, F.H., Shen, S.C., Lee, L.Y., Lee, S.H., Chan, M.T. and Lin, C.-S. 2009. Tape-Arabidopsis Sandwich - a simpler Arabidopsis protoplast isolation method. *Plant Methods* 5(1). Available at: https://doi.org/10.1186/1746-4811-5-16.

Xu, X.M., Zhao, Q., Rodrigo-Peiris, T., Brkljačić, J., He, C.S., Müller, S. and Meier, I. 2008. RanGAP1 is a continuous marker of the Arabidopsis cell division plane. Proceedings of the National Academy of Sciences of the United States of America 105(47), pp. 18637–18642. Available at: https://doi.org/10.1073/pnas.0806157105.

Yamamoto, T. et al. 2018. Improvement of the transient expression system for production of recombinant proteins in plants. *Scientific Reports* 8(1). Available at: https://doi.org/10.1038/s41598-018-23024-y.

Yang, G., Feng, D., Liu, Y., Lv, S., Zheng, M.-M. and Tan, A.-J. 2021. Research progress of a potential bioreactor: duckweed. *Biomolecules* 11(1), p. 93. Available at: https://doi.org/10.3390/biom11010093.

Yang, J. et al. 2022. Rethinking plant protein extraction: Albumin—From side stream to an excellent foaming ingredient. *Food Structure* 31, p. 100254. Available at: https://doi.org/10.1016/j.foostr.2022.100254.

Zeenko, V.V., Ryabova, L.A., Spirin, A.S., Rothnie, H.M., Hess, D., Browning, K.S. and Hohn, T. 2002. Eukaryotic Elongation Factor 1A Interacts with the Upstream Pseudoknot Domain in the 3′ Untranslated Region of Tobacco Mosaic Virus RNA. *Journal of Virology* 76(11), pp. 5678–5691. Available at: https://doi.org/10.1128/jvi.76.11.5678-5691.2002.

Zhang, X., Zhu, Y., Wu, H. and Guo, H. 2015. Post-transcriptional gene silencing in plants: a double-edged sword. *Science China-life Sciences* 59(3), pp. 271–276. Available at: https://doi.org/10.1007/s11427-015-4972-7.

Zhang, Y., Li, D., Jin, X. and Huang, Z. 2014. Fighting Ebola with ZMapp: spotlight on plant-made antibody. *Science China-life Sciences* 57(10), pp. 987–988. Available at: https://doi.org/10.1007/s11427-014-4746-7.

Zimmermann, M. et al. 2009. Metal Binding Affinities of *Arabidopsis* Zinc and Copper Transporters: Selectivities Match the Relative, but Not the Absolute, Affinities of their Amino-Terminal Domains,. Biochemistry 48(49), pp. 11640–11654. Available at: https://doi.org/10.1021/bi901573b.

Supplementary material

Supporting Figures

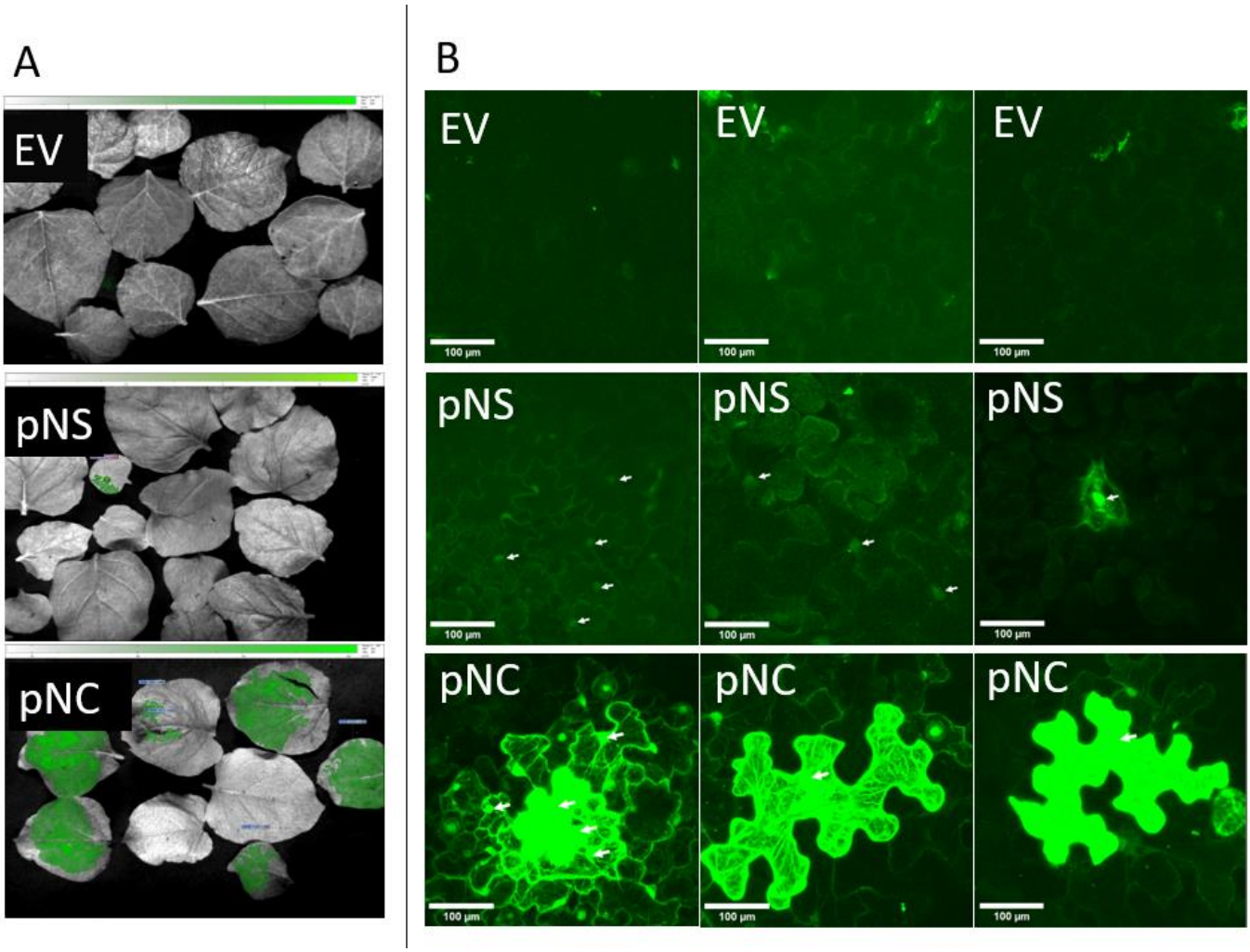


Figure S3.1: This image is identical to Figure 3.4 but with the confocal images artificially enhanced using ImageJ, increasing brightness and contrast equally across images to aid viewing of the lower expressing cells transformed with pNS-eGFP.

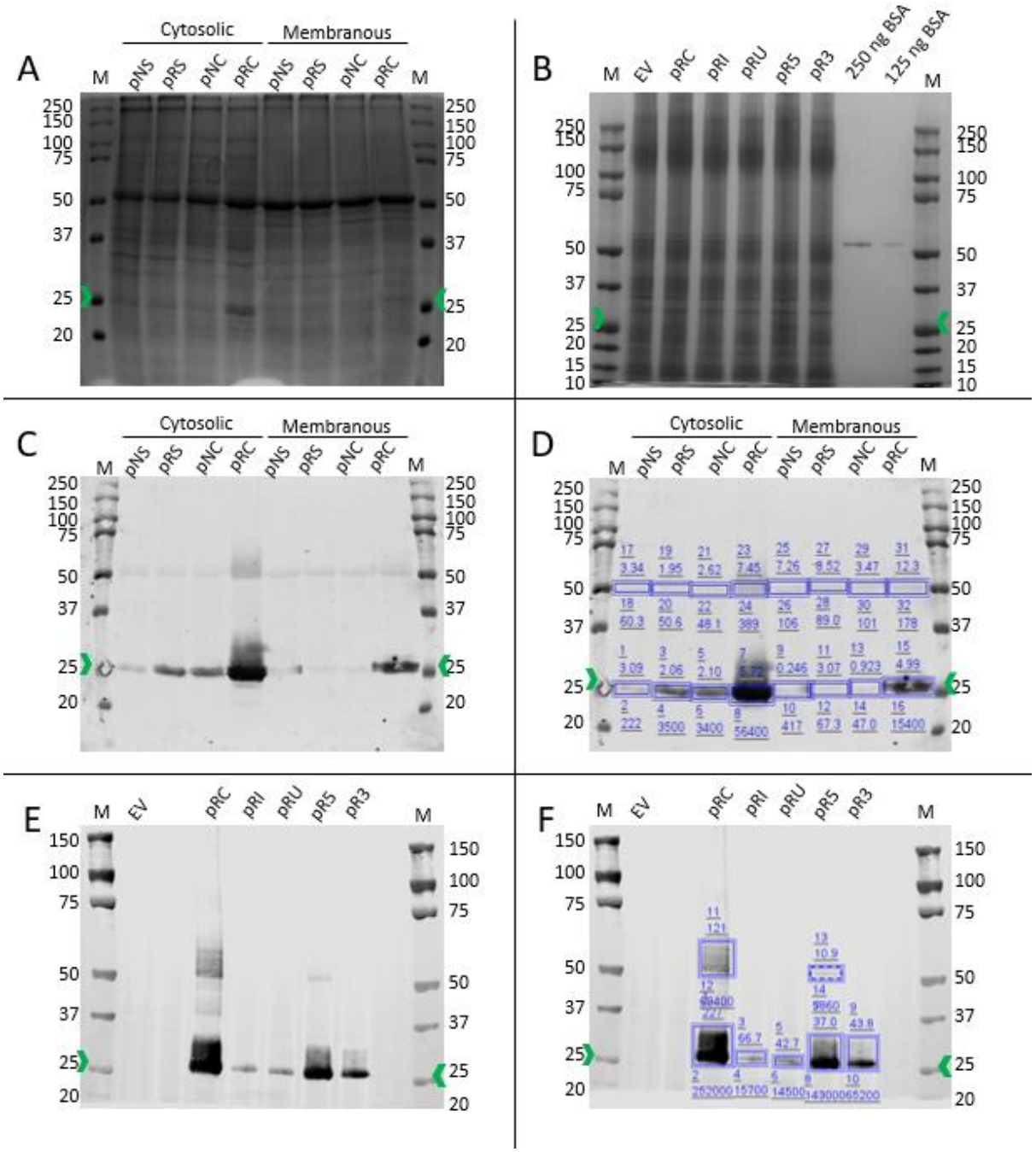


Figure S4.1: SDS-PAGE and Western Blot Analysis of Extracted Proteins from Different Samples. This figure is identical to Figure 4.11 but shows panels D and F show the raw densitometric quantification of bands in panels C and E, respectively, instead of the graphical format.

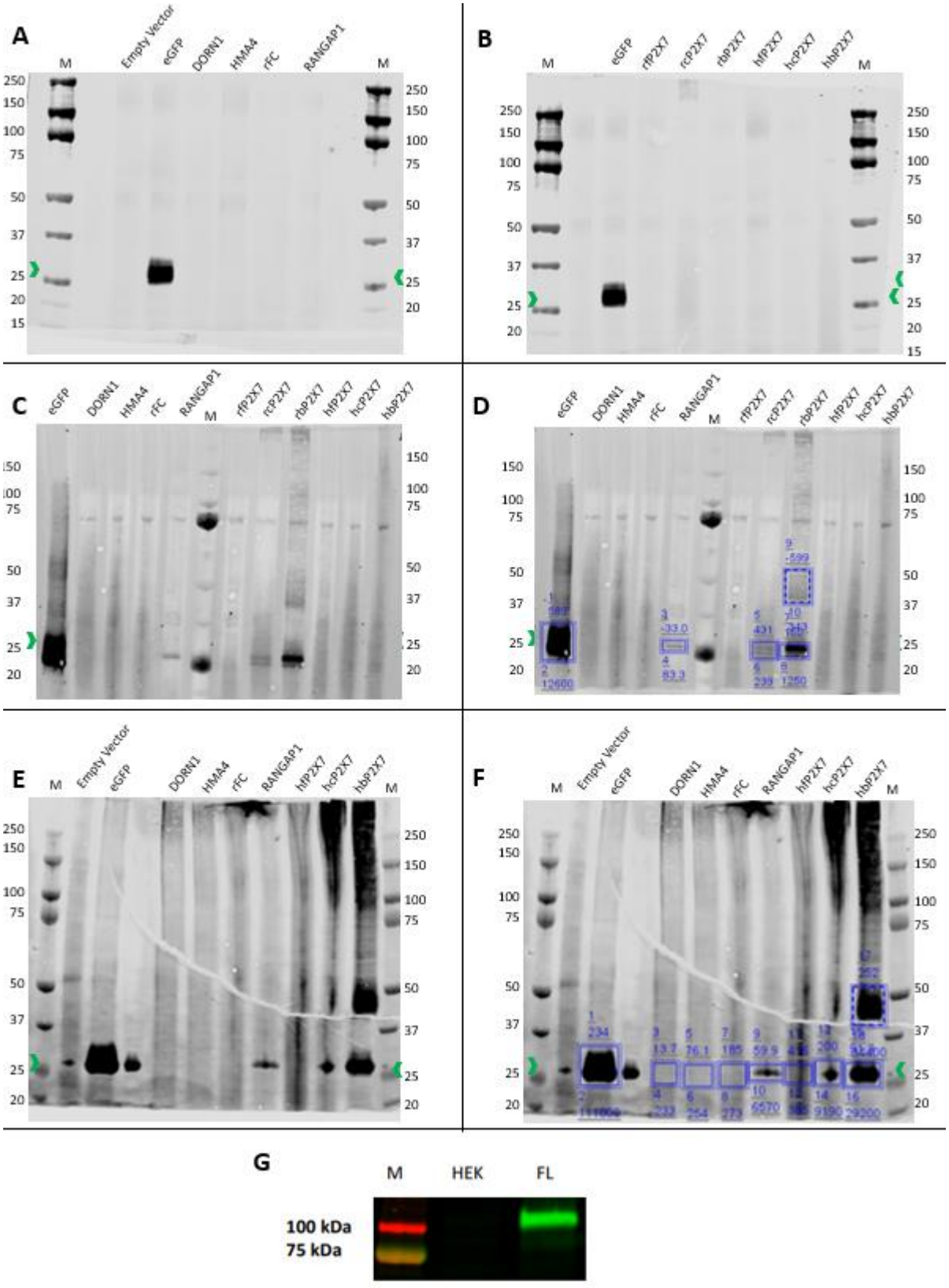


Figure S5.1: SDS-PAGE and Western Blot Analysis of Extracted Proteins from pR5 and pRC. This figure is identical to Figure 5.10 but shows panels D and F show the raw densitometric quantification of bands in panels C and E, respectively, instead of the graphical format.

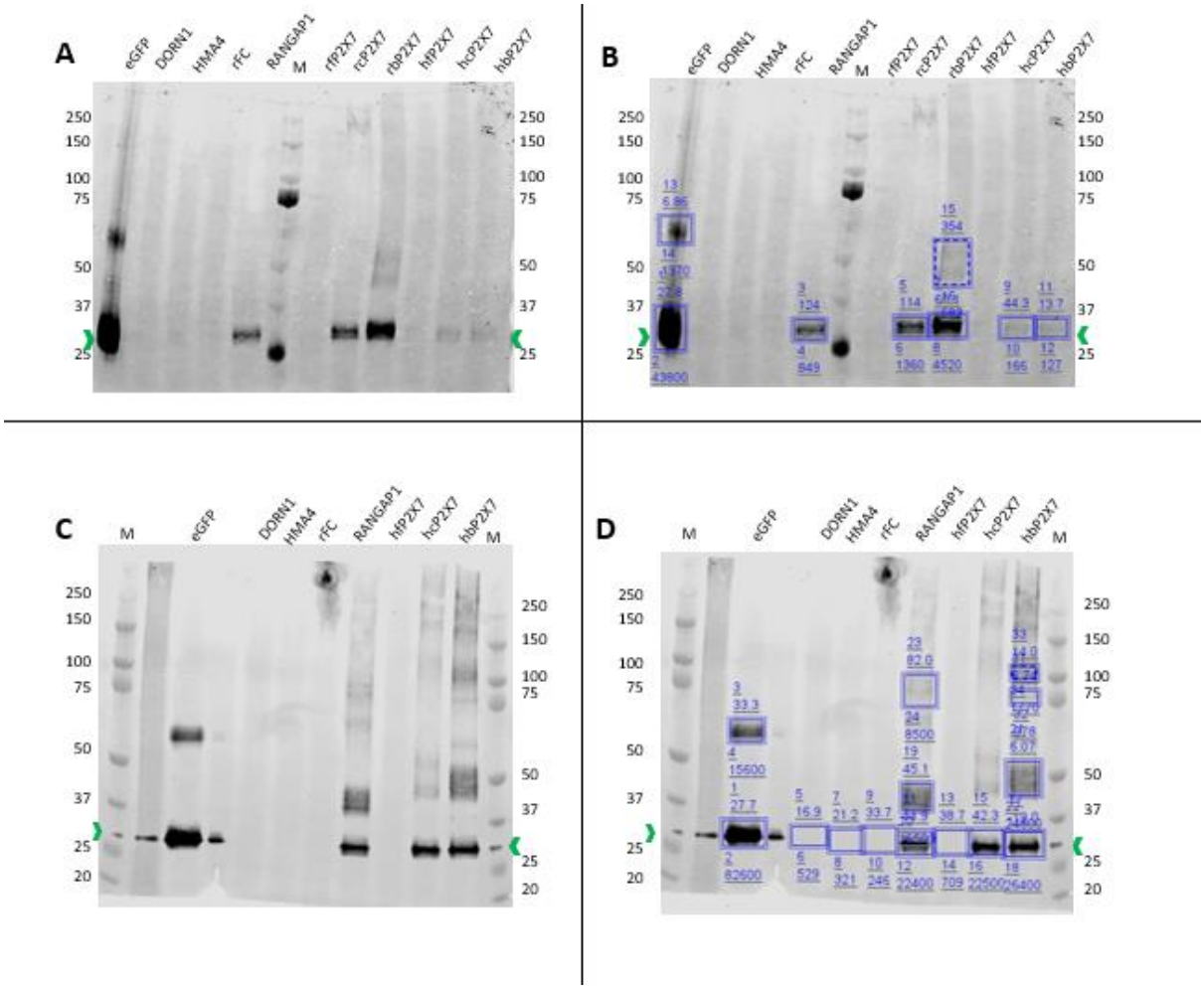


Figure S5.2: SDS-PAGE and Western Blot Analysis of Extracted Proteins from pR5 and pRC, following purification. This figure is identical to Figure 5.11 but shows panels B and D show the raw densitometric quantification of bands in panels A and C, respectively, instead of the graphical format.

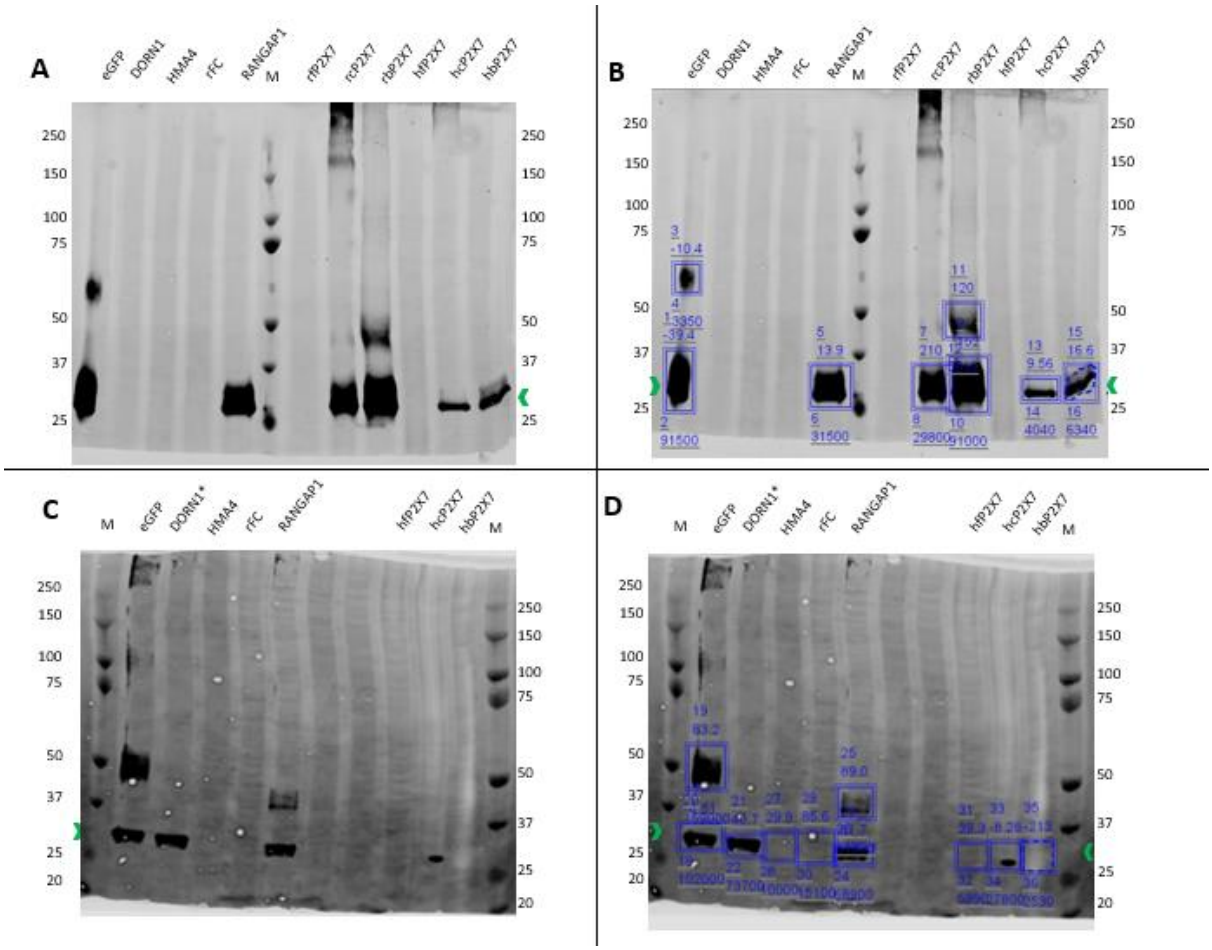


Figure S5.3: SDS-PAGE and Western Blot Analysis of Extracted Proteins from pR5 and pRC, following purification and concentration. This figure is identical to Figure 5.12 but shows panels B and D show the raw densitometric quantification of bands in panels A and C, respectively, instead of the graphical format.

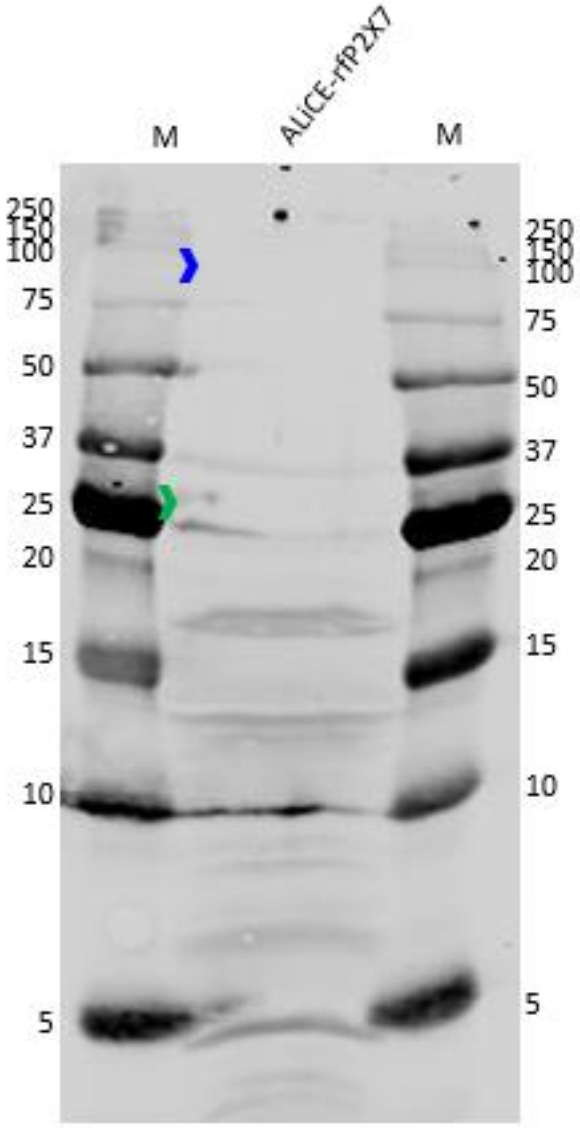
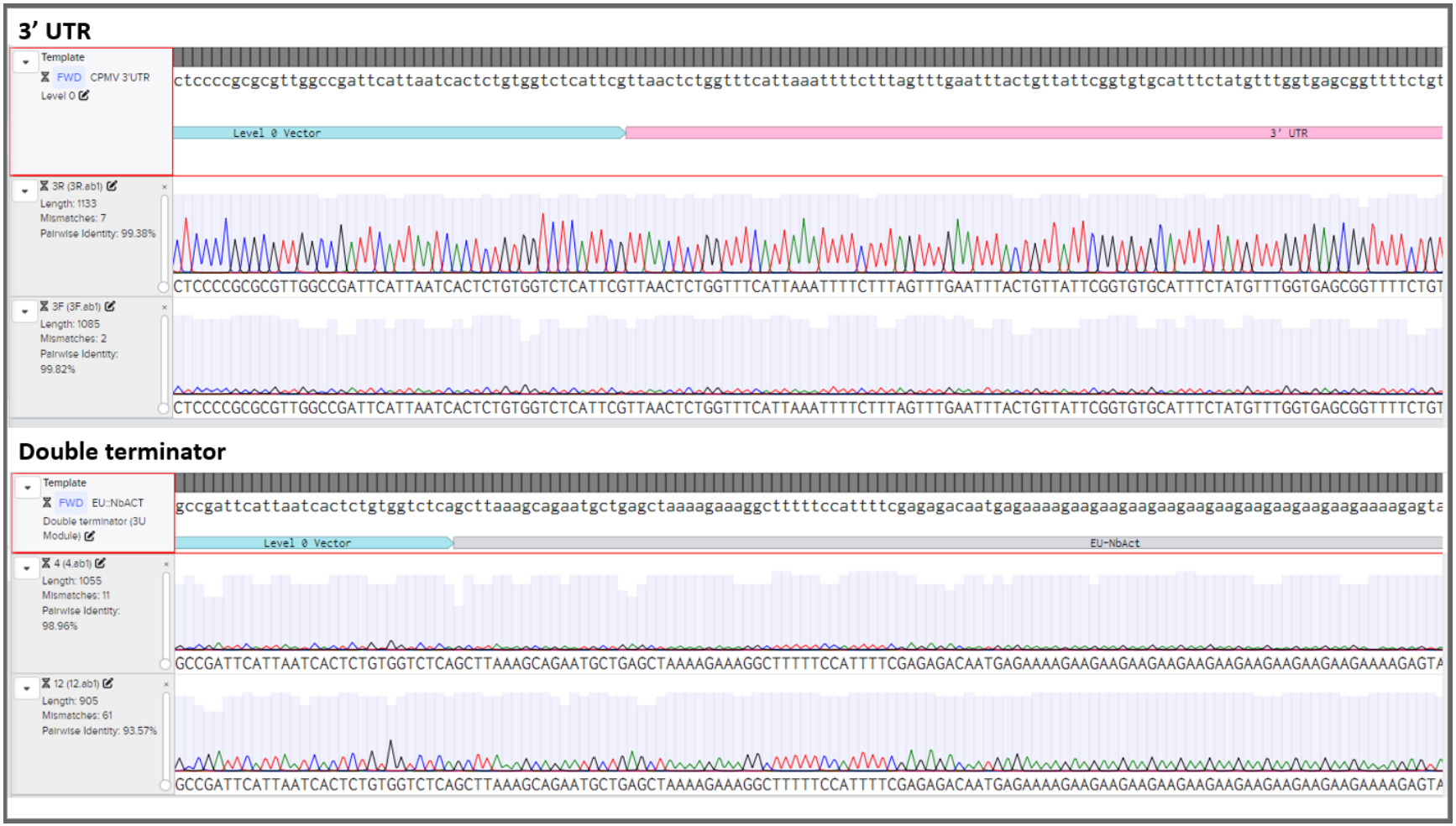


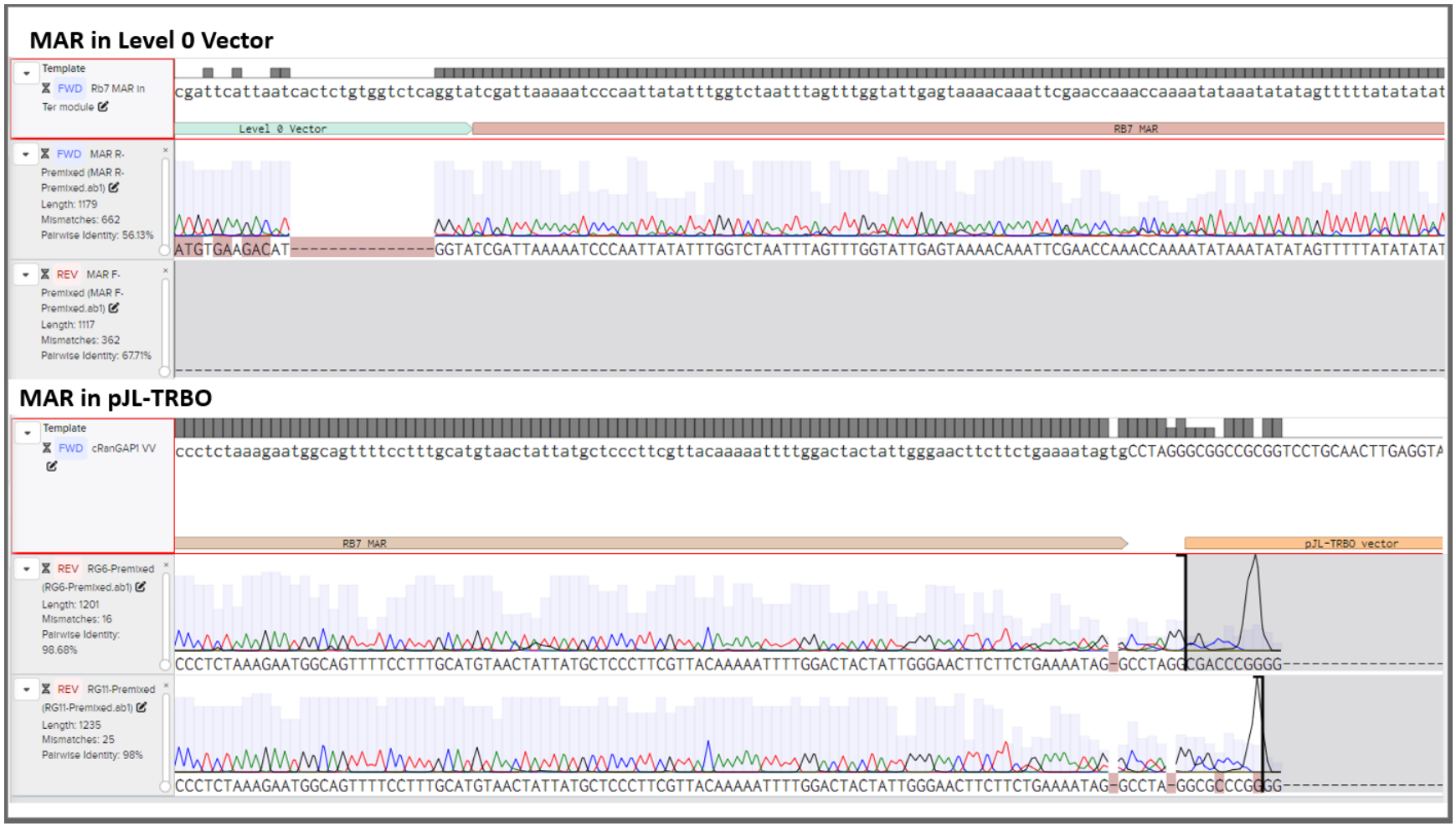
Figure S5.4: Western blot showing the protein extract of an ALiCE reaction expressing rat full length P2X7. The green arrow indicated the expected position of eGFP, and the blue arrow indicates the expected position of full length rat P2X7. Neither band is present but several non-specific bands are, suggesting that either the expression system cannot produce full length P2X7, or that the detection method is not suitable.

DNA sequencing screenshots

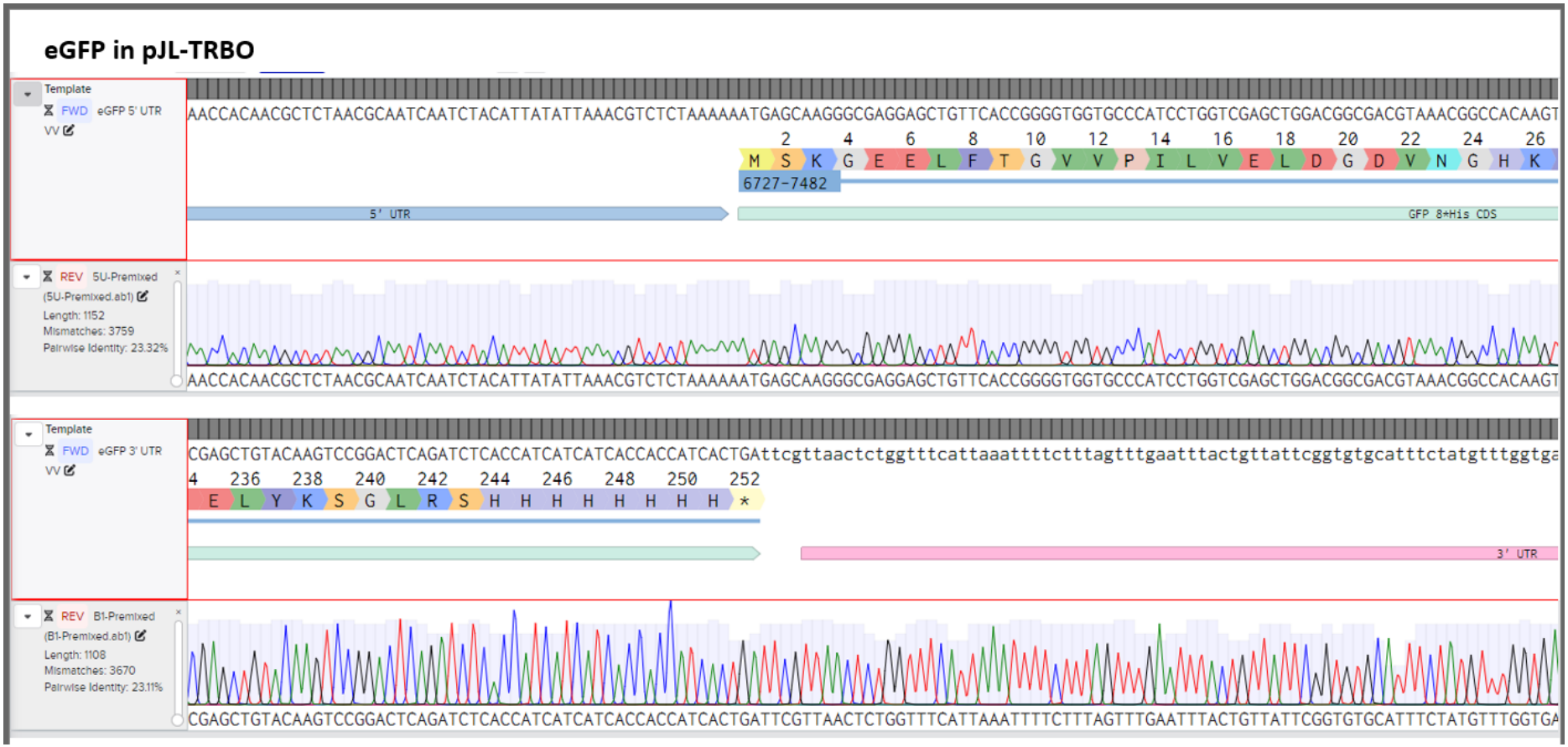
Benchling screenshot showing partial sequencing data of the CaMV 35S promoter and synthetic 5’ UTR. The chromatogram shows an imported ab1 file, with peak height corresponding to resolution, and the quality score is represented by the grey bar behind each peak.



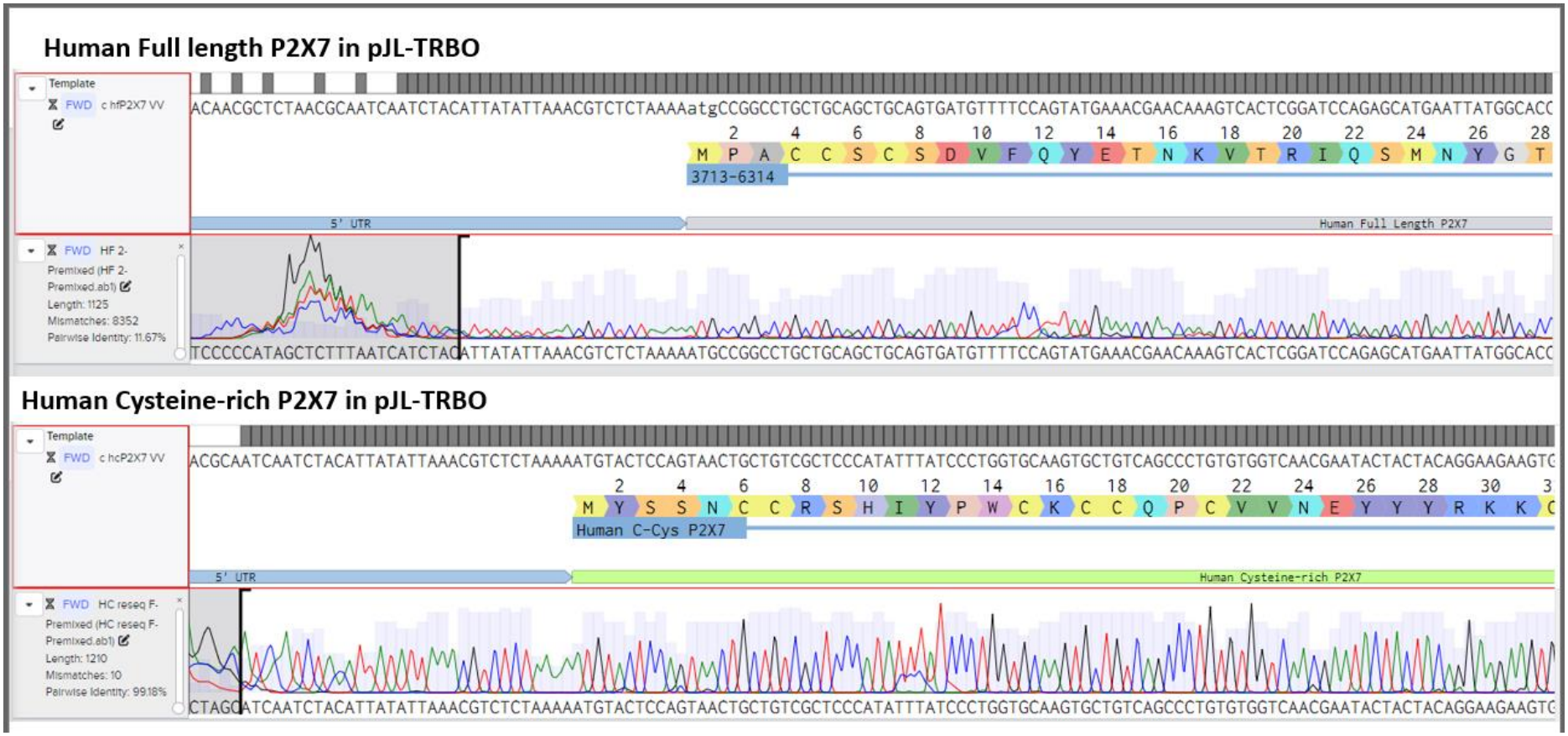
Benchling screenshot showing partial sequencing data of the CPMV 3’ UTR and double terminator. The chromatogram shows an imported ab1 file, with peak height corresponding to resolution, and the quality score is represented by the grey bar behind each peak.



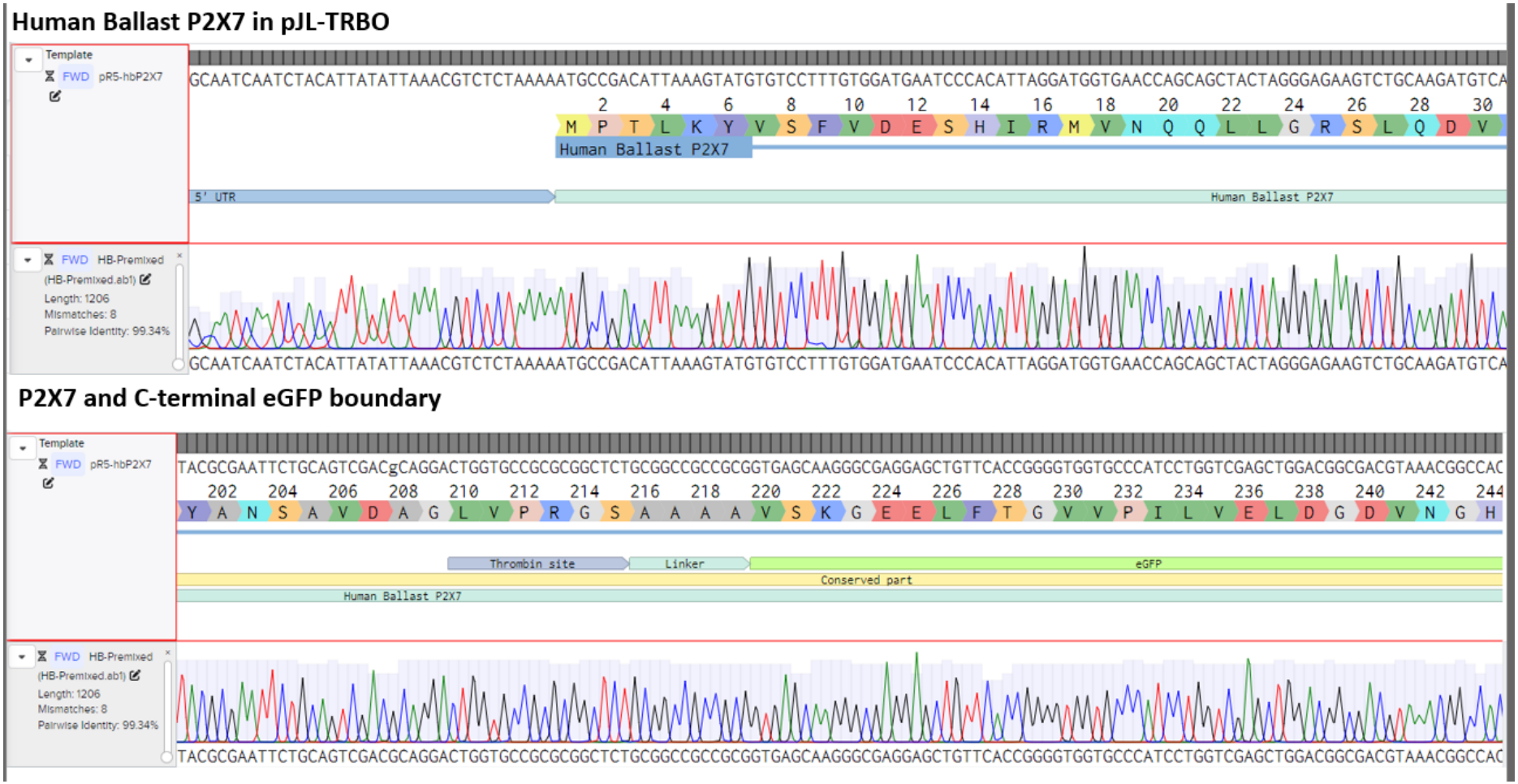
Benchling screenshot showing partial sequencing data of the RB7 matrix attachment region in the Level 0 vector (top) and one pJL-TRBO construct (bottom). The chromatogram shows an imported ab1 file, with peak height corresponding to resolution, and the quality score is represented by the grey bar behind each peak.



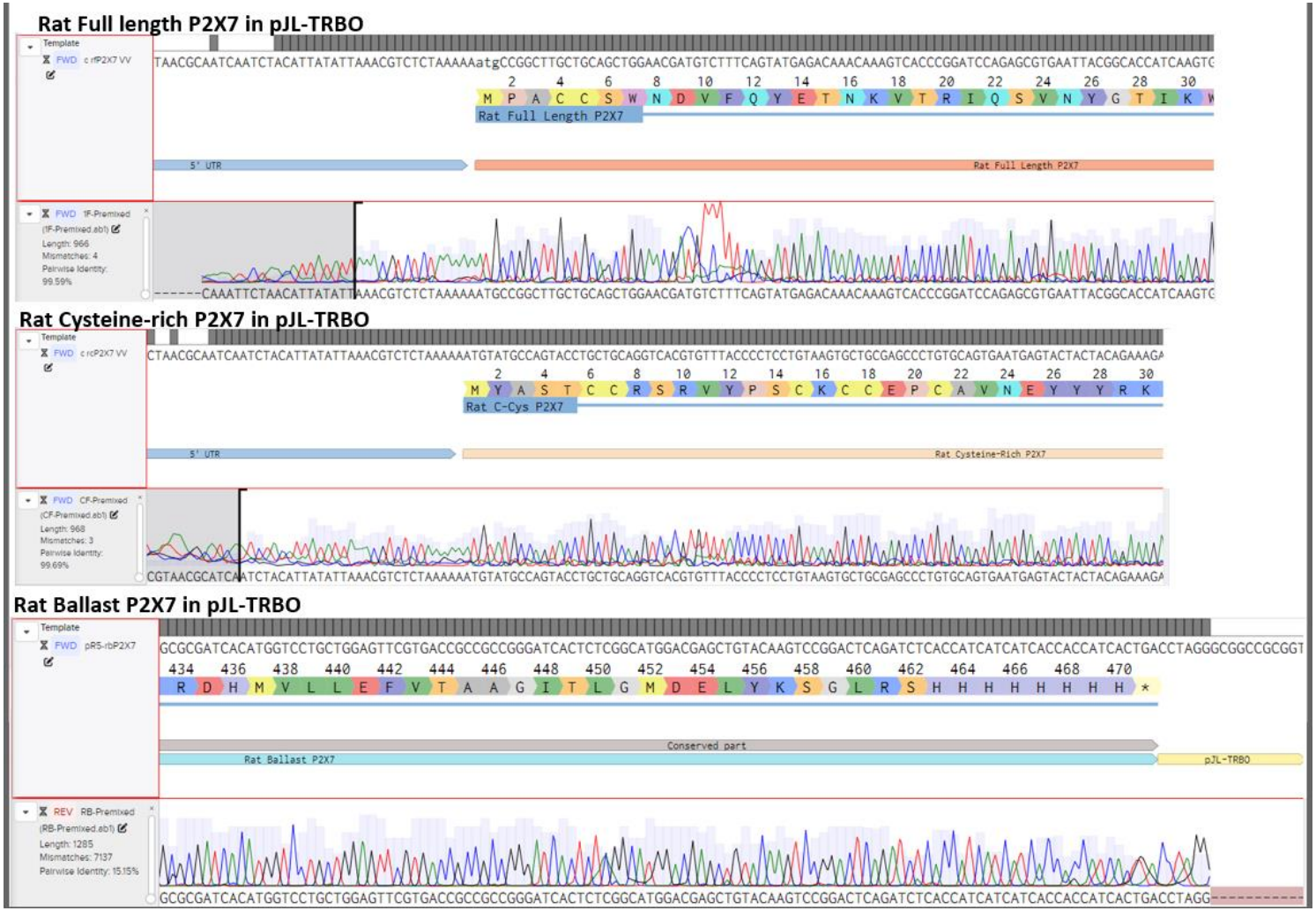
Benchling screenshot showing partial sequencing data of the eGFP coding sequence. The chromatogram shows an imported ab1 file, with peak height corresponding to resolution, and the quality score is represented by the grey bar behind each peak.



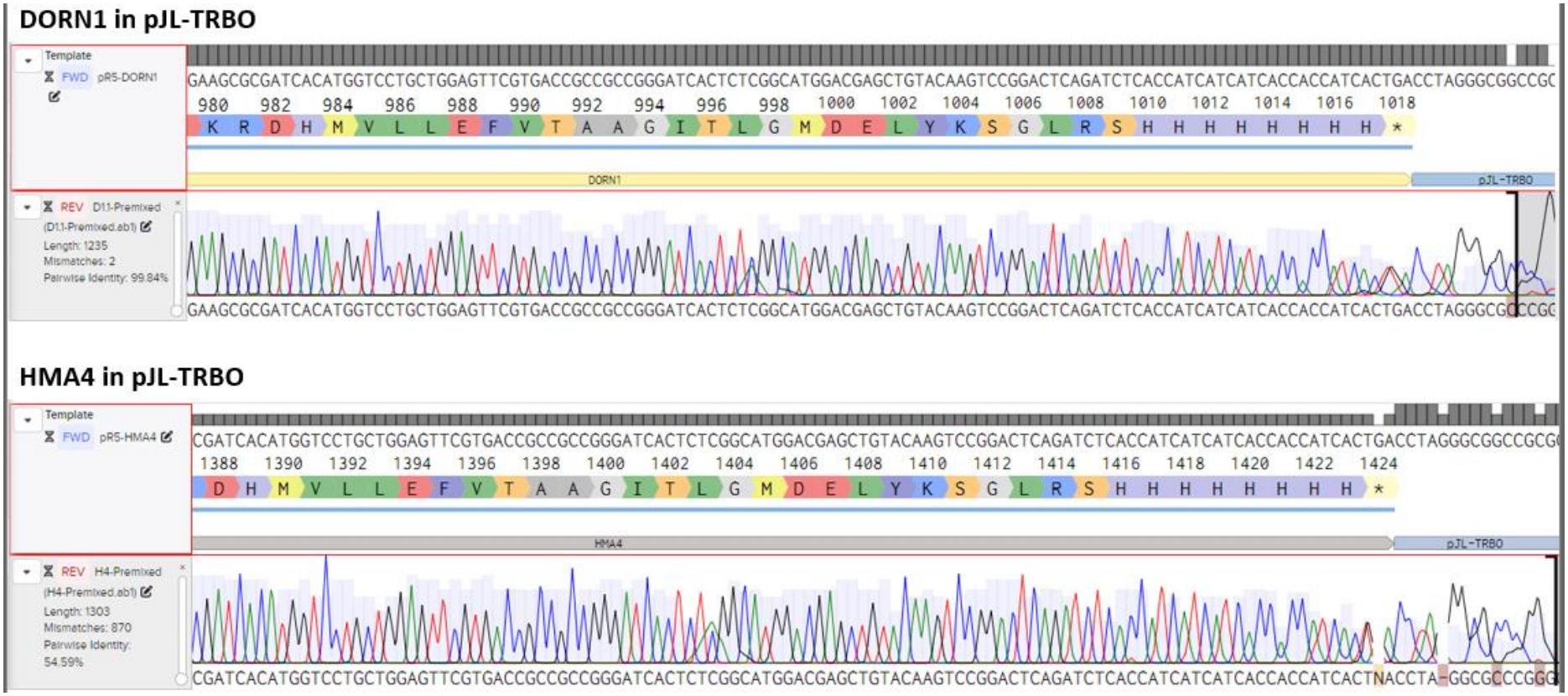
Benchling screenshot showing partial sequencing data of the human full length and cysteine rich P2X7 coding sequences. The chromatogram shows an imported ab1 file, with peak height corresponding to resolution, and the quality score is represented by the grey bar behind each peak.



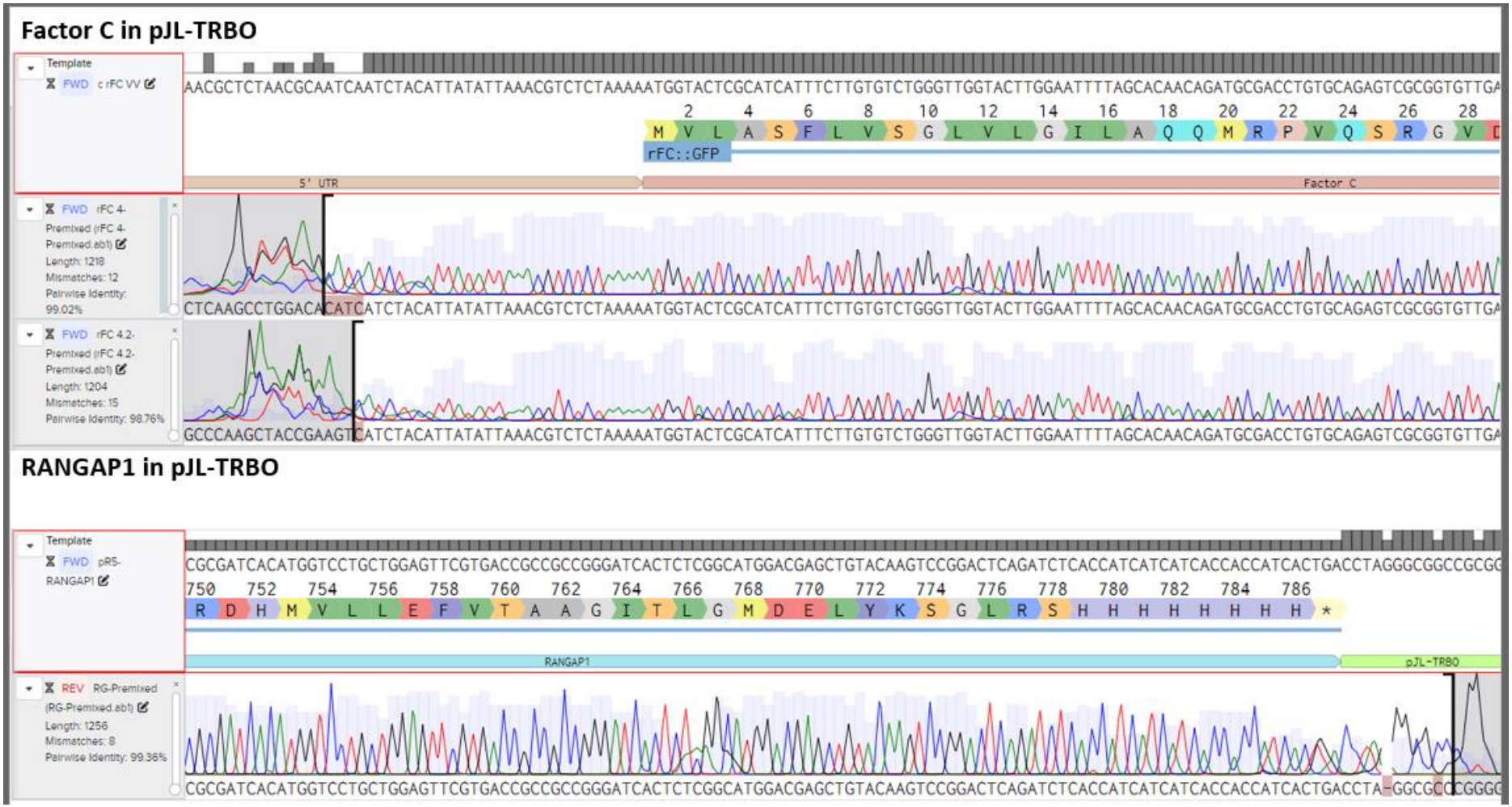
Benchling screenshot showing partial sequencing data of the human ballast P2X7 coding sequence, and showing the boundary between the P2X7 CDS and C-terminal eGFP fusion. The chromatogram shows an imported ab1 file, with peak height corresponding to resolution, and the quality score is represented by the grey bar behind each peak.



Benchling screenshot showing partial sequencing data of the rat P2X7 variants, with full length (top) cysteine rich (middle) and ballast (bottom) coding sequences. The chromatogram shows an imported ab1 file, with peak height corresponding to resolution, and the quality score is represented by the grey bar behind each peak.



Benchling screenshot showing partial sequencing data of DORN1 and HMA4 coding sequences. The chromatogram shows an imported ab1 file, with peak height corresponding to resolution, and the quality score is represented by the grey bar behind each peak.



Benchling screenshot showing partial sequencing data of Factor C and RANGAP1 coding sequences. The chromatogram shows an imported ab1 file, with peak height corresponding to resolution, and the quality score is represented by the grey bar behind each peak.

DNA Sequences

Table S1: DNA sequences used in this research. All synthesised DNA fragments, vectors, parts, oligonucleotides, RT-PCR primers, sequencing primers, and the prefix of cloning primers used in this research are shown in this table.

Name	Usage	Sequence (5' to 3')
pICH47732	Level 1 Vector	TTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGG CTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCC GTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATA CCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGA TAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGA TAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCC CAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCG TGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGA CAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAG GGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGG GTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTC AGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTT ACGGTTCCTGGCAGATCCTAGATGTGGCGCAACGATGCCGGCGAC AAGCAGGAGCGCACCGACTTCTTCCGCATCAAGTGTTTTGGCTCT CAGGCCGAGGCCCACGGCAAGTATTTGGGCAAGGGGTCGCTGGTA TTCGTGCAGGGCAAGATTCGGAATACCAAGTACGAGAAGGACGGC CAGACGGTCTACGGGACCGACTTCATTGCCGATAAGGTGGATTAT CTGGACACCAAGGCACCAGGCGGGTCAAATCAGGAATAAGGGCAC ATTGCCCCGGCGTGAGTCGGGGCAATCCCGCAAGGAGGGTGAATG AATCGGACGTTTGACCGGAAGGCATACAGGCAAGAACTGATCGAC GCGGGGTTTTCCGCCGAGGATGCCGAAACCATCGCAAGCCGCACC GTCATGCGTGCGCCCCGCGAAACCTTCCAGTCCGTCGGCTCGATG GTCCAGCAAGCTACGGCCAAGATCGAGCGCGACAGCGTGCAACTG GCTCCCCCTGCCCTGCCCGCGCCATCGGCCGCCGTGGAGCGTTCG CGTCGTCTTGAACAGGAGGCGGCAGGTTTGGCGAAGTCGATGACC ATCGACACGCGAGGAACTATGACGACCAAGAAGCGAAAAACCGCC GGCGAGGACCTGGCAAAACAGGTCAGCGAGGCCAAGCAGGCCGCG TTGCTGAAACACACGAAGCAGCAGATCAAGGAAATGCAGCTTTCC TTGTTCGATATTGCGCCGTGGCCGGACACGATGCGAGCGATGCCA

		AACGACACGGCCCGCTCTGCCCTGTTCACCACGCGCAACAAGAAA ATCCCGCGCGAGGCGCTGCAAAACAAGGTCATTTTCCACGTCAAC AAGGACGTGAAGATCACCTACACCGGCGTCGAGCTGCGGGCCGAC GATGACGAACTGGTGTGGCAGCAGGTGTTGGAGTACGCGAAGCGC ACCCCTATCGGCGAGCCGATCACCTTCACGTTCTACGAGCTTTGC CAGGACCTGGGCTGGTCGATCAATGGCCGGTATTACACGAAGGCC GAGGAATGCCTGTCGCGCCTACAGGCGACGGCGATGGGCTTCACG TCCGACCGCGTTGGGCACCTGGAATCGGTGTCGCTGCTGCACCGC TTCCGCGTCCTGGACCGTGGCAAGAAAACGTCCCGTTGCCAGGTC CTGATCGACGAGGAAATCGTCGTGCTGTTTGCTGGCGACCACTAC ACGAAATTCATATGGGAGAAGTACCGCAAGCTGTCGCCGACGGCC CGACGGATGTTCGACTATTTCAGCTCGCACCGGGAGCCGTACCCG CTCAAGCTGGAAACCTTCCGCCTCATGTGCGGATCGGATTCCACC CGCGTGAAGAAGTGGCGCGAGCAGGTCGGCGAAGCCTGCGAAGAG TTGCGAGGCAGCGGCCTGGTGGAACACGCCTGGGTCAATGATGAC CTGGTGCATTGCAAACGCTAGGGCCTTGTGGGGTCAGTTCCGGCT GGGGGTTCAGCAGCCAGCGCCTGATCTGGGGAACCCTGTGGTTGG CATGCACATACAAATGGACGAACGGATAAACCTTTTCACGCCCTT TTAAATATCCGATTATTCTAATAAACGCTCTTTTCTCTTAGGTTT ACCCGCCAATATATCCTGTCAAACACTGATAGTTTAAACCACTTC GTGCAGAAGACAATTGCAGCGTGAGACCGTCACAGCTTGTCTGTA AGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGG TGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCA GATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAG ATGCGTAAGGAGAAAATACCGCATCAGGCGCCATTCGCCATTCAG GCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCT ATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAG TTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGAC GGCCAGTGAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGTCGA CCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTC CTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAG CCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCT AACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGG CAAACCTGTCGTGCCAGCTGCGGTCTCACTCCGGATCCGAATTCG GCATTGTCTTCACAGAGTGGGGCCCACTGCATCCACCCCAGTACA TTAAAAACGTCCGCAATGTGTTATTAAGTTGTCTAAGCGTCAATT TGTTTACACCACAATATATCCTGCCACCAGCCAGCCAACAGCTCC CCGACCGGCAGCTCGGCACAAAATCACCACTCGATACAGGCAGCC

		CATCAGTCAGATCAGGATCTCCTTTGCGACGCTCACCGGGCTGGT TGCCCTCGCCGCTGGGCTGGCGGCCGTCTATGGCCCTGCAAACGC GCCAGAAACGCCGTCGAAGCCGTGTGCGAGACACCGCGGCCGCCG GCGTTGTGGATACCTCGCGGAAAACTTGGCCCTCACTGACAGATG AGGGGCGGACGTTGACACTTGAGGGGCCGACTCACCCGGCGCGGC GTTGACAGATGAGGGGCAGGCTCGATTTCGGCCGGCGACGTGGAG CTGGCCAGCCTCGCAAATCGGCGAAAACGCCTGATTTTACGCGAG TTTCCCACAGATGATGTGGACAAGCCTGGGGATAAGTGCCCTGCG GTATTGACACTTGAGGGGCGCGACTACTGACAGATGAGGGGCGCG ATCCTTGACACTTGAGGGGCAGAGTGCTGACAGATGAGGGGCGCA CCTATTGACATTTGAGGGGCTGTCCACAGGCAGAAAATCCAGCAT TTGCAAGGGTTTCCGCCCGTTTTTCGGCCACCGCTAACCTGTCTT TTAACCTGCTTTTAAACCAATATTTATAAACCTTGTTTTTAACCA GGGCTGCGCCCTGTGCGCGTGACCGCGCACGCCGAAGGGGGGTGC CCCCCCTTCTCGAACCCTCCCGGCCCGCTAACGCGGGCCTCCCAT CCCCCCAGGGGCTGCGCCCCTCGGCCGCGAACGGCCTCACCCCAA AAATGGCAGCGCTGGCCAATTCGTGCGCGGAACCCCTATTTGTTT ATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATA ACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAG TATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATT TTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAA AGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACT GGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGA ACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGC GGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCG CATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCAC AGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAG TGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCT GACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAA CATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCT GAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGT AGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACT TACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGA TAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTG GTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGTTCTCGCGG TATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGT AGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAA TAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTA

		ACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAA ACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGA TAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTG AGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCC TTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACC GCTACCAGCGGTGGTTTG
pJL-TRBO	Replicating Vector	TGGCAGGATATATTGTGGTGTAAACAAATTGACGCTTAGACAACT TAATAACACATTGCGGACGTTTTTAAAGTACTGGGGTGGTTTTGG TACCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATAT CGAATTCCTGCAGGTCAACATGGTGGAGCACGACACACTTGTCTA CTCCAAAAATATCAAAGATACAGTCTCAGAAGACCAAAGGGCAAT TGAGACTTTTCAACAAAGGGTAATATCCGGAAACCTCCTCGGATT CCATTGCCCAGCTATCTGTCACTTTATTGTGAAGATAGTGGAAAA GGAAGGTGGCTCCTACAAATGCCATCATTGCGATAAAGGAAAGGC CATCGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACC CCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCAC GTCTTCAAAGCAAGTGGATTGATGTGATAACATGGTGGAGCACGA CACACTTGTCTACTCCAAAAATATCAAAGATACAGTCTCAGAAGA CCAAAGGGCAATTGAGACTTTTCAACAAAGGGTAATATCCGGAAA CCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTTATTGTGAA GATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTGCGA TAAAGGAAAGGCCATCGTTGAAGATGCCTCTGCCGACAGTGGTCC CAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGA CGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATATCTC CACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGA CCCTTCCTCTATATAAGGAAGTTCATTTCATTTGGAGAGGGTATT TTTACAACAATTACCAACAACAACAAACAACAGACAACATTACAA TTACTATTTACAATTACAATGGCATACACACAGACAGCTACCACA TCAGCTTTGCTGGACACTGTCCGAGGAAACAACTCCTTGGTCAAT GATCTAGCAAAGCGTCGTCTTTACGACACAGCGGTTGAAGAGTTT AACGCTCGTGACCGCAGGCCCAAGGTGAACTTTTCAAAAGTAATA AGCGAGGAGCAGACGCTTATTGCTACCCGGGCGTATCCAGAATTC CAAATTACATTTTATAACACGCAAAATGCCGTGCATTCGCTTGCA GGTGGATTGCGATCTTTAGAACTGGAATATCTGATGATGCAAATT CCCTACGGATCATTGACTTATGACATAGGCGGGAATTTTGCATCG CATCTGTTCAAGGGACGAGCATATGTACACTGCTGCATGCCCAAC CTGGACGTTCGAGACATCATGCGGCACGAAGGCCAGAAAGACAGT ATTGAACTATACCTTTCTAGGCTAGAGAGAGGGGGGAAAACAGTC

		CCCAACTTCCAAAAGGAAGCATTTGACAGATACGCAGAAATTCCT GAAGACGCTGTCTGTCACAATACTTTCCAGACATGCGAACATCAG CCGATGCAGCAATCAGGCAGAGTGTATGCCATTGCGCTACACAGC ATATATGACATACCAGCCGATGAGTTCGGGGCGGCACTCTTGAGG AAAAATGTCCATACGTGCTATGCCGCTTTCCACTTCTCCGAGAAC CTGCTTCTTGAAGATTCATGCGTCAATTTGGACGAAATCAACGCG TGTTTTTCGCGCGATGGAGACAAGTTGACCTTTTCTTTTGCATCA GAGAGTACTCTTAATTACTGTCATAGTTATTCTAATATTCTTAAG TATGTGTGCAAAACTTACTTCCCGGCCTCTAATAGAGAGGTTTAC ATGAAGGAGTTTTTAGTCACCAGAGTTAATACCTGGTTTTGTAAG TTTTCTAGAATAGATACTTTTCTTTTGTACAAAGGTGTGGCCCAT AAAAGTGTAGATAGTGAGCAGTTTTATACTGCAATGGAAGACGCA TGGCATTACAAAAAGACTCTTGCAATGTGCAACAGCGAGAGAATC CTCCTTGAGGATTCATCATCAGTCAATTACTGGTTTCCCAAAATG AGGGATATGGTCATCGTACCATTATTCGACATTTCTTTGGAGACT AGTAAGAGGACGCGCAAGGAAGTCTTAGTGTCCAAGGATTTCGTG TTCACAGTGCTTAACCACATTCGAACATACCAGGCGAAAGCTCTT ACATACGCAAATGTTTTGTCCTTCGTCGAATCGATTCGATCGAGG GTAATCATTAACGGTGTGACAGCGAGGTCCGAATGGGATGTGGAC AAATCTTTGTTACAATCCTTGTCCATGACGTTTTACCTGCATACT AAGCTTGCCGTTCTAAAGGATGACTTACTGATTAGCAAGTTTAGT CTCGGTTCGAAAACGGTGTGCCAGCATGTGTGGGATGAGATTTCG CTGGCGTTTGGGAACGCATTTCCCTCCGTGAAAGAGAGGCTCTTG AACAGGAAACTTATCAGAGTGGCAGGCGACGCATTAGAGATCAGG GTGCCTGATCTATATGTGACCTTCCACGACAGATTAGTGACTGAG TACAAGGCCTCTGTGGACATGCCTGCGCTTGACATTAGGAAGAAG ATGGAAGAAACGGAAGTGATGTACAATGCACTTTCAGAATTATCG GTGTTAAGGGAGTCTGACAAATTCGATGTTGATGTTTTTTCCCAG ATGTGCCAATCTTTGGAAGTTGACCCAATGACGGCAGCGAAGGTT ATAGTCGCGGTCATGAGCAATGAGAGCGGTCTGACTCTCACATTT GAACGACCTACTGAGGCGAATGTTGCGCTAGCTTTACAGGATCAA GAGAAGGCTTCAGAAGGTGCATTGGTAGTTACCTCAAGAGAAGTT GAAGAACCGTCCATGAAGGGTTCGATGGCCAGAGGAGAGTTACAA TTAGCTGGTCTTGCTGGAGATCATCCGGAATCGTCCTATTCTAAG AACGAGGAGATAGAGTCTTTAGAGCAGTTTCATATGGCGACGGCA GATTCGTTAATTCGTAAGCAGATGAGCTCGATTGTGTACACGGGT CCGATTAAAGTTCAGCAAATGAAAAACTTTATCGATAGCCTGGTA GCATCACTATCTGCTGCGGTGTCGAATCTCGTCAAGATCCTCAAA

		GATACAGCTGCTATTGACCTTGAAACCCGTCAAAAGTTTGGAGTC TTGGATGTTGCATCTAGGAAGTGGTTAATCAAACCAACGGCCAAG AGTCATGCATGGGGTGTTGTTGAAACCCACGCGAGGAAGTATCAT GTGGCGCTTTTGGAATATGATGAGCAGGGTGTGGTGACATGCGAT GATTGGAGAAGAGTAGCTGTTAGCTCTGAGTCTGTTGTTTATTCC GACATGGCGAAACTCAGAACTCTGCGCAGACTGCTTCGAAACGGA GAACCGCATGTCAGTAGCGCAAAGGTTGTTCTTGTGGACGGAGTT CCGGGCTGTGGAAAAACCAAAGAAATTCTTTCCAGGGTTAATTTT GATGAAGATCTAATTTTAGTACCTGGGAAGCAAGCCGCGGAAATG ATCAGAAGACGTGCGAATTCCTCAGGGATTATTGTGGCCACGAAG GACAACGTTAAAACCGTTGATTCTTTCATGATGAATTTTGGGAAA AGCACACGCTGTCAGTTCAAGAGGTTATTCATTGATGAAGGGTTG ATGTTGCATACTGGTTGTGTTAATTTTCTTGTGGCGATGTCATTG TGCGAAATTGCATATGTTTACGGAGACACACAGCAGATTCCATAC ATCAATAGAGTTTCAGGATTCCCGTACCCCGCCCATTTTGCCAAA TTGGAAGTTGACGAGGTGGAGACACGCAGAACTACTCTCCGTTGT CCAGCCGATGTCACACATTATCTGAACAGGAGATATGAGGGCTTT GTCATGAGCACTTCTTCGGTTAAAAAGTCTGTTTCGCAGGAGATG GTCGGCGGAGCCGCCGTGATCAATCCGATCTCAAAACCCTTGCAT GGCAAGATCCTGACTTTTACCCAATCGGATAAAGAAGCTCTGCTT TCAAGAGGGTATTCAGATGTTCACACTGTGCATGAAGTGCAAGGC GAGACATACTCTGATGTTTCACTAGTTAGGTTAACCCCTACACCG GTCTCCATCATTGCAGGAGACAGCCCACATGTTTTGGTCGCATTG TCAAGGCACACCTGTTCGCTCAAGTACTACACTGTTGTTATGGAT CCTTTAGTTAGTATCATTAGAGATCTAGAGAAACTTAGCTCGTAC TTGTTAGATATGTATAAGGTCGATGCAGGAACACAATAGCAATTA CAGATTGACTCGGTGTTCAAAGGTTCCAATCTTTTTGTTGCAGCG CCAAAGACTGGTGATATTTCTGATATGCAGTTTTACTATGATAAG TGTCTCCCAGGCAACAGCACCATGATGAATAATTTTGATGCTGTT ACCATGAGGTTGACTGACATTTCATTGAATGTCAAAGATTGCATA TTGGATATGTCTAAGTCTGTTGCTGCGCCTAAGGATCAAATCAAA CCACTAATACCTATGGTACGAACGGCGGCAGAAATGCCACGCCAG ACTGGACTATTGGAAAATTTAGTGGCGATGATTAAAAGAAACTTT AACGCACCCGAGTTGTCTGGCATCATTGATATTGAAAATACTGCA TCTTTGGTTGTAGATAAGTTTTTTGATAGTTATTTGCTTAAAGAA AAAAGAAAACCAAATAAAAATGTTTCTTTGTTCAGTAGAGAGTCT CTCAATAGATGGTTAGAAAAGCAGGAACAGGTAACAATAGGCCAG CTCGCAGATTTTGATTTTGTGGATTTGCCAGCAGTTGATCAGTAC

		AGACACATGATTAAAGCACAACCCAAACAAAAGTTGGACACTTCA ATCCAAACGGAGTACCCGGCTTTGCAGACGATTGTGTACCATTCA AAAAAGATCAATGCAATATTCGGCCCGTTGTTTAGTGAGCTTACC AGGCAATTACTGGACAGTGTTGATTCGAGCAGATTTTTGTTTTTC ACAAGAAAGACACCAGCGCAGATTGAGGATTTCTTCGGAGATCTC GACAGTCATGTGCCGATGGATGTCTTGGAGCTGGATATATCAAAA TACGACAAATCTCAGAATGAATTCCACTGTGCAGTAGAATACGAG ATCTGGCGAAGATTGGGTTTCGAAGACTTCTTGGGAGAAGTTTGG AAACAAGGGCATAGAAAGACCACCCTCAAGGATTATACCGCAGGT ATAAAAACTTGCATCTGGTATCAAAGAAAGAGCGGGGACGTCACG ACGTTCATTGGAAACACTGTGATCATTGCTGCATGTTTGGCCTCG ATGCTTCCGATGGAGAAAATAATCAAAGGAGCCTTTTGCGGTGAC GATAGTCTGCTGTACTTTCCAAAGGGTTGTGAGTTTCCGGATGTG CAACACTCCGCGAATCTTATGTGGAATTTTGAAGCAAAACTGTTT AAAAAACAGTATGGATACTTTTGCGGAAGATATGTAATACATCAC GACAGAGGATGCATTGTGTATTACGATCCCCTAAAGTTGATCTCG AAACTTGGTGCTAAACACATCAAGGATTGGGAACACTTGGAGGAG TTCAGAAGGTCTCTTTGTGATGTTGCTGTTTCGTTGAACAATTGT GCGTATTACACACAGTTGGACGACGCTGTATGGGAGGTTCATAAG ACCGCCCCTCCAGGTTCGTTTGTTTATAAAAGTCTGGTGAAGTAT TTGTCTGATAAAGTTCTTTTTAGAAGTTTGTTTATAGATGGCTCT AGTTGTTAAAGGAAAAGTGAATATCAATGAGTTTATCGACCTGAC AAAAATGGAGAAGATCTTACCGTCGATGTTTACCCCTGTAAAGAG TGTTATGTGTTCCAAAGTTGATAAAATAATGGTTCATGAGAATGA GTCATTGTCAGGGGTGAACCTTCTTAAAGGAGTTAAGCTTATTGA TAGTGGATACGTCTGTTTAGCCGGTTTGGTCGTCACGGGCGAGTG GAACTTGCCTGACAATTGCAGAGGAGGTGTGAGCGTGTGTCTGGT GGACAAAAGGATGGAAAGAGCCGACGAGGCCATTCTCGGATCTTA CTACACAGCAGCTGCAAAGAAAAGATTTCAGTTCAAGGTCGTTCC CAATTATGCTATAACCACCCAGGACGCGATGAAAAACGTCTGGCA AGTTTTAGTTAATATTAGAAATGTGAAGATGTCAGCGGGTTTCTG TCCGCTTTCTCTGGAGTTTGTGTCGGTGTGTATTGTTTATAGAAA TAATATAAAATTAGGTTTGAGAGAGAAGATTACAAACGTGAGAGA CGGAGGGCCCATGGAACTTACAGAAGAAGTCGTTGATGAGTTCAT GGAAGATGTCCCTATGTCGATCAGGCTTGCAAAGTTTCGATCTCG AACCGGAAAAAAGAGTGATGTCCGCAAAGGGAAAAATAGTAGTAG TGATCGGTCAGTGCCGAACAAGAACTATAGAAATGTTAAGGATTT TGGGGGAATGAGTTTTAAAAAGAATAATTTAATCGATGATGATTC

		GGAGGCTACTGTCGCCGAATCGGATTCGTTTTAAATAGATCTTAC
		AGTATCACTACTCCATCTCAGTTCGTGTTCTTGTCATTAATTAAC
		GGCCTAGGGCGGCCGCGGTCCTGCAACTTGAGGTAGTCAAGATGC
		ATAATAAATAACGGATTGTGTCCGTAATCACACGTGGTGCGTACG
		ATAACGCATAGTGTTTTTCCCTCCACTTAAATCGAAGGGTTGTGT
		CTTGGATCGCGCGGGTCAAATGTATATGGTTCATATACATCCGCA
		GGCACGTAATAAAGCGAGGGGTTCGAATCCCCCCGTTACCCCCGG
		TAGGGGCCCAGGTACCCGGATGTGTTTTCCGGGCTGATGAGTCCG
		TGAGGACGAAACCTGGCTGCAGGCATGCAAGCTTGGCGTAATCAT
		GGTCATAGCCTAGAGTCCGCAAAAATCACCAGTCTCTCTCTACAA
		ATCTATCTCTCTCTATTTTTCTCCAGAATAATGTGTGAGTAGTTC
		CCAGATAAGGGAATTAGGGTTCTTATAGGGTTTCGCTCATGTGTT
		GAGCATATAAGAAACCCTTAGTATGTATTTGTATTTGTAAAATAC
		TTCTATCAATAAAATTTCTAATTCCTAAAACCAAAATCCAGTGAC
		CTGCAGCCCGGCCGGGGGATCCACTAGCAGATTGTCGTTTCCCGC
		CTTCAGTTTAAACTATCAGTGTTTGACAGGATATATTGGCGGGTA
		AACCTAAGAGAAAAGAGCGTTTATTAGAATAATCGGATATTTAAA
		AGGGCGTGAAAAGGTTTATCCGTTCGTCCATTTGTATGTGCATGC
		CAACCACAGGAGATCTCAGTAAAGCGCTGGCTGCTGAACCCCCAG
		CCGGAACTGACCCCACAAGGCCCTAGCGTTTGCAATGCACCAGGT
		CATCATTGACCCAGGCGTGTTCCACCAGGCCGCTGCCTCGCAACT
		CTTCGCAGGCTTCGCCGACCTGCTCGCGCCACTTCTTCACGCGGG
		TGGAATCCGATCCGCACATGAGGCGGAAGGTTTCCAGCTTGAGCG
		GGTACGGCTCCCGGTGCGAGCTGAAATAGTCGAACATCCGTCGGG
		CCGTCGGCGACAGCTTGCGGTACTTCTCCCATATGAATTTCGTGT
		AGTGGTCGCCAGCAAACAGCACGACGATTTCCTCGTCGATCAGGA
		CCTGGCAACGGGACGTTTTCTTGCCACGGTCCAGGACGCGGAAGC
		GGTGCAGCAGCGACACCGATTCCAGGTGCCCAACGCGGTCGGACG
		TGAAGCCCATCGCCGTCGCCTGTAGGCGCGACAGGCATTCCTCGG
		CCTTCGTGTAATACCGGCCATTGATCGACCAGCCCAGGTCCTGGC
		AAAGCTCGTAGAACGTGAAGGTGATCGGCTCGCCGATAGGGGTGC
		GCTTCGCGTACTCCAACACCTGCTGCCACACCAGTTCGTCATCGT
		CGGCCCGCAGCTCGACGCCGGTGTAGGTGATCTTCACGTCCTTGT
		TGACGTGGAAAATGACCTTGTTTTGCAGCGCCTCGCGCGGGATTT
		TCTTGTTGCGCGTGGTGAACAGGGCAGAGCGGGCCGTGTCGTTTG
		GCATCGCTCGCATCGTGTCCGGCCACGGCGCAATATCGAACAAGG
		AAAGCTGCATTTCCTTGATCTGCTGCTTCGTGTGTTTCAGCAACG
		CGGCCTGCTTGGCCTCGCTGACCTGTTTTGCCAGGTCCTCGCCGG

		CGGTTTTTCGCTTCTTGGTCGTCATAGTTCCTCGCGTGTCGATGG TCATCGACTTCGCCAAACCTGCCGCCTCCTGTTCGAGACGACGCG AACGCTCCACGGCGGCCGATGGCGCGGGCAGGGCAGGGGGAGCCA GTTGCACGCTGTCGCGCTCGATCTTGGCCGTAGCTTGCTGGGCCA TCGAGCCGACGGACTGGAAGGTTTCGCGGGGCGCACGCATGACGG TGCGGCTTGCGATGGTTTCGGCATCCTCGGCGGAAAACCCCGCGT CGATCAGTTCTTGCCTGTATGCCTTCCGGTCAAACGTCCGATTCA TTCACCCTCCTTGCGGGATTGCCCCGACTCACGCCGGGGCAATGT GCCCTTATTCCTGATTTGACCCGCCTGGTGCCTTGGTGTCCAGAT AATCCACCTTATCGGCAATGAAGTCGGTCCCGTAGACCGTCTGGC CGTCCTTCTCGTACTTGGTATTCCGAATCTTGCCCTGCACGAATA CCAGCGACCCCTTGCCCAAATACTTGCCGTGGGCCTCGGCCTGAG AGCCAAAACACTTGATGCGGAAGAAGTCGGTGCGCTCCTGCTTGT CGCCGGCATCGTTGCGCCACATCTAGGTACTAAAACAATTCATCC AGTAAAATATAATATTTTATTTTCTCCCAATCAGGCTTGATCCCC AGTAAGTCAAAAAATAGCTCGACATACTGTTCTTCCCCGATATCC TCCCTGATCGACCGGACGCAGAAGGCAATGTCATACCACTTGTCC GCCCTGCCGCTTCTCCCAAGATCAATAAAGCCACTTACTTTGCCA TCTTTCACAAAGATGTTGCTGTCTCCCAGGTCGCCGTGGGAAAAG ACAAGTTCCTCTTCGGGCTTTTCCGTCTTTAAAAAATCATACAGC TCGCGCGGATCTTTAAATGGGGTGTCTTCTTCCCAGTTTTCGCAA TCCACATCGGCCAGATCGTTATTCAGTAAGTAATCCAATTCGGCT AAGCGGCTGTCTAAGCTATTCGTATAGGGACAATCCGATATGTCG ATGGAGTGAAAGAGCCTGATGCACTCCGCATACAGCTCGATAATC TTTTCAGGGCTTTGTTCATCTTCATACTCTTCCGAGCAAAGGACG CCATCGGCCTCACTCATGAGCAGATTGCTCCAGCCATCATGCCGT TCAAGGTGCAGGACCTTTGGAACAGGCAGCTTTCCTTCCAGCCAT AGCATCATGTCCTTTTCCCGTTCCACATCATAGGTGGTCCCTTTA TACCGGCTGTCCGTCATTTTTAAATATAGGTTTTCATTTTCTCCC ACCAGCTTATATACCTTAGCAGGAGACATTCCTTCCGTATCTTTT ACGCAGCGGTATTTTTCGATCAGTTTTTTCAATTCCGGTGATATT CTCATTTTAGCCATTTATTATTTCCTTCCTCTTTTCTACAGTATT TAAAGATACCCCAAGAAGCTAATTATAACAAGACGAACTCCAATT CACTGTTCCTTGCATTCTAAAACTTTAAATACCAGAAAACAGCTT TTTCAAAGTTGTTTTCAAAGTTGGCGTATAACATAGTATCGACGG AGCCGATTTTGAAACCACAATTATGGGTGATGCTGCAACTCGAGC GGGCCGGGAGGGTTCGAGAAGGGGGGGCACCCCCCTTCGGCGTGC GCGGTCACGCGCACAGGGCGCAGCCCTGGTTAAAAACAAGGTTTA

		TAAATATTGGTTTAAAAGCAGGTTAAAAGACAGGTTAGCGGTGGC CGAAAAACGGGCGGAAACCCTTGCAAATGCTGGATTTTCTGCCTG TGGACAGCCCCTCAAATGTCAATAGGTGCGCCCCTCATCTGTCAG CACTCTGCCCCTCAAGTGTCAAGGATCGCGCCCCTCATCTGTCAG TAGTCGCGCCCCTCAAGTGTCAATACCGCAGGGCACTTATCCCCA GGCTTGTCCACATCATCTGTGGGAAACTCGCGTAAAATCAGGCGT TTTCGCCGATTTGCGAGGCTGGCCAGCTCCACGTCGCCGGCCGAA ATCGAGCCTGCCCCTCATCTGTCAACGCCGCGCCGGGTGAGTCGG CCCCTCAAGTGTCAACGTCCGCCCCTCATCTGTCAGTGAGGGCCA AGTTTTCCGCGAGGTATCCACAACGCCGGCGGCCGGCCGCGGTGT CTCGCACACGGCTTCGACGGCGTTTCTGGCGCGTTTGCAGGGCCA TAGACGGCCGCCAGCCCAGCGGCGAGGGCAACCAGCCCGGTGAGC GTCTCTAGTGGACTGATGGGCTGCCTGTATCGAGTGGTGATTTTG TGCCGAGCTGCCGGTCGGGGAGCTGTTGGCTGGCTGG
CaMV 35S Long promoter	GoldenGate part	TTAATTAAGGAGGAATTCCAATCCCACAAAAATCTGAGCTTAACA GCACAGTTGCTCCTCTCAGAGCAGAATCGGGTATTCAACACCCTC ATATCAACTACTACGTTGTGTATAACGGTCCACATGCCGGTATAT ACGATGACTGGGGTTGTACAAAGGCGGCAACAAACGGCGTTCCCG GAGTTGCACACAAGAAATTTGCCACTATTACAGAGGCAAGAGCAG CAGCTGACGCGTACACAACAAGTCAGCAAACAGACAGGTTGAACT TCATCCCCAAAGGAGAAGCTCAACTCAAGCCCAAGAGCTTTGCTA AGGCCCTAACAAGCCCACCAAAGCAAAAAGCCCACTGGCTCACGC TAGGAACCAAAAGGCCCAGCAGTGATCCAGCCCCAAAAGAGATCT CCTTTGCCCCGGAGATTACAATGGACGATTTCCTCTATCTTTACG ATCTAGGAAGGAAGTTCGAAGGTGAAGGTGACGACACTATGTTCA CCACTGATAATGAGAAGGTTAGCCTCTTCAATTTCAGAAAGAATG CTGACCCACAGATGGTTAGAGAGGCCTACGCAGCAAGTCTCATCA AGACGATCTACCCGAGTAACAATCTCCAGGAGATCAAATACCTTC CCAAGAAGGTTAAAGATGCAGTCAAAAGATTCAGGACTAATTGCA TCAAGAACACAGAGAAAGACATATTTCTCAAGATCAGAAGTACTA TTCCAGTATGGACGATTCAAGGCTTGCTTCATAAACCAAGGCAAG TAATAGAGATTGGAGTCTCTAAAAAGGTAGTTCCTACTGAATCTA AGGCCATGCATGGAGTCTAAGATTCAAATCGAGGATCTAACAGAA CTCGCCGTCAAGACTGGCGAACAGTTCATACAGAGTCTTTTACGA CTCAATGACAAGAAGAAAATCTTCGTCAACATGGTGGAGCACGAC ACTCTGGTCTACTCCAAAAATGTCAAAGATACAGTCTCAGAAGAT CAAAGGGCTATTGAGACTTTTCAACAAAGGATAATTTCGGGAAAC CTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTCATCGAAAGG

		ACAGTAGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTGCGAT AAAGGAAAGGCTATCATTCAAGATCTCTCTGCCGACAGTGGTCCC AAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGAG GTTCCAACCACGTCTACAAAGCAAGTGGATTGATGTGACATCTCC ACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGAC CCTTCCTCTATATAA
GeneWiz 5' UTR	Synthesised DNA Fragment	TTTAAGAGACGCAACCACAACGCTCTAACGCAATCAATCTACATT ATATTAAACGTCTCTAAAA
GeneWiz 3' UTR	Synthesised DNA Fragment	TTAACTCTGGTTTCATTAAATTTTCTTTAGTTTGAATTTACTGTT ATTCGGTGTGCATTTCTATGTTTGGTGAGCGGTTTTCTGTGCTCA GAGTGTGTTTATTTTATGTAATTTAATTTCTTTGTGAGCTCCTGT TTAGCAGGTCGTCCCTTCAGCAAGGACACAAAAAGATTTTAATTT TATT
GeneWiz EU terminator	Synthesised DNA Fragment	AAAGCAGAATGCTGAGCTAAAAGAAAGGCTTTTTCCATTTTCGAG AGACAATGAGAAAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG AAAAGAGTAAATAATAAAGCCCCACAGGAGGCGAAGTTCTTGTAG CTCCATGTTATCTAAGTTATTGATATTGTTTGCCCTATATTTTAT TTCTGTCATTGTGTATGTTTTGTTCAGTTTCGATCTCCTTGCAAA ATGCAGAGATTATGAGATGAATAAACTAAGTTATATTATTATACG TGTTAATATTCTCCTCCTCTCTCTAGCTAGCCTTTTGTTTTCTCT TTTTCTTATTTGATTTTCTTTAAATCAATCCATTTTAGGAGAGGG CCAGGGAGTGATCCAGCAAAACATGAAGATTAGAAGAAACTTCCC TCTTTTTTTTCCTGAAAACAATTTAACGTCGAGATTTATCTCTTT TTGTAATGGAATCATTTCTACAGTTATGAC
GeneWiz Actin terminator	Synthesised DNA Fragment	ATACAGCATTCCCAGAAAGAGAAACAGAAGAAATATACAAACTTT CATTTTGAGAGCAGCACCTCGTCTATTGATTGCAGATAATATGCT TCTCATTTGTATTTCCTTTTGATTATTTTTGTTTCTATCCCTTTG TTTGAGTCAATCTCAAATATTCGGTCATTGTTGGTATGAAAAATC AAGCAGTTCATGTTAAGAGTCAATTTAAAATTAATATTTTTATAT AGAGTTGTATGTGAAATGATGTTGTGATTTGGTATATATGGATAA AGAGCTTGTCAGTTCATTTTGGTCTCATTTTTTTGGTATCCAAAT AAGAAACACAAAAGGGATATGTCCCTCTACTATCAAATATTAGTT ATAAGTATTCATGTTATACTATTCGATATTTTCTACCCCAATCGT TACCTATTTAAAAGTATTTACCCCTCCATCTATCAAACCCCTGGA CCCAGCTTTCCTATTACATGTGGCTTCATCTTAAGCCCCCAAACC TTTTTCTTATTTTTGATTTTTAAAGGCTCATCTTAAAATTTATTA CTCAAATTAATACCTCTTAATAACCCACCTCAAGGACCCAGTAAT

		TAAATATCCAATTAGCTCCAGTAATTGGGGTTCATATTAGCTCCA GTCTTAAATTTTAAAGGCGATGATCGTATTCCTCCACTTGGTTCA TTTATACTCAAAGAATACTCAATGTCTTTAGTGTTTAGATAACTT TTTGTAAATCATATAGATTGTTTTAACAAAAAACAATTCAATAGT AGATTTTCACATGAAAGTTACATAAAAATTCTTTAAAATTACTTT CTCAAAAAATTGTTCCAAACATATTATCCCACAATTAAACTCAAT CTGTTTTTCGAAACCTAAATCAAAACCAATCCAACTACCTTATAT AATATATAATCAATACATTGTAAAGAACTGCATGTTCTTTTAAAT TTTGGGGGCAAAGTTATTCCGTACGTTCACACATGTACTAATAGG AGGTAATAAATGATATGTGAAACAATCGAGGTGTAAACAAGCTAG CAT
GeneWiz RB7 MAR	Synthesised DNA Fragment	GTGAAGACATGGTATCGATTAAAAATCCCAATTATATTTGGTCTA ATTTAGTTTGGTATTGAGTAAAACAAATTCGAACCAAACCAAAAT ATAAATATATAGTTTTTATATATATGCCTTTAAGACTTTTTATAG AATTTTCTTTAAAAAATATCTAGAAATATTTGCGACTCTTCTGGC ATGTAATATTTCGTTAAATATGAAGTGCTCCATTTTTATTAACTT TAAATAATTGGTTGTACGATCACTTTCTTATCAAGTGTTACTAAA ATGCGTCAATCTCTTTGTTCTTCCATATTCATATGTCAAAATCTA TCAAAATTCTTATATATCTTTTTCGAATTTGAAGTGAAATTTCGA TAATTTAAAATTAAATAGAACATATCATTATTTAGGTATCATATT GATTTTTATACTTAATTACTAAATTTGGTTAACTTTGAAAGTGTA CATCAACGAAAAATTAGTCAAACGACTAAAATAAATAAATATCAT GTGTTATTAAGAAAATTCTCCTATAAGAATATTTTAATAGATCAT ATGTTTGTAAAAAAAATTAATTTTTACTAACACATATATTTACTT ATCAAAAATTTGACAAAGTAAGATTAAAATAATATTCATCTAACA AAAAAAAAACCAGAAAATGCTGAAAACCCGGCAAAACCGAACCAA TCCAAACCGATATAGTTGGTTTGGTTTGATTTTGATATAAACCGA ACCAACTCGGTCCATTTGCACCCCTAATCATAATAGCTTTAATAT TTCAAGATATTATTAAGTTAACGTTGTCAATATCCTGGAAATTTT GCAAAATGAATCAAGCCTATATGGCTGTAATATGAATTTAAAAGC AGCTCGATGTGGTGGTAATATGTAATTTACTTGATTCTAAAAAAA TATCCCAAGTATTAATAATTTCTGCTAGGAAGAAGGTTAGCTACG ATTTACAGCAAAGCCAGAATACAAAGAACCATAAAGTGATTGAAG CTCGAAATATACGAAGGAACAAATATTTTTAAAAAAATACGCAAT GACTTGGAACAAAAGAAAGTGATATATTTTTTGTTCTTAAACAAG CATCCCCTCTAAAGAATGGCAGTTTTCCTTTGCATGTAACTATTA TGCTCCCTTCGTTACAAAAATTTTGGACTACTATTGGGAACTTCT TCTGAAAATAGTGCCTAGGCGCTTTGTCTTCAC

GeneWiz recombinant Factor C CDS	Synthesised DNA Fragment	GTGAAGACATAATGGTACTCGCATCATTTCTTGTGTCTGGGTTGG TACTTGGAATTTTAGCACAACAGATGCGACCTGTGCAGAGTCGCG GTGTTGATTTGGGCTTGTGCGACGAAACAAGGTTTGAGTGTAAAT GCGGGGATCCAGGGTACGTATTCAACGTTCCAATGAAGCAATGTA CTTACTTTTATCGTTGGCGACCTTACTGCAAGCCTTGCGACGATC TTGAGGCTAAAGACATTTGTCCTAAGTATAAGAGATGTCAGGAAT GCAAAGCTGGTTTGGATTCCTGCGTTACTTGTCCACCTAACAAAT ACGGAACATGGTGCTCGGGGGAATGCCAATGTAAGAACGGTGGAA TATGTGACCAAAGAACAGGTGCATGTACATGTAGAGATAGGTATG AGGGAGCGCATTGTGAGATCCTAAAGGGGTGCCCACTCTTGCCCT CCGATTCACAGGTACAAGAGGTGAGAAACCCTCCTGATAACCCAC AAACAATCGACTATAGTTGCTCACCTGGATTTAAACTAAAAGGGG TTGCTAGAATCTCCTGTCTGCCAAATGGACAATGGAGTAGCTTCC CTCCGAAGTGTATCAGAGAATGCGCCAAGGTGTCCAGCCCCGAAC ACGGCAAGGTAAACGCTCCATCTGGCAACATGATTGAAGGTGCGA CCTTACGATTCTCTTGCGATAGCCCGTACTATCTAATAGGCCAGG AAACTCTAACTTGTCAAGGCAACGGGCAGTGGTCGGGACAAATAC CACAGTGTAAGAAGCTCGTTTTCTGTCCAGACCTCGACCCGGTTA ATCACGCAGAACATCAAGTTAAGATAGGTGTAGAGCAGAAATATG GACAATTTCCACAGGGGACGGAAGTTACATATACCTGTAGTGGAA ACTACTTTTTAATGGGTTTTAATACACTTAAATGTAACCCTGATG GATCGTGGTCCGGATCTCAACCCAGCTGTGTCAAGGTTGCTGATA GAGAAGTTGACTGTGATTCAAAAGCTGTTGACTTCCTAGATGACG TCGGCGAACCGGTAAGAATCCACTGTCCTGCCGGATGTAGCTTAA CTGCAGGAACGGTCTGGGGTACTGCGATCTACCATGAGCTTTCCA GTGTTTGTCGTGCCGCTATCCATGCAGGAAAACTGCCAAATTCGG GTGGTGCAGTTCATGTTGTGAATAACGGACCTTACTCTGATTTTC TTGGAAGCGATTTGAATGGAATTAAGAGTGAAGAACTGAAATCTT TAGCACGGTCTTTTAGATTCGATTATGTGAGTTCAAGCACTGCTG GAAGGTCTGGATGTCCTGACGGGTGGTTCGAAGTGGAGGAAAATTG CGTTTACGTTACATCAAAACAACGGGCCTGGGAGCGTGCGCAGGGT GTGTGCACCAATATGGCAGCCAGACTTGCCGTTTTGGATAAGGATC TTATTCCCAGTTCACTCACCGAAACTTTGCGCGGAAAGGGATTAAC TACAACATGGATTGGTTTGCACAGGTTGGATGCCGAAAAACCGTTT GTGTGGGAGCTGATGGATAGGTCTAATGTCGTTTTAAATGATAATC TCACCTTCTGGGCCAGCGGAGAACCTGGAAATGAGACGAACTGTGT CTACTTGGATATTCGCGATCAACTCCAACCAGTGTGGAAAACTAAG TCGTGTTTTCAGCCTTCTTCATTCGCATGTATGATGGATTTGAGTG

		ATCGCAACAAAGCTAAATGCGATGATCCGGGTCCTCTTGAGAACGG CCATGCTACACTTCATGGTCAGAGTATCGACGGTTTCTATGCGGGA TCTTCAATAAGATATAGTTGCGAAGTCCTTCATTATCTGTCCGGAA CTGAGACGGTTACATGCACAACTAATGGTACTTGGTCTGCTCCTAA ACCTAGATGTATTAAGGTTATTACTTGTCAAAACCCTCCTGTTCCC TCTTACGGTTCCGTTGAAATAAAGCCACCGTCACGTACCAACTCGA TATCTAGAGTGGGATCGCCTTTTCTGCGGCTTCCGAGACTCCCACT TCCGTTGGCTAGGGCTGCTAAGCCGCCACCTAAACCTCGGTCGTCC CAGCCCAGTACCGTCGATCTCGCTTCTAAGGTCAAGTTACCAGAAG GTCATTATAGGGTGGGGAGTAGGGCAATTTATACCTGTGAATCTCG TTATTATGAGCTTCTTGGTTCTCAAGGACGTAGATGTGACTCCAAT GGCAACTGGTCTGGACGACCAGCAAGCTGCATTCCAGTGTGTGGTC GGTCAGACAGCCCGCGTAGCCCTTTTATCTGGAACGGGAATTCAAC AGAGATCGGTCAGTGGCCTTGGCAAGCGGGGATTTCTAGGTGGTTG GCTGACCACAATATGTGGTTTTTGCAATGTGGCGGTAGTCTACTTA ACGAGAAGTGGATCGTGACAGCAGCGCACTGCGTGACTTATTCCGC TACCGCAGAGATTATCGATCCATCTCAATTCAAAATTTACCTCGGA AAATATTACCGAGACGATTCAAGAGATGATGATTATGTCCAGGTTC GTGAAGCCCTCGAAATCCATGTCAATCCAAATTACGACCCAGGAAA TCTTAATTTCGATATCGCTCTAATTCAACTTAAAACGCCAGTTACC CTGACGACTCGAGTCCAACCTATCTGCCTCCCTACTGATATAACTA CAAGGGAGCACTTAAAGGAGGGGACATTGGCTGTAGTAACTGGGTG GGGCCTTAATGAGAACAATACATACTCAGAGATGATACAACAGGCA GTCCTACCGGTTGTAGCTGCTTCTACCTGTGAGGAAGGCTACAAAG AAGCGGATTTGCCCTTAACGGTGACGGAAAACATGTTCTGCGCTGG CTATAAAAAAGGAAGATACGATGCTTGTTCGGGTGATTCTGGTGGT CCACTGGTGTTTGCTGATGACTCTAGGACTGAGAGACGTTGGGTCC TCGAGGGTATTGTTTCATGGGGTTCACCCTCAGGTTGCGGAAAGGC TAACCAGTACGGTGGTTTTACGAAGGTGAATGTTTTCCTTTCTTGG ATTAGACAGTTCATATGATTCGTTGTCTTCAC
GeneWiz DORN1 CDS	Synthesised DNA Fragment	GTGAAGACATAATGGCTCGTTGGTTGCTTCAGATCCTGATCATCTC TTCTCTTCATCTGAGCTCTGTATCAAGTCAACAAGAGACAAGCTTT GTCTATGAAAGCTTCCTCGACCGACAAAATCTTTATCTTGACAAAT CTGCAATAGTACTTCCCAGTGGATTATTGCAGTTGACAAATGCTTC AGAGCACCAGATGGGTCATGCTTTCCACAAGAAACCAATTGAGTTC AGTTCATCTGGACCACTCTCCTTCTCAACACACTTTGTGTGTGCTT TGGTGCCTAAGCCAGGTTTTGAAGGTGGCCATGGTATTGTCTTTGT ATTGTCTCCTTCTATGGACTTTACTCACGCAGAATCAACTAGATAC

		TTGGGGATTTTCAATGCTTCAACAAATGGATCTTCCTCTTATCACG
		TACTTGCTGTTGAGCTTGACACTATTTGGAACCCAGATTTCAAAGA
		TATTGACCACAATCATGTGGGGATTGATGTGAACAGTCCTATATCT
		GTCGCAATAGCTTCAGCATCTTACTATTCCGACATGAAAGGGAGTA
		ATGAAAGCATAAACCTCTTGAGTGGAAACCCTATACAGGTCTGGGT
		GGATTATGAAGGCACTTTGCTCAATGTTTCGGTTGCTCCTCTTGAA
		GTCCAGAAGCCAACTCGACCTCTTTTATCACACCCCATCAACCTTA
		CAGAACTTTTTCCAAATAGATCAAGTCTGTTTGCTGGGTTCTCTGC
		AGCAACAGGGACAGCGATCAGTGATCAATATATTCTCTGGTGGAGT
		TTTAGCATAGACAGAGGATCGCTGCAGCGACTTGATATCTCAAAAC
		TTCCTGAAGTTCCTCATCCTAGAGCTCCACATAAGAAAGTATCTAC
		ACTGATTATTCTCCTACCCGTTTGTCTGGCTATTTTGGTGTTGGCT
		GTTCTTGCAGGACTTTATTTCCGCAGGAGACGCAAGTATTCAGAAG
		TATCTGAAACATGGGAAAAGGAGTTTGATGCGCATCGGTTCTCTTA
		CAGGTCCTTGTTCAAAGCAACAAAGGGGTTCAGCAAAGATGAATTT
		CTGGGAAAAGGAGGTTTTGGAGAAGTTTATAGAGGAAATCTCCCAC
		AAGGCAGAGAAATAGCAGTGAAAAGAGTGTCCCATAATGGTGATGA
		AGGTGTGAAGCAGTTTGTGGCTGAAGTCGTGAGCATGAGATGTCTG
		AAACACAGGAATCTTGTACCACTGTTTGGGTATTGCAGAAGAAAAC
		GAGAACTTCTTCTTGTCTCAGAGTACATGCCCAATGGTAGTCTTGA
		CGAGCATTTGTTTGATGACCAGAAACCGGTTCTTTCTTGGTCACAG
		AGACTTGTGGTCGTTAAGGGTATAGCCTCTGCTCTCTGGTACCTGC
		ATACAGGTGCTGACCAAGTTGTTCTGCACCGAGATGTCAAAGCTTC
		CAACATTATGTTAGACGCGGAGTTCCATGGAAGGTTAGGGGATTTT
		GGCATGGCCAGGTTTCATGAGCACGGAGGCAATGCAGCTACAACAG
		CAGCTGTTGGGACTGTAGGGTACATGGCCCCGGAGCTAATCACAAT
		GGGAGCTTCCACTGGAACTGATGTTTACGCATTTGGTGTTTTCATG
		CTTGAAGTAACCTGTGGGAGGAGGCCCGTGGAACCGCAGTTGCAAG
		TTGAGAAACGACATATGATCAAATGGGTTTGTGAGTGCTGGAAAAA
		GGATTCTTTGCTTGATGCTACTGACCCGAGATTGGGAGGTAAATTT
		GTAGCTGAGGAAGTCGAGATGGTTATGAAACTGGGTCTGCTCTGTT
		CAAATATAGTGCCAGAATCCAGACCCACAATGGAACAAGTGGTCCT
		GTACCTAAACAAAAACCTACCTTTGCCAGATTTCTCACCATACACT
		TTGGGGATTGGCACATTTGCTCCGGTTCTGGTAGATGCATCCTCCC
		TTGTAGTCTCTTCCGCTTCATGGAGTTTGTCAGGTCCATCGATGTC
		ATCATCCTCGCCCAACCACTCACCTTACGCTTGGCAGTCAACAGAT
		CAGCCATGGGGACAGACAATAGACACTAAGAACTCTCTTCACATAG
		TTGCAGAACCAGAGAAGCCCTCACCCGCTGTCAAAATGGTCACATT

		GCCTGCAGAAGATCCTCAGTCCAATCATTCCAGCATCAGCAGTCAG AGAGTCCAGCCTGTAAAGAGGGAAAAACGGCGACTGCATCAGATAC TGGTGGCATTCCCTTGGATAAACAAACAATACTTTAAACTTGGATT ACCAAAACATATTGTACATGTTTCTTTGTTCTTTTTTCTTCAGCTT GCGAGGTTATGATTCGTTGTCTTCGT
GeneWiz BAG4 CDS	Synthesised DNA Fragment	GTGAAGACATAATGATGCATAATTCAACCGAAGAATCGGAATGGGA GGTTAGACCTGGTGGTATGCTCGTCCAACGAAGAGACGACGCCGCT TCCTCCGACCACAAACCTCTCCAGGATCCCGACTCTGCTTCCGCCG CTTTTGCTCAAACCATCAGAATCACTGTTTCTCATGGCTCATCGCA CCACGATCTTCATATTTCTGCTCACGCCACTTTCGGGGATGTGAAG AAAGCTCTTGTTCAGAAAACTGGATTGGAAGCTAGTGAATTGAAGA TCTTGTTCAGAGGAGTTGAGAGAGATGATGCTGAACAATTGCAAGC TGCTGGTGTTAAGGATGCGTCTAAGCTTGTTGTTGTTGTTGAGGAT ACGAATAAGAGAGTGGAACAACAGCCTCCTGTGGTAACTAAAGAGA TGGAAAAAGCTATTGCTGCTGTTAACGCGGTTACAGGAGAGGTCGA TAAGCTCTCGGATAGAGTTGTTGCTTTAGAAGTTGCTGTGAATGGA GGGACGCAAGTTGCGGTGCGGGAGTTTGACATGGCTGCAGAGCTTC TTATGAGGCAGCTGCTCAAATTGGATGGCATTGAGGCTGAAGGGGA CGCTAAAGTACAGCGTAAGGCTGAGGTACGTAGAATCCAAAACTTG CAGGAGGCTGTGGATAAGTTGAAGGCAAGATGTTCAAATCCGTTTG TGGATCAGAGCAAAGCTGCAGCTGTAAGCACTGAGTGGGAATCGTT CGGAAACGGTGTCGGAAGCTTGAACCCTCCTCCGCCAGCTTCTCCT TCGGCCAATGTAACTCAAGATTGGGAGAAATTTGACTGAGCTTTTG TCTTCTG
GeneWiz REP CDS	Synthesised DNA Fragment	GTGAAGACATAATGCCTTCAGCTCCACAGAAAACAAAATCGTTTCG ACTCCAGACCAAGTATGTTTTCCTTACTTACCCCAGATGTTCATCT TCTGCTGAAAATCTCCGTGATTTTCTGTGGGATAAGCTCTCAAGAT TCGCTATCTTTTTTATAGCAATCGCAACTGAACTTCATCAGGACGG GACACCTCACTTGCATTGTCTAATTCAATTGGATAAAAGATCTAAT ATAAGAGATCCTTCGTTCTTCGACCTAGAGGGAAACCACCCTAACA TCCAACCAGCTAAAAACAGTGAACAGGTGTTGGAGTATATAAGCAA GGACGGTAATGTCATAACAAAGGGTGAGTTTAAGAAGCATAGAGTA TCTCCTAGTAAATCTGATGAACGTTGGAGAACTATCATTCAAACTG CTACAAGCAAAGAAGAGTACCTGGACATGATTAAGGATGAGTTCCC CCACGAGTGGGCGACAAAATTGCAGTGGCTAGAGTACAGTGCCAAT AAACTTTTTCCACCGCAACCTGAAATATACCAAGCTACGTTCACGG AGGAAGATCTTCAATGCCATGAGGATCTTCAGTTATGGAGGGATCA ACATCTGTACCACGAGCCAAGGCGTGCAGGTACACGAATTCCTTCC

		CTTTACATCTGTGGACCATCAAGGACTGGAAAGACTACCTGGGCTC GGTCCTTGGGGCGCCATAATTATTGGAATGGCACCATTGATTTCAC CGTCTATGATGATCACGCGACGTATAATGTGATCGACGACATACCG TTCAAGTTCGTTCCACTTTGGAAACAACTCATCGGCTGTCAATCAG ACTTTACAGTTAACCCAAAGTATGGAAAGAAGAAAAAAATTAAAGG AGGTGTTCCGTGCATTATTTTATGTAACGATGATGAAGATTGGTTA AAGAACATGTCCCCGGCACAGATCGAGTACTTTGAAGCAAACTGCA TTACTCATTTTATGTATGCTGCCGAAACTTTCTTTGCCCCTGAAAG CAGTTCTCATTGAGCTTTTGTCTTCTG
GeneWiz REPA CDS	Synthesised DNA Fragment	GTGAAGACATAATGCCGTCAGCACCGCAAAAAACTAAATCGTTCAG GCTCCAGACAAAGTACGTATTCCTAACCTACCCTCGGTGTTCATCT TCAGCCGAGAACCTGAGGGATTTCCTTTGGGACAAACTGTCTAGAT TTGCGATCTTTTTTATAGCAATTGCTACAGAGTTACATCAGGATGG TACACCACATTTGCACTGTCTCATCCAACTTGACAAGAGGTCGAAC ATTCGCGACCCGAGTTTCTTTGATCTCGAAGGTAATCATCCAAATA TTCAACCTGCTAAGAATTCTGAACAGGTTCTTGAGTATATTAGTAA AGATGGGAATGTAATCACAAAAGGTGAGTTTAAGAAGCATCGTGTT AGCCCTTCAAAGTCCGATGAGCGATGGAGAACAATCATACAAACCG CTACAAGCAAAGAAGAATATTTGGATATGATTAAGGATGAGTTCCC ACATGAATGGGCAACGAAACTGCAATGGCTCGAATATTCAGCAAAT AAGCTTTTCCCACCTCAACCAGAAATATATCAAGCTACTTTTACGG AAGAGGATCTTCAGTGCCATGAGGATCTACAGTTGTGGAGAGATCA GCATTTATACCACGTGAGTGTCGACGCCTACAGATTGGTTCACAAC GTCACGTTGGTGGAGGCTCACTCTGACCTTGTTTGGATGGACGATA TCTCCAGAAACTTAGAGGGACTAGAACCAGGATCACCTCCCAGCAC TTCAGCGGATCAGGTTGTGCCCGAACGACAACACGGTCCTGAGGCT TCCGAAGGCACCATTACTGGAATGGGACCTTCGACTTCTCTTTCTA TGATGACTACACGTCCTACTACCTCCAGTACCACTTCTCCAAGCAA CAGTTCTCATTCGGGGTCTAATTGAGCTTTTGTCTTCTG
GeneWiz CMV2b CDS	Synthesised DNA Fragment	GTGAAGACATAATGGCTTTCCCCGCCCCCGCTTTCTCACTAGCCAA TCTTTTGAACGGCAGTTACGGTGTCGACACTCCCGAGGAAGTGGAA CGTTTGCGATCTGAGCAACGTGAAGAGGCTGCTGCGGCCTGTCGTA ACTACAGGCCCCTACCCGCTGTGGATGTCAGCGAGAGTGTCACAGA GGATGCGCATTCCCTCCGAACTCCTGACGGAGCTCCCGCTGAAGCG GTGTCAAGTGAGTTTGTTACTTATGGTGCTGAAGATTACCTTGAAA AATCTGATGATGAGCTCTAGGCTTTTGTCTTCTG

GeneWiz HPB tag	Synthesised DNA Fragment	TTGGTCTCTACATGTCCGGATATCCTTACGACGTTCCAGACTATGC TAGTGGGGGATTGGAAGTTCTCTTTCAGGGACCACTAGGATCAGCC GATCCTACCAGCTCCGAAACTTCGAGTCACCCGCCCGGTACAATTG TAGCACCTATGGCTGGTTTGGTCGTTAAGGTCCTCGTGGAAAATGA GGCTAAGGTTGATCAGGGTCAGCCAATACTTGTTTTAGAGGCGATG AAAATGGAGCATGTGGTGAAAGCCCCGAGCTCTGGCTCTATCCAAG ATCTGAAGGTAAAAGCAGGGCAACAAGTCTCTGATGGATCAGCTCT TTTCAGAATTAAGGGTTAGTTCGTTGTCGAGACCCC
GeneWiz CBD tag	Synthesised DNA Fragment	TTGGTCTCTACATGTCCGGATGCATTACCGGAGATGCGCTGGTCGC GCTTCCAGAAGGGGAATCGGTTAGGATCGCAGATATAGTTCCTGGG GCCCGACCTAATAGCGATAACGCCATAGACCTAAAGGTGTTGGACC GTCACGGTAATCCAGTGCTTGCAGATCGACTCTTCCATTCAGGTGA ACATCCGGTTTATACGGTCAGAACAGTGGAAGGATTGCGGGTGACT GGCACTGCTAATCATCCTTTACTGTGTTTGGTCGATGTTGCCGGGG TACCAACCTTACTTTGGAAATTGATTGACGAGATTAAGCCCGGTGA TTATGCAGTTATTCAAAGATCTGCATTTTCTGTAGACTGTGCTGGT TTCGCAAGAGGAAAACCTGAGTTTGCTCCAACAACATATACTGTAG GCGTGCCGGGACTTGTGAGGTTTCTCGAGGCCCACCATAGAGATCC CGATGCTCAAGCTATCGCGGATGAGCTCACGGATGGACGTTTCTAC TACGCTAAGGTTGCATCCGTTACCGACGCTGGTGTTCAGCCTGTCT ACAGTCTACGCGTTGATACTGCTGATCATGCTTTTATCACAAACGG ATTCGTATCACACGCTACAGCTGAGTATGGAGGTACTCCATTGGCT AATTTTCCGCCTCCCGAGCAAAGAGATGATGAATTCTTCGTGGAAG CCGCAATTAATCAGGCTAGCGATCATTTTACAGAGATCAAGGCTTT ACTAAATAACCGTTCGTCATGGCCTGCGAGGCTGATAAAGGATCTC AGTTACAATTATTATATGGACCTTACTGAGGTCTTTGAAGCAGGTT ATTCAGTAGACGACATCAAAGTTACCATCGGCTACTGTGAATCCGG GATGGATGTTGAAATCTCTCCTATAACGCACCTTTACGATAACATT TACTACATTAAAATATCTTATATTGATGGAACCAACATTTGCCCAT AGTTCGTTGTCGAGACCCC
GeneWiz FLAG tag	Synthesised DNA Fragment	TTGGTCTCTACATGTCCGGAGATTATAAAGACGACGATGATAAGTA GTTCGTTGTCGAGACCCC
Benchling eGFP RT-PCR primer 1F	Forward RT-PCR primer	TACGGCAAGCTGACCCTGAAGT

Benchling eGFP RT-PCR primer 1R	Reverse RT-PCR primer	AAGAAGATGGTGCGCTCCTGGA
Benchling eGFP RT-PCR primer 2F	Forward RT-PCR primer	TACGGCAAGCTGACCCTGAAGT
Benchling eGFP RT-PCR primer 2R	Reverse RT-PCR primer	GTAGGTCAGGGTGGTCACGAGG
Benchling eGFP RT-PCR primer 3F	Forward RT-PCR primer	TCCAGGAGCGCACCATCTTCTT
Benchling eGFP RT-PCR primer 3R	Reverse RT-PCR primer	TTGCCGTCCTCCTTGAAGTCGA
Benchling eGFP RT-PCR primer 4F	Forward RT-PCR primer	TACGGCAAGCTGACCCTGAAGT
Benchling eGFP RT-PCR primer 4R	Reverse RT-PCR primer	CGTAGGTCAGGGTGGTCACGAG
Benchling eGFP RT-PCR primer 5F	Forward RT-PCR primer	CCGACAAGCAGAAGAACGGCAT
Benchling eGFP RT-	Reverse RT-PCR primer	ATGTGATCGCGCTTCTCGTTGG

PCR primer 5R		
Primer3 eGFP RT-PCR primer 1F	Forward RT-PCR primer	GGTGAACTTCAAGATCCGCC
Primer3 eGFP RT-PCR primer 1R	Reverse RT-PCR primer	CTTGTACAGCTCGTCCATGC
Primer3 eGFP RT-PCR primer 2F	Forward RT-PCR primer	GCAAGCTGACCCTGAAGTTC
Primer3 eGFP RT-PCR primer 2R	Reverse RT-PCR primer	TCCTTGAAGTCGATGCCCTT
Primer3 eGFP RT-PCR primer 3F	Forward RT-PCR primer	GACGACGGCAACTACAAGAC
Primer3 eGFP RT-PCR primer 3R	Reverse RT-PCR primer	GGCGGATCTTGAAGTTCACC
Primer3 eGFP RT-PCR primer 4F	Forward RT-PCR primer	GCAAGCTGACCCTGAAGTTC
Primer3 eGFP RT-PCR primer 4R	Reverse RT-PCR primer	GTCTTGTAGTTGCCGTCGTC

Primer3 eGFP RT-PCR primer 5F	Forward RT-PCR primer	TAAACGGCCACAAGTTCAGC
Primer3 eGFP RT-PCR primer 5R	Reverse RT-PCR primer	AAGTCGTGCTGCTTCATGTG
PacI sticky end primer	To add PacI to the 5' end of a target sequence	TTAATTAAN$^{(20)}$ where N$^{(20)}$ is 20 nucleotides of target sequence
AvrII sticky end primer	To add AvrII to the 3' end of a target sequence	CCTAGGN$^{(20)}$ where N$^{(20)}$ is 20 nucleotides of target sequence
C1 - Level -1 and Level 0 forward primer	Sequencing primer to check Level -1 and Level 0 parts	CGTTATCCCCTGATTCTGTGGATAAC
C2 – Level -1 and Level 0 reverse primer	Sequencing primer to check Level -1 and Level 0 parts	GTCTCATGAGCGGATACATATTTGAATG
C3 - Level 1 forward primer	Sequencing primer to check Level 1 parts	GAACCCTGTGGTTGGCATGCACATAC

C4 - Level 1 reverse primer	Sequencing primer to check Level 1 parts	CTGGTGGCAGGATATATTGTGGTG
C5 - Level 2 forward primer	Sequencing primer to check Level 2 parts	GTGGTGTAAACAAATTGACGC
C6 - Level 2 reverse primer	Sequencing primer to check Level 2 parts	GTGGTGTAAACAAATTGACGC

Table S2: Tissue masses used for the final protein extractions of the pRC-CDS and pR5-CDS constructs. Leaves from four transformed plants were removed, ground and harvested, with the final collected weight measured. Analysis of total protein extracts was standardised to 20 µg of protein as measured by Bradford assay, but for the purified and concentrated samples, western blot signal must be normalised against the starting mass of material.

Tissue	Mass (g)
pRC-eGFP	9
pRC-DORN1	5.4
pRC-HMA4	6.2
pRC-FactorC	5.7
pRC-RANGAP1	5.5
pRC-hfP2X7	11
pRC-hcP2X7	10.2
pRC-hbP2X7	9.2
pR5-eGFP	2.2
pR5-DORN1	5.5
pR5-HMA4	3.8
pR5-FactorC	7.2
pR5-RANGAP1	5.7
pR5-rfP2X7	4.3
pR5-rcP2X7	7.5
pR5-rbP2X7	4.8
pR5-hfP2X7	4
pR5-hcP2X7	5.3
pR5-hbP2X7	5.7

List of Materials

Gene synthesis

There were no physical materials related to gene synthesis, but the DNA sequences ordered can be found in the supplementary material (Table S1).

DNA purification

Plasmid DNA extraction

Resuspension Buffer P1 – 50 mM Tris·HCl, 10 mM EDTA, 100 µg/mL RNase A, pH 8.0

Lysis Buffer P2 - 200 mM NaOH, 1% w/v SDS

Neutralisation Buffer N3 – 4.2 M Gu-HCl, 0.9M KOAc, pH 4.2

Wash Buffer PE – 10 mM Tris·HCl, 80% v/v ethanol, pH 7.5

DNA extraction from an agarose gel

Zymoclean Gel DNA recovery kit (Zymoresearch, catalogue number: D4001/D4002)

Cloning

Q5 Amplification of DNA fragments

Q5 High-fidelity DNA polymerase and buffer (NEB, catalogue number: M0491S)

10 mM Deoxynucleotide (dNTP) Solution Mix (NEB, catalogue number: N0447S)

10 µM each of forward and reverse primer (see Table S1 for sequences)

Golden Gate cloning

Type IIS restriction endonuclease, Eco31I (BsaI; Thermofisher, catalogue number: ER0291) or BpiI (BbsI; Thermofisher, catalogue number: ER1012)

T4 DNA ligase and T4 DNA ligase buffer (NEB, catalogue number: M0202S)

Golden Gate vector and insert in a 1:3 molar ratio

Conventional cloning

Restriction enzymes AvrII (NEB, catalogue number: R0174S) and PacI (NEB, catalogue number: R0547S) with rCutSmart buffer

T4 DNA ligase and buffer (NEB, catalogue number: M0202S)

Vector and insert in a 1:3 molar ratio

Screening of colonies by colony PCR

PCRBIO Taq Mix Red (PCRBiosystems, catalogue number: PB10.13-02)

10 µM of each forward and reverse primer (see Table S1 for sequences)

Agarose gel electrophoresis

Agarose

Tris-acetate-EDTA (TAE) buffer

SafeView

Gel tank and leads

Power pack

DNA size marker (1Kb plus; NEB, catalogue number: N0469S; SmartLadder, catalogue number: MW-1700-02)

Bacterial cells

Growth of bacteria

E. coli DH5 Alpha (Zymoresearch, catalogue number: T3007)

E. coli DH10 Beta (NEB, catalogue number: C3019H)

tumefaciens GV3101 (VWR, catalogue number: 103753-234)

Luria Broth, 20 g/L, pH 7.2 (Melford; catalogue number: L24060)

Agar (Duchefa Biochemie; catalogue number: M1002)

Antibiotics at the following concentrations: Kanamycin (50 µg/ml), Spectinomycin (100 µg/ml), Carbenicillin (100 µg/ml), Rifampicin (30 µg/ml), Gentamycin (20 µg/ml), Cefotaxime (100 µg/ml)

1M IPTG and 20 mg/mL X-gal stocks

Transformation of E. coli

E. coli DH5 Alpha (Zymoresearch, catalogue number: T3007)

E. coli DH10 Beta (NEB, catalogue number: C3019H)

Luria Broth, 20 g/L, pH 7.2 (Melford; catalogue number: L24060)

A shaking incubator set to 200RPM and 37°C

A water bath set to 42°C (DH10 Beta only)

Transformation of A. tumefaciens

tumefaciens GV3101 (VWR, catalogue number: 103753-234)

Luria Broth, 20 g/L, pH 7.2 (Melford; catalogue number: L24060)

Gene Pulser/MicroPulser Electroporation Cuvettes (BioRad, catalogue number: 165-2086)

MicroPulser Electroporator (BioRad, catalogue number: 1652100)

A shaking incubator set to 200RPM and 28°C

Plants

Growth of Arabidopsis and Nicotiana plants

Soil and sand in a 3:1 ratio (Levington advance professional growing media M2 grade; Melcourt horticultural sharp sand)

Microagar (Duchefa Biochemie; catalogue number: M1002)

Antibiotics at the following concentrations: Kanamycin (50 µg/ml), Spectinomycin (100 µg/ml), Carbenicillin (100 µg/ml), Rifampicin (30 µg/ml), Gentamycin (20 µg/ml), Cefotaxime (100 µg/ml)

Ethanol

Bleach

Triton-X 100 (SigmaAldrich; catalogue number: T8787)

Growth cabinets or facilities

Bright Yellow-2 tobacco cell cultures

mLS medium (4.3 g/L MS salt mix, 30 g/L sucrose 200 mg/L KH_2PO_4, 1 mg/L Thiamine·HCl, 100 mg/L *myo*-Inositol, and 0.236 mg/L 2,4-D, pH 5.8)

A shaking incubator set to 27°C and 130 RPM in darkness

Stable transformation of A. thaliana by floral dipping

Four-week old *A. thaliana* plants

Transformed Agrobacterium culture

Sucrose (Fisher; catalogue number: S/560/65)

Silwet L-77 (De Sangosse; 0640)

Transient transformation of A. thaliana protoplasts

Mature *A. thaliana* plants

Time-tape

Magic-tape

Enzyme solution (1% cellulase, 0.25% macerozyme, 0.4 M mannitol, 10 mM $CaCl_2$, 20 mM KCl, 0.1% BSA and 20 mM MES, pH 5.7)

Modified W5 solution (154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, 5 mM glucose, and 2 mM MES, pH 5.7)

Modified MMg solution (0.4 M mannitol, 15 mM $MgCl_2$, and 4 mM MES, pH 5.7)

PEG solution (40% (w/v) PEG, 0.1 M $CaCl_2$, 0.2 M mannitol)

Shaking platform

6-well plate

Centrifuge

Stable transformation of N. benthamiana leaf disks

Bleach

Ethanol

1 cm Potato-borer

Rooting media (MS medium, 50 mg/L Kanamycin, 100 mg/L Cefotaxime (Claforan, Duchefa C0111)

Shooting media (MS medium, 1 mg/L 6-Benzylaminopurine (BAP), 0.1 mg/L Naphthalene acetic acid (NAA), 50 mg/L Kanamycin, 100 mg/L Cefotaxime (Claforan, Duchefa C0111)

Transient transformation of N. benthamiana

Seven-week old *N. benthamiana* plants

Transformed Agrobacterium culture

MMA infiltration buffer (10 mM $MgCl_2$, 10 mM MES 200 µM acetosyringone, pH 5.6)

Needleless 10 mL syringe

Vacuum pump

Transformation of BY-2 plant cell packs

BY-2 cell cultures

PCP infiltration buffer (0.5 g L^{-1} Murashige-Skoog type medium, 146 mM sucrose, 10 mM glucose monohydrate 200 µM acetosyringone, 15 mM MES, pH 5.6)

Macherey-Nagel™ Receiver Plates (Fisher, Catalogue number: 16183547)

Ultracentrifuge with 96-well plate attachment

100 mL deionised water

Functional assays in plant cell packs

100 mM Yo-Pro-1 or ethidium bromide

100 mM JNJ-47965567

100 mM BzATP

PhotonIMAGER Optima

Transformed plant cell packs

RNA and protein extraction

RNA extraction

Plant tissue to be extracted

Liquid nitrogen

Pestle and mortar

TRI reagent (Zymo Research, Catalogue Number R2050-1-200)

Chloroform

Isopropanol

75% Ethanol

RNase-free water at 65°C

Protein extraction

Liquid nitrogen

Pestle and mortar

One of the following protein extraction buffers (PEBs):

PEB 1.0 (Phosphate-buffered saline, pH 7.4, and 1 x Roche cOmplete protease inhibitor)

PEB 2.0 (50 mM HEPES-KOH, pH 7.6, 150 mM NaCl, 10% v/v glycerol, 0.25% FC-12, 1 mM PMSF, 1 x cOmplete protease inhibitor (Sigma Aldrich, Catalogue number: P9599))

PEB 3.0 (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, and 10 mM EDTA, pH 8.0)

PEB 4.1 (50 mM Tris-HCl, pH 7.4, 330 mM sucrose, 5 mM EDTA, 5 mM DTT and 1 x sigma plant protease inhibitor cocktail)

PEB 4.2 (50 mM Tris-HCl, pH 7.4, 165 mM NaCl, 5 mM DTT and 1 x sigma plant protease inhibitor cocktail)

PEB 4.3 (50 mM phosphate, pH 8.0, 10 mM Tris, pH 8.0, 500 mM NaCl, 0.1% Tween 20, 0.1% FC-12, and 1 x sigma plant protease inhibitor)

PEB 5.0 (20 mM sodium phosphate, pH 7.5, 150 mM NaCl, 1% (v/v) Triton X-100, 5 mM DTT, 1 mM PMSF and 1 x sigma plant protease inhibitor cocktail)

Miracloth

Ultracentrifuge capable of reaching 100,000 *x g*

Solubilisation buffer (PEB containing 0.5% w/v FC-12)

Expression analyses

Analysis of in planta fluorescence by confocal microscopy

Axio Observer.Z1 with appropriate excitation laser and emission bandpass filters

Glass slides and cover slips

FM™ 4-64 (ThermoFisher, catalogue number: T3166)

confocal dishes (VWR, catalogue number: 734-2905)

Analysis of in planta fluorescence using photon imaging

PhotonIMAGER Optima™

M3 Vision software

Transcript detection using RT-PCR

DNA-*free*™ DNA Removal Kit (Invitrogen Catalogue Number: AM1906)

Water bath set to 37°C

ProtoScript® II First Strand cDNA Synthesis Kit (NEB, Catalogue number: E6560S)

Thermocycler

PCRBIO Taq Mix Red (PCRBiosystems, catalogue number: PB10.13-02)

Forward and reverse primer mix at 10 µM each (see Table S1 for sequences)

Extraction of total plant protein and quantification of protein concentration by Bradford assay and fluorescence

96-well plate

Clariostar plate reader

Bio-Rad Protein Assay solution (Bio-Rad, Catalogue number: 5000006)

BSA standards

SDS-PAGE

NuPAGE™ 4 to 12%, Bis-Tris gels (ThermoFisher, catalogue number: NP0321BOX)

Gel tank

NuPAGE™ MOPS SDS Running Buffer (20X; ThermoFisher, catalogue number: NP0001)

6 x SDS loading dye 0.375 M Tris pH 6.8, 12% SDS, 60% glycerol, 0.6 M DTT, 0.06% bromophenol blue; 1:5 v/v

Coomassie blue staining of protein gels

InstantBlue® Coomassie Protein Stain (Abcam, catalogue number: ab119211)

Quantification of extracted GFP by western blot

Mini nitrocellulose membrane (ThermoFisher, catalogue number: LC2000)

Trans-Blot® Turbo™ Transfer System (BIO-RAD, catalogue number: 1704150

Intercept® (TBS) Blocking Buffer (LI-COR, catalogue number: 927-60001)

Primary antibody solution (10 mL Blocking buffer containing 2% Tween (v/v) and 1/5000 v/v primary antibody (Anti-GFP (Plant Specific) Antibody; antibodies.com, catalogue number: A50024))

TBS containing 2% v/v Tween

Secondary antibody solution (10 mL Blocking buffer containing 2% Tween (v/v) and secondary antibody (Goat anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ Plus 800; Invitrogen, catalogue number: A32730) at a concentration of 1/13500 v/v)

Odyssey® DLx Imaging System.

Protein purification

His-pulldown using Ni-Sepharose beads with PEB 4.2

Ni Sepharose® High Performance (SigmaAldrich, Catalogue number: GE17-5268-01)

Solubilisation buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 1% (w/v) dodecyl maltoside, 1 x Roche cOmplete protease inhibitor)

Binding and wash buffer (Solubilisation buffer with 25 mM imidazole)

Elution buffer (Solubilisation buffer with 500 mM imidazole)

His-pulldown using Ni-Sepharose beads with PEB 5.0

Ni Sepharose® High Performance (SigmaAldrich, Catalogue number: GE17-5268-01)

Solubilisation buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 1% (v/v) Triton X-100)

Wash buffer (Solubilisation buffer with 20 mM imidazole)

Elution buffer (Solubilisation buffer with 500 mM imidazole)

Binding buffer (20 mM sodium phosphate, pH 7.4, 925 mM NaCl, 40 mM imidazole)

His-pulldown using Ni-NTA magnetic beads

HisPur™ Ni-NTA Magnetic Beads (ThermoFisher, Catalogue number: 88831)

Solubilisation buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 1% (v/v) Triton X-100)

Wash buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 20 mM Imidazole)

Elution buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 500 mM Imidazole)

x Binding buffer (20 mM sodium phosphate, pH 7.4, 925 mM NaCl, 40 mM imidazole)

Manual HisTrapHP column purification

1 mL HisTrapHP column (Cytiva, Catalogue number: 29051021)

Equilibration buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 20 mM Imidazole) filtered using a 0.2 µM filter

Elution buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 500 mM Imidazole) filtered using a 0.2 µM filter

Needless syringes

20% ethanol filtered using a 0.2 µM filter

AKTA HisTrapHP column purification

HisTrapHP column (1 mL or 5 mL, Cytiva, Catalogue numbers: 29051021 and 17524801, respectively)

Equilibration buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 20 mM Imidazole)

Elution buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 500 mM Imidazole)

AKTA Prime or AKTA Start

HPB-tag

Dynabeads® M-280 beads (ThermoFisher, Catalogue number: 11205D)

Radio Immunoprecipitation Assay (RIPA) buffer 1 (50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, pH 7.4, 0.5% sodium-deoxycholate)

RIPA buffer 2 (50 mM Tris-HCl, 20 mM NaCl, 1 mM EDTA, pH 7.4, 0.5% sodium-deoxycholate)

CBD-tag

Microcrystalline cellulose powder (SigmaAldrich, Catalogue number: 435236)

dH_2O

Centrifuge

Wash buffer (Tris-HCl, pH 7.5)

2× SDS-PAGE buffer (100 mM Tris-HCl, pH 6.8, 4% (w/v) SDS, 0.1% (w/v) bromophenol blue, 20% (v/v) glycerol 200 mM β-mercaptoethanol)

FLAG-tag (Fab-Trap)

DYKDDDDK Fab-Trap

RIPA buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 0.1 % SDS, 1% Triton™ X-100, 1% sodium deoxycholate)

Dilution buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 80% v/v Elix H_2O added immediately prior to use)

Wash buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05 % Sodium deoxycholate, 0.5 mM EDTA, 80% v/v Elix H_2O added prior to use)

2× SDS-PAGE buffer (100 mM Tris-HCl, pH 6.8, 4% (w/v) SDS, 0.1% (w/v) bromophenol blue, 20% (v/v) glycerol 200 mM β-mercaptoethanol)

Centrifuge

Anti-GFP antibody (GFP-Trap)

GFP-Trap Magnetic Particles M-270

RIPA buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 0.1 % SDS, 1% Triton™ X-100, 1% sodium deoxycholate)

Dilution buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 0.018% sodium azide)

Wash buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05 % Sodium deoxycholate, 0.5 mM EDTA)

2× SDS-PAGE buffer (100 mM Tris-HCl, pH 6.8, 4% (w/v) SDS, 0.1% (w/v) bromophenol blue, 20% (v/v) glycerol 200 mM β-mercaptoethanol)

Magnetic microtube stand

Protein concentration

Vivaspin 500 centrifugal concentrator columns (3000 MWCO; Merck, catalogue number: Z629367)

Zeiss LSM 880 confocal microscope settings

BitsPerPixel = 8

DimensionOrder = XYCZT

IsInterleaved = false

IsRGB = false

LittleEndian = true

PixelType = uint8

Series 0 Name = Cytoplasm settings #1

SizeC = 2

SizeT = 1

SizeX = 1024

SizeY = 1024

SizeZ = 41

Appliance|Data|ShuttleAndFindData|Calibration|Marker|FocusPosition #1 = 0

Appliance|Data|ShuttleAndFindData|Calibration|Marker|FocusPosition #2 = 0

Appliance|Data|ShuttleAndFindData|Calibration|Marker|FocusPosition #3 = 0

Appliance|Data|ShuttleAndFindData|Calibration|Marker|Id #1 = Marker:1

Appliance|Data|ShuttleAndFindData|Calibration|Marker|Id #2 = Marker:2

Appliance|Data|ShuttleAndFindData|Calibration|Marker|Id #3 = Marker:3

Appliance|Data|ShuttleAndFindData|Calibration|Marker|StageXPosition #1 = 0

Appliance|Data|ShuttleAndFindData|Calibration|Marker|StageXPosition #2 = 0

Appliance|Data|ShuttleAndFindData|Calibration|Marker|StageXPosition #3 = 0

Appliance|Data|ShuttleAndFindData|Calibration|Marker|StageYPosition #1 = 0

Appliance|Data|ShuttleAndFindData|Calibration|Marker|StageYPosition #2 = 0

Appliance|Data|ShuttleAndFindData|Calibration|Marker|StageYPosition #3 = 0

Appliance|Data|ShuttleAndFindData|Calibration|MicroscopeType = LM

Appliance|Data|ShuttleAndFindData|Calibration|StageOrientation|X = -1

Appliance|Data|ShuttleAndFindData|Calibration|StageOrientation|Y = 1

Appliance|Id = ShuttleAndFind:1

DisplaySetting|Channel|ChannelUnit|ChannelType = Unspecified

DisplaySetting|Channel|ChannelUnit|FactorI = 1

DisplaySetting|Channel|ChannelUnit|OffsetI = 0

DisplaySetting|Channel|ChannelUnit|UnitI = Unknown

DisplaySetting|Channel|Color = #FFFFFF

DisplaySetting|Channel|ColorMode = Color

DisplaySetting|Channel|DyeName = Alexa Fluor 488

DisplaySetting|Channel|Gamma = 1

DisplaySetting|Channel|High = 1

DisplaySetting|Channel|Id = 356421557011386296631411222403159241255 2

DisplaySetting|Channel|IsSelected = true

DisplaySetting|Channel|Low = -0

DisplaySetting|Channel|Name = T PMT

Experiment|AcquisitionBlock|AcquisitionModeSetup|AcquisitionMode = StackFocus

Experiment|AcquisitionBlock|AcquisitionModeSetup|ApertureDiameter = 4.4099999999999993

Experiment|AcquisitionBlock|AcquisitionModeSetup|BiDirectional = true

Experiment|AcquisitionBlock|AcquisitionModeSetup|BiDirectionalZ = false

Experiment|AcquisitionBlock|AcquisitionModeSetup|BitsPerSample = 8

Experiment|AcquisitionBlock|AcquisitionModeSetup|CameraBinning = 1

Experiment|AcquisitionBlock|AcquisitionModeSetup|CameraFrameHeight = 1030

Experiment|AcquisitionBlock|AcquisitionModeSetup|CameraFrameOffsetX = 0

Experiment|AcquisitionBlock|AcquisitionModeSetup|CameraFrameOffsetY = 0

Experiment|AcquisitionBlock|AcquisitionModeSetup|CameraFrameWidth = 1300

Experiment|AcquisitionBlock|AcquisitionModeSetup|CameraSuperSampling = 0

Experiment|AcquisitionBlock|AcquisitionModeSetup|DimensionT = 10

Experiment|AcquisitionBlock|AcquisitionModeSetup|DimensionX = 1024

Experiment|AcquisitionBlock|AcquisitionModeSetup|DimensionY = 1024

Experiment|AcquisitionBlock|AcquisitionModeSetup|DimensionZ = 41

Experiment|AcquisitionBlock|AcquisitionModeSetup|FilterMethod = Average

Experiment|AcquisitionBlock|AcquisitionModeSetup|FilterMode = Line

Experiment|AcquisitionBlock|AcquisitionModeSetup|FilterSamplingNumber = 2

Experiment|AcquisitionBlock|AcquisitionModeSetup|FitFramesizeToRoi = false

Experiment|AcquisitionBlock|AcquisitionModeSetup|FocalDistanceLsmTubeLense = 150

Experiment|AcquisitionBlock|AcquisitionModeSetup|FocalDistanceScanLense = 52.5

Experiment|AcquisitionBlock|AcquisitionModeSetup|FocalDistanceStandardTubeLense = 164.5

Experiment|AcquisitionBlock|AcquisitionModeSetup|FocusStabilizer = false

Experiment|AcquisitionBlock|AcquisitionModeSetup|HdrEnabled = false

Experiment|AcquisitionBlock|AcquisitionModeSetup|HdrImagingMode = 0

Experiment|AcquisitionBlock|AcquisitionModeSetup|HdrIntensity = 1

Experiment|AcquisitionBlock|AcquisitionModeSetup|HdrNumFrames = 1

Experiment|AcquisitionBlock|AcquisitionModeSetup|InterpolationY = 1

Experiment|AcquisitionBlock|AcquisitionModeSetup|Objective = Plan-Apochromat 20x/0.8 M27

Experiment|AcquisitionBlock|AcquisitionModeSetup|OffsetX = 0

Experiment|AcquisitionBlock|AcquisitionModeSetup|OffsetY = 0

Experiment|AcquisitionBlock|AcquisitionModeSetup|OffsetZ = 5.9993000000000393e-005

Experiment|AcquisitionBlock|AcquisitionModeSetup|PixelPeriod = 2.0606060606060607e-006

Experiment|AcquisitionBlock|AcquisitionModeSetup|PreScan = false

Experiment|AcquisitionBlock|AcquisitionModeSetup|ReferenceZ = 0.0095856859999999995

Experiment|AcquisitionBlock|AcquisitionModeSetup|RequestedScanSpeed = 1

Experiment|AcquisitionBlock|AcquisitionModeSetup|Rotation = 0

Experiment|AcquisitionBlock|AcquisitionModeSetup|RtBinning = 1

Experiment|AcquisitionBlock|AcquisitionModeSetup|RtFrameHeight = 512

Experiment|AcquisitionBlock|AcquisitionModeSetup|RtFrameWidth = 512

Experiment|AcquisitionBlock|AcquisitionModeSetup|RtLinePeriod = 2.4242424242424244e-007

Experiment|AcquisitionBlock|AcquisitionModeSetup|RtOffsetX = 0

Experiment|AcquisitionBlock|AcquisitionModeSetup|RtOffsetY = 0

Experiment|AcquisitionBlock|AcquisitionModeSetup|RtRegionHeight = 512

Experiment|AcquisitionBlock|AcquisitionModeSetup|RtRegionWidth = 512

Experiment|AcquisitionBlock|AcquisitionModeSetup|RtSuperSampling = 1

Experiment|AcquisitionBlock|AcquisitionModeSetup|RtZoom = 1

Experiment|AcquisitionBlock|AcquisitionModeSetup|SamplingEnd = 1

Experiment|AcquisitionBlock|AcquisitionModeSetup|SamplingStart = 1

Experiment|AcquisitionBlock|AcquisitionModeSetup|ScalingX = 4.1513291048191956e-007

Experiment|AcquisitionBlock|AcquisitionModeSetup|ScalingY = 4.1513291048191956e-007

Experiment|AcquisitionBlock|AcquisitionModeSetup|ScalingZ = 1.5e-006

Experiment|AcquisitionBlock|AcquisitionModeSetup|ScannerOnlineCorrection = true

Experiment|AcquisitionBlock|AcquisitionModeSetup|SimRotations = 3

Experiment|AcquisitionBlock|AcquisitionModeSetup|TimeSeries = false

Experiment|AcquisitionBlock|AcquisitionModeSetup|TrackMultiplexType = Line

Experiment|AcquisitionBlock|AcquisitionModeSetup|UseRois = false

Experiment|AcquisitionBlock|AcquisitionModeSetup|WaitState = true

Experiment|AcquisitionBlock|AcquisitionModeSetup|ZoomX = 1

Experiment|AcquisitionBlock|AcquisitionModeSetup|ZoomY = 1

Experiment|AcquisitionBlock|Laser|LaserName = ArgonRemote

Experiment|AcquisitionBlock|Laser|LaserPower = 0.025000000000000001

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Attenuator|ExcitationIntensity = 0

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Attenuator|Laser = ArgonRemote

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Attenuator|LaserSuppression = false

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Attenuator|Polarization_Stokes0 = 1

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Attenuator|Polarization_Stokes1 = -1

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Attenuator|Polarization_Stokes2 = 0

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Attenuator|Polarization_Stokes3 = 0

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Attenuator|Transmission = 0.01

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Attenuator|Wavelength = 4.8800000000000003e-007

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|BeamSplitter|BeamSplitterServoPositio n #1 = 0

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|BeamSplitter|BeamSplitterServoPositio n #2 = 0

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|BeamSplitter|BeamSplitterServoPositio n #3 = 0

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|BeamSplitter|BeamSplitterServoPositio n #4 = 0

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|BeamSplitter|Filter #1 = MBS 488

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|BeamSplitter|Filter #2 = Plate

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|BeamSplitter|Filter #3 = Mirror

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|BeamSplitter|Filter #4 = NoneLSM

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|BeamSplitter|Identifier #1 = MainBeamSplitterDescanned1

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|BeamSplitter|Identifier #2 = MainBeamSplitterDescanned2

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|BeamSplitter|Identifier #3 = DichroicBeamSplitterDescanned1

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|BeamSplitter|Identifier #4 = MainBeamSplitterNonDescanned

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|CameraIntegrationTime = 2.0606060606060607e-006

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|CenterWavelength = 5.50618398e-007

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Collimator|Name #1 = LSMCOLLI_V

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Collimator|Name #2 = LSMCOLLI_V2

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Collimator|Position #1 = 888

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Collimator|Position #2 = 766

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|CondensorAperture = -1

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|CondensorFrontlensPosition = -1

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|AiryScanMagnification #1 = 1

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|AiryScanMagnification #2 = 1

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|AiryScanMode #1 = Off

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|AiryScanMode #2 = Off

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|AiryScanTransformationXX #1 = 0

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|AiryScanTransformationXX #2 = 0

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|AiryScanTransformationXY #1 = 0

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|AiryScanTransformationXY #2 = 0

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|AiryScanTransformationYX #1 = 0

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|AiryScanTransformationYX #2 = 0

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|AiryScanTransformationYY #1 = 0

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|AiryScanTransformationYY #2 = 0

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|AiryScanVirtualPinholeSize #1 = 1

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|AiryScanVirtualPinholeSize #2 = 1

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|AmplifierGain #1 = 1

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|AmplifierGain #2 = 1

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|AmplifierOffset #1 = 0

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|AmplifierOffset #2 = 0

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|Color #1 = #00FF00

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|Color #2 = #FFFFFF

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|DetectorIdentifier #1 = SpectralImager1

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|DetectorIdentifier #2 = NonDescannedTransmissionPmt1

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|DetectorMode #1 = Integration

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|DetectorMode #2 = Integration

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|DetectorWavelengthRange|WavelengthEnd = 6.2179838199999998e-007

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|DetectorWavelengthRange|WavelengthStart = 4.9723340999999994e-007

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|DigitalGain #1 = 1

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|DigitalGain #2 = 1

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|DigitalOffset #1 = 0

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|DigitalOffset #2 = 0

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|Dye = Alexa Fluor 488

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|Folder = Alexa Fluor 488

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|Id #1 = 183321911911716654151887566583014201 34

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|Id #2 = 35642155701138629663141122240315924125 52

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|ImageChannelName #1 = ChS1

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|ImageChannelName #2 = T PMT

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|LaserSuppression #1 = true

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|LaserSuppression #2 = true

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|Palette #1 = LsmDetectorPalette_0_0

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|Palette #2 = LsmDetectorPalette_0_1

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|PinholeDiameter #1 = 3.09344e-005

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|PinholeDiameter #2 = 0

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|PureRatioSource #1 = false

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|PureRatioSource #2 = false

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|SpectralScanChannels #1 = 32

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|SpectralScanChannels #2 = 32

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|Voltage #1 = 700

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Detector|Voltage #2 = 389.61038961038952

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|DeviceMode = LSM_ChannelMode

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|FieldStopPosition = 0.6348448687350835

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|FilterTransmission = -1

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|LaserSuppressionMode = None

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|Name = Track1

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|PalmSlider = true

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|ReflectedLightLampIntensity = 0

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|SimGratingPeriod = 0

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|TirfAngle = 0

Experiment|AcquisitionBlock|MultiTrackSetup|TrackSetup|TransmittedLightLampIntensity = 0

Experiment|AcquisitionBlock|TilesSetup|PositionGroup|AutoFocusMode = Off

Experiment|AcquisitionBlock|TilesSetup|PositionGroup|AutoFocusOffset = 0

Experiment|AcquisitionBlock|ZStackSetup|Extrapolate = false

Experiment|AcquisitionBlock|ZStackSetup|Interpolation = Cubic

Experiment|AcquisitionBlock|ZStackSetup|StackBrightnessCorrection = false

Experiment|ExperimentBlockIndex = 0

Information|Application|Name = AIMApplication

Information|Application|Version = 14.0.27.201

Information|Document|CreationDate = 2023-12-13T10:11:56

Information|Document|Name = Cytoplasm settings

Information|Document|Rating = 0

Information|Document|SubType = Image

Information|Document|Title = Cytoplasm settings

Information|Document|UserName = LSM User

Information|Image|Channel|AcquisitionMode #1 = LaserScanningConfocalMicroscopy

Information|Image|Channel|AcquisitionMode #2 = LaserScanningConfocalMicroscopy

Information|Image|Channel|Attenuation #1 = 0.98999999999999999

Information|Image|Channel|Attenuation #2 = 0.98999999999999999

Information|Image|Channel|Binning #1 = 1x1

Information|Image|Channel|Binning #2 = 1x1

Information|Image|Channel|ChannelType #1 = Unspecified

Information|Image|Channel|ChannelType #2 = Unspecified

Information|Image|Channel|ContrastMethod #1 = Fluorescence

Information|Image|Channel|ContrastMethod #2 = Other

Information|Image|Channel|DetectionWavelength|Ranges = 497.23340999999994-621.79838199999995

Information|Image|Channel|Detector|Id #1 = Detector:0:0

Information|Image|Channel|Detector|Id #2 = Detector:0:1

Information|Image|Channel|DigitalGain #1 = 1

Information|Image|Channel|DigitalGain #2 = 1

Information|Image|Channel|EmissionWavelength = 559.515896

Information|Image|Channel|ExcitationWavelength = 488.00000000000006

Information|Image|Channel|Fluor = Alexa Fluor 488

Information|Image|Channel|Gain #1 = 700

Information|Image|Channel|Gain #2 = 389.61038961038952

Information|Image|Channel|Id #1 = 183321911911716654151887566583014201 34

Information|Image|Channel|Id #2 = 356421557011386296631411222403159241 2552

Information|Image|Channel|IlluminationType #1 = Epifluorescence

Information|Image|Channel|IlluminationType #2 = Epifluorescence

Information|Image|Channel|LaserScanInfo|Averaging #1 = 2

Information|Image|Channel|LaserScanInfo|Averaging #2 = 2

Information|Image|Channel|LaserScanInfo|FrameTime #1 = 5.0641454545454545

Information|Image|Channel|LaserScanInfo|FrameTime #2 = 5.0641454545454545

Information|Image|Channel|LaserScanInfo|LineTime #1 = 2.4242424242424244e-007

Information|Image|Channel|LaserScanInfo|LineTime #2 = 2.4242424242424244e-007

Information|Image|Channel|LaserScanInfo|PixelTime #1 = 2.0606060606060607e-006

Information|Image|Channel|LaserScanInfo|PixelTime #2 = 2.0606060606060607e-006

Information|Image|Channel|LaserScanInfo|SampleOffsetX #1 = 0

Information|Image|Channel|LaserScanInfo|SampleOffsetX #2 = 0

Information|Image|Channel|LaserScanInfo|SampleOffsetY #1 = 0

Information|Image|Channel|LaserScanInfo|SampleOffsetY #2 = 0

Information|Image|Channel|LaserScanInfo|SampleRotation #1 = 0

Information|Image|Channel|LaserScanInfo|SampleRotation #2 = 0

Information|Image|Channel|LaserScanInfo|ScanningMode #1 = LineSequential

Information|Image|Channel|LaserScanInfo|ScanningMode #2 = LineSequential

Information|Image|Channel|LaserScanInfo|ZoomX #1 = 1

Information|Image|Channel|LaserScanInfo|ZoomX #2 = 1

Information|Image|Channel|LaserScanInfo|ZoomY #1 = 1

Information|Image|Channel|LaserScanInfo|ZoomY #2 = 1

Information|Image|Channel|LightSource|Id #1 = LightSource:0

Information|Image|Channel|LightSource|Id #2 = LightSource:0

Information|Image|Channel|Magnification #1 = 1

Information|Image|Channel|Magnification #2 = 1

Information|Image|Channel|Mode #1 = Off

Information|Image|Channel|Mode #2 = Off

Information|Image|Channel|Name #1 = ChS1

Information|Image|Channel|Name #2 = T PMT

Information|Image|Channel|Offset #1 = 0

Information|Image|Channel|Offset #2 = 0

Information|Image|Channel|PhotonConversionFactor #1 = 816.62732252395824

Information|Image|Channel|PhotonConversionFactor #2 = 816.62732252395824

Information|Image|Channel|PinholeSizeAiry = 0.72508583237879609

Information|Image|Channel|Polarization|StokesParameter1 #1 = 1

Information|Image|Channel|Polarization|StokesParameter1 #2 = 1

Information|Image|Channel|Polarization|StokesParameter2 #1 = -1

Information|Image|Channel|Polarization|StokesParameter2 #2 = -1

Information|Image|Channel|Polarization|StokesParameter3 #1 = 0

Information|Image|Channel|Polarization|StokesParameter3 #2 = 0

Information|Image|Channel|Polarization|StokesParameter4 #1 = 0

Information|Image|Channel|Polarization|StokesParameter4 #2 = 0

Information|Image|Channel|TransformationXX #1 = 0

Information|Image|Channel|TransformationXX #2 = 0

Information|Image|Channel|TransformationXY #1 = 0

Information|Image|Channel|TransformationXY #2 = 0

Information|Image|Channel|TransformationYX #1 = 0

Information|Image|Channel|TransformationYX #2 = 0

Information|Image|Channel|TransformationYY #1 = 0

Information|Image|Channel|TransformationYY #2 = 0

Information|Image|Channel|VirtualPinholeSize #1 = 1

Information|Image|Channel|VirtualPinholeSize #2 = 1

Information|Image|Channel|Wavelength #1 = 488.00000000000006

Information|Image|Channel|Wavelength #2 = 488.00000000000006

Information|Image|ComponentBitCount = 8

Information|Image|Medium = Air

Information|Image|MicroscopeRef|Id = Microscope:0

Information|Image|ObjectiveRef|Id = Objective:0

Information|Image|OriginalScanData = true

Information|Image|PixelType = Gray8

Information|Image|RefractiveIndex = 1

Information|Image|SizeC = 2

Information|Image|SizeX = 1024

Information|Image|SizeY = 1024

Information|Image|SizeZ = 41

Information|Image|S|Scene|Index = 0

Information|Image|S|Scene|Position|X = 13430.6

Information|Image|S|Scene|Position|Y = -8566.8

Information|Image|S|Scene|Position|Z = 9525.69

Information|Image|Track|ChannelRef|Id #1 = 183321911911716654151887566583014201 34

Information|Image|Track|ChannelRef|Id #2 = 356421557011386296631411222403159241 2552

Information|Image|Track|Id = Track:0

Information|Image|T|BinaryList|AttachmentName = TimeStamps

Information|Image|Z|Interval|Increment = 1.5

Information|Image|Z|Interval|Start = 0

Information|Image|Z|StartPosition = 0

Information|Instrument|Detector|AmplificationGain #1 = 1

Information|Instrument|Detector|AmplificationGain #2 = 1

Information|Instrument|Detector|Gain #1 = 1

Information|Instrument|Detector|Gain #2 = 1

Information|Instrument|Detector|Id #1 = Detector:0:0

Information|Instrument|Detector|Id #2 = Detector:0:1

Information|Instrument|Detector|Zoom #1 = 1

Information|Instrument|Detector|Zoom #2 = 1

Information|Instrument|Id = Instrument:0

Information|Instrument|Microscope|Id = Microscope:0

Information|Instrument|Microscope|System = LSM 880 IndiMo, Axio Imager 2

Information|Instrument|Objective|Id = Objective:0

Information|Instrument|Objective|Immersion = Air

Information|Instrument|Objective|LensNA = 0.80000000000000004

Information|Instrument|Objective|Manufacturer|Model = Plan-Apochromat 20x/0.8 M27

Information|Instrument|Objective|NominalMagnification = 20

Information|User|DisplayName = LSM User

Layer|Circle|AcquisitionFlags = 0x80000006

Layer|Circle|CharacterSet = ANSI

Layer|Circle|FontFamily = Arial

Layer|Circle|FontSize = 16

Layer|Circle|FontStyle = Normal

Layer|Circle|FontWeight = Normal

Layer|Circle|Foreground = #FF0000

Layer|Circle|Geometry|CenterX = 609.65418894830657

Layer|Circle|Geometry|CenterY = 735.60071301247763

Layer|Circle|Geometry|Radius = 104.29865025464197

Layer|Circle|Id = 429193156713135684052537903789223236641 7

Layer|Circle|MeasureEnabled = false

Layer|Circle|MouseCatchArea = 6

Layer|Circle|Stroke = #FF0000

Layer|Circle|StrokeThickness = 1

Layer|Circle|TextElement|Id = Annotation

Layer|Circle|TextElement|IsEditable = false

Layer|Circle|TextElement|Position = 706.395722,801.311943

Layer|Circle|Valid = true

Layer|ClosedPolyline|AcquisitionFlags = 0x80000006

Layer|ClosedPolyline|CharacterSet = ANSI

Layer|ClosedPolyline|FontFamily = Arial

Layer|ClosedPolyline|FontSize = 16

Layer|ClosedPolyline|FontStyle = Normal

Layer|ClosedPolyline|FontWeight = Normal

Layer|ClosedPolyline|Foreground = #FF0000

Layer|ClosedPolyline|Geometry|Points = 74.837790,626.081996 83.964349,598.702317 65.711230,574.973262 32.855615,596.877005 40.156863,618.780749 51.108734,644.335116

Layer|ClosedPolyline|Id = 27811853791294510285248300972213962298 77

Layer|ClosedPolyline|MeasureEnabled = false

Layer|ClosedPolyline|MouseCatchArea = 6

Layer|ClosedPolyline|Stroke = #FF0000

Layer|ClosedPolyline|StrokeThickness = 1

Layer|ClosedPolyline|TextElement|Id = Annotation

Layer|ClosedPolyline|TextElement|IsEditable = false

Layer|ClosedPolyline|TextElement|Position = 80.313725,637.033868

Layer|ClosedPolyline|Valid = true

Layer|CoordinateUnitX #1 = Pixel

Layer|CoordinateUnitX #2 = Pixel

Layer|CoordinateUnitY #1 = Pixel

Layer|CoordinateUnitY #2 = Pixel

Layer|CoordinateUnitZ #1 = Pixel

Layer|CoordinateUnitZ #2 = Pixel

Layer|Ellipse|AcquisitionFlags = 0x80000006

Layer|Ellipse|CharacterSet = ANSI

Layer|Ellipse|FontFamily = Arial

Layer|Ellipse|FontSize = 16

Layer|Ellipse|FontStyle = Normal

Layer|Ellipse|FontWeight = Normal

Layer|Ellipse|Foreground = #FF0000

Layer|Ellipse|Geometry|CenterX = 710.04634581105165

Layer|Ellipse|Geometry|CenterY = 857.89661319073082

Layer|Ellipse|Geometry|RadiusX = 34.439916859985672

Layer|Ellipse|Geometry|RadiusY = 32.498755198430516

Layer|Ellipse|Id = 60082145511446049344131144105213316949 0

Layer|Ellipse|MeasureEnabled = false

Layer|Ellipse|MouseCatchArea = 6

Layer|Ellipse|Rotation = 302.00538320808317

Layer|Ellipse|Stroke = #FF0000

Layer|Ellipse|StrokeThickness = 1

Layer|Ellipse|TextElement|Id = Annotation

Layer|Ellipse|TextElement|IsEditable = false

Layer|Ellipse|TextElement|Position = 719.172906,867.023173

Layer|Ellipse|Valid = true

Layer|OpenArrow|AcquisitionFlags = 0x80000006

Layer|OpenArrow|CharacterSet = ANSI

Layer|OpenArrow|FontFamily = Arial

Layer|OpenArrow|FontSize = 16

Layer|OpenArrow|FontStyle = Normal

Layer|OpenArrow|FontWeight = Normal

Layer|OpenArrow|Foreground = #FF0000

Layer|OpenArrow|Geometry|X1 = 0

Layer|OpenArrow|Geometry|X2 = 1023

Layer|OpenArrow|Geometry|Y1 = 512

Layer|OpenArrow|Geometry|Y2 = 512

Layer|OpenArrow|Id = 347964401910836391252901371321696613040

Layer|OpenArrow|LineEndStyle = OpenArrow

Layer|OpenArrow|MeasureEnabled = false

Layer|OpenArrow|MouseCatchArea = 6

Layer|OpenArrow|Stroke = #FF0000

Layer|OpenArrow|StrokeThickness = 1

Layer|OpenArrow|TextElement|Id = Annotation

Layer|OpenArrow|TextElement|IsEditable = false

Layer|OpenArrow|TextElement|Position = 0.000000,0.000000

Layer|OpenArrow|Valid = true

Layer|Usage #1 = Annotation

Layer|Usage #2 = Linescan

Location = E:\Confocal data\13122023\Cytoplasm settings.czi

Scaling|Distance|Id #1 = X

Scaling|Distance|Id #2 = Y

Scaling|Distance|Id #3 = Z

Scaling|Distance|Value #1 = 4.1513291048191956e-007

Scaling|Distance|Value #2 = 4.1513291048191956e-007

Scaling|Distance|Value #3 = 1.5e-006

Version = 1.0

(Fiji Is Just) ImageJ 2.14.0/1.54f; Java 1.8.0_322 [64-bit]; Windows 10 10.0; 150MB of 12208MB (1%)

Title: Cytoplasm settings.czi

Width: 425.0961 microns (1024)

Height: 425.0961 microns (1024)

Depth: 61.5 microns (41)

Size: 82MB

Resolution: 2.4089 pixels per micron

Voxel size: 0.4151x0.4151x1.5 micron^3

ID: -4

Bits per pixel: 8 (color LUT)

Display ranges

 1: 0-255

 2: 8-255

Image: 1/82 (c:1/2 z:1/41 - Cytoplasm settings #1)

 Channel: 1/2

 Slice: 1/41

 Composite mode: "composite"

No threshold

ScaleToFit: false

Uncalibrated

Path: E:\Confocal data\13122023\Cytoplasm settings.czi

Screen location: 161,25 (1920x1080)

Coordinate origin: 0,0,0

No properties

No overlay

No selection