DOCTOR OF PHILOSOPHY

DogTag – a novel genetically encoded proximity labeling strategy for problematic protein-protein interactions

Myles, Sally Mary

Award date:
2023

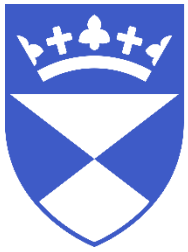
DogTag – a novel genetically encoded proximity labeling strategy for problematic protein-protein interactions

Sally Mary Myles

Doctor of Philosophy

November 2023

The Division of Plant Sciences, School of Life Sciences, University of Dundee.



University of Dundee

Contents

List of Figures

List of Tables

List of abbreviations

APS: Ammonium persulfate

BCA: Bicinchoninic acid assay

DMSO: Dimethyl sulfoxide

DNA: Deoxyribonucleic acid

EDTA: Ethylenediaminetetraacetic acid

LB: Lysogeny broth

LBG: Lysogeny broth plus glucose

NEM: N-ethylmaleimide

PafA: Proteosome accessory factor A

PAT: Protein S-Acyl Transferase

PBS: phosphate-buffered saline

PCR: Polymerase chain reaction

Pup: Prokaryotic ubiquitin-like protein

RLK: Receptor-like kinase

RNA: Ribonucleic acid

SDS: Sodium dodecyl sulfate

SDS-PAGE: Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SOC: Super optimal broth with catabolite repression

TBS: Tris-buffered saline

TEMED: Tetramethylethylenediamine

TLR: Toll-Like Receptor

TM: Transmembrane

Tris: tris(hydroxymethyl)aminomethane

Acknowledgments

I would first like to thank all the members of the Hemsley lab past and present, in particular Piers and Charlotte for the cups of tea, chats, and just the massive amount of help you have both been to me. From supporting me scientifically and personally I'm not sure I could have done it without you both. I hope despite the various storms we have had to weather it has all been worth it Piers!

A big thank you also goes it everyone on the ground floor of B block corridor for your cheery smiles, positive atmosphere, and helpful words of encouragement. When you love what you do and where you're doing it, it really helps. Thanks also go to Kerry and Kara for being there and fully understanding the EastBio experience, the science and logistic chats were crucial to my sanity.

Thanks to all my rugby teammates for keeping me grounded, without the run around and a bit of non-science banter it is quite possible I would have gone stir crazy. In particular Emma, Mandy, and Janet. Thanks for being simultaneously supportive whilst maintaining a "just get on with it "attitude. Thank you to my parents and sisters, a laugh, joke, and a hot meal really can-do wonders.

Lastly thank go to my partner Sandy and daughter Ella, thank you both for allowing me the time and support I needed to do this and for being there during the hard times. Thanks especially to Ella for listening to every presentation I needed to practice.

And lastly a little shout-out to Marvin, thanks for not arriving early. It's made things significantly more straightforward.

Declaration

I declare that I am the author of this thesis and the research described in this work is my own. Any contributions from others are clearly stated and acknowledged. All references cited have been consulted by myself. I confirm that this work has not been previously accepted for any other higher degree.

Sally M. Myles

We certify that Sally M. Myles has fulfilled the relevant ordinance and regulations of the University court and is qualified to submit this thesis for the degree of Doctor of Philosophy.

Dr Piers A. Hemsley

School of Life Sciences,
Dundee University.

Summary of main results

Receptor-Like kinases are responsible for the perception of almost all extracellular stimuli in plants. Two well characterized receptor like kinases are Flagellin Sensitive 2, a receptor involved in plant pathogen immune responses and Brassinosteroid Insensitive 1, a receptor involved in hormone perception and plant growth. Both of these plant proteins have been shown to possess the post translational S-acylation, a dynamic fatty acid modification catalysed by protein S-acyl transferase (PAT) enzymes.

Protein-protein interactions are a key component of cellular processes. Including the interaction between receptor like kinases and PATs. These interactions have been difficult to characterize using traditional protein-protein interaction techniques on account of weak transient interactions between membrane proteins. Through this work I have developed a new proximity labelling technique, DogTag, based on the Pup-IT system that is suitable for tricky membrane interactions in plants. I have optimised aspects of the machinery to enhance sensitivity and improve accuracy.

I have been successful in demonstrating that the DogTag Technique can be used for the study of pairwise interactions, with the successful identification of some promising receptor like kinase and PAT interactions. Including PATs 8, 9 and 12. I have also demonstrated that DogTag works for protein discovery with mass spectrometry. Successfully identifying a number of interesting novel potential BRI1 interacting proteins, including members of the HSP90 family.

Chapter 1

Introduction

1.1 Receptor-like kinase superfamily

In plants the ability to perceive extracellular stimuli is provided by Receptor-like kinases (RLKs). Boasting more than 610 members in Arabidopsis, nearly 2.5% of Arabidopsis protein coding genes, RLKs represent one of the largest gene families in plants (Shiu & Bleeker, 2001).

When compared to mammalian systems, RLKs have the most similarities to Toll-like receptor (TLRs) and interleukin receptor (IL-Rs) families. Due to their involvement in innate immunity by the recognition of pathogen-associated molecular patterns and cytokines they can be seen as analogous in many ways to RLKs. They are also, like RLKs, localised at the cell surface and possess extra cellular and transmembrane domains but, unlike members of the RLK family, do not have intercellular kinase domains. Instead TLRs and IL-Rs rely on a subfamily of cytosolic adaptor proteins that mediate signalling by acting as a kinase domain (Kenny *et al*.,2008)

In comparison to the more limited protein families seen in mammalian systems, the plant RLKs are highly diverse and cover a wide array of functions and as such can also be compared to the range of functions performed by G-protein-coupled receptors (GPCRs) in animals. These transmembrane receptors are responsible for the perception of a ligand resulting in an ATP-dependant cellular response (Hanlon & Andrew, 2015). Plants appear to lack canonical GPCRs and RLKS appear to have filled this niche, highlighting how broadly ranging and important to plant function RLKs are.

RLKs share a conserved structure including an extracellular ligand binding domain, a single pass transmembrane domain and an intercellular serine/threonine kinase domain (Walker,1994, Torii,2000). As a family they possess a wide array of extracellular domains, further indicating a likely wide involvement in plant signal transduction and perception of varied molecules (Shiu & Bleeker,2001)

Only a small portion of RLKs have actually been studied in any detail, but this work does indicate that RLKs are involved in a wide variety of processes, with RLKs being reportedly involved in cell wall sensing, biotic and abiotic stress signalling, plant microbe interactions and plant growth and development. Flagellin Sensitive 2 (FLS2), an RLK involved in the perception of plant pathogens and Brassinosteroid Insensitive 1 (BRI1), an RLK involved in hormone perception and plant growth and development, are both examples of well characterized RLKs.

Despite their vast array of functions, their seemingly highly conserved structure, particularly within the transmembrane and kinase domains, means that there are a lot of functional commonalities between RLKs. One of these is a post-translational modification called S-acylation.

1.1.1 Flagellin Sensitive 2 (FLS2)

Like innate immunity in mammalian systems, plants also possess defence systems against pathogens (Hammond-Kosack *et al*., 1997). Plants do this in part using receptor-like kinases such as Flagellin Sensitive 2 (FLS2) (Gómez-Gómez *et al*., 2000). In Arabidopsis thaliana the ligand for FLS2 is the bacterial

protein flagellin, specifically a 22 amino acid motif within flagellin called flg22. Flagellin is a highly conserved bacterial protein; recognition of flg22 by FLS2 therefore provides the plant with a broad range of resistance to diverse pathogens. Upon binding flagellin, FLS2 rapidly dimerizes with its co-receptor, BRI1-associased kinase1 (BAK1) (Chinchilla *et al*., 2007). This dimerization then results in a series of phosphorylation events triggering downstream signalling and, ultimately, an immune response.

FLS2 has been demonstrated to be both constitutively and dynamically S-acylated, with a rapid increase in S-acylation occurring upon ligand binding in a BAK1 dependant manner (Hurst *et al*., 2019, Hurst *et al*., 2021). This indicates that FLS2 must interact with one or more PATs for this S-acylation to occur. However, interacting PATs for FLS2 have not been identified or confirmed despite a range of genetic and biochemical approaches being tried.

As a well-studied RLK, FLS2 does have a number of identified interacting partners, however the vast majority have been identified as affecting FLS2 related function first with physical interaction confirmed later. One of the best characterised interactions is with its co-receptor BAK1. This interaction was first identified by the use of reverse genetics approach in insertional mutants in order to identify insertions that possessed reduced flg22 sensitivity. Interaction was then confirmed using a co-immunoprecipitation approach (Chinchilla *et al*., 2007). Another confirmed interactor is a member of the MORC superfamily, CRT1, that appears to be involved in plant immunity through DNA recombination and repair as well as chromatin structure. (Kang *et al*., 2012). As with BAK1 this interaction was first suspected after genetic screenings that were then confirmed by Co-Immunoprecipitation. (Kang *et al*., 2012). Other proteins such as BIK1, PBL1, PBL2 and PBS2 involved in PAMP-triggered immunity

have also been shown to interact with FLS2. Again, these were discovered using transcriptomic approaches, with candidates being identified by upregulation upon flg22 treatment. Interaction was then confirmed using co-immunoprecipitation (Zhang *et al*., 2010). This suggests that direct identification of protein-protein interactions of FLS2 is a non-trivial problem.

1.1.2 Brassinosteroid Insensitive 1 (BRI1)

Another important and essential role of RLKs is in hormone perception and plant growth and development. Brassinosteroid Insensitive 1 (BRI1), a brassinosteroid receptor, is one of the receptors responsible for this. The brassinosteroid phytohormones are crucial in the growth and development of plants, controlling a vast array of responses in response to changing environments. This is demonstrated by the array of extreme phenotypes seen in mutants unable to produce or detect brassinosteriods. These phenotypes include dwarfed stature, a reduction in male fertility, round leaves and photomorphogenetic defects (Lozano-Elena & Can˜ o-Delgado 2019). As with FLS2 upon ligand binding, BRI1 forms a heterodimer with its co-receptor BAK1, resulting in an intracellular phosphorylation cascade. This then increases the activity and stability of specific transcription factors, such as BZR1 and BES1 (Li *et al*., 2018), responsible for the regulation of thousands of brassinosteroid responsive genes, involved in many developmental processes within the plant.

Like FLS2, BRI1 has also been shown to be S-acylated and again seems to undergo constitutive and ligand responsive S-acylation (Charlotte Hurst & Silu Li., Unpublished). A family of protein S-acyl transferase (PAT) enzymes are known to be involved in the catalysis of S-acylation. This again makes it likely that BRI1 interacts with one or more currently unidentified PATs.

As another well characterized RLK BRI1 also has a number of well-established interactors. Firstly, it shares the co-receptor BAK1 with FLS2. For the discovery of this interaction the BRI1 kinase domain was used as bait in a yeast two-hybrid screen and the BAK1 kinase domain was identified. No transmembrane or extracellular regions were found in the BAK1 interacting clones, suggesting that these regions of BAK1 cannot be expressed in yeast or prevent the yeast-2hybrid system from functioning. This makes sense as its unlikely the yeast-based system would possess required factors present in the native plant environment to allow for correct expression. BRI1 and BAK1 were also shown to have similar gene expression levels as well as protein localisation. Interaction was again confirmed via co-immunoprecipitation (Nam & Li., 2002). The wider family of somatic embryogenesis receptor kinases (SERKs; BAK1 is SERK3) has also been genetically implicated in mediating early brassinosteroid signalling events, with the interaction suggested using crystal structure analysis. (Santiago *et al*., 2013).

For both FLS2 and BRI1, most well characterised interactions have relied upon a genetic screening approach followed by hypothesis testing using established techniques such as co-immunoprecipitation, and more recently by FRET based assays. This has likely missed many interactors that may act redundantly, show mild or unexpected phenotypes, do not form stable complexes or are inactive in heterologous systems. Despite FLS2 and BRI1 being well characterized this leaves large scope for the identification of interacting proteins likely involved in their function. It is thought that the addition of S-acylation to RLKs may be involved in the activation of RLKs by controlling the localisation of the receptors. Changing the state could signal for trafficking to the membrane, as well as for

degradation to allow for fresh trafficking of receptors, resulting is sustained signalling.

1.2 S-acylation

1.2.1 S-acylation of receptor-like kinases

S-acylation is a dynamic post-translational protein modification whereby fatty acids such as palmitate or stearate are covalently linked to cysteine residues by a thioester linkage (Drisdel & Green,2004). Despite being first identified in the 1970s it remains a poorly understood protein modification (Schmidt & Schlesinger, 1979). It also appears to be highly abundant, with an estimated 40% of all membrane proteins being affected by S-acylation (Hemsley *et al*.,2013). S-acylation, and unlike other known lipid protein modifications, is also reversible, potentially providing means to act as a regulator of protein function in a similar manner to phosphorylation or ubiquitination.

Unlike other lipid modifications however, little is known about S-acylation's role in protein function. It has been described as a membrane anchoring modification but its presence on integral membrane proteins that do not require membrane anchors to associate with membrane environments, and dynamic nature, suggest a greater role in protein function. S-acylation plays possible roles in a number of things including, protein localisation, stability, protein interactions as well as protein function.

In mammalian systems this dynamic regulation of proteins has been more widely characterized with around 80 proteins identified that undergo reversible S-acylation, including roles in protein localisation, dynamic regulation, cell signalling, protein trafficking and metabolic, immune and cancer responses

(Martin *et al*., 2011). In comparison, much less is known about dynamic S-acylation in plants, with work limited to SGN1 and ROP6 showing activation state dependant S-acylation (Alassimone *et al*., 2016, Sorek *et al*., 2017) and ligand mediated S-acylation described for FLS2, with S-acylation being demonstrated as an important step in functional immune signalling (Hurst *et a*l., 2021).

1.2.2 Protein S-acyl transferase (PAT) enzymes

Although free thiols can become S-acylated non enzymatically *in vitro, in vivo* S-acylation occurs more efficiently due to catalysis by protein S-acyl transferase (PAT) enzymes. Originally identified in yeast, PATs are generally considered eukaryotic-specific proteins due to the complexity of the s-acylation process and a lack of prokaryotic structural homologs (Röttig *et al.,* 2013). PATs possess a conserved active site along with 4-6 transmembrane domains. Their N- and C-terminals, however, show large variability, likely involved with substrate specificity (Roth *et al*., 2002, Huang *et al*., 2009, Greaves *et al*., 2010, Gonzalez Montoro *et al*., 2011, Howie *et al*.,2014, Rana *et al*., 2018)

In Arabidopsis 24 PATs have been identified that add S-acyl groups to proteins (Hemsley & Grierson 2008, Batistic, 2012). In mammalian systems a group of acyl-protein thioesterases/palmitoyl protein thioesterases (APTs/PPTs) are responsible for the removal of acyl groups from proteins, making them essential for dynamic S-acylation. These are as yet uncharacterized in Arabidopsis. The combined, opposing, action of PATs and APTs results in the dynamic nature of S-acylation.

Not much is known about plant PATs in comparison to other eukaryotic systems. Transient expression of Arabidopsis PAT-GFP fusions in *N. benthamina* demonstrates a variety of localisations including the plasma membrane, ER, Golgi, intracellular vesicles, and the tonoplast (Batistic 2012). Characterization of a few PATs indicates a wide-ranging effect on plant function, with loss of these enzymes sometimes even resulting in non-viable plants or seeds (Li Y. *et al*., 2016) It is also highly likely given the pleiotropic phenotypes observed in PAT knockout plants that each PAT has a large number of substrates, meaning the loss of one could have wide ranging effects on plant functionality. This idea is supported by the small number of PATs identified in plants compared to the large number of S-acylated proteins discovered. It still remains unclear how PAT substrate specificity is determined.

As S-acylated proteins RLKs must interact with one or more PAT enzymes, but as both are integral membrane proteins this interaction is hard to identify by traditional means. Currently only one study has identified PAT interactions with RLKs; the lectin-RLK subfamily, P2K1 (DORN1) reportedly interacts with PATs 5 and 9 (Chen *et al.,* 2021), but whether this involves S-acylation of P2K1 by PATs 5 and 9 or phosphorylation of PATs 5 and 9 by P2K1 is open to debate.

1.2.3 Membrane environment

With RLKs and many PATs both found in the plasma membrane, understanding and accounting for this environment is crucial to understanding and studying their function. Not much is known about RLK organisation in the membrane, however it has been shown that BRI1 And FLS2 seem to partition into distinct

microdomains within the plasma membrane, with BRI1 also associating with cortical microtubules (Bücherl *et al*., 2017) It also remains unclear as to how FLS2 and BRI1 change in their mobility following activation, with some data suggesting increased mobility and others reporting reduced mobility (Ali *et al*., 2007, Bücherl *et al.,* 2017). PATs have been observed in various membranes of the plant cell including plasma membrane, ER, Golgi, intracellular vesicles, and the tonoplast (Batistic 2012). The defined localisation of these proteins likely indicates that the localisation of PATs is crucial to function.

Many established discovery PPI techniques either involve the disruption of the membrane environment (e.g., detergent solubilisation prior to affinity capture) or are carried out completely removed from the physiological environment in which these membrane proteins interact with each other (e.g., heterologous expression in yeast). This makes it incredibly difficult to detect true interactions or avoid disrupting interactions and may even increase both false negative and false positive rates of identification. As a result, many established RLK interactions are between highly stable interacting partners.

In the case of the RLK P2K1 (DORN1), it is reported to interact with PATs 5 and 9. In a similar way to other RLK interactors they were first identified indirectly, in this case via a screen for P2K1 kinase substrates. This was done using a mass spectrometry-based in vitro phosphorylation strategy; kinase client assay (KiC assay). This initially identified PAT5 and PAT9 as potential substrate interactors. This interaction was then directly tested by techniques such as split-luciferase complementation imaging and bimolecular fluorescence complementation (Chen *et al*., 2021). This approach allowed for the minimisation of membrane disruption, but again was confirmatory, rather than discovery of a PPI.

1.3 Techniques for studying protein-protein interactions.

To understand how cellular processes work it is crucial as part of the discovery process to identify PPIs. Everything from protein synthesis to catabolism relies on a complex range of PPIs (Braun & Gingras 2012). There are many different forms and modes of protein-protein interactions, including: hydrophobic, electrostatic, ligand/co-factor mediated, activation state dependant, bridged, transient and permanent. Most interactions will be a combination of these forms/modes.

Due to the diversity of protein-protein interactions, many different techniques have been developed that are each suited to study different types of interaction. Although some very strong PPIs were discovered in the 1960s it wasn't until the development of gene fusions with epitope tags such as HA, GFP, FLAG and MYC, as well as improved microscopy techniques such as fluorescently tagged proteins, that more in-depth studies of PPIs became possible (Braun & Gingras 2012).

PPI discovery involves identifying potential interactions on a large scale, while PPI validation involves confirming the biological relevance and functional significance of these interactions. Both phases are essential for building a comprehensive understanding of the complex network of interactions that drive cellular processes. With this in mind, many different techniques exist for studying protein-protein interactions, each with its own advantages and disadvantages (Table 1.1). These techniques span Co-Immunoprecipitation (Co-IP), Yeast-2-Hybrid (Y2H), microscopy approaches and proximity labelling techniques.

For proteins that form stable complexes Co-IP approaches can work well once optimised. The use of multiple tags in tandem affinity purification can allow for a more stringent pull down of interacting proteins within complexes compared to the use of a single affinity tag (Beck *et al*., 2014). In terms of higher throughput methods, techniques such as Y2H exist that allow for screening of a library of proteins against a bait protein of interest. Additionally, microscopy-based techniques like florescence resonance emission transfer, bioluminescence resonance energy transfer and bimolecular fluorescence complementation technology are all ways in which PPIs may be studied (Li *et al*., 2017).

However, there are a number of different drawbacks associated with all of these approaches. Co-Immunoprecipitation generally requires a stable interaction, making it unsuitable for identifying weak or transient interactions. Microscopy techniques work well to validate and quantify PPIs but, generally, do not allow the discovery of novel interactors as they require genetic fusion of potential interactors to fluorescent proteins for detection, precluding large scale library screens. Y2H can report on transient or weak interactors but relies upon a high-quality library and functional gene fusions. Y2H, or other heterologous systems, also has a high false positive/negative rate associated with them as they are performed in the absence of a physiologically relevant environment, which is often lacking things like interacting partners, co-factors, appropriate sub-cellular targeting, or membrane organisation (Li *et al*., 2017). Even by combining these approaches, it is still difficult to identify true or transient PPIs and care must be taken to validate findings.

Technique	**Organism**	**Throughput**	**discovery or confirmatory**	**Advantages**	**Disadvantages**
Yeast2Hybrid	Yeast	High throughput	Discovery and Confirmatory	Easy to perform, inexpensive	False positives, not physiologically relevant
Co-IP	plants	Low throughput	Discovery and Confirmatory	Easy to perform, theoretically possible to perform without tags	Not suitable for transient interactions, difficult to use membrane proteins
FRET	plants	Low throughput	Confirmatory	Interaction site visualized, PPI independent of concentration, dynamic analysis possible, quantitative	Expensive equipment and considerable training necessary
Split YFP	plants	Low Throughput	Confirmatory	Not a biochemical approach. Shows PPIs in a living system.	High rate of false positives

Table 1.1 - Table of techniques currently used to study PPIs between plant proteins

1.4 Proximity labelling

To combat the various limitations, present in other PPI analysis techniques, particularly for membrane proteins, in vivo proximity dependent labelling techniques such as BioID, Apex and PUP-IT have been developed. These methods work by proximity and not by interaction and are able to identify weak and transient PPIs, by attaching a protein of interest onto something that allows you to identify what proteins your protein of interest has been proximal to. However, as these systems report on proximity rather than direct interaction, identified interactors may be multiple steps removed from the bait protein in terms of direct PPIs. Depending on circumstance this can be either a help or hindrance, but these methods do allow for discovery work not possible using other strategies.

1.4.1 APEX

An engineered ascorbate peroxidase (APEX) fused to a protein of interest can be used in combination with mass spectrometry as a discovery tool for interactions. APEX works by oxidizing biotin-pheonols into biotin-phenoxyl radicals that react with proximal proteins which become biotinylated, allowing for enrichment by streptavidin and identification by mass spectrometry (Nguyen *et al*., 2019). Some drawbacks associated with APEX include low cellular activity and sensitivity possibly due to poor folding and stability (Lam *et al*., 2015). To combat this APEX2 was developed; this enzyme has a higher catalytic activity and sensitivity (Lam *et al*., 2015). Despite this, the requirement of sufficient biotoxic hydrogen peroxide and provision of exogenous biotin-

phenol for reliable APEX activity still strongly limits its applications *in vivo*, particularly outside of tissue culture (Xu *et al*., 2021).

1.4.2 BioID

Another proximity labelling technique that uses biotinylation is BioID. The basis of BioID is a biotin ligase from E. *coli*, BirA. BirA is highly specific and catalyses the biotinylation of its substrate protein, Biotin Carboxyl Carrier Protein (BCCP), at a specific lysines (Li *et al*., 2017). In order to obtain a more promiscuous biotin ligase a mutant BirA* was developed that lacked the specificity found in the wild type of enzyme. BirA* activates biotin to biotin-AMP, but rather than be retained in the active site, the activated biotin is released, forming a cloud of biotin-AMP able to label free lysines on nearby proteins. Although the BirA* protein of interest (POI) fusion is produced within the cell the addition of biotin to the media is required for the BioID system to work (Li *et al*., 2017). These biotinylated proteins can then be easily captured, exploiting the strong interaction between biotin and streptavidin (Li *et al*., 2017). As biotinylation is a rare protein modification in vivo, most biotinylated proteins will result from proximity to BirA*. Due to of the promiscuity of BirA*, the cloud of activated biotin and the need for a free lysine residue, the radius of biotinlyated proteins around the POI can contain both false negatives and false positives with respect to components of a complex. Because biotinylation in BioID is based on proximity and not interaction, true interactors can be missed (no free lysine) and false interactors can be labelled (free lysine and in close proximity). In order to establish true protein-protein interactions from random labelling well designed background controls and numerous repeats can help. In principle BioID is one of the best current methods for establishing a “snapshot” of transient PPIs and

its sensitivity allows for the identification of hundreds of interacting proteins for a single POI (Li *et al*., 2017). However, in practice BioID requires long periods ~24 hours to achieve detectable labelling, making it more suited for mapping large stable complexes, such as the nuclear pore, rather than transient interactions (Roux *et al*.,2012).

1.4.3 Proximity Labelling improvements

To overcome some of these draw backs various improvements to the system have been made such as BioID2, TurboID and miniTurbo. BioID2 utilised a smaller biotin ligase which allowed for more-selective targeting of fusion proteins, a lower biotin requirement and improved labelling of proximal proteins (Kim *et al*., 2016). TurboID, and a smaller version termed miniTurbo, were developed using a directed evolution approach. These biotin ligases were capable of significantly faster labelling and functioned at lower temperatures than the original BioID (May *et al*., 2020). Although this provided some significant advantages over the original system as a technique there still remain some issues. BioID detects proximity and doesn't necessarily detect directly interacting proteins, particularly in the case of freely diffusible BirA-variant tagged proteins that leave a trail of activated biotin as they move. Because of this it still identifies mostly abundant proteins in stable complexes and can't be used as a true quantitative approach (Lin *et al.*, 2017). One of the biggest problems that exists with all current techniques is the ability to study PPIs in vivo in intact organisms such as plants. In order to identify physiologically relevant PPIs an ideal system would need to meet a number of criteria:

1. Can be used in a whole organism context

2. Can identify and record PPIs without outside intervention

3. Be a genetically encoded system

4. Can identify a breadth of PPIs including weak and transient interactions

5. Is independent of protein type (soluble, membrane, etc.)

An ideal system that meets these requirements would have a genetically encoded protein component that would be capable of marking or capturing an interacting protein.

1.5 Pupylation

Prokaryotic ubiquitin-like protein (Pup) is a protein that is post translationally added to proteins by a peptide ligase enzyme PafA in a process known as pupylation. This system, analogous to the eukaryotic ubiquitin ligase system, is found in a number of prokaryotes but is best studied in *Mycobacterial* and *Corynebacterial* species. Similarly, to ubiquitination, pupylation also acts to target proteins for degradation. Proteasome accessory factor A (PafA) catalyzes the pupylation of lysines on target proteins. Pup encoding genes are found in two forms depending on the bacterium; one encodes Pup with a C-terminal glutamine (PupQ) and the other Pup with a C-terminal glutamate PupE. PupE is the catalytically relevant substrate and PupQ is converted to PupE post-translationally via glutamine deamidation by the action of Deamidase of Pup (DOP). PafA interacts with the glutamate form of Pup (PupE) and is the only enzyme that carries out pupylation. Unlike some other post translational modifications PafA doesn't seem to recognise a motif and instead seems to be capable of pupylating any available lysine present on target proteins. Pupylation is reversible in bacteria, with de-pupylation performed by DOP, the enzyme that deamidates PupQ.

Despite working in a similar way to ubiquitination in eukaryotes neither PupE nor PafA possess any homology to any ubiquitin ligases or ubiquitin found in eukaryotes. This means it's unlikely to interact with any eukaryotic proteins. This, combined with the lack of a recognisable Pupylation motif, may indicate that PafA would be a good tool for the study of PPIs possessing high activity, low specificity, and negligible interaction with the host biochemistry.

1.5.1 PUP-IT

Despite working in a similar way to ubiquitination in eukaryotes, neither PupE nor PafA possess any homology to any ubiquitin ligases or ubiquitin. This means it's unlikely to interact with any eukaryotic proteins. Expression of free PafA and PupE in *Escherichia coli* (Liu *et al*., 2018), yeast and plants (Hemsley, unpublished) result in pupylation of host proteins suggesting that recombinant PafA shows high activity in heterologous systems. Prior to the start of the project work in the Hemsley lab has shown expression of PupE with PafA fused bait proteins in yeast results in pupylation of distinct subsets of host proteins, indicating that this system to some extent works in yeast and plants and that fusion to bait proteins can impart specificity to the PafA pupylation reaction. It has been shown that only the conserved C-terminal helix of PupE is actually required for conjugation onto proteins by PafA. This lack of conserved sequence outside of this helix makes the conserved C-helix an attractive scaffold for attaching affinity tags for purification and identification of Pupylated proteins.

The pupylation system has the potential to be developed into an attractive new technique for the study of PPIs and this formed the basis of my PhD project. However, in August 2018 Liu et al. published the basic characterization and

application of the PUP-IT system (Liu *et al*., 2018). The principles of the system aligned with the principles of the system we were trying to develop; however, no work was carried out in plants. This system can potentially be successfully developed into a more versatile and accurate PPI technique having a number of advantages over existing PPI methods. Due to PafA's high apparent activity the system has the potential to identify weak and transient interactions with reduced random labelling. This is due to PafA's natural low absolute specificity being spatially restrained by fusion to a protein of interest, thereby limiting labelling to proteins interacting or closely associated with the bait protein (figure 1.1). Because all the components can be genetically encoded and assembled within the cell it can be done in a closed system such as whole organisms. This opens the possibility for examining interactions in a tissue or cell type of choice during developmental stages, stimulus, or challenge in any genetically tractable organism such as most plants, vertebrate model systems like mice or zebra fish or simpler systems such as drosophila or C. *elegans*. Compared to existing methods in a similar vein, such as BioID that involves the addition of biotin, the pupylation system avoids the need for the use of exogenous substrates and allows for truly physiologically relevant results.

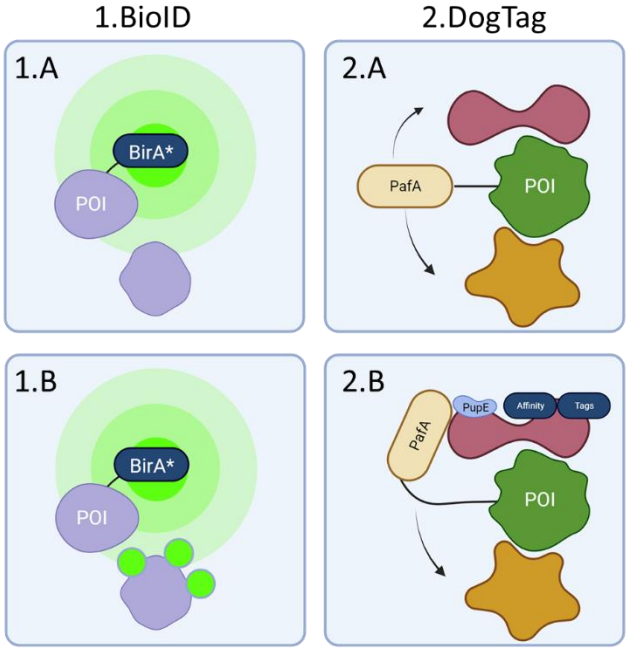


Figure 1.1. BioID and DogTag work differently. Although both proximity labelling techniques, the mechanisms of BioID and DogTag do differ.1.A BioID, protein of interest is fused to the BirA* enzyme , resulting in a cloud of activated biotin around BirA. 1.B any proximal proteins within this radius of biotin are biotinylated and can then be identified. 2.A DogTag, PafA is fused to a protein of interest on a flexible linker. 2.B for pupylation of proximal proteins PafA, unlike BirA*, must make direct contact.

1.5.2 Potential impact of Pupylation-based proximity labelling in biological systems

PPIs are crucial in every aspect of cellular plant processes and the further development of pupylation as a PPI tool will be especially effective in the context of whole multicellular organisms. Because of its encoded nature, and its high natural specificity, pupylation has the potential to give physiological results from whole multicellular organisms without random labelling, having a huge impact on the study of PPIs.

One area in which this technology could be of interest would be drug development. Currently the therapeutic alteration or disruption of PPIs is an important aspect in drug discovery. At present 60% of all approved drug targets are integral membrane proteins despite membrane proteins only making up a quarter of the total proteome. Interactions involving membrane proteins can often be weak or transient additionally membrane protein interactions ideally need to be studied in a physiologically relevant environment. This highlights the need for a whole organism system that can successfully identify these types of PPIs.

Another area in which they are of particular interest is there involvement in how plants perceive and respond to pathogens. An interaction between a plant and pathogen involves two major groups of proteins. Receptor-like kinases which detect external pathogenic stimuli and R- proteins which detect intercellular stimuli such as changes in cellular activity due to pathogen-derived proteins called effectors. Existing methods for the study of PPIs have been used in this area but because of the limited nature of these techniques aspects pant pathogen interactions are still not fully understood. A strong understanding of all the various PPIs involved in this is crucial to fully understand processes

underlying plant resistance or susceptibility to pathogens. Developing a new more flexible and sensitive tool to study these PPIs invaluable.

S-acylation, the addition of fatty acyl groups to cysteine residues, is an emerging post-translational regulatory mechanism that is carried out by protein S-palmitoyl transferases (PATs). It has been demonstrated that both BRI1 and FLS2 appear to be dynamically S-acylated (Hurst *et al*., 2021, Hurst & Li unpublished). Due to BRI1 and FLS2 being membrane proteins and PATs being redox sensitive integral membrane enzymes and the apparent transient weak nature of their interaction it makes them poor candidates for current established approaches. With many techniques requiring harsh conditions resulting in a disruption of the native membrane environment and reduced physiologically relevant results. With PUP-ITs ability to detect weak and transient interactions as well as its potential to work for membrane proteins it presents an opportunity to further understand the interaction between BRI1, FLS2 and their associated PATs.

The goal of this project is therefore the development of a new proximity-based protein-protein interaction technique based on PUP-IT in plants using receptor like kinases (RLK) FLS2 and BRI1 as test proteins. Ideally this technique would have various benefits over current techniques including use in a whole organism context, no requirement for outside intervention, being a genetically encoded system, applicable to a variety of PPIs including weak and transient interactions and independent of protein type (suitable for soluble and membrane proteins).

I aim to use the system to identify pairwise interactions, the S-palmitoyl transferase that S- acylates FLS2 and BRI1, and as a new discovery tool to identify other proteins that interact with RLKs.

Chapter 2

Materials and methods

2.1 Plant culture, growth, and transformation.

2.1.1 *Nicotiana benthamiana* growth conditions

Nicotiana benthamiana Domin (*N. benthamiana*) seeds were germinated and grown on soil with one plant per 10 cm pot under 16h light: 8h dark cycles with illumination of 130-150 µE.m^{-2}. Plants were grown under controlled conditions in a greenhouse for 4-5 weeks at 24 °C and 40-65 % relative humidity prior to use.

2.1.2 Transient transformation of *N. benthamiana*

A single *Agrobacterium tumefaciens* colony was used to inoculate 5 mL LB containing appropriate antibiotic selection. The culture was incubated at 28 °C overnight with vigorous shaking (225 rpm). 2 mL culture was pelleted and resuspended in 500 µL infiltration buffer (10 mM MES pH 5.7, 10 mM $MgCl_2$, 10 µM acetosyringone) and incubated at room temperature for 3 h with gentle mixing. Bacteria were diluted to a final OD_{600} 0.1 with the gene silencing viral suppressor, P19 (Voinnet *et al*., 2003) at a final OD_{600} 0.1 and syringe infiltrated into 5-week-old *N. benthamiana* leaves. Leaves were harvested after 48 h, flash frozen in liquid nitrogen and stored at -80 °C until use.

2.1.3 Vacuum infiltration of chemicals and inhibitors into *N. benthamiana*

Constructs containing the gene of interest were transiently expressed in *N. benthamiana* by syringe infiltration. 48 h post-infiltration, leaf disks were cut from each leaf using a 5 mm cork borer. Leaf disks were washed gently in

infiltration buffer for 10 minutes with continuous mixing to remove DAMPs. Leaf disks were vacuum infiltrated 2x 1 minute at 10 kPa with fresh infiltration buffer and transferred to multi well plates and treated with N-Ethylmaleimide. They were then left to incubate at 20 °C in illuminated Panasonic MLR-350 growth chambers for 6h. Leaf disks were then harvested and flash frozen.

2.2 Bacterial & Yeast strains

2.2.1 E. *coli* genotypes

NEB5α: *fhuA2 Δ(argF-lacZ) U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*

DH10B: F^- *endA1 deoR*$^+$ *recA1 galE15 galK16 nupG rpsL* Δ*(lac)X74* φ80*lacZΔM15 araD139* Δ*(ara, leu)7697 mcrA* Δ*(mrr-hsdRMS-mcrBC)* Str^R λ^-

2.2.2 A. *Tumefaciens* genotypes

GV3101 pMP90: Contains streptomycin and rifampicin selectable markers. The pMP90 helper plasmid confers gentamycin resistance and contains a deletion of pTiC58 TDNA (Van Larebeke *et al*., 1974).

2.2.3 *S. cerevisiae* cell genotypes

BY4742: MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0

2.2.4 Antibiotic selection media

Lysogeny Broth (LB) media with 0.8 % agar was melted and allowed to cool to ~55 °C to prevent heat-labile antibiotics degrading. Under sterile conditions, appropriate antibiotics were then added to the cooled media and mixed thoroughly. Under sterile conditions, 25 mL media was poured into 9 cm diameter petri dishes and allowed to set. Plates were stored at 4 °C until use for a maximum of 7 days.

2.2.5 *E. coli* transformations

For transformations, competent cells were defrosted on ice for 30 minutes. 150-400 ng of DNA was then added, and cells incubated for another 30 minutes to ensure proper mixing before the transformations were carried out according to table 2.2. After transformation cells were put on ice for 2 minutes to recover before the adding 5 volumes of SOC media (2 % w/v tryptone, 0.5 % w/v yeast extract, 10 mM NaCl, 2.5 mM KCL, 10 mM $MgCL_2$, 10 mM $MgSO_4$ and 20 mM glucose pH7.0 with NaOH). Cells were then incubated at 37 °C in an orbital shaker at 225 rpm for 1 hour before being plated out on LB agar with the appropriate antibiotics. Plates were dried and incubated for 16 hours at 37 °C in a static incubator.

Cell line	Competency	Used for	Cell volume	DNA	Transformation method	Native antibiotic resistance
NEB5α	Chemical	Site-directed mutagenesis	25 µL	1 µL	42°C for 30 s	none
DH10B	Electro	Various	50 µL	2 µL	2800 V 0.1 cm plate gap	Streptomycin
GV3101 pMP90	Electro	Expression vectors	50 µL	1 µL	2800 V 0.1 cm plate gap	Rifampicin & gentamycin

Table 2.2 Procedures for bacterial transformation

2.2.6 A. *tumefaciens* transformation

For transformations competent cells were defrosted on ice for 30 minutes. 50 µL of GV3101 pMP90 were transformed with 150-400 ng of DNA and were incubated for 30 minutes to ensure proper mixing and then transformed by electroporation at 2800 V (Table 2.2). Cells were then left to recover on ice for 2 minutes before 1 mL of LB media was added. Cells were then incubated on an orbital shaker at 225 rpm at 28 °C for 3 hours before being plated on LB agar and appropriate antibiotics and incubated at 28 °C in a static incubator for 48 hours.

2.2.7 *S. cerevisiae* Transformation

For the transformation of yeast, the Lithium aetate/SS carrier DNA/PEG method of Gietz and Woods (2006) was used. A YPAD plate (1 % yeast extract, 1 % peptone, 2 % glucose, 0.004 % adenine hemisulfate solidified with 2 % agar) was streaked with a yeast glycerol stock stored at -80 °C for single colonies and incubated at 28 °C for 72 hours in a static incubator. A single colony was picked to a starter culture in YPAD liquid media in 5 mL volume and incubated overnight at 28 °C in an orbital shaker at 225 rpm. The OD600 was then measured after 16 hours and used to inoculate 25-50 mL of pre-warmed YPAD to a starting OD600 of 0.5. the culture was then incubated for a further 3 hours at 28 °C in an orbital shaker at 225 rpm until OD600 2.0 (late-logarithmic phase

of growth). Once in late-log phase the culture was transferred into a clean 50 mL centrifuge tube and cells pelleted at 3000 xg for 5 minutes.

The yeast were then washed once with sterile distilled water and spun again in the same conditions. The water was then removed, and the yeast pellet resuspended in 1 mL of 100 mM lithium acetate (LiAc) and transferred into a 1.5 mL microfuge tube. The yeast was then pelleted again with a brief spin and resuspended in 100 mM LiAc achieving a final volume of 250-500 µL (dependant on size of original culture). The suspension was then aliquoted out into 50 µL parts in microfuge tubes for individual transformations. Before transformation, the cells were briefly pelleted with a 30 second centrifugation at 6,000 rpm and the LiAc removed. The following was then layered on top of the pelleted yeast in order: 240 µL PEG3300 50 % (w/v), 36 µL 1 M LiAC, 50 µL pre-boiled and snap-cooled salmon sperm DNA (2 mg/mL in TE), and a mix of linearised plasmid DNA, PCR product and water for a combined volume of 34 µL. The tubes were then vortexed to resuspend the yeast and mix the components fully before being placed in a 30 °C water bath for 30 minutes followed by a 30 minute 42 °C heat shock. Transformations were then pelleted briefly by 30 second centrifugation at 6,000 rpm. The transformation mix was carefully pipetted off and the pellet resuspended gently in 1mL of sterile distilled water. Between 50 µL- 200 µL was then plated on appropriate drop out selection media and incubated at 28 °C in a static incubator for 72 hours.

2.3 Nucleic acid manipulation

2.3.1 Standard PCR

Standard polymerase chain reaction (PCR) was carried out using gene-specific oligonucleotide primers (Sigma, UK) and Quick-load One*Taq* DNA polymerase

with standard buffer (NEB, Ipswich, USA) using the cycling conditions shown in table 2.3 Unless otherwise stated, the final primer concentration was 0.2 µM and the annealing temperature was 55 °C. PCRs were carried out in a G-storm GS1 thermal cycler (G-Storm, Somerset, UK). PCR product was separated by agarose gel electrophoresis.

Taq Enzyme	Initial denaturing temp and time	Denaturing temp and time	Extension temp and time	Final extension	N° cycles
OneTaq	94 °C for 2 min	94 °C for 30 s	1 min/kb 68 °C	68 °C 5 minutes	35
Q5	98 °C for 30 s	98 °C for 30 s	30 s/kb 72 °C	72 °C 10 minutes	25
Phusion	98 °C for 30 s	98 °C for 30 s	30 s/kb 72 °C	72 °C 10 minutes	20

Table 2.3 Cycling conditions for PCR

2.3.2 High fidelity PCR

Gene-specific oligonucleotide primers (Sigma) were used to amplify DNA fragments from 1 µL template genomic DNA or cDNA. Plasmid DNA was diluted 1:10 - 1:100 in water and 1 µL of the dilution used as template in the PCR reaction.

Primers were designed using the Primer3 programme (http://primer3.ut.ee/). Listed on pages 151 to 158. Genes were amplified using the proof reading Q5 enzyme (NEB) or the proof reading Phusion enzyme (Thermo), using the high GC enhancer buffer if required for template with high GC content.

2.3.3 Colony PCR of *E. coli*

E. coli colonies were screened for correct vector recombination or ligation by standard PCR but 1 µL volume used for template DNA was replaced with 1 µL water in the PCR master mix. Each colony was numbered, carefully selected with a different sterile pipette tip and restreaked on LB agar containing the appropriate antibiotics. Each tip was then used to spike a PCR reaction. The restreak plate was incubated at 37 °C overnight.

For pENTR D-TOPO vector screening, a gene specific forward primer and the M13 reverse primer were used to screen for colonies with inserts in the correct orientation. For expression vectors, a gene-specific forward primer and a vector-specific reverse primer were used to screen for colonies containing the insert in the correct orientation.

2.3.4 Colony PCR of *S.cerevisiae*

To screen for correct vector recombination or ligation *S.cerevisiae* colonies were screened by standard PCR. Transformants were initially re-steaked on selective plate prior to picking. This allowed for a larger streak of the colony to be used in the PCR reactions. Colonies were resuspended in 20 µL of 0.02 M NaOH and heated at 95 °C for 1 hour. 1-2 µL of this boiled yeast suspension was used as the template for the PCR after clarification by centrifugation. Gene specific primers were used alongside yeast housekeeping gene primers to test for DNA extraction and desired recombinations.

2.3.5 Site Directed Mutagenesis (SDM)

SDM was carried out using the Q5® Site-Directed Mutagenesis Kit. Primers were designed using the NEBaseChanger™ online NEB primer design software.

For PCR the following reagents were assembled in a thin-walled PCR tube (Table 2.4).

	25 µl RXN	**FINAL CONC.**
Q5 Hot Start High-Fidelity 2X Master Mix	12.5 µl	1X
10 µM Forward Primer	1.25 µl	0.5 µM
10 µM Reverse Primer	1.25 µl	0.5 µM
Template DNA (1–25 ng/µl)	1 µl	1-25 ng
Nuclease-free water	9.0 µl	

Table 2.4 – SDM PCR reagents

The reagents were mixed completely and then transferred to a thermocycler performing the following cycling conditions (Table 2.5) .

Step	Temp	Time
Initial Denaturation	98°C	30 seconds
25 Cycles	98°C	10 seconds
	50–72°C*	10–30 seconds
	72°C	20–30 seconds/kb
Final Extension	72°C	2 minutes
Hold	4–10°C	∞
Table 2.5- SDM PCR cycling conditions		

Q5-optimized annealing temperatures for mutagenic primers were determined using the NEBaseChanger™ online NEB primer design software.

Kinase, Ligase & DpnI (KLD) treatment was done by assembling the following reagents (Table 2.6).

	Volume	Final Conc.
PCR Product	1 µl	
2X KLD Reaction Buffer	5 µl	1X
10X KLD Enzyme Mix	1 µl	1X
Nuclease-free Water	3 µl	
Table 2.6 – KLD reagents		

This was mixed well by pipetting up and down and incubated at room temperature overnight (12 -16 hours). This was stored at -20 °C until use for transformation of *E. coli*.

2.3.6 *S.cerevisiae* plasmid extraction

Correct colonies identified by colony PCR were picked and used in the inoculation of 5 mL overnight cultures in drop out selection media. These were incubated for 16 hours at 28 °C in an orbital shaker at 225 rpm. 2-3 mL of the overnight culture was then centrifuged at 10,000 rpm for 10minutes and the pellet resuspended in 100 µL of yeast protoplasting buffer (60 µL 2 M sorbitol, 50 µL 1 M Tris-HCL pH7.5, 20 µL 0.5 M EDTA and 0.7 µL beta-mercaptoethanol per mL in sterile distilled water). 10 µL lyticase (50 U/µL) was added and incubated at 37 °C in a water bath for 1 hour in order to digest the

yeast cell walls. After protoplasting, yeast cells were collected by centrifugation at 900 rpm. Cells were then processed as for bacterial minipreps.

2.3.7 Plasmid isolation from *E. coli* (Miniprep)

The Wizard® Plus SV Miniprep kit was used for small scale DNA preparation.

Cell Lysis Solution	**Cell Resuspension Solution**
0.2M NaOH 1% SDS	50mM Tris-HCl (pH 7.5) 10mM EDTA 100µg/mL RNase A
Neutralization Solution	**Column Wash Solution**
4.09M guanidine hydrochloride 0.759M potassium acetate 2.12M glacial acetic acid Final pH is approximately 4.2.	60mM potassium acetate 8.3mM Tris-HCl (pH 7.5) 0.04mM EDTA (pH 8.0) 60% EtOH

Table 2.7 Miniprep solutions

A single, well isolated colony from a fresh LB plate was inoculated in 4-5 mL of LB medium with appropriate antibiotics and grown overnight (12 -16 hours) at 37 °C with shaking. 3 mL of culture was harvested by centrifugation in a microfuge tube at 6,000 rpm for 5 minutes. Supernatant was poured away, and inverted tubes blotted on a paper towel to remove excess media. 250 µL of cell resuspension solution was then added and the cell pellet was completely resuspended by pipetting. 250 µL of cell lysis solution was then added and the tubes mixed by inversion four times. The tubes were then incubated at room temperature for 5 minutes until the suspension cleared. 10 µL of Alkaline Protease Solution was then added before being mixed by inversion four times and incubating at room temperature for 5 minutes. After incubation 350 µL of neutralization solution was added and immediately mixed by inversion. The bacterial lysate was then centrifuged at maximum speed in a microfuge for 10 minutes at room temperature. Spin columns were inserted into 2mL collection tubes. The cleared lysate was then decanted into the spin column without

disturbing or transferring any of the white precipitate with the supernatant. The supernatant was then centrifuged at maximum speed using a microfuge for 1 minute at room temperature. The spin column was removed and the flow through discarded. The spin column was then reinserted into the collection tube and 750 µL of column wash solution added. Again, this was centrifuged at maximum speed in a microfuge for 1 minute at room temperature. The spin column was removed to allow the discard of flow through before being reinserted. The wash procedure was then repeated using 250 µL of column wash solution. It was then centrifuged at maximum speed in a microfuge for 2 minutes at room temperature. The spin column was then transferred to a new sterile 1.5 mL microcentrifuge tube taking care not to transfer any of the column wash solution. The DNA was then eluted by adding 50 µL of nuclease-free water to the spin column and centrifuging at max speed in a microfuge for 1 minute at room temperature. After elution the spin column was discarded, and the eluted DNA stored at –20 °C.

2.3.8 Gel electrophoresis of DNA

PCR products were separated by agarose gel electrophoresis in TBE buffer (220 mM Tris-HCl, 180 mM boric acid, 5 mM EDTA, pH 8.3, 0.01 % v/v SYBR safe [Thermo, Paisley, UK]). HyperLadder 1 (Bioline, London, UK) or 50 bp DNA ladder (NEB) were used as molecular weight markers. Bands were visualised using a UVP BioDoc-It™ system equipped with a blue light Safe Imager (Invitrogen, Paisley, UK), a Cannon J6x11 11-70mm f1.4 TV zoom lens (Cannon, Tokyo, Japan) and documented using a UPP-110S video graphic printer.

2.3.9 Gel extraction of DNA

Products were separated by agarose gel electrophoresis. Bands were visualised using an S37102 SafeImager transilluminator (Invitrogen) equipped with a blue safelight and the band of interest excised from the gel. DNA was extracted using the QIAquick gel extraction kit (QIAGEN). Briefly, the gel slice was weighed and dissolved in 3 volumes (where 100 mg = 100 µL) of QG buffer with gentle mixing at room temperature. 1 volume of 100 % isopropanol was added, and the tubes mixed by inversion. The solution was passed through a nucleic acid binding column by centrifugation for 1 minute at 10,000 rpm and the flow-through discarded. The column was washed with 500 µL QG buffer and centrifuged as before. 750 µL buffer PE reconstituted with 100 % ethanol was added to the column and left to stand for 5 minutes to remove excess salt. The PE buffer was removed by centrifugation for 1 minute at 10,000 rpm. 20 µL buffer EB (10 mM Tris-HCl pH 8.5) was added directly to the membrane and left to stand for 3 minutes. The column was transferred to a new 1.5 mL microfuge tube and centrifuged as before to elute the DNA. Recovery of PCR product was measured using a NanoDrop ND-1000 UV-Vis spectrophotometer (Thermo) at an absorbance of 260 nm. Recovery of DNA was also measured by agarose gel electrophoresis.

2.3.10 Plasmid modification by oligonucleotide annealing

Self-complimentary oligonucleotides (Sigma) were designed with 5’ and 3’ overhangs complimentary to a restriction site of choice in the base vector. Oligonucleotides were combined to a final concentration of 25 µM in 50 µL 2 M NaCl solution. Oligonucleotides were incubated at 95 °C for 5 minutes in a hot

block and then annealed by allowing the hot block to slowly cool to room temperature.

2.3.11 Oligonucleotide ligation

2 µg vector was digested for at least 6 h at 37 °C with the appropriate enzyme(s) in a 20 µL reaction volume. Digested vector was purified by gel extraction as previously described. Freshly annealed oligonucleotides (at final concentration of 25 µM. See 2.3.11 oligonucleotide annealing) were diluted 1:100 in water. 1 µL of diluted annealed oligonucleotides were ligated into 250-500 ng digested vector using 400 units of T4 DNA ligase (NEB) in 1 x T4 ligase buffer (NEB) in a 20 µL reaction volume. The reaction was incubated in a water bath at 10 °C and allowed to warm up to room temperature overnight. After ~16 h, 10 units of the original restriction enzyme used to digest the vector was added to the reaction and incubated for a further 8 h at 37 °C to linearise any original vector and reduce non-recombinant background. The ligation was transformed into electrocompetent DH10B cells.

Page Break

2.3.12 LR recombination reaction

Gateway® LR reactions were used for recombination between an attL-containing entry clone and an attR-containing destination vector to generate an expression clone.

The following components (Table 2.8) were added to a thin walled 0.2 mL PCR tube at room temperature and mixed.

Entry Clone (supercoiled, 100-300ng)	0.5ul
Destination vector (supercoiled 150ng/ul)	0.5ul
5X LR Clonase™ reaction buffer	1ul
TE Buffer pH8.0	3ul

Table 2.8 – LR reagents

LR Clonase™ enzyme was vortexed briefly to mix and 1 µL added to the above components and mixed well by vortexing briefly twice. The reaction was then incubated at room temperature overnight (12 -16 hours). After incubation 1 µL of 2 µg/µL of Proteinase K solution was added and incubated at 37 °C for 10 minutes. This was then transformed into competent *E. coli* and selected for using appropriate antibiotic-resistance for expression clones.

2.3.13 Diagnostic restriction digests

Restriction digests of plasmid DNA were carried out in a 15 µL reaction volume containing 10 units of each enzyme (NEB), 2 µL plasmid (150-400 ng) DNA and 1x Cutsmart buffer (NEB).

Restriction digests of PCR products were carried out in a 10 µL reaction volume containing 5 µL PCR product, 10 units of each enzyme (NEB) and 1x Cutsmart buffer (NEB).

Samples were incubated at 37 °C for 1 h. After 1 h, tubes were briefly centrifuged to collect the sample and 6 x loading dye (NEB) was added to each sample and mixed vigorously. Restriction digests were separated by gel electrophoresis.

2.4 Protein analysis methods

2.4.1 Direct extraction of proteins into SDS-PAGE loading buffer

Three *N. benthamiana* leaf discs, transiently expressing the protein of interest, were ground to fine powder in liquid nitrogen. 150 µL 1 x Laemmli sample buffer (with 2.5 % v/v β-mercaptoethanol) was then added. Samples were ground a second time in Laemmli sample buffer to ensure total tissue disruption. Samples were centrifuged for 1 minute at 10,000 rpm and 100 µL supernatant was removed to a fresh tube. Supernatant was stored at -20 °C until use.

2.4.2 Crude Protein extraction

Nicotiana benthamiana leaf tissue was ground to a fine powder in liquid nitrogen. 5 mL of extraction buffer containing 0.5% Sarkosyl was then added along with protease inhibitors. This was then solubilised for 10 minutes with occasional mixing on ice. It was then centrifuged at max rpm in a swing bucket rotor for 5 minutes before the supernatant was gently passed through 2 layers of miracloth into a fresh 15mL tube. Samples were then centrifuged at 12,000 rpm for 10 minutes in a fixed angle rotor. The protein concentration was determined by BCA assay., and stored at -80 °C.

2.4.3 BCA assay

BCA assays were carried out on total protein extract from *N. benthamiana*. Extracted proteins were diluted 1:10 in extraction buffer. 0, 0.125, 0.25, 0.5, 1, 2 $mg.mL^{-1}$ BSA standards were made up in extraction buffer.

A master mix of 10 µL reagent B was mixed with 500 µL reagent A per sample and standard in a 15 mL falcon tube and mixed vigorously. 500 µL of the master mix was dispensed into the diluted samples and standards and mixed vigorously. Samples were incubated for 30 minutes in a 37 °C water bath. Absorbance was measured using a biophotometer spectrophotometer UV/Vis reader (Eppendorf) at 562 nm. Protein concentration was calculated using a linear regression line through the standard 0.0125 – 2 $mg.mL^{-1}$ points.

2.4.4 Immunoprecipitation of proteins with streptavidin

Immunoprecipitation of proteins was done using TwinStrep enrichment. This was performed using 3 mg of total protein extracted from *N. benthamiana*. 10 µL equivalent (200µL of supplied suspension) of magstrep type 3 XT beads (IBA life sciences) per sample were first added to microfuge tube and placed on

a magnetic separator to separate beads to remove supernatant. The beads were then resuspended in extraction buffer to wash. This was repeated 3 times. After removal of the final wash the magnetic beads were resuspended in 3mg of protein extract. This was then incubated at room temperature with gentle mixing for 120 minutes. Reaction tube was then placed back in the magnetic separator and the supernatant carefully removed. The beads were then subsequently washed three times with wash buffer before either being eluted in 1 x Laemmli sample buffer (with 2.5 % v/v β-mercaptoethanol) for analysis by Western blot or resuspended in 18 ul of 20 mM Tris-HCl pH 8.0 for analysis by mass spectrometry

2.4.5 SDS-PAGE separation of proteins

Samples in SDS-PAGE sample buffer were all heated in a shaking hot block (1500 rpm) to 65 °C for 5 minutes, briefly collected by centrifugation and loaded into wells of the stacking gel (Table 2.9) 2 µL pre-stained protein PageRuler Plus (Thermo) was used as a molecular weight marker.

	7.5 % resolving gel	**15% resolving gel**	**Stacking gel**
Resolving buffer pH 8.8	4 mL	4 mL	-
Stacking buffer pH 6.8	-	-	1.25 mL
30 % bisacrylamide	4 mL	8 mL	0.75 mL
Sterile distilled water	8 mL	4 mL	3 mL
Ammonium persulfate (10% in water)	80 µL	80 µL	30 µL
TEMED	8 µL	8 µL	10 µL

Table 2.9 Recipe for SDS-PAGE gels

Proteins were separated using the Laemmli system (Laemmli, 1970) by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) in SDS-PAGE running buffer (25 mM Tris-HCl, 192 mM glycine, 10 % v/v SDS). 1.0 mm thick gels of the appropriate percentage acrylamide (see Table 2.9) for resolving the proteins of interest were used. Empty wells were filled with blank loading buffer to maintain the ionic gradient across the gel. 10 mA current per gel was applied while proteins migrated through the stacking gel. Once the loading dye entered the resolving gel, the current was increased to 20 mA per gel until the dye front eluted from the gel.

2.4.6 Tris-Tricine SDS PAGE

Samples were all heated in a shaking hot block (1500 rpm) to 95 °C for 5 minutes, briefly collected by centrifugation and loaded into wells of the stacking gel. 2 µL Ultra Low Molecular Weight Marker (Sigma-Aldrich) was used as a molecular weight marker.

	Resolving	**Stacking**
2.5M Tris buffer(pH 8.8)	5.6mL	0.76 mL
30 % bisacrylamide	3.33 mL	0.66 mL
Sterile distilled water	3.42 mL	0.9 mL
Ammonium persulfate (10%)	150 µL	150µL
TEMED	6µL	5µL

Table 2.10 recipe for Tris-Tricine SDS PAGE

Proteins were separated using the Schägger system (Laemmli, 1970) by Tris-Tricine sodium dodecyl sulphate-polyacrylamide gel electrophoresis (Tris-Tricine SDS-PAGE) in Tris-Tricine SDS-PAGE running buffer (100 mM Tris, 100 mM Tricine 0.1% (w/v) SDS). 1.0 mm thick gels of the appropriate percentage acrylamide for resolving the proteins of interest were used. Empty wells were filled with blank loading buffer to maintain the ionic gradient across the gel. 10 mA current per gel was applied while proteins migrated through the stacking gel. Once the protein entered the resolving gel, the current was increased to 20 mA per gel until dye front elution from the gel.

2.4.7 Protein transfer to PVDF membrane

Following SDS-PAGE separation of proteins, proteins were transferred to PVDF membrane (immobilon-P as standard or immobilon-P SQ for low MW proteins). PVDF was activated by briefly soaking in methanol followed by equilibration in ice-cold Towbin transfer buffer (25 mM Tris-HCl, 192 mM glycine, 0.04 % w/v SDS, 20 % v/v methanol) for 15 minutes. Following SDS-PAGE protein separation, gels were equilibrated in ice cold Towbin transfer buffer for 15 minutes. For initial detection of PupE constructs, proteins were transferred at 45 V for 30 minutes in a Bio-Rad transblot SD Semi-Dry Transfer cassette. For all other experiments, proteins were transferred in ice cold standard Towbin transfer buffer at 65 V for 2.5 h in a Bio-Rad mini Trans-Blot cell tank at 4 °C.

After transfer, the PVDF membrane was washed twice with sterile distilled water, once with 100 % methanol and once with sterile distilled water to remove residual SDS from the membrane. Membranes were stained with Ponceau S stain (0.2 % w/v Ponceau S, 5 % v/v acetic acid in water) for 5 minutes to visualise protein transfer. Membranes were then washed twice with sterile

distilled water to remove excess Ponceau stain and finally in 100 % methanol to dehydrate the membrane for storage.

2.4.8 Western blotting

After transfer for detecting proteins of interest, membranes were reactivated in 100 % methanol for 2 minutes and washed once with sterile distilled water. Membranes were blocked for 1 h in 5 % skimmed milk powder (Marvel) dissolved in phosphate-buffered saline or Tris-buffered saline with 0.05 % tween-20 (PBS-T or TBS-T, respectively) depending on the primary antibody used. The membranes were washed once for 5 minutes with TBS-T or PBS-T and probed with the primary antibody (Table 2.11) in blocking solution. The membrane was washed 3 times for 5 minutes with TBS-T or PBS-T to remove any unbound primary antibody. The membrane was probed with appropriate secondary antibody diluted in blocking solution. Membranes were washed 3 times in TBS-T or PBS-T, followed by 2 washes in TBS or PBS.

Antibody	Supplier	Blocking	Dilution	Secondary
GFP	Roche	5 % milk PBS-T	1: 2000	Anti-Mouse
HA	Santa Cruz	5 % milk TBS-T	1: 2000	Anti-Rat
FLAG (M2)	Sigma	5 % milk TBS-T	1: 2000	Anti-Mouse
Anti-Rat HRP	Santa Cruz	5 % milk TBS-T	1: 20000	-
Anti-mouse HRP	Sigma	5 % milk TBS-T	1: 20000	-

Table 2.11 Antibodies used

Proteins were visualised by incubating membranes with a 3:1 mix of pico: femto SuperSignal™ West chemiluminescent substrate (Pierce) at room temperature for 5 minutes. Membranes were imaged using a G:boxx F3 gel doc system equipped with a f/1.2 zoom lens and analysed with GeneSys (Syngene) software.

Chapter 3

Development and implementation of a potential genetically encoded PafA-based proximity labelling system in yeast.

3.1 Introduction

This chapter is a collection of work to develop and optimise a pupylation based proximity labelling method performed prior to the publication of the PUP-IT system by (Liu, Zheng *et al*. 2018). Due to the work already done towards the characterization and development of the method it contains incomplete work or work that was not pursued any further as the focus of the project changed to the application of the technique in plants.

PUP-IT is built on the native activity of two proteins, PupE and PafA, involved in a proteasome-like protein degradation pathway of some Actinomycete Gram-positive bacteria. Prokaryotic ubiquitin-like protein (PupE) is post-translationally added to proteins by a peptide ligase enzyme, Proteasome accessory factor A (PafA), in a process known as pupylation. This system, functionally analogous to the eukaryotic ubiquitin proteasomal degradation system, is found in several Actinomycete groups such as the Mycobacteria and Corynebacteria, and directly targets pupylated proteins for degradation. To pupylate proteins PafA does not appear to recognise an amino acids or structural motif within a substrate, but instead seems to be capable of pupylating any available surface lysine present on target proteins (Barandun *et al*. 2012) (Liu, Zheng *et al*. 2018).

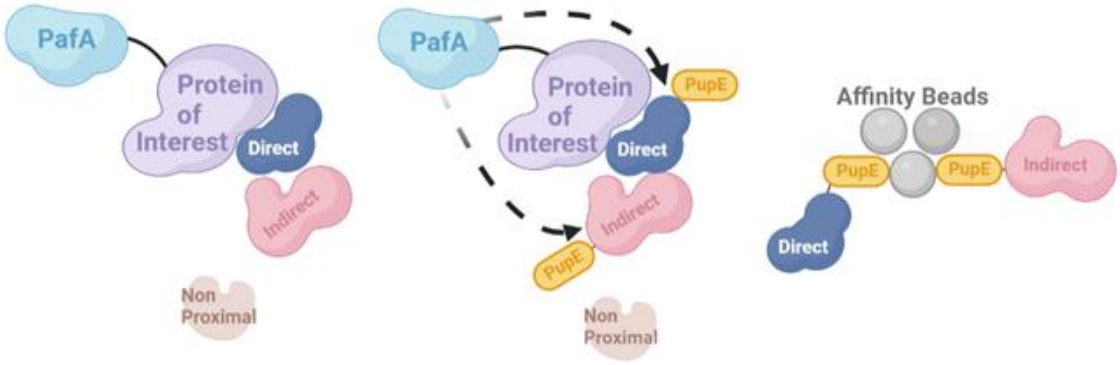
Interestingly, despite working in a similar way to ubiquitination in eukaryotes, neither PupE nor PafA possess any homology to any ubiquitin ligases or ubiquitin, meaning once reconstituted in eukaryotic systems it is highly unlikely that pupylation will result in degradation of proteins through the classical eukaryotic proteasomal route (Özcelik *et al.* 2012). The lack of a recognisable pupylation motif seems to indicate that PafA likely has low specificity coupled with high activity. This promiscuity indicates that PafA could be an effective potential tool for the study of PPIs (Liu, Zheng *et al*. 2018).

For the work in this thesis PafA and PupE constructs were derived from *Streptomyces turgidiscabies* a soil dwelling plant pathogenic bacterium causing scab in potatoes. By using this species, it improved the likelihood of functionality of the system in plant species, as the pupylation machinery would be functional at plant growth temperatures and conditions as opposed to the 37°C temperatures favoured by many of the animal pathogenic and commensal Actinobacterial orders such as Mycobacteria and Corynebacteria.

Preliminary work in the Hemsley lab has shown that expression of PupE with PafA fused bait proteins results in pupylation of distinct subsets of host proteins in yeast, indicating that this system is functional in yeast, and that fusion to bait proteins can impart specificity to the PafA pupylation reaction. Furthermore, it has been shown that only the conserved c-terminal helix of PupE is required for conjugation onto proteins by PafA (Sutter *et al*. 2010). This makes the

conserved c-helix an attractive minimal scaffold for attaching affinity tags for purification and identification of pupylated proteins.

Figure 3.1. Proposed DogTag system
PafA fused to a protein of interest would puppylate proximal interacting protein with freely expressed PupE-Affinity tag fusions. These proteins could then be captured by affinity bead and identified by proteomics of Western blot



. Before the initial publication of the PUP-IT system (Liu, Zheng *et al*. 2018) the project aims focused more heavily on the characterization of a new pupylation based PPI system. As a well characterised eukaryotic model, able to grow at similar temperatures to plants but enabling faster working, *Saccharomyces cerevisiae* was used for the initial expression of PafA and PupE. This was to confirm that the constructs could be expressed, and to establish if there was any toxicity associated with expression of either PafA or PupE in a eukaryotic system. Next the aim was to co-express the PafA and PupE proteins to check for labelling activity in vivo before designing and implementing a system that would allow for inducible protein-protein interactions couple with proximity

labelling to confirm the specificity, mechanism, and functionality of pupylation based labelling.

This chapter is therefore a collection of the work that was completed in the first 18 months of the PhD, with considerable progress made towards the original project aims, including the expression and detection of PafA in *Saccharomyces cerevisiae* and the development and groundwork laid for the inducible proximity labelling of proteins by PafA. However, as the basic characterization of the technique was published while this work was in progress, the overall aims of the project were shifted from development to implementation, with the other results chapters focusing on these new aims (figure 3.1).

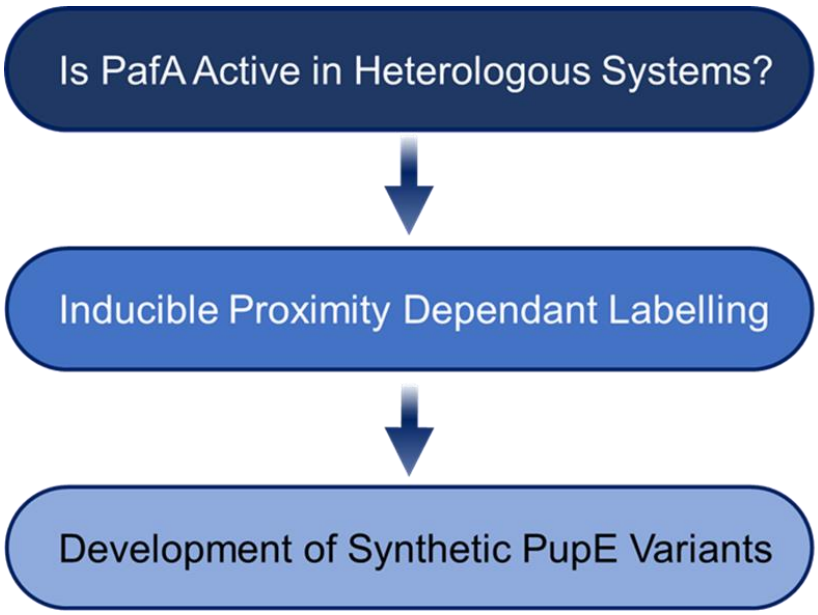


Figure 3.2. Initial workflow

A graphic illustrating the original planned workflow for system optimisation to meet the projects initial aims

3.2 Results

3.2.1 Generation of PupE expression constructs

In order to begin the development of a new Pupylation based proximity labelling system PupE and PafA components were designed and created with our original aims in mind (figure 3.2). While the best characterised PupE and PafA proteins are found in Mycobacterial and Corynebacterial species the aim of this projects was to demonstrate the system in a plant model. Therefore, the soil resident plant pathogenic bacterium *Streptomyces turgidiscabi*es was chosen as the source organism for the constructs. The main reasons for this are because the optimal growth temperature for plants and a soil dwelling plant pathogen are similar, increasing the likelihood of optimal PafA enzyme activity and folding in plants.

In order to find the most effective composition of PupE for labelling in plants various PupE constructs were designed with variations to allow for improved expression, detection by Western blot and the ability to pulldown on affinity matrices. All constructs were initially designed with a ubiquitin leader to support expression of small peptides in plants and that would be cleaved off in vivo (Hondred D *et al*., 1999) by deubiquitinating enzymes, a TwinStrep tag for quick, easy, and cheap pulldown of PupE tagged proteins and a 3xFLAG tag for detection by Western blot.

For the optimisation of the system, it was decided to test three different lengths of the PupE protein; the minimal length - the shortest c-terminal helix sequence reportedly required for PafA interaction covering $PupE^{46-72}$ (Sutter, *et al*. 2010), the conserved length - a slightly longer length including the c-terminal helix

sequence with the addition of five conserved amino acids covering $PupE^{41-72}$, and the Full Length - the full length PupE sequence covering $PupE^{2-72}$.

Studies into the mechanism and function of native PafA and PupE have shown that only 25 residues in the c-terminal helix are the minimal recognition motif for PafA interaction (Sutter *et al.*, 2010). By utilising this with $PupE^{46-72}$, allows for the smallest possible PupE tag which might prove the most efficient for a proximity labelling system. The slightly longer $PupE^{41-72}$ allows for a five amino acid spacer just in case issues arise with the direct fusion of PupE to the tags resulting in issues in function

and activity and the full length $PupE^{1-72}$ gives me the opportunity to see if the native PupE, rather than truncated PupE versions, is actually the most active and efficient for the system (figure 3.4).

Work already done in the lab had produced an initial $PupE^{46-71}$ construct in the orientation Ub-3xFLAG-TwinStrep-$PupE^{46-71}$. To hopefully improve functionality by making the Twin STREP more accessible, this construct was rearranged using site directed mutagenesis into the orientation Ub-TwinStrep-3xFLAG-$PupE^{46-72}$. The conserved, $PupE^{41-72}$, and full length PupE, $PupE^{1-72}$ constructs were then all subsequently made in the same orientation using site directed mutagenesis. Negative, non-ligatable, PupA controls of each length were also made using site directed mutagenesis by substituting the PupE C-terminal glutamate with alanine (Özcelik *et al*. 2012). PupE constructs shown in figure 3.3.

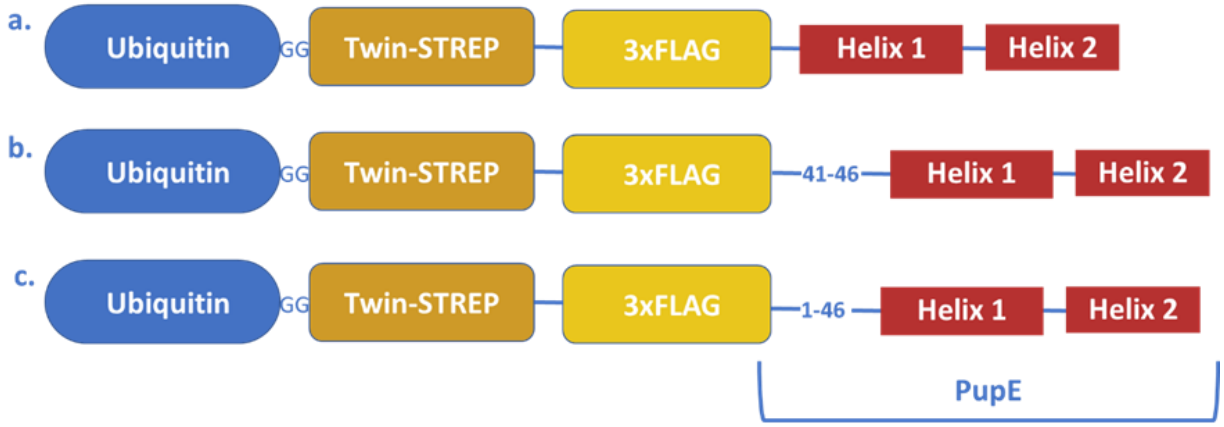


Figure 3.3. PupE constructs

Schematic showing the structure of the PupE constructs created. All constructs were made including a ubiquitin leader to drive expression, a TwinStrep tag for enrichment and a 3xFLAG tag for Western blot detection. The portion of PupE included varied across constructs and all constructs were made with a Negative PupA counterpart. A. Ub-TwinStrep-3xFLAG-PupE$^{46-72}$ B. Ub-TwinStrep-3xFLAG-PupE$^{41-72}$ C. Ub-TwinStrep-3xFLAG-PupE$^{1-72}$.

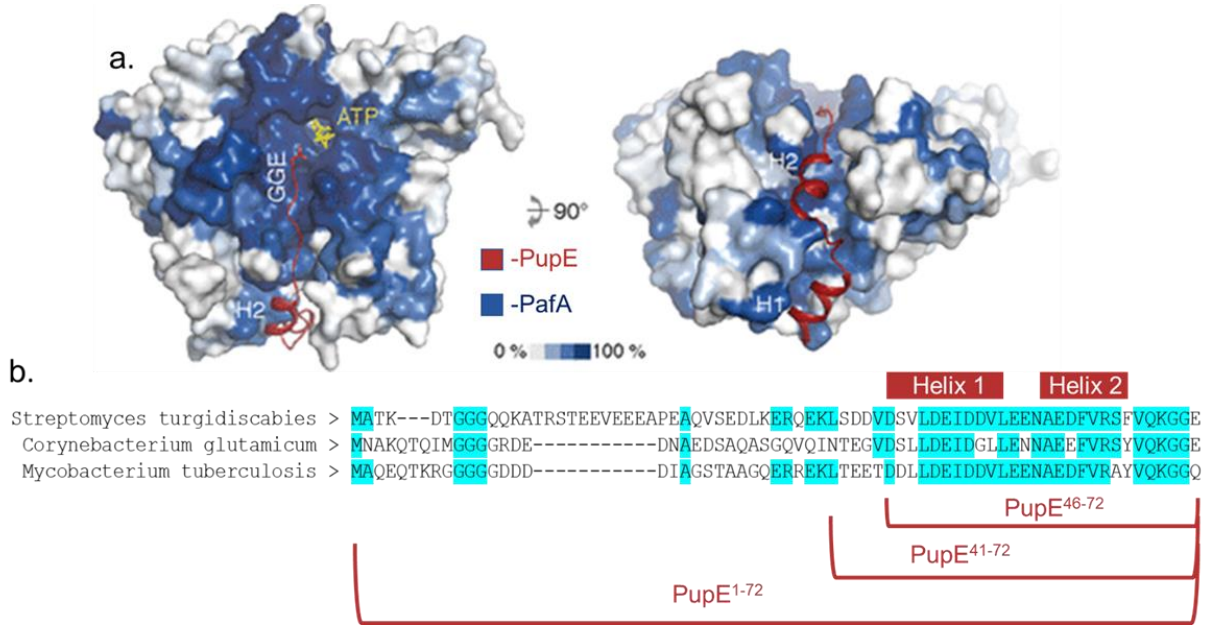


Figure 3.4. The structure of the PafA/ PupE Complex.

a. PafA acts as scaffold to induce folding of Pup (red) into two helices (H1and H2) connected by a short linker. The C-terminal glutamate and ATP (yellow) are shown in stick representation. The ligase is shown in surface representation coloured according to conservation: from no conservation (white) to highly conserved (blue). (Barandun et al. 2013)

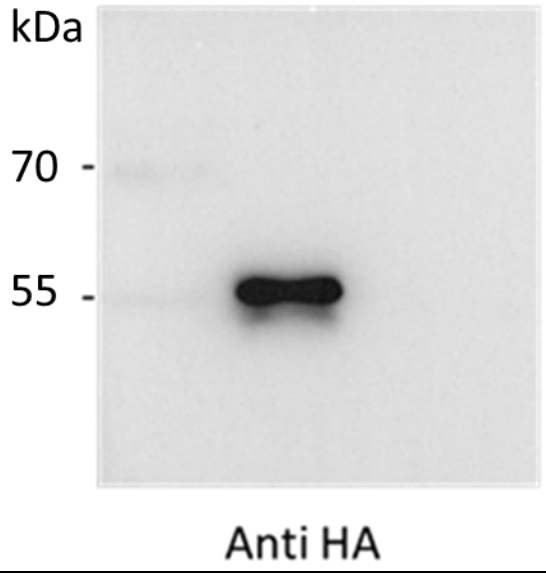
b. Alignment of mature PupEs from different actinobacterial species. The regions involved in the interactions with the Pup ligase PafA (Helix1 and Helix2). Residues conserved with *S. turgidiscabies* highlighted in blue. Various construct lengths indicated below in Red.

3.2.2 Expression of PafA in yeast

To test the pupylation system in vivo, co-expression of both PupE and PafA is required. To determine the ability of PafA fusions to express without impacting cellular function in *Saccharomyces cerevisiae* PafA from *Streptomyces turgidiscabies* was expressed using a pRS315 based vector and the Akr1 promoter with a 3xHA C-terminal tag for detection by Western blot (proAkr1: PafA-3HA) (Sikorski *et al*. 1989) (Roth *et al*. 2002). Expression of PafA in *S. cerevisiae* strain BY4742 was confirmed by anti-HA Western blot (figure 3.5).

As with the PupE constructs a relevant negative control of PafA (PafA*) was made by substitution of the active site of PafA using an Asp57Asn mutation. This was done using site directed mutagenesis (Guth *et al*. 2011).

Figure 3.5. PafA is expressed in yeast. Anti-HA Western blot of PafA-3xHA (proAkr1: PafA-3HA) ,54kDa, to check for expression in *Saccharomyces cerevisiae* strain BY4742.



3.2.3 Toxicity testing in yeast of PupE and PafA expression

As the system was being developed for use in eukaryotic models, after successfully expressing and detecting PafA in *S. cerevisiae,* construct toxicity in heterologous systems needed to be tested. To test whether expression of these constructs in *S. cerevisiae* has an adverse effect on growth, a time course experiment was carried out where growth was measured by OD_{600} every hour. PafA was constitutively expressed from the AKR1 promoter while PupE and PupA variants were driven by the Gal1/10 promoter in pYES-DEST52. Induction of PupE and PupA expression was achieved by the addition of galactose to raffinose grown cultures inoculated to OD_{600} 0.2.

This time course showed that expression of PafA and the various PupE constructs did not appear to have an obvious impact on growth, indicating that PafA and PupE are not individually toxic in *S. cerevisiae* (figure 3.6).

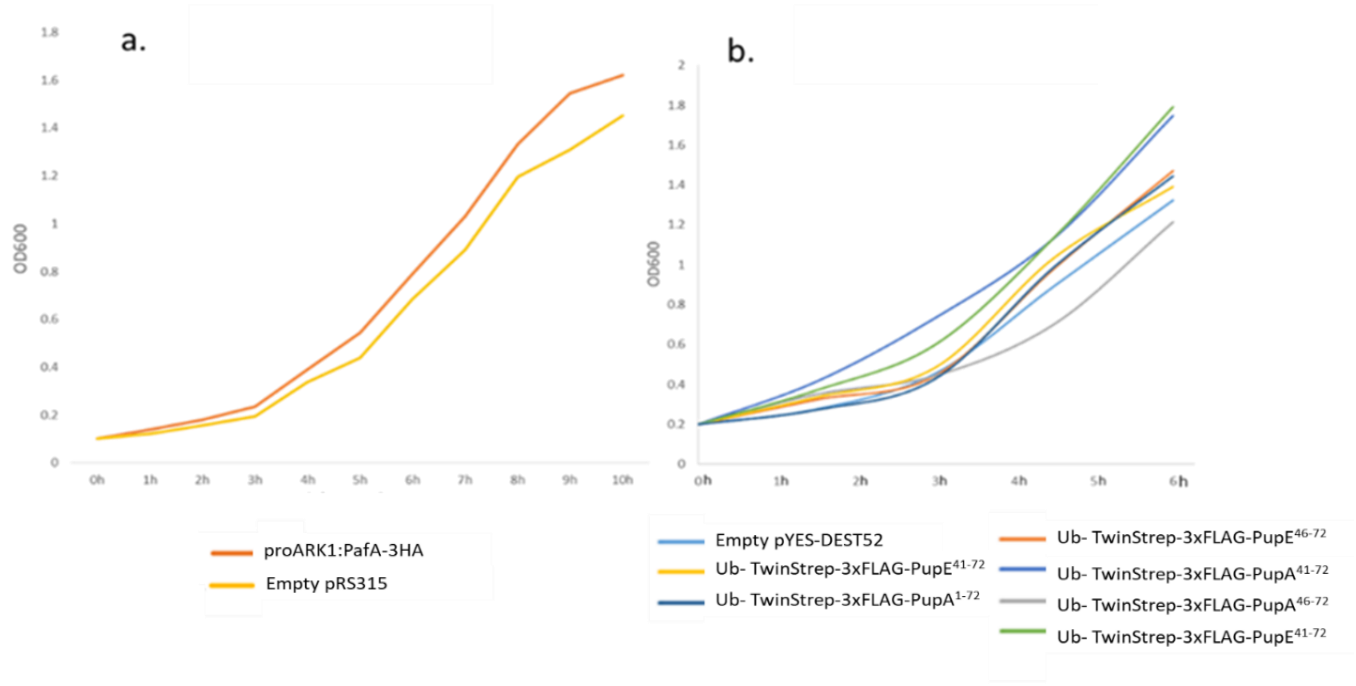


Figure 3.6. PafA & PupE are not toxic in yeast.
Time course of BY4742 *Saccharomyces cerevisiae* growth. **A**. OD600 of proAkr1: PafA-3HA alongside empty pRS315 grown in -Leu drop out media and innoculated to OD600 0.1. OD600 then measured and recorded every hour. **B**. PupE constructs including PupE$^{46-72}$, PupE$^{41-72}$, and PupE$^{1-72}$ alongside negative PupA counterparts and an empty vector pYES2 control. Grown in -Ura drop out media with 2% Raffinose and 0.1% Glucose and induced with Galactose at OD600 0.2, with OD600 then measured and recorded every hour.

3.2.4 Design of an artificial interaction system to test for PafA proximity labelling

After establishing successful non-toxic expression of components of the PupE/PafA system in *S. cerevisiae* the next step in initial characterization is the generation of an artificial interaction system to validate proximity dependant rather than general PafA/PupE activity towards other proteins in yeast.

The published GAI/GID system (Miyamoto *et al*., 2012) is ideal for this purpose in yeast and comprises the plant Gibberellin hormone (GA) nuclear receptor

Gibberellin Insensitive Dwarf 1 (GID1) and the GID1/GA dimer binding region of the DELLA domain of Gibberellic Acid Insensitive, GAI^{1-92} (GAI1*). Gibberellin induced dimerization between GAI^{1-92} and GID1 can be exploited to force an interaction between PafA and a protein of interest, allowing us to test for proximity dependant PupE labelling activity. Before application this inducible interaction would initially be confirmed by microscopy using a reporter tags in place of PafA (figure 3.7).

To generate constructs, each fragment (figure 3.7) was PCR amplified using primers with 30 nucleotide homologous overlaps and, alongside linearized vector, was transformed into *S. cerevisiae* where homologous recombination then generated each vector from the supplied fragments. GAI^{1-92} and GID1 were amplified from *A. thaliana* cDNA. The promoters $ALD6_{pro}$ and $RPL18B_{Pro}$, the terminators $ENO2_{term}$ and $TDH1_{term}$ and $STE2^{1-73}$ signal peptide and 1^{st} transmembrane span were all amplified from *S. cerevisiae* genomic DNA. Colonies were screened by yeast colony PCR to identify likely correct recombinant plasmids followed by sequencing of recovered plasmid DNA.

In the planned work, after validation with the mCherry and GFP reporter tags, mCherry would be replaced by PafA while GFP would either act as a test for proximity dependant labelling or be replaced by a known substrate of PafA as a positive control.

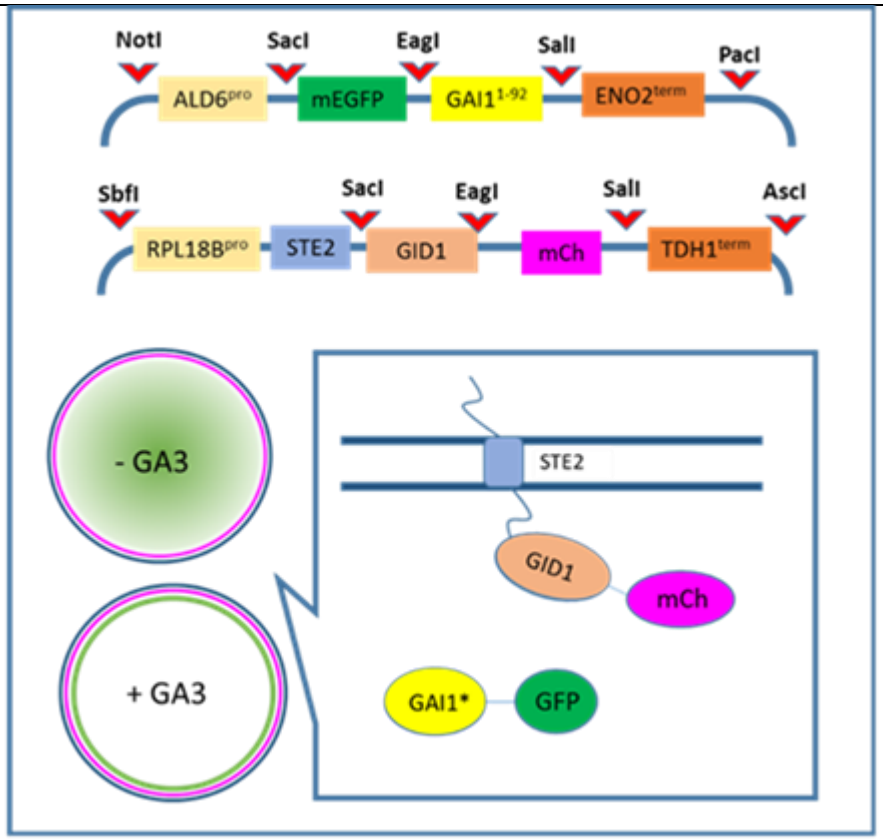


Figure 3.7. GID GAI1 interaction characterization upon GA3 treatment GID1 and GAI1* dimerize. GID1 was fused to the 1st transmembrane domain of STE2(a yeast membrane protein) and the reporter mCherry (mCh). GAI1* was fused to the reporter GFP. Upon GA3 treatment GAI1* would dimerize with GID1 resulting in membrane localisation and the presence of both mCherry and GFP at the cell membrane

3.3 Conclusion

The work described was designed for the initial creation, validation, and optimisation of parts of a novel proximity labelling system. This work successfully demonstrated non-toxic expression of PafA in eukaryotic systems but stopped short of demonstrating enzymatic activity. As, unlike PafA, PupE was never tested for expression by Western blot in *S. cerevisiae*, its suspected effects on growth can only be inferred from OD600 measurements. As the direction of the project changed after the publication of PUP-IT (Liu, *et al.* 2018) this work was stopped, and work initially planned for yeast further explored in planta.

Various adaptations of the components of the PupE/PafA machinery were successfully designed and made, including the generation of an artificial interaction system for activity validation utilizing the GAI/GID logic gate system (Miyamoto *et al.* 2012), but as this was no longer required for basic characterization and demonstration of interaction it was never used.

After the demonstration and characterization of a pupylation based proximity labelling PPI technique with the PUP-IT system (Liu, *et al.* 2018) the aims of the project were significantly changed to optimising the application of a PUP-IT-like system in plants as a discovery tool for problematic PPIs.

Chapter 4

Optimisation for the application of the DogTag proximity labelling system in plants

4.1 Introduction

Despite the publication and basic characterization of Pupylation as a proximity labelling technique in mammalian systems (Liu, Zheng *et al*. 2018) there was no data to suggest whether the system would also be effective in planta. As a result of the published competing work (Liu, Zheng *et al*. 2018) the overall aims of the project shifted to optimization and application of pupylation in plants (figure 4.1). This chapter focuses on the initial parameterisation and optimisation of this technology in plants.

PupE reportedly only requires a small portion of the protein containing the c-terminal helices for interaction with PafA. As the effective length of PupE, and ability of PafA to utilise N-terminally tagged forms of PupE, during labelling has not been explored I wanted to investigate what constituted the most effective PupE for proximity labelling and purification of tagged proteins. Ideally, to minimise disruption of tagged protein function and maintain its physiological function, the tag would be as small as possible whilst still being efficiently ligated.

Unlike the BirA derived proximity labelling systems that produce a cloud of activated biotin-AMP that label lysines within a ~25 nm radius, PafA requires direct protein contact to label a protein. While this will likely reduce non-specific background in the PafA system it may limit labelling of some proteins. As Pupylation works by the labelling of a proteins free surface lysines it is therefore possible the system will not be effective for labelling all proteins due to lack of a

sterically accessible, unmodified lysine. In the case of using the PafA system for investigating problematic pair-wise interactions with limited rotational mobility, such as between membrane proteins constrained in the plane of the membrane, this may lead to false negatives. GFP is a commonly used tag for proteins that has a high proportion of free surface lysines and is therefore a good candidate to try as a generic pupylation acceptor for use in pairwise interaction studies. This will be tested more thoroughly in future chapters where I aim to identify the Protein S-Acyl Transferase (PAT) enzymes that S-acylate receptor kinases.

To address these questions, in this chapter I focus on investigating three fundamental areas underpinning the successful implementation of proximity dependant labelling of the proteome by PafA/PupE:

1. Is PafA stable and active in plants?
2. Which amino acids and features of PupE are required for effective labelling of proteins in plants?
3. Is a GFP tag effective as a “generic acceptor” for pupylation?

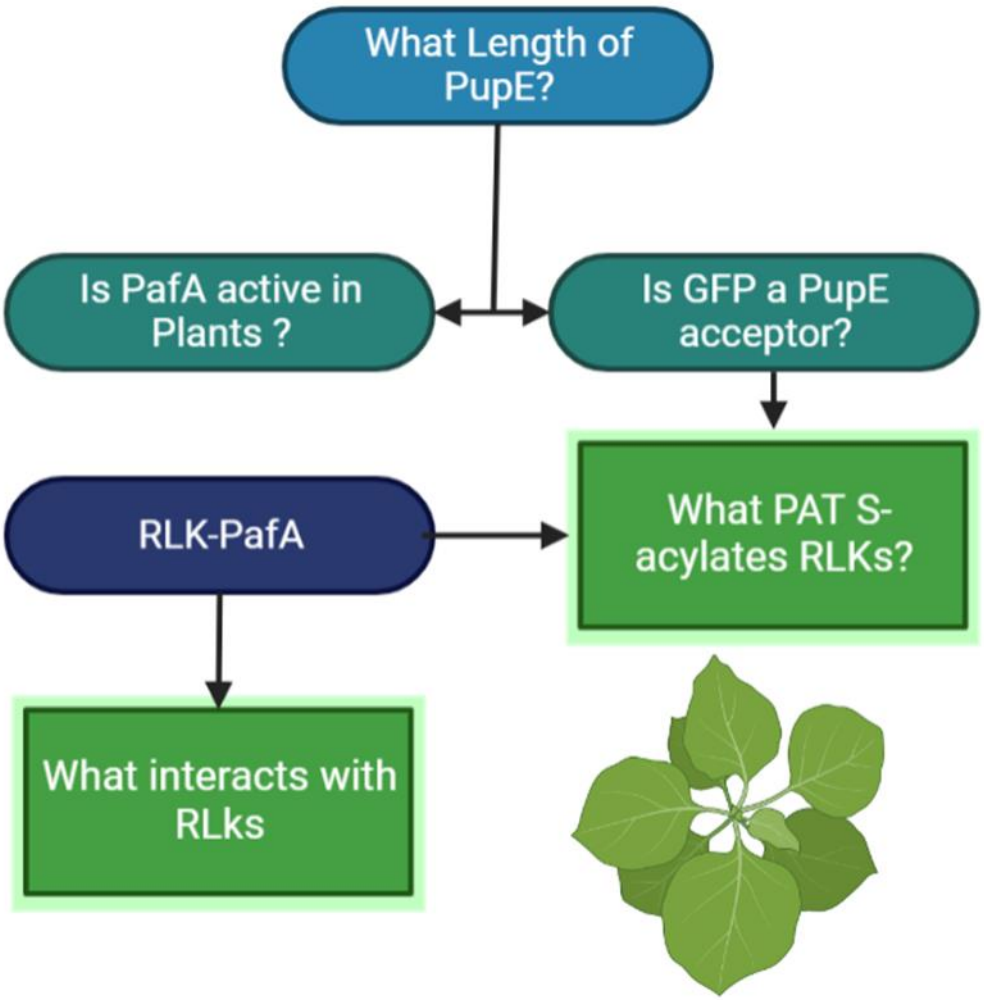


Figure 4.1. Proposed workflow.

First steps are to optimise the system by identifying if PafA is active in plants, what length of PupE is the most effective and whether GFP can be a pupylation acceptor. After this is established, the system can then be applied to study some PPIs of RLKs in both pairwise and discovery approaches.

4.2 Results

4.2.1 Optimisation of PupE expression in plants

PupE is a small peptide (~9kDa) with a highly disordered N-terminus; this would hopefully be an ideal scaffold for various affinity and epitope tags to enable detection and purification of tagged proteins. To define the required features of PupE for PafA mediated conjugation to proteins in plants a number of different length constructs have been designed (described in previous chapter). For affinity purification all constructs include TwinStrep as it is cheap, fast, and typically demonstrates low background/non-specific binding. Initial constructs also contain a 3xFLAG tag for visualisation of labelling by Western blot.

Small peptides are typically hard to express in plants (Hondred *et al.,* 1999) but this can be mitigated by fusing the peptide of interest to the C-terminus of ubiquitin. PupE expression constructs have therefore been made including a ubiquitin leader for improved plant expression with the idea that the ubiquitin would be cleaved post-translationally from PupE by ubiquitin proteases to liberate PupE. However, in initial expression experiments minimal PupE appeared to run closer to its expected Ub-PupE fusion size (~17kDa) than its expected cleaved size (~9kDa) based on primary amino acid sequence.

To establish if ubiquitin-PupE fusion cleavage was occurring correctly, the minimal PupE construct, Ub-TwinStrep-3xFLAG-PupE$^{46\text{-}72}$ (PupE$^{46\text{-}72}$) was transiently expressed in *Nicotiana benthamiana.* Leaf discs were then vacuum infiltrated with or without *N*-ethylmaleimide to inhibit ubiquitin cleavage by deubiquitinating enzymes (DUBs) (Tian *et al*., 2011) and incubated for 6 hours. Protein was extracted, separated by SDS PAGE and PupE identified by anti-FLAG Western blot (Figure 4.2). In the sample treated with N-ethylmaleimide

there seems to be a larger quantity of the ~17 kDa band compared to the untreated sample. This may suggest that the ~17kDa band is uncleaved Ub-PupE and that in non-N-ethylmaleimide treated plants a proportion of Ubq-PupE is being cleaved and the liberated PupE was just not successfully resolved or captured using the Laemmli buffer system and/or 0.45 µm pore PVDF membrane. However, even in the untreated sample there does seem to be a proportion of un-cleaved Ub-PupE, indicating that cleavage may not be as efficient as it could be. Based on the literature (Gilchrist *et al*., 1997) Ub fusion cleavage approaches 100% efficiency, so this result was surprising.

Next, I wanted to check if the expected small size of cleaved PupE$^{46-72}$ combined with standard SDS-PAGE and Western blotting techniques were not sensitive enough or appropriate for resolution of TwinStrep-3xFLAG-PupE$^{46-72}$. To address this, I optimised the use of Tris-Tricine SDS-PAGE system in order to better separate proteins of low molecular weight. Using this system, along with 0.2 µm PVDF membrane, I successfully resolved a low molecular weight ladder (figure 4.2) and optimised detection of protein samples containing TwinStrep-3xFLAG-PupE$^{46-72}$. Despite these efforts I was still unable to detect TwinStrep-3xFLAG-PupE$^{46-72}$ at the expected mass and so revisited the possibility of failed Ubiquitin cleavage another way.

To explore if the construct was being cleaved in planta by DUBs, the cleavage site between Ubiquitin and TwinStrep-3xFLAG-PupE$^{46-72}$ was mutated using site directed mutagenesis (C-terminal gly-gly of ubiquitin changed to ala-ala) to prevent cleavage. However, expression of this construct was never detected, and I hypothesise that the non-cleavable Ub-PupE was degraded as if it were a ubiquitinated polypeptide. This hypothesis could have been explored more by blocking this potential degradation using protease inhibitors. This would suggest

that my original Ub-TwinStrep-3xFLAG-PupE$^{46-72}$ fusion, hypothetically able to be processed by DUBs, would not accumulate if DUB cleavage had not occurred and instead would be degraded by the proteasome. These data together suggest that the Ub-TwinStrep-3xFLAG-PupE$^{46-72}$ fusion is likely being processed effectively and that PupE runs at a higher-than-expected molecular weight by SDS-PAGE than would be expected. This is further supported by other work in the lab where a PupE form based on my construct, but without ubiquitin fusion, was expressed in bacteria. I used this as a positive control to determine the running size of TwinStrep-3xFLAG-PupE$^{46-72}$ and was able to resolve a band using my optimised Tris-Tricine SDS-PAGE system and Western blot at ~15 KDa (Figure4.2).

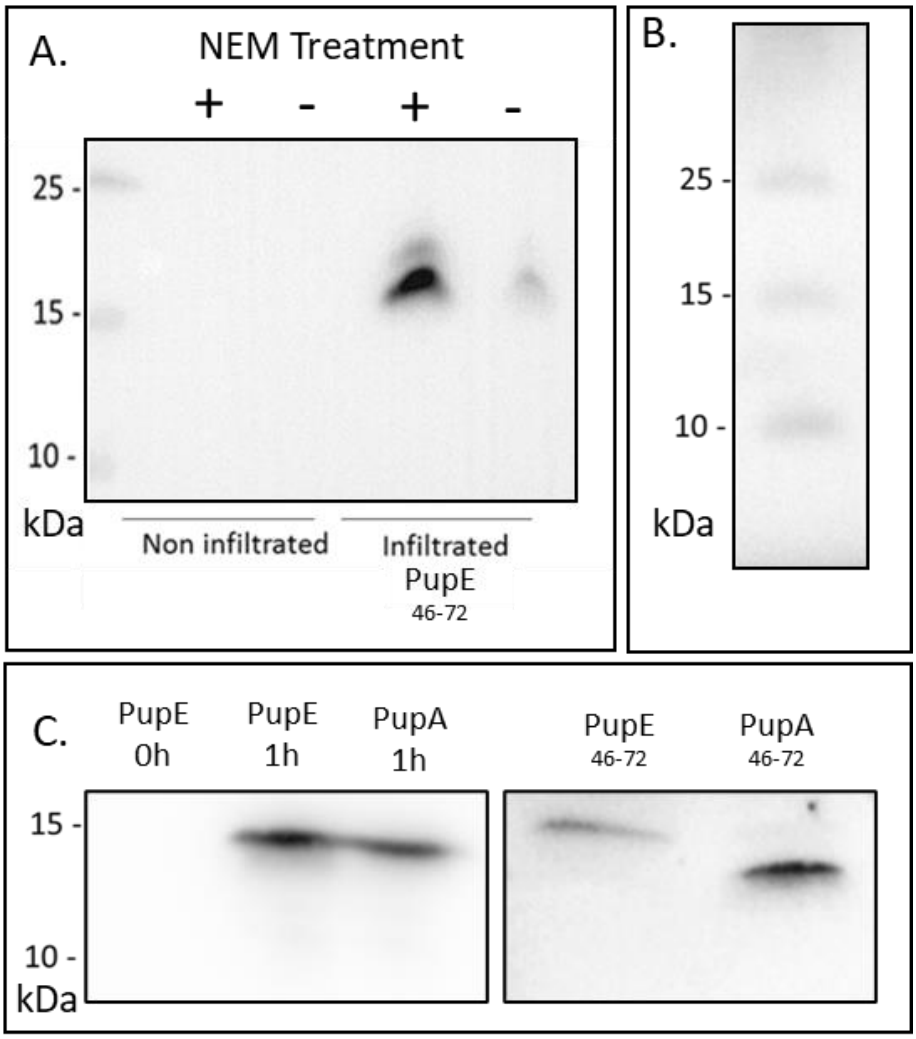


Figure 4.2 optimisation of PupE detection

A. Infiltrated with Ub-TwinStrep-3xFLAG-PupE46-72 and non-infiltrated leaf disks were vacuum infiltrated with or without N-Ethylmaleimide to inhibit ubiquitin cleavage, protein was extracted, separated by SDS PAGE, and identified by anti-FLAG Western Blot. Un-cleaved Ub-TwinStrep-3xFLAG-PupE46-72 expected size of ~17kDa. **B**. A low molecular weight ladder was resolved on a Tris-Tricine gel and transferred onto a 0.2 µm PVDF membrane. **C.** Ubiquitin free induced bacterial PupE and PupA construct expression compared with transient expression of Ub-TwinStrep-3xFLAG-PupE46-72 (Left) and Ub-TwinStrep-3xFLAG-PupA46-72 in *N.benthaminana* resolved on SDS PAGE, detected by Western blot, anti-FLAG (Right).

After establishing that TwinStrep-3xFLAG-PupE$^{46-72}$ runs at ~15kDa and not the expected ~9kDa I then transiently expressed all three lengths of PupE alongside their respective PupA negative controls in *N. benthamiana.* These were successfully resolved using the optimised Tris-Tricine SDS-PAGE system in order to better separate proteins of low molecular weight and detected by Western blot using 0.2 µm PVDF membrane (figure 4.3).

As seen on the Western blot (figure 4.3) for TwinStrep-3xFLAG-PupE$^{46-72}$ and TwinStrep-3xFLAG-PupE$^{41-72}$ the PupA Negatives, made by substituting the PupE C-terminal glutamate with a physiologically inactive alanine, run at a smaller size than the functional PupE. I hypothesise that this is due to a change in the charge of the protein seen with the substitution of glutamate to alanine altering mobility in SDS-PAGE.

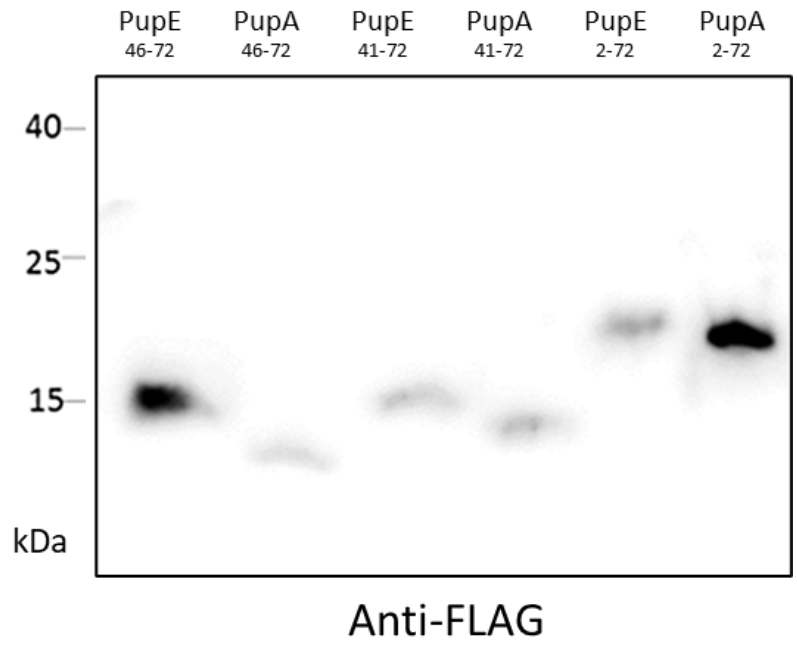


Figure 4.3 Detection of PupE
Transient expression of TwinStrep-3xFLAG-PupE$^{46-72}$ and TwinStrep-3xFLAG-PupA$^{46-72}$, TwinStrep-3xFLAG-PupE$^{41-72}$ and TwinStrep-3xFLAG-PupA$^{41-72}$ and TwinStrep-3xFLAG-PupE$^{2-72}$ and TwinStrep-3xFLAG-PupA$^{2-72}$ lengths in *N.benthamina* . Resolved on a Tris-Tricine SDS PAGE. Protein detected by anti-FLAG Western blot.

4.2.2 PafA is expressed and active in plants

To determine whether PafA is stable and active in plants, I expressed HA tagged PafA and my minimal TwinStrep-3xFLAG-PupE$^{46-72}$ or TwinStrep-3xFLAG-PupA$^{46-72}$ constructs transiently in *N. benthamiana.* In making the PafA constructs I encountered difficulties with vector stability in *E. coli* that was overcome by the addition of the *Legumin J* Intron (Bäumlein *et al*., 1986) to the PafA coding sequence. This stabilised the construct and allowed for successful growth and propagation of the vectors in bacteria.

After *Legumin J* intron mediated stabilisation, 3xHA-PafA was successfully transiently expressed in *N. benthamiana* and co-expressed alongside the

minimal length TwinStrep-3xFLAG-PupE$^{46-72}$ and TwinStrep-3xFLAG-PupA$^{46-72}$. A ladder of anti-FLAG reactive bands, indicative of TwinStrep-3xFLAG-PupE$^{46-72}$ conjugation, was seen only when TwinStrep-3xFLAG-PupE$^{46-72}$ and 3xHA-PafA were co expressed (figure 4.4). This indicates that PafA is both stable and active in plants, that TwinStrep-3xFLAG-PupE$^{46-72}$ is functional for conjugation by PafA to proteins, and that labelling only occurs when both 3xHA-PafA and TwinStrep-3xFLAG-PupE$^{46-72}$, but not TwinStrep-3xFLAG-PupA$^{46-72}$, are present.

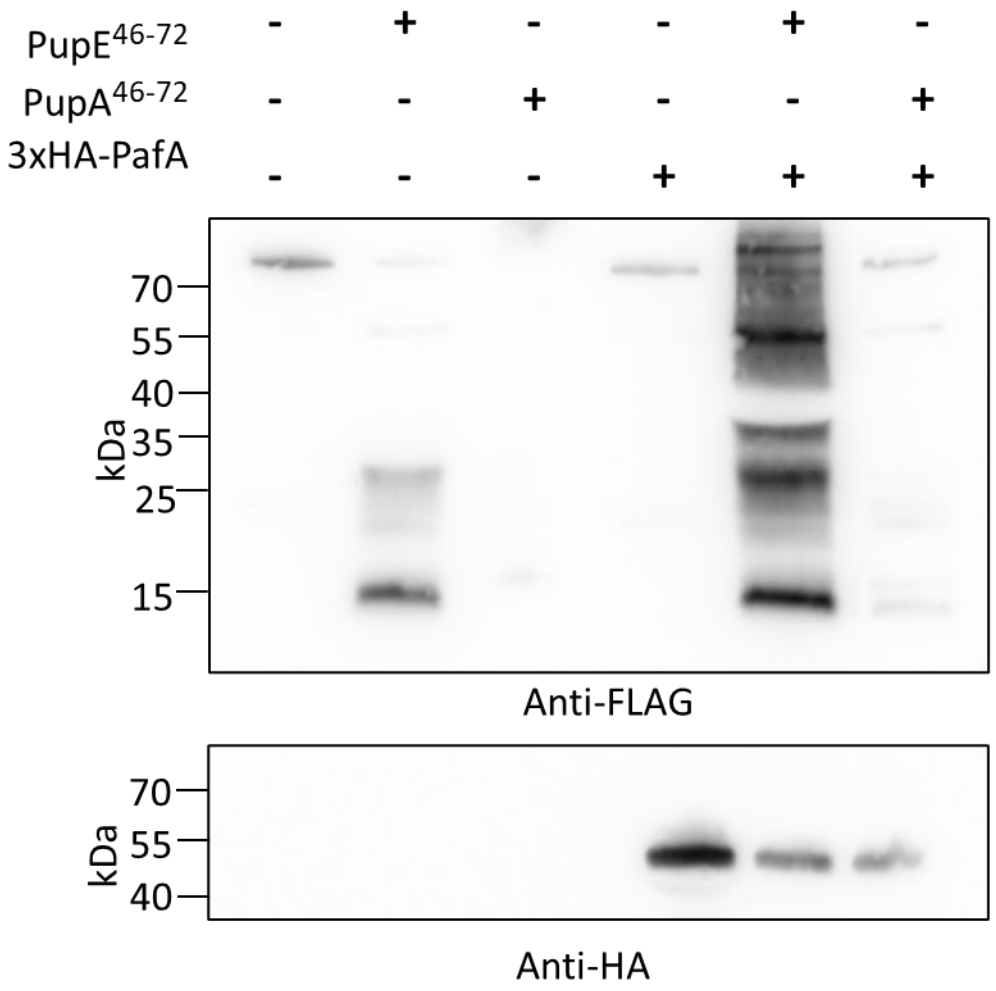


Figure 4.4 Reconstitution of the PafA/PupE system in plants

TwinStrep-3xFLAG-PupE$^{46-72}$, TwinStrep-3xFLAG-PupA$^{46-72}$ and 3xHA-PafA were individually and co expressed transiently in *N. benthamiana*, separated by SDS PAGE, and detected by Western Blot. PupE ~15KDa (Anti FLAG) PafA ~54KDa (Anti HA)

4.2.3 Is GFP a PupE acceptor?

The mechanism of pupylation relies on free surface exposed lysines; it is therefore important to establish a validated substrate for pupylation in pair-wise interaction assays as it is unknown how many available surface lysines on a protein are required for effective labelling. As GFP is a commonly used tag and is the current tag present on the PAT constructs I intend to apply the system to later, it is a good first candidate as a generic acceptor.

As GFP contains 20 surface exposed lysines, is soluble, has a globular structure and has no reported post-translational modifications affecting lysine residues, it is highly likely that it will be a good pupylation substrate. However, if GFP is not an effective pupylation substrate the possibility of using a bacterial pupylated protein or domain as a PupE acceptor that could be fused to proteins in the place of GFP could be explored.

To check GFPs suitability as a substrate I created a GFP-PafA fusion, GFP-5xGGSGG-PafA-3xHA. The construct contained a 3xHA tag for detection by Western blot and a 5xGGSGG linker between PafA and GFP to reduce potential steric interference between these two globular proteins while ensuring that PafA is maintained in close proximity to GFP.

GFP-5xGGSGG-PafA-3xHA was transiently expressed in *N. benthamiana* alongside TwinStrep-3xFLAG-PupE$^{46-72}$ and successfully identified by Western blot (figure 4.5). As previously seen, labelling is only detected in the presence of both GFP-5xGGSGG-PafA-3xHA and TwinStrep-3xFLAG-PupE$^{46-72}$, with no labelling seen with only GFP-5xGGSGG-PafA-3xHA or TwinStrep-3xFLAG-PupE$^{46-72}$ expressed alone or co-expressed with the non-conjugable TwinStrep-3xFLAG-PupA$^{46-72}$. This data also shows that the PafA protein remains active as

a simultaneous N- and C-terminal fusion, an important factor when considering implementation of the system.

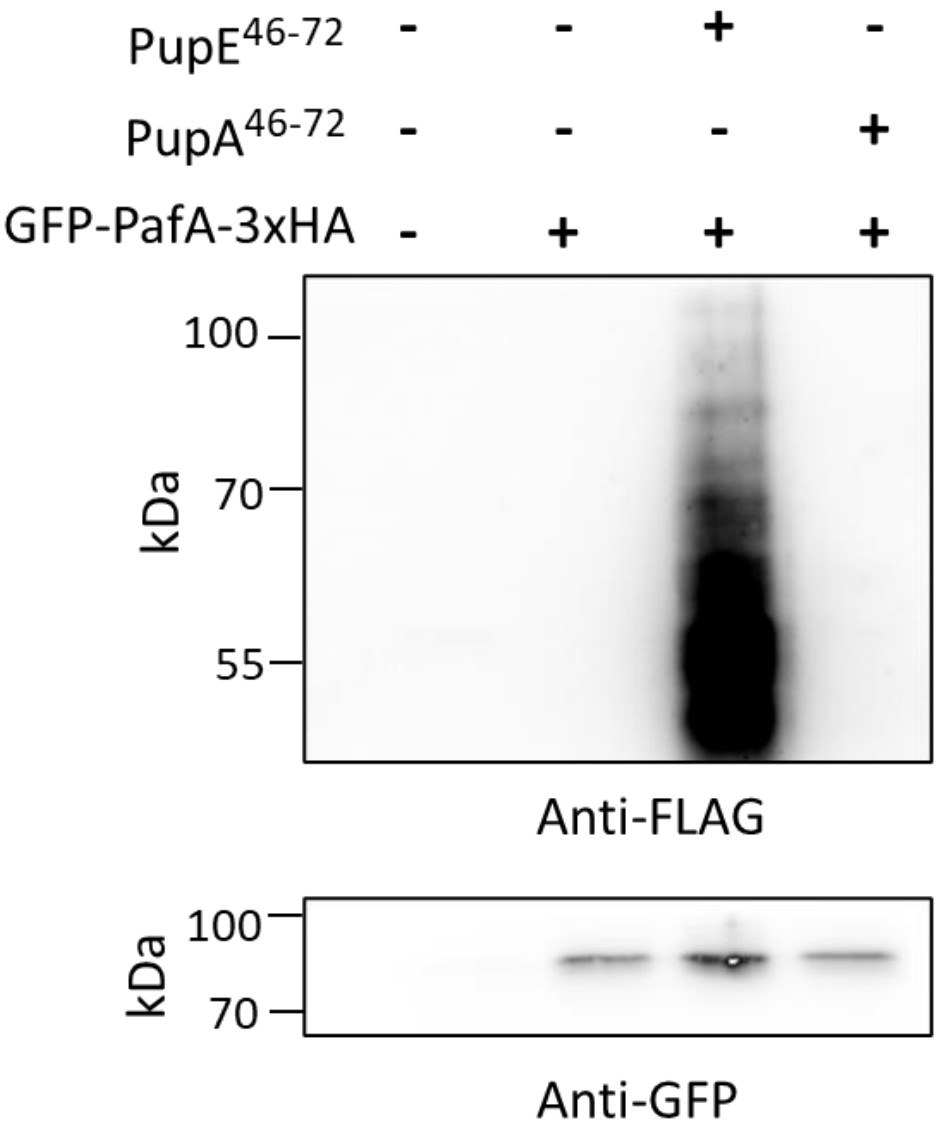


Figure. 4.5 GFP-PafA fusions are active in plants

TwinStrep-3xFLAG-PupE$^{46-72}$, TwinStrep-3xFLAG-PupA$^{46-72}$ and GFP-5xGGSGG-PafA-3xHA (GFP-PafA) were individually and co-expressed transiently in *N. benthamiana* separated by SDS PAGE, and detected by Western Blot. PupE/PupA Anti FLAG GFP-5xGGSGG-PafA-3xHA Anti GFP.

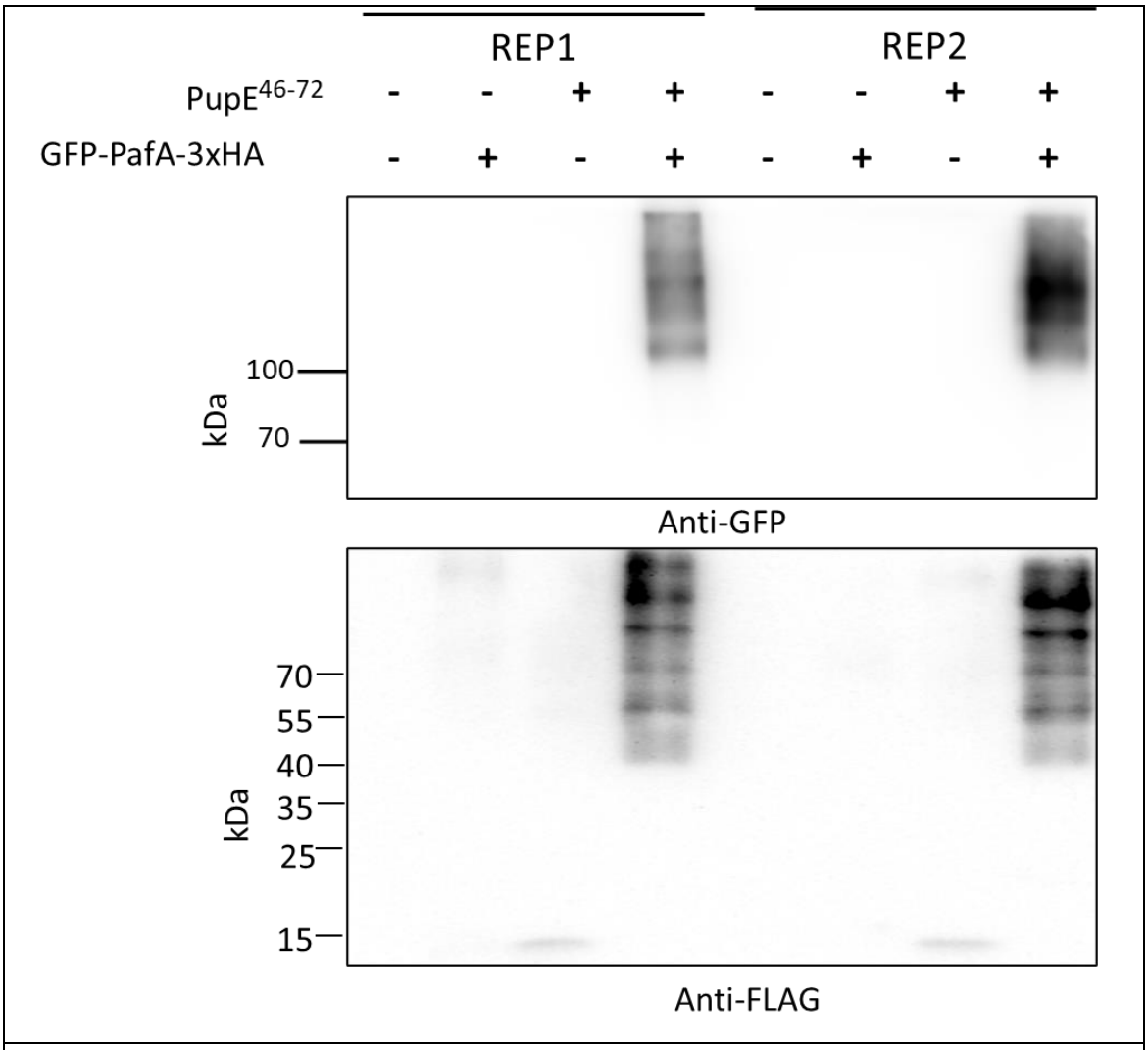


Figure 4.6. TwinStrep enrichment of pupylated proteins

TwinStrep-3xFLAG-PupE$^{46-72}$ and GFP-5xGGSGG-PafA-3xHA (GFP-PafA) were individually and co expressed transiently in *N. benthamiana.* Enriched using type3 MagSTREPbeads eluted and separated by SDS PAGE and pupylated 5xGGSGG-PafA-3xHA detected by anti-GFP Western blot. Active labelling and PupE expression was detected by anti-FLAG Western blot.

Next, to directly demonstrate pupylation of GFP by PafA, the GFP-PafA fusion, GFP-5xGGSGG-PafA-3xHA, was again transiently expressed in *N. benthamiana* alongside TwinStrep-3xFLAG-PupE$^{46-72}$ and TwinStrep-3xFLAG-PupA$^{46-72}$. On this occasion two technical replicates were performed. Protein was extracted and enriched using the TwinStrep tag present on TwinStrep-3xFLAG-PupE$^{46-72}$ to pulldown pupylated proteins. Samples are enriched using TwinStrep and GFP-5xGGSGG-PafA-3xHA identified by Western blot. A smear is visible above the molecular weight of GFP-5xGGSGG-PafA-3xHA. I hypothesise this represents a mixed population of TwinStrep-3xFLAG-PupE$^{46-72}$ labelled GFP-5xGGSGG-PafA-3xHA as with the addition of each PupE tag the protein will increase in mass by approximately ~9 kDa. These TwinStrep enriched proteins were digested with trypsin and analysed by mass spectrometry. Overall coverage of GFP-5xGGSGG-PafA-3xHA was high (~59% of full construct, ~61% of GFP, figure 4.7), and 13 of the 22 lysines within GFP and the linker were contained within identified peptides. However, analysis only detected the tryptic remnant of PupE modification (GGE) on two lysine containing peptides derived from GFP (figure 4.7). This may indicate that there are PafA specificity determinants making pupylation of some GFP lysines more probable, or that other lysines in GFP were pupylated in a stochastic manner and were therefore too low abundance compared to counterpart non-modified peptides to be detected by mass spectrometry. There is also evidence in the literature to support the existence of poly-PupE chains on a lysine residue; if this is occurring on PupE added to GFP then polypupylation may block cleavage at the lysine immediately before the GGE motif used here as a marker for PupE remnants leading to modified peptides being missed due to the steric hinderance introduced by the poly-PupE branching. In addition, pupylation of

lysines with PupE leads to blocked trypsin cleavage. This may lead to peptides being too large for reliable detection in standard MS/MS proteomic set-ups. Coupled with the potential polypupylation issue highlighted above, many pupylated peptides may just not have been suitable for detection under our processing and analysis workflow. Certainly, the Western blotting data suggests that the pupylation profile of GFP-5xGGSGG-PafA-3xHA is more complex than the mass spectrometry data indicates. The use of alternative proteases may alleviate some of these issues but in all cases examined (Chymotrypsin, Asp-N, Glu-C) substantial (> 10 amino acid) PupE remnants would be left on peptides, likely confounding attempts at analysis. As the presence of these PupE remnants would make it difficult to correctly identify proteins. Confirming the PafA dependant nature of the labelling these GGE modified lysines were only detected from GFP-5xGGSGG-PafA-3xHA and PupE co-expressing samples.

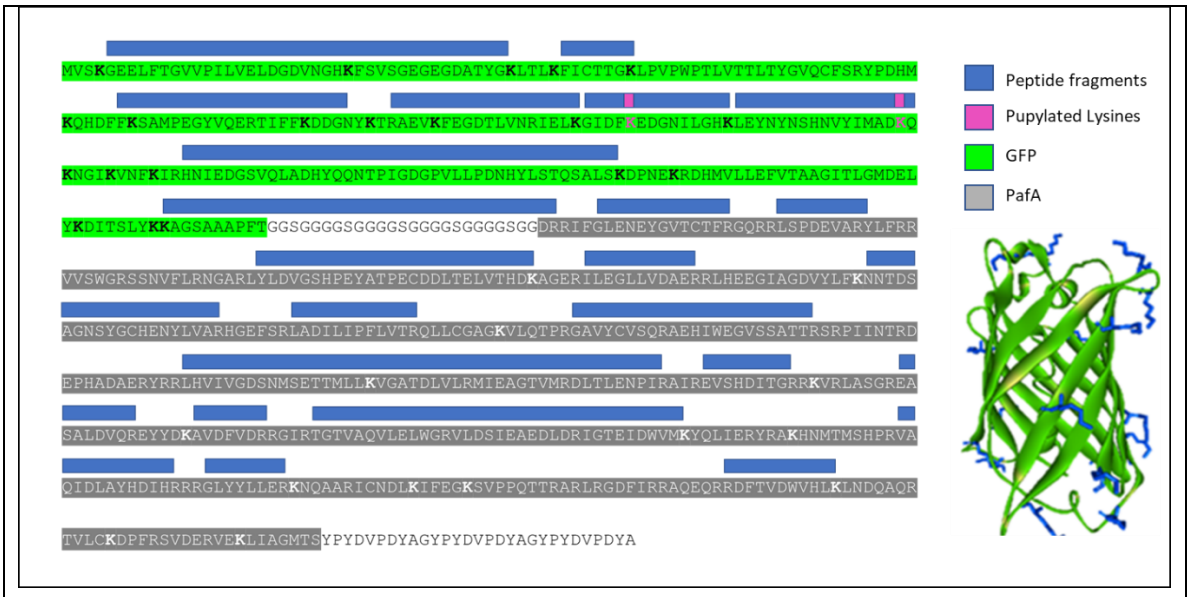


Figure 4.7 illustration of peptide detected from GFP

GFP-5xGGSGG-PafA-3xHA, was transiently expressed in *N. benthamiana* alongside TwinStrep-3xFLAG-PupE$^{46\text{-}72}$ and TwinStrep-3xFLAG-PupA$^{46\text{-}72}$ This was performed twice. Protein was extracted and enriched using the TwinStrep, digested with trypsin and analysed by mass spectrometry. Peptide fragments of the GFP-5xGGSGG-PafA-3xHA construct are shown above sequence as blue bars with identified pupylated lysine's shown in magenta

4.2.4 How does length of PupE affect labelling efficiency?

After establishing that GFP is an effective pupylation substrate it was used as a way to test the effectiveness of pupylation using the different lengths of PupE. It is expected that all of the PupE constructs will interact with PafA resulting in labelling, but it was possible that the slight variations in size might have an impact on efficiency.

GFP-5xGGSGG-PafA-3xHA was transiently expressed in *N. benthamiana* alongside the three PupE constructs, TwinStrep-3xFLAG-PupE$^{46-72,}$ TwinStrep-3xFLAG-PupE$^{41-72,}$ TwinStrep-3xFLAG-PupE$^{1-72,}$ and their non-functional PupA counterparts. Protein was extracted and labelling detected and analysed by Western blot from crude lysates (figure 4.8).

All lengths of PupE, TwinStrep-3xFLAG-PupE$^{46-72,}$ TwinStrep-3xFLAG-PupE$^{41-72,}$ TwinStrep-3xFLAG-PupE$^{1-72}$ were active and capable of labelling across the proteome with TwinStrep-3xFLAG-PupE$^{1-72}$ being most active. Unexpectedly, it also indicated that the full length, TwinStrep-3xFLAG-PupA$^{1-72}$ construct appeared to be active. Given the known prevention of PafA ligase ability by a C-terminal alanine on PupA it is likely that this apparent result stems from mislabelling or mixing of cultures. For the purposes of this work TwinStrep-3xFLAG-PupE$^{46-72}$ and TwinStrep-3xFLAG-PupE$^{41-72}$ were effective and this apparent anomaly with full length TwinStrep-3xFLAG-PupE/A$^{1-72}$ was not investigated further.

We had planned to transiently express GFP-5xGGSGG-PafA-3xHA in *N. benthamiana* alongside the three PupE constructs TwinStrep-3xFLAG-PupE$^{46-72,}$ TwinStrep-3xFLAG-PupE$^{41-72,}$ TwinStrep-3xFLAG-PupE$^{2-72}$ and their non-functional PupA counterparts and then to enrich with TwinStrep and detect

GFP-5xGGSGG-PafA-3xHA by Western blot to establish which PupE seemed to most effectively label GFP. However, with its smaller size, effective labelling visible in crude lysates, and work already having been done using TwinStrep-3xFLAG-PupE$^{46-72}$and TwinStrep-3xFLAG-PupA$^{46-72}$ as the PafA substrate, it was decided to move forward using this PupE construct without further optimisation or analysis.

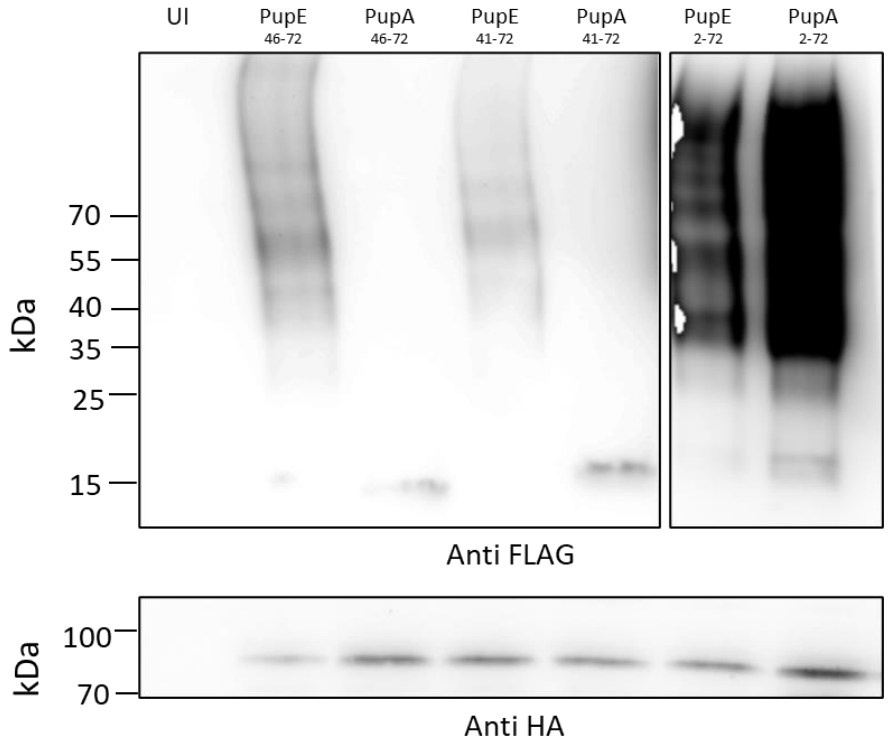


Figure 4.8. Expression of PupE lengths with GFP-PafA

Transient expression of TwinStrep-3xFLAG-PupE$^{46-72}$ and TwinStrep-3xFLAG-PupA$^{46-72}$, TwinStrep-3xFLAG-PupE$^{41-72}$ and TwinStrep-3xFLAG-PupA$^{41-72}$ and TwinStrep-3xFLAG-PupE$^{2-72}$ and TwinStrep-3xFLAG-PupA$^{2-72}$ lengths co-expressed with GFP-5xGGSGG-PafA-3xHA in *N. benthamina*. Resolved by SDS PAGE. PupE detected by anti-FLAG and GFP-5xGGSGG-PafA-3xHA anti-HA Western blot. Both panels represent two different exposures of the same Western blot due to detection levels.

4.3 Conclusion

I have successfully optimised the expression and detection of PupE in plants and in doing so confirmed its true running size of 15kDa and demonstrated that the Ubiquitin leader is successfully cleaved off when expressed in planta. All lengths of PupE are expressed and detectible by Western blot, although it is unclear which length is truly the most effective at labelling. I cannot draw any definitive conclusions from the different length expressions alongside GFP-5xGGSGG-PafA-3xHA as all lengths appeared to label the proteome, but as other work has been successful with $PupE^{46-72}$ I decided to proceed with the more minimal length. If I had more time, I would have repeated the co expression of the different PupE/A length constructs alongside GFP-5xGGSGG-PafA-3xHA as $PupA^{2-72}$ seemed to remain active, once the reason for this apparent anomaly has been identified it may be possible to draw further conclusions.

Initially 3xHA-PafA appeared unstable in bacteria, preventing its expression in plants. However, after the addition of the legumin J intron the construct stabilised, and transient expression was achieved alongside PupE, demonstrating pupylation activity in plants. Labelling only occurring in the presence of both PupE and PafA indicating that there is no endogenous peptide ligase activity in plants capable of using PupE as a substrate. This is a crucial point for minimising background and successful implementation of the system to detect PPIs.

I demonstrated that GFP was a good pupylation acceptor. Not only was the GFP-5xGGSGG-PafA-3xHA fusion still active, but also capable of self-pupylation, indicated by Western blot and shown by the pupylation of two separate lysine’s by mass spectrometry. In doing this I also showed that it was

possible to enrich pupylated proteins using the TwinStrep tag present on my PupE constructs.

This work shows steps taken to optimise the PafA PupE system for the use in plants. I have successfully reconstituted the system in planta and demonstrated its activity. I have also shown its ability to label the commonly used tag GFP and demonstrated activity, labelling, enrichment, and detection of protein using my optimised system. This can now be applied as a Protein-Protein interaction technique to study transient interactions involving membrane proteins in plants. Through the strategies already set out this project aims to apply this technique for both discovery and pairwise interaction analysis; specifically, to define RLK complex composition and RLK-PAT pairs.

Chapter 5

The Application of DogTag as a novel PPI Technique

5.1 Introduction

After successful optimisation and reconstitution of the PafA/ PupE DogTag system in plants the next step is to try and apply it as a protein-protein proximity or interaction detection technique. This new system has the potential to be a more versatile and accurate PPI technique in comparison to existing methods. This includes a potential reduction in false positive interactions due to the affinity label being non-diffusible, and an increased potential for physiological relevance over other similar techniques such as Bio and Turbo ID due to being a closed system that can be fully genetically encoded in an organism. As the system does not require exogenous application of any substance (e.g., biotin) it is also potentially far more suited to whole organism investigation than techniques such as the many BioID variations or APEX (Lam *et al*., 2014, Li *et al.*, 2017).

The PafA/PupE proximity labelling system has the potential to be applied in one of two ways; either as a pairwise tool to test potential interactions between two putative interacting proteins via Western blotting, or as a discovery tool to identify a range of novel interactors of a protein of interest utilizing mass spectrometry for identification (figure 5.1).

In this chapter I demonstrate proof-of-concept work showing the application of the technique for pairwise and discovery work.

5.1.1 Pairwise application

For the pairwise work the experimental goal was to identify the protein S-acyl transferase (PAT) or PATs involved in the S-acylation (a fatty acid-based post translational modification) of the Receptor-like Kinases (RLKs) Flagellin Sensitive 2 (FLS2) and Brassinosteroid Insensitive 1 (BRI1). Previous work has shown that both FLS2 (Hurst *et al*.,2021) and BRI1 (Silu Li, unpublished) are S-acylated both in response to ligand binding and constitutively at a separate site, but it is still unknown which PATs are involved in these processes. Proximity labelling is an ideal method to study interactions between PATS and RLKs because:

- Interaction between PATs and RLKs, being enzyme and substrate, are likely transient and therefore not amenable to methods such as co-immunoprecipitation.
- Interactions with PATs are frequently mediated via a 3rd party adaptor protein (Salaun, *et al*. 2020) making non-plant expression-based systems less likely to work.
- PATS are poly-topic, redox sensitive, integral membrane proteins (Rana, *et al*. 2018) making in vitro methods more difficult to perform, requiring balancing and careful of redox and buffer components (e.g., salts, detergents, reducing agents) making these experiments difficult to optimise. In addition, the use of some detergents can also introduce additional confounding factors, such as inducing detergent-resistant membrane formation, that can artificially stabilise a false positive interaction.

- RLKs are integral membrane proteins contained within large integral membrane complexes.
- Non-plant expression of full length RLKs has not been successful (Silke Robatzek, personal communication), preventing the ability to utilise yeast-based screening systems such as yeast-2-hybrid.
- The whole RLK complex is stabilised by the lipid bilayer (Hurst, *et al.* 2021) so disruption by detergents for e.g., co-IP will disrupt some or all of the interactions.

In keeping with these ideas, protein-substrate interactions with PATs have proved difficult to demonstrate using other established PPI techniques.

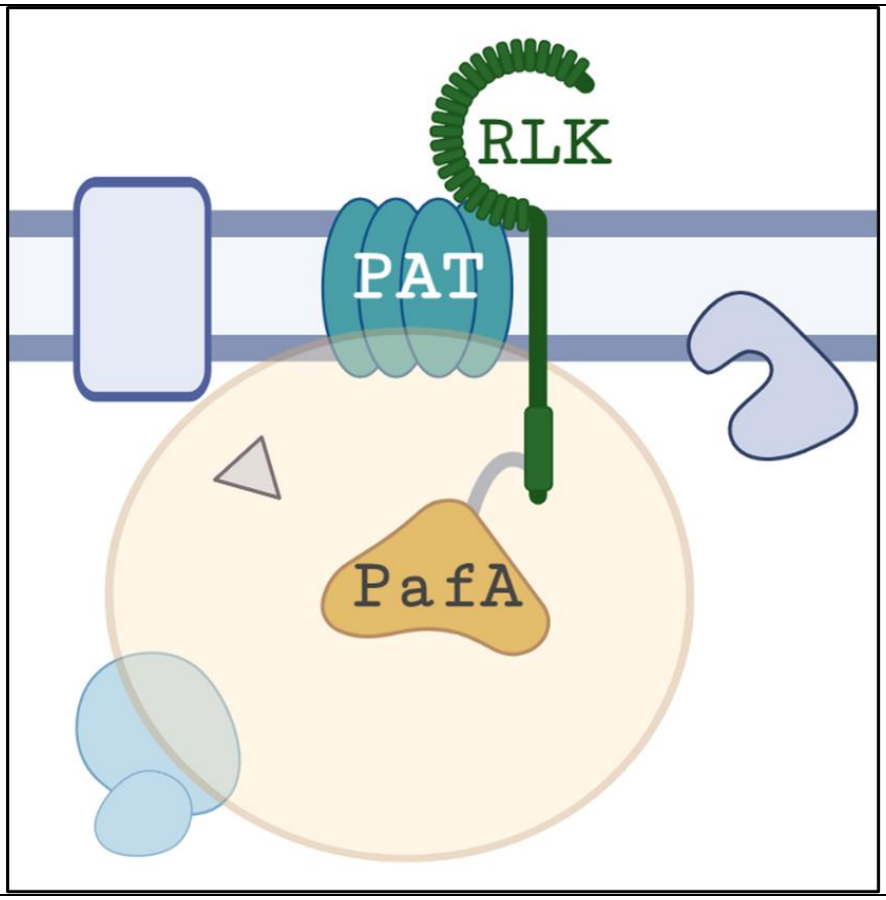


Figure 5.1 DogTag as a discovery tool
Illustration of how a PafA(yellow) /RLK (green) fusion would label proximal proteins within labelling radius (light yellow)

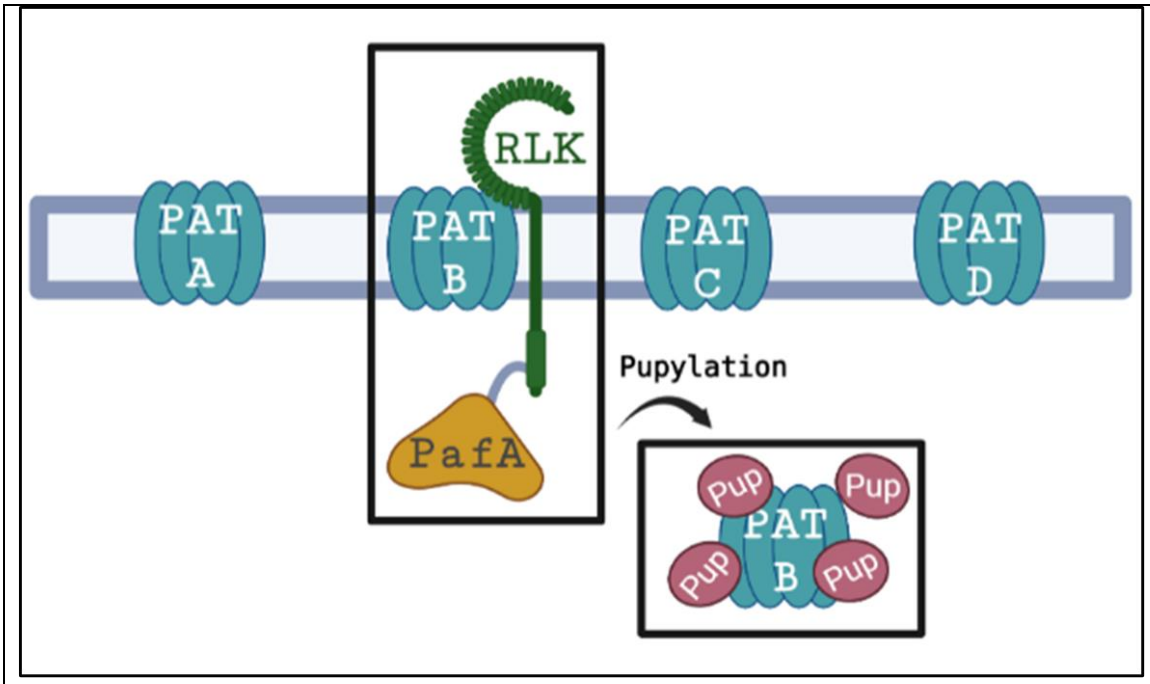


Figure 5.2 DogTag as a pairwise tool
Illustration of how PafA (yellow)/ I/RLK (green) fusion would label proximal PAT proteins. Allowing for identification.

5.1.2 Discovery application

In plants, receptor-like kinases are a large family of integral membrane proteins responsible for perception of a diverse array of extracellular stimuli, including hormone signals, symbiotic interactions, cell wall status and pathogens. Two well characterised RLKs are Flagellin Sensitive 2 (FLS2), involved in immune responses, and Brassinosteroid Insensitive 1 (BRI1), involved in plant development. As these proteins are integral membrane proteins, they are poor candidates for current established PPI approaches.

With both FLS2 and BRI1 being involved in such important plant processes, it is highly likely that there is a large number of PPIs occurring as part of the signalling process that have previously gone undetected by established PPI techniques. Being able to identify these interactions in a physiologically relevant

environment would therefore be a huge advantage to further understand these crucial plant processes.

The PafA/PupE system offers the opportunity, as a discovery too, for the identification of these previously unknown RLK interactors. By simply attaching PafA to the RLK of interest and co-expressing PupE, any proteins in proximity to PafA will be pupylated, allowing for capture and subsequent identification by mass spectrometry.

In this chapter I demonstrate some proof-of-concept work showing the application of the technique for pairwise and discovery work. Including some optimisation towards improved detection and the identification of some interesting potential interactors.

5.2 Results

5.2.1 Cloning of RLK-PafA fusions

One of the ultimate goals of the project is to identify the protein S-acyl transferase(s) (PATs) that S-acylates RLKs. I have therefore created FLS2-PafA and a BRI1-PafA fusions as case study proteins. These fusions will allow for the pupylation of any interacting proteins of FLS2 and BRI1 by PafA.

Previous work in the lab indicated that FLS2 and BRI1 were difficult to work with in *E. coli* in our typical Gateway entry vectors due to limited restriction cloning sites and observed instability of the vectors themselves. I therefore decided to adapt the entry vectors for use with recombinatorial cloning in *Saccharomyces cerevisiae* yeast to minimise passage through *E. coli* and allow for restriction free cloning. The Gateway entry vector pENTR-D-TOPO containing FLS2 and BRI1 were successfully modified for transformation into *S. cerevisiae* by the recombination of a PCR product containing the yeast 2µ origin of replication and URA3 selection marker amplified from pYES2 into the *BsaHI* site of the entry

vector backbone. Subsequently a PCR fragment containing PafA-3xHA was recombined into the yeast adapted entry vectors containing FLS2 or BRI1 using the *AscI* site at the 3' end of FLS2 or BRI1. An *AccI* site was also introduced into this PCR product and separates FLS2 or BRI1 from PafA. A 5x GGSGG linker created by oligo annealing was ligated into the *AccI* site in order to offset any steric hindrance in the resulting protein products.

FLS2-5xGGSGG-PafA-3xHA and BRI1-5xGGSGG-PafA-3xHA were then recombined into the plant expression vector pK7WG2 under control of the 35S promoter. proFLS2:FLS2-5xGGSGG-PafA-3xHA and proBRI1:BRI1-5xGGSGG-PafA-3xHA were also made in a similar manner to that described above as an option to explore more native expression levels if desired or to transform Arabidopsis.

5.2.2 Expression tests of RLK-PafA fusions

FLS2-5xGGSGG-PafA-3xHA and BRI1-5xGGSGG-PafA-3xHA were transiently expressed in *N. benthamiana* using agrobacterium infiltration. Protein was extracted and expression determined by Western blot from crude lysates (Figure 5.3 & 5.4). Both FLS2 and BRI1 PafA fusions express at the expected size, with both running heavier than their amino acid sequence would suggest, likely due to the reported effects of glycosylation (Häweker *et al*., 2010). Interestingly BRI1-5xGGSGG-PafA-3xHA fusion consistently showed a second anti-HA-reactive band, absent from non-infiltrated controls, at approximately 350-400 kDa that was resistant to heating at 65°C in reducing SDS-PAGE sample buffer. The identity and composition of this second band was not determined but may be a constitutive BRI1 homodimer as reported previously (Wang *et al*., 2005).

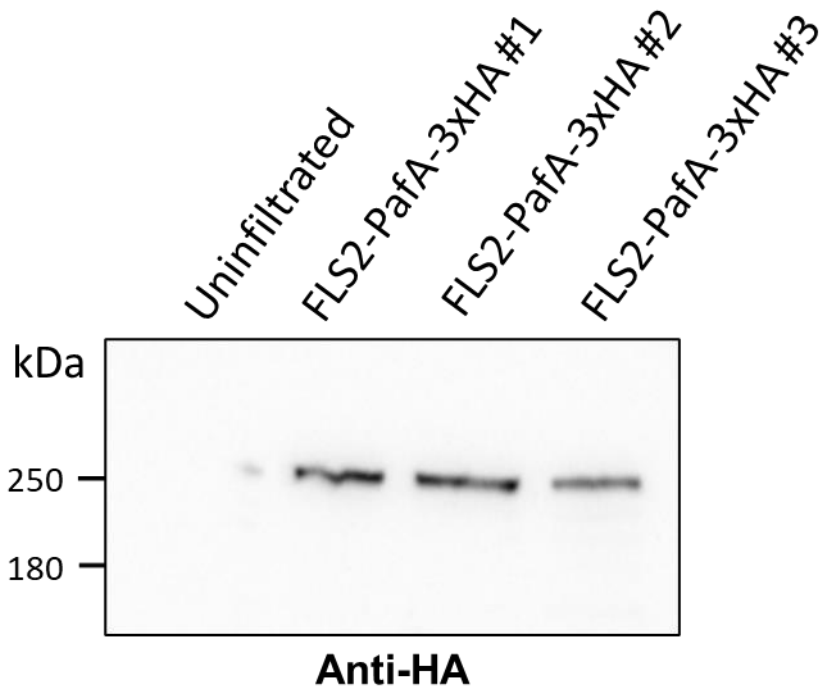


Figure 5.3 Transient expression of FLS2/ PafA fusions
Three colonies of FLS2-5xGGSGG-PafA-3xHA (FLS2-PafA) were individually and transiently in *N. benthamiana.* separated by SDS PAGE and detected by anti-HA Western blot.

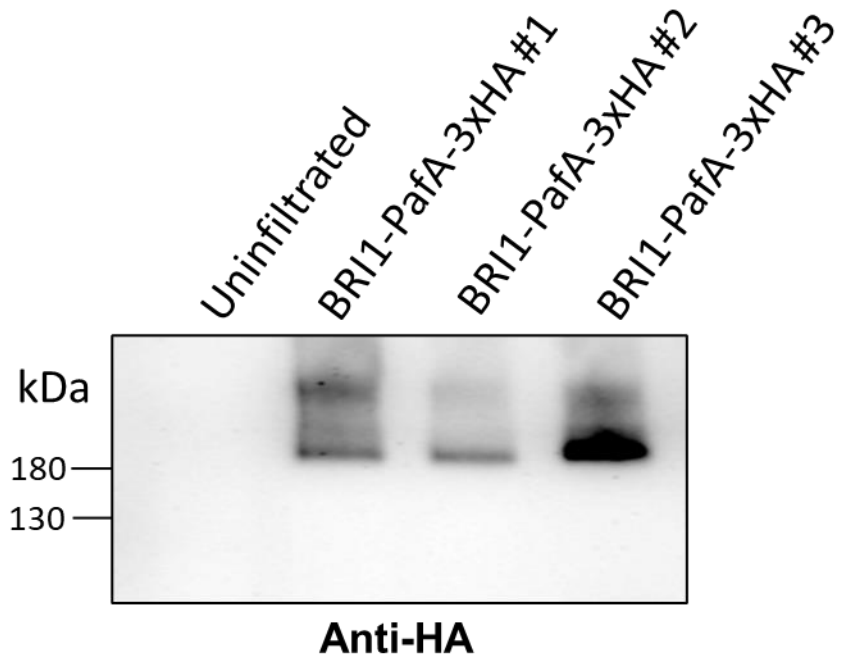


Figure 5.4 Transient expression of BRI1/ PafA fusions
Three colonies BRI1-5xGGSGG-PafA-3xHA (BRI1-PafA) were individually and transiently in *N. benthamiana.* separated by SDS PAGE and detected by anti-HA Western blot.

5.2.3 Implementing DogTag as a pairwise proximity detection method

After the successful expression and detection of FLS2-5xGGSGG-PafA-3xHA and BRI1-5xGGSGG-PafA-3xHA the constructs were judged ready to use in in pairwise proximity labelling experiments alongside PATs to attempt to identify which PATs S-acylate these RLKs. A limited number of PATs were selected for this to reduce the number being considered down from 24 to make experimental work more manageable. Based on previous work we know that S-acylation of RLKs occurs at the plasma membrane (Hurst *et al.* 2021) in response to ligand binding, we therefore excluded all PATs reported to localise to non-plasma membrane compartments within the cell (Batistic 2012). This reduced the number of PATs to examine down to 9; PATs 4, 5, 6, 7, 8, 9, 12, 19 and 21.

Based on my previous work GFP appears to be a useful tag for pupylation, as well as being a good epitope for Western blotting and visual marker for protein expression. To perform these pairwise experiments plasma membrane localised PATs fused to GFP (Batistic 2012) were transiently co-expressed alongside TwinStrep-3xFLAG-PupE$^{46-72}$ and FLS2-5xGGSGG-PafA-3xHA or BRI1-5xGGSGG-PafA-3xHA in *N. benthamiana*. Protein was extracted and enriched using the TwinStrep tag present on TwinStrep-3xFLAG-PupE$^{46-72}$ to pulldown pupylated proteins, representing proteins that had been in close proximity to either FLS2 or BRI1. Anti-GFP Western blotting was then performed to estimate enrichment or depletion of PATs compared to input levels.

For FLS2-5xGGSGG-PafA-3xHA one pilot replicate was performed looking for interaction with PAT6, PAT9 and PAT12. Previous work in the lab using an S-acylation boost assay indicated they may be good potential interactors (Charlotte Hurst, unpublished). As FLS2 is involved in immune responses it will

not be activated in *N. benthamiana* without treatment with the FLS2 ligand peptide flg22. 48 hours after infiltration with FLS2-5xGGSGG-PafA-3xHA, TwinStrep-3xFLAG-PupE$^{46-72}$ and the relevant PATs, half the leaves were treated with 10 µM flg22 while the other was treated with water only and samples harvested after 2h. Therefore, for each PAT pairwise test there was an active and inactive FLS2 sample, hopefully enabling identification of the enzyme underlying the dynamic acylation that occurs due to FLS2 activation. Protein was then extracted and inputs for FLS2-PafA, the PATs and total pupylated proteome labelling visualised by Western blot. PAT-GFP expression was also confirmed before lysis by fluorescence microscopy. After confirmation of expression, pupylated proteins were enriched using the TwinStrep tag. Enriched samples were then resolved by SDS-PAGE and PAT recovery determined by anti-GFP Western blot. (Figure 5.5)

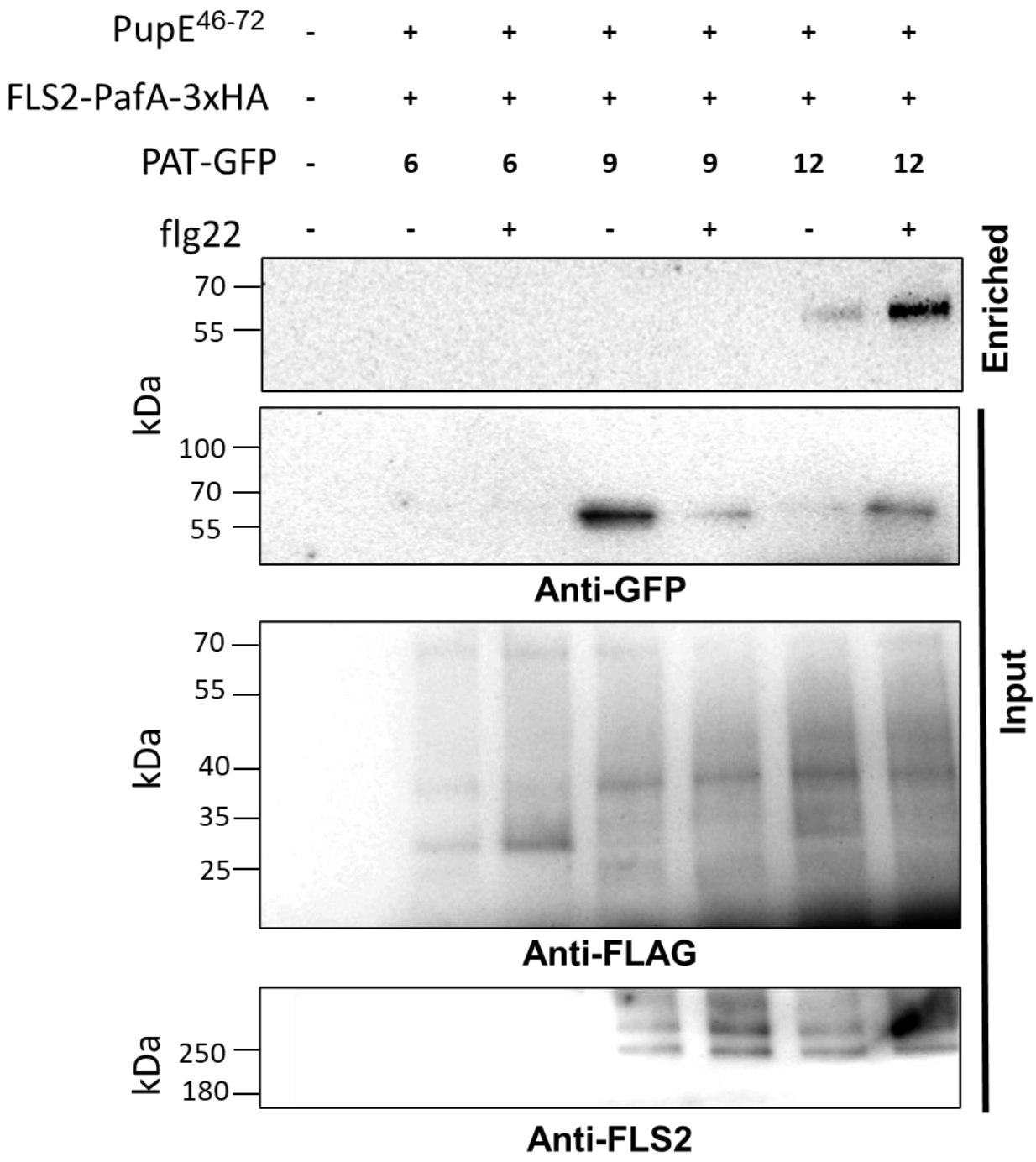


Figure 5.5 FLS2 pairwise PAT identification

FLS2-5xGGSGG-PafA-3xHA (FLS2-PafA), TwinStrep-3xFLAG-PupE$^{46-72}$ (PupE) and the relevant PATs were transiently co-expressed in *N. benthamiana.* 48 hours after infiltration half the leaves were treated with 10 µM flg22 while the other was treated with water only and samples harvested after 2h. Inputs to check for FLS2, PupE and PAT expression were done by anti-FLS2, anti-FLAG and anti-GFP Western blots respectively. Samples were then enriched using TwinStrep and pupylated PATs detected by anti-GFP Western blot.

Based on the one FLS2-5xGGSGG-PafA-3xHA pairwise experiment there is the possibility that PAT12 may be an interactor, however it was also detected in the non activated FLS2 sample. This may indicate that PAT12 is always present in the FLS2 complex regardless of activation state. Given the qualitative nature of these data, and lack of a no PupE sample as a further negative control for non-specific binding, it is hard to make further conclusions about relative enrichment. PAT9 was detected in the input, but not in the enriched sample, strongly suggesting no interaction with active or inactive FLS2. Unfortunately, despite repeated attempts, PAT6 appeared to not express. Despite some interesting indications, because this was only done once, it is impossible to draw clear conclutions from this data beyond indicating that PAT12 appears to be close to FLS2 while PAT9 isn't. Due to the extra difficulty and technical steps invovled in activating FLS2, unlike BRI1 which would posess some native activity, for these pairwise experients it was decided to focus the rest of the pairwise work on BRI1 where an endogenous ligand (brassinosteroids) is present in plants.

For BRI1-5xGGSGG-PafA-3xHA multiple replicates were performed looking for interaction with all known plasma membrane PATS; 4, 5, 6, 7, 8, 9, 12, 19 and 21, along with the tonoplast membrane PAT, PAT10, as a negative control. Unlike FLS2 there is expected to be some constitutive activation of BRI1 in *N. benthamiana* due to its role as the principle brassinosteroid receptor involved in plant development and the universal presence of brassinosteriods in plants. It therefore shouldn't be necessary to activate BRI1 samples with ligand. As before, protein was extracted and inputs for the PATs, BRI1 and pupylated proteome labelling visualised by Western blot. PAT-GFP expression was also confirmed before lysis by fluorescence microscopy. After confirmation of

expression pupylated proteins were enriched using the TwinStrep tag. These enriched samples were then resolved by SDS-PAGE and PAT enrichment detected by anti-GFP Western blot (Figure 5.6, 5.7 & 5.8).

Despite being a negative control some PAT10 does seem to have been identified in the enriched sample. This indicates that either PAT10 and BRI1 encounter each other at some point, possibly in the ER/Golgi during protein synthesis and transport or at the vacuole when activated BRI1 is degraded,(Martins *et al*., 2015, Russinova et al., 2004) or due to compartment “leakage” of over-expressed protein. We also cannot discount the possibility of free PafA in the cell following BRI1-PafA turnover or proteolysis. Due to this, the enrichment signal of PAT10 is considered to be the baseline signal above which a plasma membrane PAT must reach for consideration as a potential BRI1 interactor.

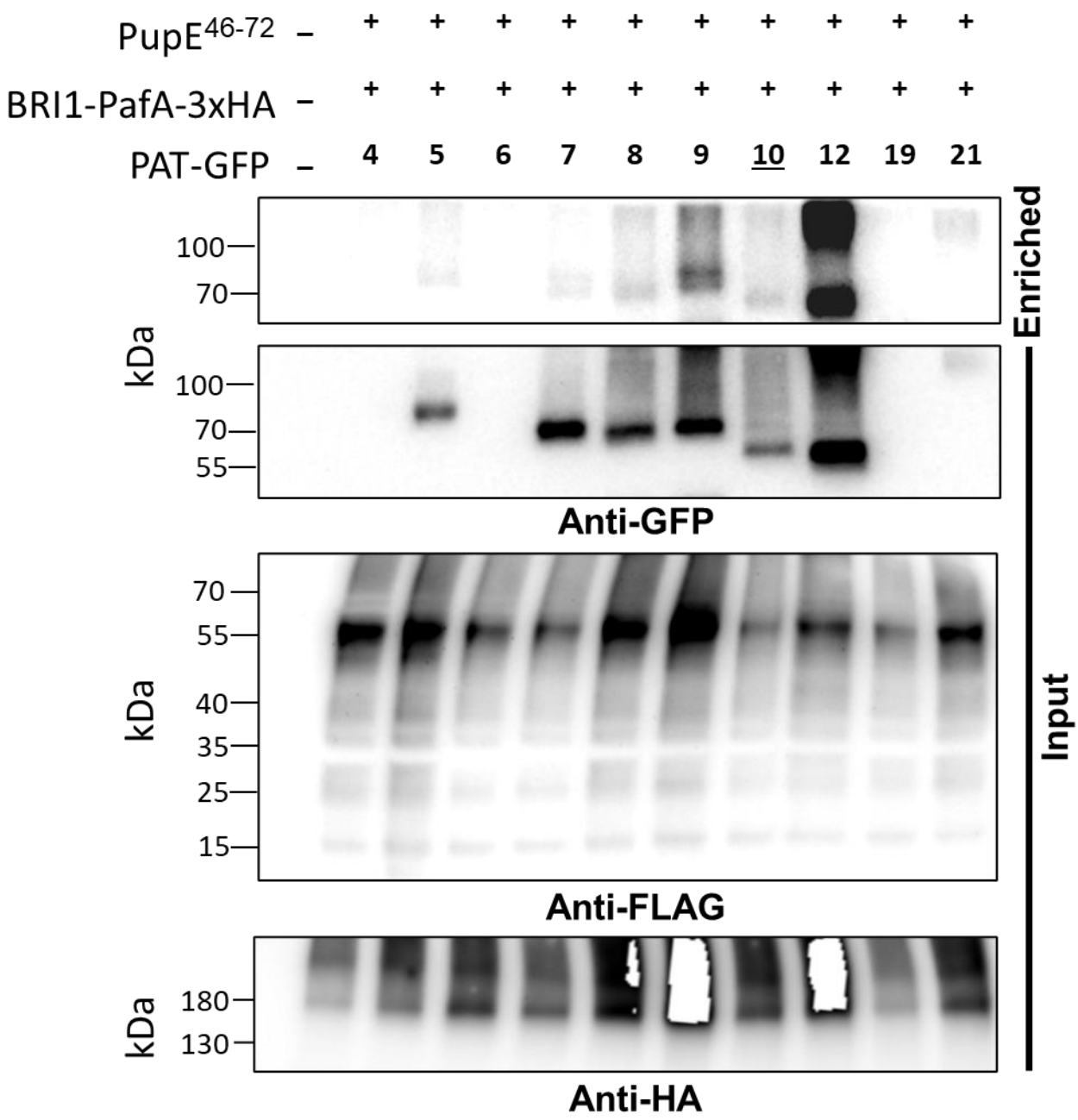


Figure 5.6 BRI1 pairwise PAT identification REP1

BRI1-5xGGSGG-PafA-3xHA (BRI1-PafA), TwinStrep-3xFLAG-PupE$^{46-72}$ (PupE) and the relevant PATs were transiently co-expressed in *N. benthamiana*. Inputs to check for BRI1, PupE and PAT expression were done by anti-HA, anti-FLAG and anti-GFP Western blots respectively. Samples were then enriched using TwinStrep and pupylated PATs detected by anti-GFP Western blot. PAT10 (underlined) was used as a negative control as it is not a plasma membrane PAT like the others chosen.

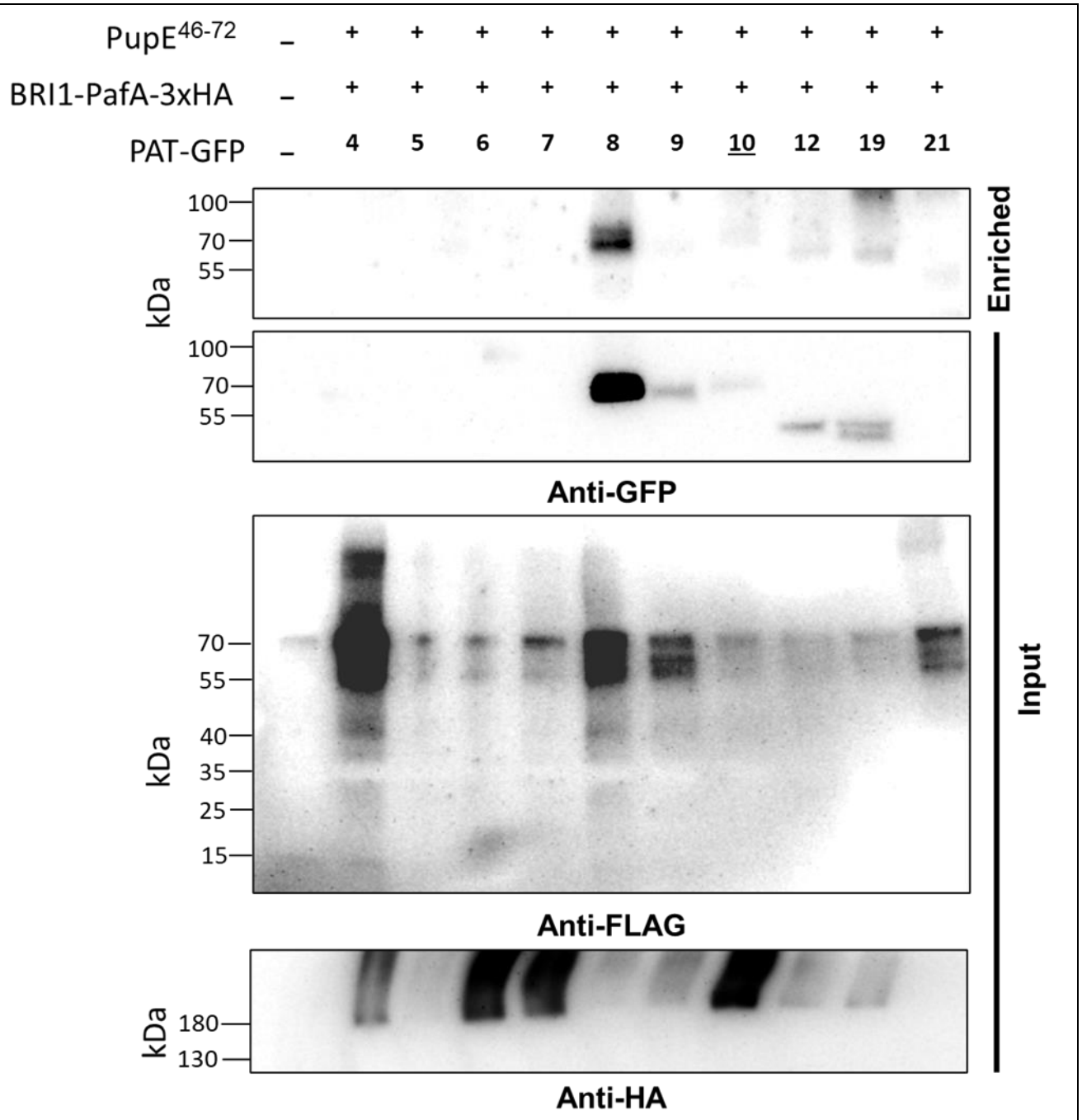


Figure 5.7 BRI1 pairwise PAT identification REP2

BRI1-5xGGSGG-PafA-3xHA (BRI1-PafA), TwinStrep-3xFLAG-PupE$^{46-72}$ (PupE) and the relevant PATs were transiently co-expressed in *N. benthamiana*. Inputs to check for BRI1, PupE and PAT expression were done by anti-HA, anti-FLAG and anti-GFP Western blots respectively. Samples were then enriched using TwinStrep and pupylated PATs detected by anti-GFP Western blot.

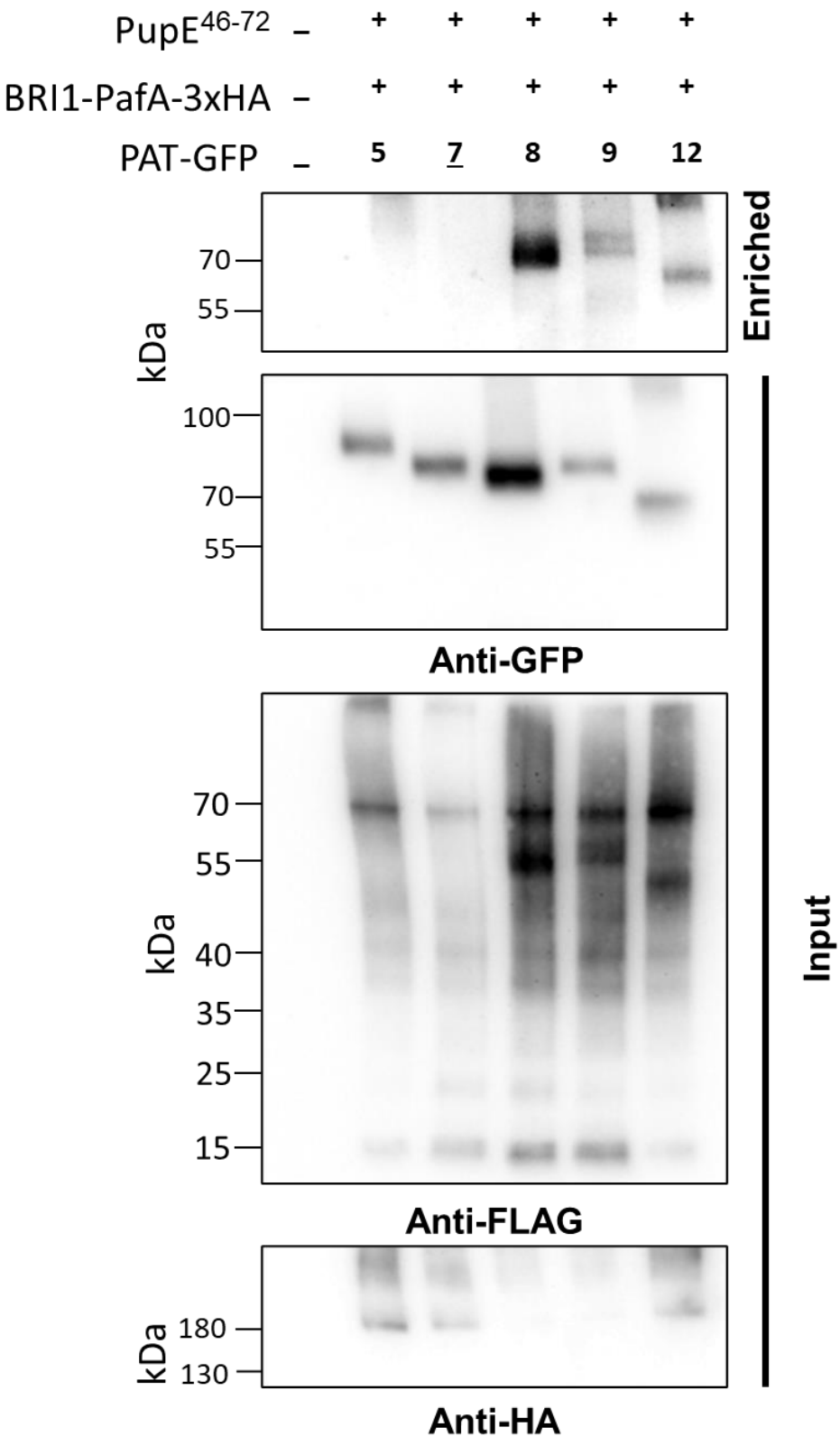


Figure 5.8 focused BRI1 pairwise PAT identification
BRI1-5xGGSGG-PafA-3xHA (BRI1-PafA), TwinStrep-3xFLAG-PupE$^{46-72}$ (PupE) and the relevant PATs were transiently co-expressed in *N. benthamiana*. Inputs to check for BRI1, PupE and PAT expression were done by anti-HA, anti-FLAG and anti-GFP Western blots respectively. Samples were then enriched using TwinStrep and pupylated PATs detected by anti-GFP Western blot.

Two replicates including all plasma membrane PATS and PAT10 were carried out (Figures 5.6 and 5.7) and one additional, slimmed down, replicate (Figure 5.8) looking at some potentially positive PATs, along with an unlikely interactor (PAT7).

The two full replicates indicate that PATs,8, 9 and 12 are BRI1 interacting candidates, and this was confirmed in the more slimmed down replicate where they came out strongly in the enriched samples. PAT5 appeared to be on the borderline for selection as a BRI1 interactor, however, in the focussed repeat PAT5 enrichment was not particularly strong. PAT7 was expressed well in the first replicate and is found at the PM but not detected in the enriched sample, it was therefore used as an unlikely interactor in the more focused replicate as a better negative control than PAT10, where it continued to go undetected in the enriched sample. Thus, it seems likely that PATs 5 and 7 are not interactors of BRI1.

Unfortunately, PATs 4 and 6 were not detected by Western blot, even in the input samples, suggesting that they did not express. Despite repeated attempts, PATs4 and 6 could not be reliably expressed. However, as some interesting potential plasma membrane PAT interactors were detected without them, it wasn’t deemed necessary in the time available to attempt to successfully express them.

5.2.4 DogTag as a discovery tool for protein-protein interactions

BRI1 forms multi-protein complexes in the plasma membrane (Aker *et al.,*2008). However, it is highly likely that BRI1 has a large range of PPIs and protein-proximal interactions that have not been detected by traditional methods. This is likely due to the fact that traditional or established PPI detection techniques are not suited to RLKs as integral membrane proteins that are suspected to have many transient and weak PPIs, many of which will be stabilised by the membrane environment. Due to these issues the DogTag system is a good option, orthogonal to prior methods, for discovery work by mass spectrometry, as it may be able to detect these as yet unidentified weak, proximal or membrane stabilised interactions.

BRI1-5xGGSGG-PafA-3xHA and TwinStrep-3xFLAG-PupE$^{46-72}$ were individually or co-expressed transiently in *N. benthamiana.* Proteins were extracted and expression and degree of pupylation determined by Western blot (figure 6.9). Pupylated proteins from each sample were then enriched by magstrep type 3 XT affinity purification. Samples of just TwinStrep-3xFLAG-PupE$^{46-72}$, BRI1-5xGGSGG-PafA-3xHA and TwinStrep-3xFLAG-PupE$^{46-72}$ and BRI1-5xGGSGG-PafA-3xHA were then eluted from the beads with 1x reducing SDS-PAGE sample buffer. An aliquot was analysed by SDS-PAGE and Western blotting for the extent of labelling and capture. The remaining sample was separated by SDS-PAGE, excised as one piece of acrylamide, in-gel trypsin digested and cleaned up by C18 tips, followed by analysis using nano-liquid chromatography mass spectrometry (Thermo Q-Exactive HF, 120-minute gradient; 5-60% acetonitrile). Three repeats were performed.

Raw data files were analysed using MaxQuant (v1.6.17.0; default settings) apart from the following user determined settings: protease: Trypsin/P (Specific, max 2 missed cleavages); variable modifications: Oxidation (M); Acetyl (Protein N-term); Glu->pyro-Glu; fixed modifications: Carbamidomethyl (C); min peptide length: 7 amino acids; max peptide mass: 7000 Da; LFQ enabled; .NET core disabled. A subsequent search, taking into account the post-Trypsin GGE remnant of PupE conjugated to lysines by PafA, was also performed as above, but allowing for 3 missed cleavages and including GGE (K) as an additional variable modification. In both cases the NbDE proteome dataset was used to search against (Kourelis, *et al*. 2019).

Subsequent inspection of the protein Groups output file indicated that only 51 peptides had been identified across all 3 samples. As the system was over expressed, it was expected that some known BRI1 interacting proteins would be identified in all 3 samples. This was remarkably low based on past experience from the lab, but reanalysis to discount technical errors in MaxQuant set up did not result in increased numbers of proteins. Due to the low numbers of identified peptides and proteins, and limited LFQ identifications across biological repeats, a full statistical analysis would be impossible to perform in a meaningful manner as the data was so limited. A semi-quantitative analysis based on spectral counting was therefore performed. This revealed a number of proteins that were identified solely or primarily in the BRI1-PafA/PupE samples, but not in the negative controls, and could therefore be BRI1 interactors (table 6.1). In particular the identification of 10 peptides mapping to HSP80/90 family members in all 3 experimental repeats, but in none of the controls, suggests that this is a genuine association. However, the semi-quantitative nature of this

analysis, coupled with the relatively limited numbers of proteins identified, precludes major conclusions being made.

After this first exploration for candidate interactors by mass spectrometry revealed some potential interactors, but a high degree of signal from likely contaminants and low total identifications, the next step was to attempt some optimisations in order to improve the sensitivity of using the DogTag system as a discovery tool. Three approaches were tested:

1. Protamine sulfate treatment. Despite bead-based enrichment, the sheer abundance of some proteins in plants, such as RuBisCo, components of the light harvesting complexes and RuBisCo activase, present a possible issue when attempting to detect less abundant but interesting BRI1 interactors by mass spectrometry as the intensity of peptides from these proteins are often 1-3 orders of magnitude higher and will therefore be preferentially selected for analysis during mass spectrometry runs. To combat this, protamine sulfate precipitation (Kim, Lee et al. 2013) was used to try and deplete these highly abundant, but likely irrelevant, proteins from input samples, potentially allowing for a clearer picture of other proteins present.
2. Addition of a BRI1 brassinosteroid binding mutant as a negative control. This mutant copies the Gly644 to Asp mutation observed in the severe *bri1-6* loss of function allele (Noguchi, *et al*. 1999), and is proposed to disrupt binding of brassinosteriods to the island domain of BRI1 within leucine-rich repeat 21 without disrupting protein folding (Hothorn, *et al.* 2011). This mutant will therefore not form a brassinosteroid induced dimer with BAK1 and will not initiate signalling downstream of BRI1. This should allow me to distinguish between proteins that interact with BRI1 as part of

brassinosteroid induced activation, from those that are stable interactors or form part of BRI1 synthesis, maturation, and recycling.

3. On-bead digestion. To further improve sensitivity, samples will be processed by on-bead digestion before being sent for mass spec analysis. Other work in the lab comparing in-gel with on-bead digestion for S-acyl proteomic studies indicates that on bead digestion appears to give lower background and greater sensitivity in mass spectrometry-based analyses.

To try and identify novel BRI1 interactors the following combinations were transiently co-expressed in *N. benthamiana* (two replicates performed).

1. BRI1-5xGGSGG-PafA-3xHA and TwinStrep-3xFLAG-PupE$^{46-72}$ (experimental).
2. BRI1 G^{644}D-5xGGSGG-PafA-3xHA (denoted as BRI1*) clones #3 and #5 and TwinStrep-3xFLAG-PupE$^{46-72}$ (control for brassinosteroid inducible interactions).
3. BRI1-5xGGSGG-PafA-3xHA and non-conjugable TwinStrep-3xFLAG-PupA$^{46-72}$ (control for background binding).

Tissue was harvested and lysed from both replicates, one replicate was treated with protamine sulfate and the other replicate without. Samples were then resolved by SDS-PAGE and BRI1 and PupE labelling detected by Western blot (figure 6.10). BRI1, BRI1* and PupE labelling expression was only seen in the samples without the addition of protamine sulfate. This indicates that, despite published evidence suggesting otherwise (Kim, *et al.* 2013), protamine sulfate also depleted BRI1 and BRI1* levels in the sample rendering it undetectable. As a result, further processing of samples without the addition of protamine sulfate was performed.

Protein was extracted and enriched using the TwinStrep tag present on TwinStrep-3xFLAG-PupE$^{46-72}$ by magstrep type 3 XT beads, to pulldown pupylated proteins representing proteins that had been in close proximity to BRI1. After enrichment samples were reduced with DTT and cysteine sulfhydryls blocked with iodoacetamide. Samples were then digested using trypsin/Lys-C on bead. Digested peptides were then separated from the beads and send for analysis by Mass Spectrometry.

As before Raw data files were analysed using MaxQuant (v1.6.17.0 default settings) apart from the following user determined settings: protease: Trypsin/P (Specific, max 2 missed cleavages); variable modifications: Oxidation (M); Acetul (Protein N-term); Glu->pyro-Glu; fixed modifications: Carbamidomethyl (C); min peptide length: 7 amino acids; max peptide mass: 7000 Da.; LFQ enabled; .NET core disabled. A subsequent search, considering the post-Trypsin GGE remnant of PupE conjugated to lysines by PafA, was also performed as above, but allowing for 3 missed cleavages and including GGE (K) as an additional variable modification. In both cases the NbDE proteome dataset was used to search against (Kourelis *et al*. 2019).

After inspection of the protein groups output file significantly more proteins were identified following on-bead digestion compared to initial screens, 261 compared to 51, respectively. Although overall protein numbers still did seem low, possibly due to a less than complete reference proteome. Making it possible that more peptides were detected but could not be identified. As well as this many peptides identified seemed to appear in low numbers, with the vast majority appearing only once, but often across more than one of the samples. Compared to the initial protein discovery candidates identified also differed, likely due to increased peptides being captured and the addition of the BRI1*

controls. Due to only one replicate of this optimised mass spectrometry discovery, it was not possible to do any analysis or draw any major conclusions, but the results to allude to further interesting potential interactions and show that the changes made do seem to have improved outputs.

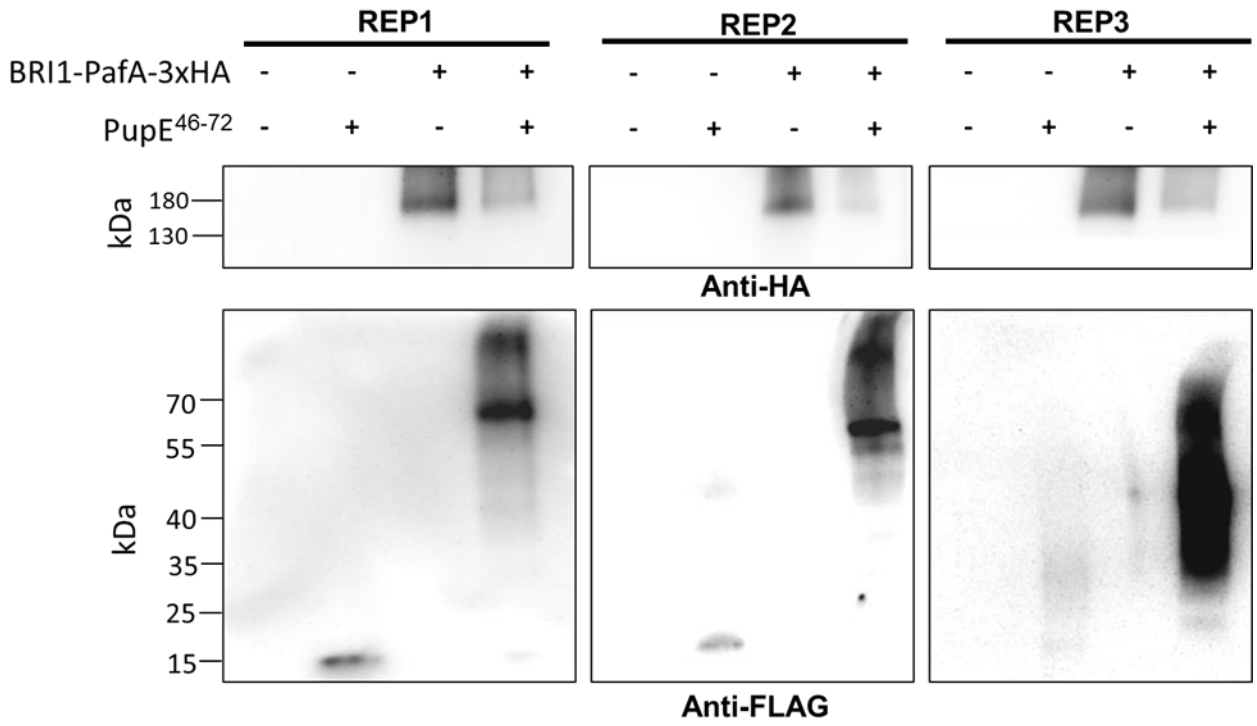


Figure 5.9 PupE and BRI1-PafA inputs for Mass Spectrometry
Three replicates of BRI1-5xGGSGG-PafA-3xHA (BRI1-PafA), TwinStrep-3xFLAG-PupE$^{46-72}$ (PupE) were transiently co-expressed and individually expressed in *N. benthamiana.* Expression of BRI1-PafA was determined by anti-HA Western blot and pupylation levels and PupE expression determined by anti-FLAG Western blot before sample enrichment by TwinStrep.

Primary Protein ID	Description	1	2	3	Sum	1	2	3	Sum
NbD003943.1	heat shock cognate protein 80-like (XP_009763345.1)	0	0	0	0	6	11	1	18
NbD037404.1	heat shock protein 82 (XP_019232766.1)	0	0	0	0	0	3	0	3
NbD042270.1	Partial, ribulose bisphosphate carboxylase/oxygenase	0	0	0	0	0	3	0	3
NbE44071750.1	HSP70 (ARH56399.1)	0	0	0	0	0	3	0	3
NbD000081.1	ribulose-1,5-bisphosphate carboxylase/oxygenase small	0	0	0	0	1	1	0	2
NbD045567.1	biotin carboxyl carrier protein of acetyl-CoA carboxylase-like isoform X2	0	0	0	0	0	2	0	2
NbD048983.1	fructose-bisphosphate aldolase 1, chloroplastic (XP_016505617.1)	0	0	0	0	0	2	0	2
NbD001885.1	ATP synthase CF1 alpha subunit (plastid) (NP_054481.2)	0	0	0	0	0	2	0	2
NbE05067621.1	biotin carboxyl carrier protein of acetyl-CoA carboxylase, chloroplastic-like (XP_016439803.1)	0	0	0	0	0	0	1	1
NbE05068898.1	Partial, uncharacterized protein LOC107762784 (XP_016436653.1)	0	0	0	0	0	1	0	1
NbE03062097.1	BRI1 protein (ABO27628.1)	0	0	0	0	0	1	0	1
NbD029427.1	Partial, heat shock cognate 70 kDa protein (XP_016481229.1)	0	0	0	0	0	1	0	1
NbD036057.1	transketolase, chloroplastic-like (XP_019244962.1)	0	0	0	0	0	1	0	1
NbD046047.1	uncharacterized protein LOC107824885 (XP_016507178.1)	0	0	0	0	0	1	0	1
NbD025147.1	catalase isozyme 3 (XP_016446003.1)	0	0	0	0	0	1	0	1
NbD046317.1	S-adenosylhomocysteine hydrolase (BAR72292.1)	0	0	0	0	0	1	0	1
NbE03058101.1	ruBisCO large subunit-binding protein subunit alpha-like (XP_016503051.1)	0	0	0	0	0	1	0	1
NbD051479.1	ras-related protein Rab11D/RabA5E (XP_019264623.1)	0	0	0	0	0	1	0	1
NbD004137.1	microtubule-associated protein 70-5-like (XP_016498926.1)	0	0	0	0	0	0	1	1
NbD026078.1	pectinesterase/pectinesterase inhibitor U1 (XP_019234085.1)	0	0	0	0	0	0	1	1
NbE44069874.1	nodulin-related protein 1-like isoform X2 (XP_018631997.1)	0	0	0	0	0	1	0	1
NbD009244.1	cytochrome P450 78A5-like (XP_019245900.1)	0	0	0	0	0	1	0	1
NbE44073124.1	BEACH domain-containing protein B isoform X1 (XP_019227812.1)	0	0	0	0	0	1	0	1
NbE05062980.1	craniofacial development protein 2-like, partial (XP_016433452.1)	0	0	0	0	0	1	0	1
NbD008367.1	fructose-bisphosphate aldolase 1, chloroplastic-like (XP_019254826.1)	0	0	0	0	0	1	0	1
NbD042178.1	peroxisomal (S)-2-hydroxy-acid oxidase (XP_019251654.1)	0	0	0	0	0	1	0	1
NbD023654.1	phosphatidate cytidylyltransferase 4, chloroplastic-like isoform X1 (XP_016485932.1)	0	0	0	0	0	1	0	1

Table 5.1 attractive BRI1 interactor candidates.
Candidates identified by spectral count based on mass spectrometry data as good potential interactors of BRI.

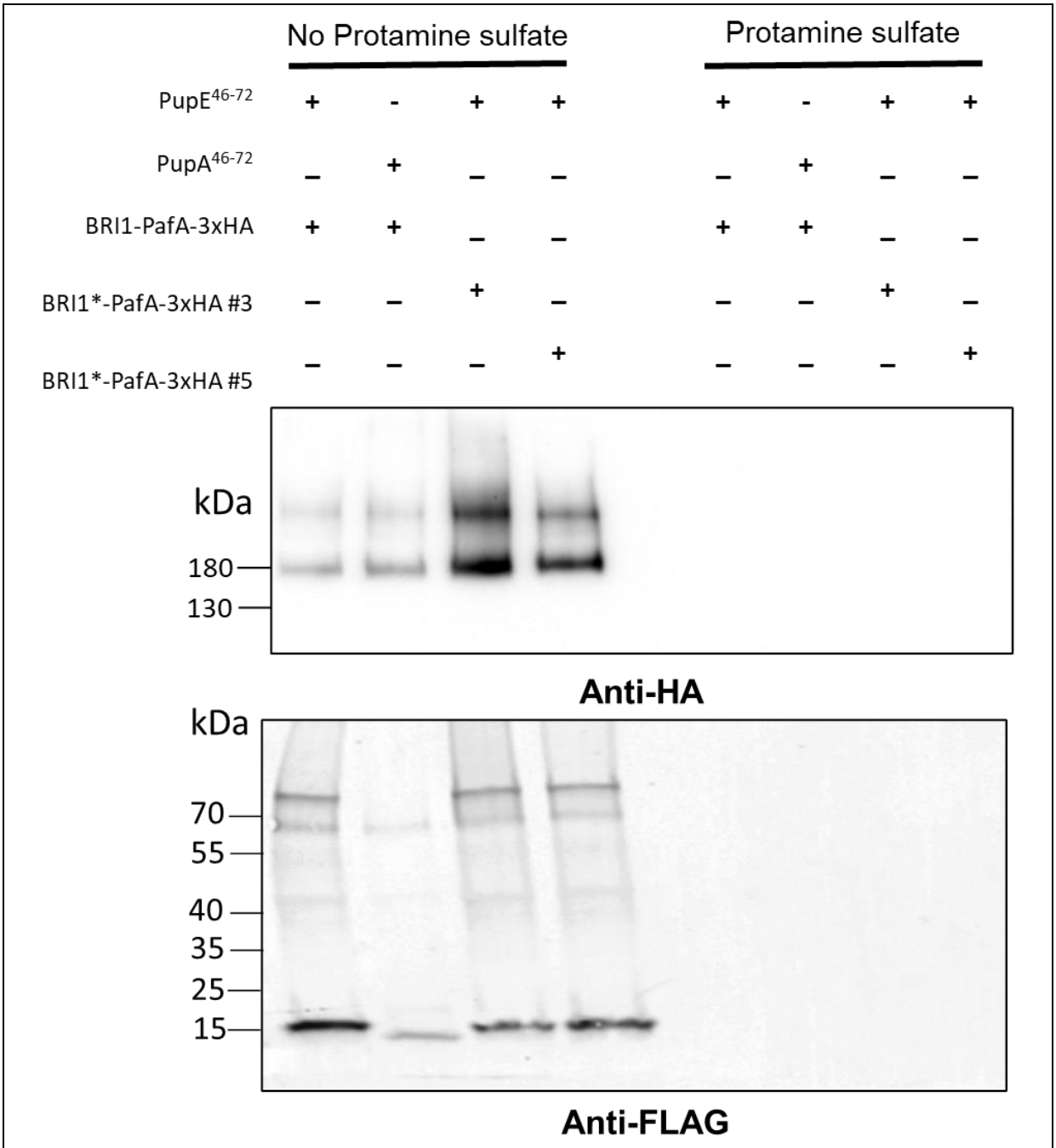


Figure 5.10 PupE and BRI1-PafA input for optimised Mass Spectrometry

Two replicates of BRI1-5xGGSGG-PafA-3xHA (BRI1-PafA), TwinStrep-3xFLAG-PupE$^{46-72}$ (PupE), TwinStrep-3xFLAG-PupA$^{46-72}$ (PupA)and BRI1 G644D-5xGGSGG-PafA-3xHA (denoted as BRI1*) clones #3 and #5 were transiently co-expressed and individually expressed in *N. benthamiana.* During processing one replicate was treated with protamine sulfate whilst the other was left untreated. Expression of BRI1-PafA was determined by anti-HA Western blot and pupylation levels and PupE and PupA expression determined by anti-FLAG Western blot before sample enrichment by TwinStrep.

5.3 Conclusion

I have successfully started the application and demonstrated functionality of the PupE/ PafA DogTag system as a discovery and pairwise proximity labelling technique for traditionally difficult proteins, using RLKs and PATs as case studies.

I have shown successful expression of my BRI1 and FLS2 PafA fusions transiently in *N. benthamiana* by Western blot and subsequently used these constructs to validate the PafA/PupE system to identify interactors. For the identification of any RLK interacting PATs I have shown that co-expression of RLK-PafA fusions alongside TwinStrep-3xFLAG-PupA$^{46-72}$ and the PATs of interest seems to result in a level of selective labelling and isolation by TwinStrep enrichment. This can be seen by the consistent identification of PATs 8, 9 and 12 as being proximal to BRI1 and stands in contrast to the consistent lack of PAT7 in enriched samples, despite seemingly strong expression before enrichment. This demonstrates a level of specificity within the system, even when done transiently with 35S promoter driven expression. Interestingly, despite being initially used as a negative control, PAT10 showed weak presence in pull-down samples. This could be due to non-specific binding to the magstrep type 3 XT beads or background labelling. Background labelling could be caused by free PafA from limited proteolysis of BRI1-PafA fusions in the cell. However, this seems unlikely given the consistent negative behaviour of PAT7. Alternatively, BRI1 and PAT10 may share sub-compartments during synthesis (ER), maturation (ER or Golgi), trafficking (TGN or vesicles) or BRI1 degradation (MVB or the tonoplast - PAT10 proposed location) resulting in labelling occurring at non-plasma membrane sites. Finally, PAT10 may show

inappropriate localisation when over-expressed in our system resulting in inappropriate labelling; a similar phenomenon was reported during cell biological studies where PAT10 was also observed on Golgi-like vesicles when expressed at high levels (Batistic 2012).

I have also shown that the system can be applied for discovery science, with constructs expressing well and labelled proteome being successfully enriched before being sent for mass spectrometry analysis. Proteins identified by the first round include BRI1 itself, indicating either auto-labelling or that BRI1 monomers occur in close proximity to each other. Previous work supports BRI1 occurring in homo-multimers (Wang *et al*., 2005). and my Western blotting data supports some form of high-MW BRI1 containing complex that is stable under the SDS-PAGE conditions used. Some credible candidates that could conceivably be in proximity with BRI1 (e.g., HSPs, Actin, MAP70-5, CYP78A5 Nodulin-related protein 1, RabA5E, BCHB) were also identified albeit with reduced confidence and similar abundance to unlikely interactors. Many of these proteins are involved in folding or trafficking of synthesised proteins and may well be in proximity with BRI1 for long periods of time, explaining their detection. HSP90 family members have been identified as co-factors in the folding and plasma membrane activity of CERK1, providing a precedent for HSP90 interaction with BRI1 (Chen *et al.* 2010) Overall, however, results from this experiment were remarkably poor, with very few peptides and proteins identified, despite the extent of labelling visible by Western blot and samples passing all mass spectrometry QC tests.

As this is a new technique it is important to identify known interactors as well as some candidates of novel interactions. Due to the results from this first run being surprisingly poor, with very few identified proteins, I next optimised the

experimental set-up, trying to mitigate for the potential masking of low abundance true positive proteins by abundant non-specific false-positive proteins. The use of protamine sulfate to selectively precipitate RuBisCo has been proposed (Kim*, et al.*2013). However, in my hands, protamine sulfate also removed BRI1 from samples, suggesting that protamine sulfate may not be as selective for RuBisCo as proposed. An additional control, using a non-brassinosteroid binding $G^{644}D$ mutant form of BRI1 (nominally recapitulates the weak *bri1-6* EMS-induced allele) (Noguchi *et al*. 1999) was also developed in order to distinguish proteins associated with BRI1 following brassinosteroid binding from proteins that constitutively interact. Finally, based on previous work in the lab, on bead digestion was used to hopefully increase the sensitivity of analysis. The observed beneficial nature of on-bead digestion in other projects in the lab likely stems from unbiased digestion of non-specific and specifically bound proteins when compared to eluting with SDS-PAGE sample buffer, where weak interactions between the matrix and non-specifically bound proteins are likely preferentially disrupted compared to the strong specific interaction between TwinStrep tagged proteins and magstrep type 3 XT beads (K_D in low pM range and very thermally stable).

Results from this improved mass spectrometry were more promising, with the number of peptides identified increasing to 261. Although still low it is a significant increase from the 51 identified in the three earlier replicates. It is also quite possible that more peptides were present in the sample but were unable to be defined due to a less than complete reference proteome. Although this was a pilot experiment so no real conclusions can be drawn peptides identified did differ from the initial mass spectrometry experiments. In comparison to the first experiments members of the HSP90 family did not come out as strongly, as

they were observed in the negative BRI1-PafA/PupA sample. Indicating possible non-specific binding during enrichment. Although it is of note that in the positive samples HSP90 family members were also present in increased numbers compared to the negative sample. Due to the nature of this pilot experiment it is hard to determine if further replicates would come out stronger for HSP90s being good candidate interactors supporting the earlier mass spectrometry.

This work has successfully begun to demonstrate and optimise the application of the PafA/PupE technique as a novel PPI method for traditionally difficult plant proteins. For both pairwise analysis and discovery.

Chapter 6

Discussion

6.1 Room for improvement in the study of Protein-Protein interactions

As a crucial mechanism in the control of cellular processes, understanding and identifying protein-protein interactions (PPIs), including weak, transient and interactions involving traditionally difficult proteins are key to getting a full understanding of the cell. From the formation of structural protein complexes to enzyme substrate interactions, PPIs really are involved in everything.

Due to this, there are a breadth of well-established ways in which to study PPIs. Some of these methods include yeast-2-hybrid, protein reconstitution assays such as split-ubiquitin and split-YFP, co-immunoprecipitation coupled with either mass spectrometry or Western blot and pull-downs. These techniques have played an invaluable role in our understanding of many processes and well characterised PPIs, but these methods suffer from some or all of the following drawbacks.

- being low throughput,
- indirect,
- poor sensitivity,
- high background,
- use of heterologous systems (non-native environment),
- library bias (size, complexity, diversity),
- tag effects on protein function,

- disruptive (requires cell lysis and/or use of detergents before capture of interactions).

This can result in a less than physiologically relevant result, with a high chance of false positives and false negatives. This has the biggest effect on weak or transient protein interactions, as the described techniques favour more stable interactions between non-membrane associated proteins. This results in the majority of PPIs discovered in these ways being stable, soluble and/or abundant complexes. This, however, does not represent the true variety of interactions occurring in a cell, disproportionately disregarding many potentially important and interesting weak, transient or membrane protein interactions that may occur.

To overcome these issues one of the best options appears to be the use of a proximity labelling technique. This is more likely to provide a relatively unbiased snapshot of the cellular environment and give a clearer picture as the variety of PPIs that are occurring. One of the main current examples of this is BioID (and subsequent variations) (Kim et al., 2016) which has been shown to be a successful way to map PPIs. BioID works by utilizing a promiscuous bacterial biotin ligase BirA*; by attaching BirA* to a protein of interest a cloud of active biotin is produced, labelling any surface lysine of proximal proteins. These labelled proteins can then be captured using streptavidin and identified by mass spectrometry. This technique has proved promising for initial candidate identification but, like other established techniques, does have drawbacks. These limitations include:

- A large labelling circumference- With the diffusion diameter of activated biotin (estimated to be between 25 and 50nm) being larger than many protein

complexes it is difficult to know if labelled proteins are true interactors, or nearby but irrelevant proteins.

- Movement of BirA* tagged proteins will leave a trail of activated biotin resulting in non-specific labelling outside of complexes of interest. This is especially true for cytosolic or other highly mobile proteins.
- Long incubation times - BioID labelling typically requires long incubations with exogenous biotin resulting in non-specific background and a bias towards detection of stable long-term interactions over weak and transient ones. However, this limitation has largely been overcome through the use of TurboID. (Kim *et al.,* 2016)
- Additional Biotin required - High concentrations of exogenous biotin needs to be introduced. This typically requires tissue culture-like conditions, reducing the physiological relevance, types of tissue that can be analysed and largely precludes whole organism analysis.

An ideal technique for the identification of physiologically relevant PPIs would possess a number of attributes:

- Useful in a whole organism context in physiological conditions
- Identify PPIs without outside intervention or supplementation
- Be entirely genetically encoded
- Able to detect weak and/or transient interactions

Whilst BioID is close to this its unable to meet all of these requirements. Most of the drawbacks associated with BioID could be overcome by using an alternative enzyme that could add a genetically encoded tag. This would remove the need for exogenous inputs and allow for whole organism experiments in a closed system thereby increasing the physiological relevance.

Actinobacteria natively possess a system that could be exploited for this.

6.2 Development and validation of a novel genetically encoded proximity labelling system in yeast

The first steps towards establishing this new PupE/PafA PPI technique were carried out in *Saccharomyces cerevisiae*. However, whilst carrying out this initial proof of principle work, the PUP-IT system using PupE and PafA from different bacteria was published (Liu *et al.* 2018) making completion of this initial characterization unnecessary for implementation of the system.

The work described in chapter 3 was designed in order to create, validate, and optimise aspects of a new proximity labelling system with aims including:

- Demonstrating PafA activity in heterologous systems.
- Demonstration of (inducible) proximity dependant labelling.
- The development of synthetic PupE variants optimised for proximity labelling and affinity capture.

Initial exploration towards these aims focussed on the expression and detection of the two parts of the system PafA and PupE in yeast and showing that PafA was active in heterologous systems. Non-toxic expression of PafA in eukaryotic systems was successfully demonstrated but unfortunately, I was unable to demonstrate any enzymatic activity for PafA in yeast or possible effects of PupE expression and subsequent conjugation on growth. This could be for a number of reasons; it is possible that the ubiquitin leader on the PupE constructs was not being cleaved correctly, marking the proteins for degradation. However, this is unlikely as when later expressed in plants PupE detection indicated cleavage was occurring correctly. It is also possible that as PupE expression was driven

by an inducible promoter, Gal1/10, that it was just never successfully activated and expressed under the conditions tested

Had the system remained unpublished the demonstration and detection of PafA and PupE labelling by Western blot in *S. cerevisiae* would have been optimised and pursued. However, given that the basic principles were now in the public domain this work was stopped.

Next, an artificial interaction system to validate proximity dependant PafA/PupE activity towards other proteins was designed. It utilised the GAI/GID gibberellin inducible dimerization system that would work well in yeast and can be exploited to force an interaction between PafA and a protein of interest, providing a way to demonstrate PafA dependant PupE labelling of proximal proteins.

Initial components of this machinery were generated with the intention of validating the GAI/GID system by microscopy using reporter tags in the place of PafA such as mCherry or GFP. However, as basic characterization and demonstration of interaction was no longer required this was never used.

The publication of the PUP-IT technique provided a large amount of characterisation of the PafA/PupE system. This included demonstration that PafA maintains enzymatic activity as N-terminus or C-terminus fusion to a protein of interest, and that PafA is able to self pupylate and pupylate a variety of free lysines on the protein its fused to. This indicated that that PupE labelling is promiscuous and independent of primary sequence. They also were successful in demonstrating that the N-terminus of PupE can be modified with the addition of tags without a loss of function and that PupE can be truncated to 28 amino acids whilst remining active.

After these developments it therefore made sense to apply the knowledge shown by PUP-IT and combine them with the resources I had developed towards the optimisation and application of PafA/PupE in planta. In addition, I aimed to improve upon the described PUP-IT system by making sure that all components were genetically encoded; PUP-IT requires post-labelling and lysis application of biotin ligase to biotinylate an encoded biotin acceptor peptide on their version of PupE to enable enrichment and purification. My adaptation, using the biotin mimicking TwinStrep tag peptide, removes the need for this additional handling step and makes the system truly fully genetically encoded and self-contained, with reduced potential for technical error. However there is of course the chance for novel error from the capture of biotinylated proteins by the streptavidin beads.

6.3 DogTag: A new way to look at protein protein interactions in Planta

Despite publication of a PafA/PupE based PPI system by (Liu et al 2018) there was still no data to show the system would work in planta and remained promising and worth exploring further. So, the focus moved from basic characterisation of the principles of the system to its optimisation and application in plants. To do this, aspects of the literature needed validated in plants such as:

- PupE being an effective scaffold for affinity tags
- PafA is enzymatically active in plants and capable of pupylating proteins
- PafA and PupE can be detected by Western blot
- No motif is required for PafA mediated pupylation

Based on validating the literature, and on optimizing the use of the techniques in plants, new aims were set out.

- **What length of PupE?** Which length of PupE is the most effective and what combination of affinity tags work the best for both detection by Western blot and affinity purification?
- **Is PafA active in plants?** When transiently expressed does PafA maintain activity and successfully label the free lysines of proteins with PupE?
- **Is GFP a PupE acceptor?** As a commonly used tag does GFP have enough free lysines to allow for pupylation by PafA in a pair-wise interaction assay?

6.4 PupE: Small but mighty

Based on the literature its reported that only the c-terminal helix of PupE is required for interaction with PafA, this was also reported by the PUP-IT paper (Liu *et al* 2018) but there was also potential that the full length PupE or a slightly longer PupE may actually be more effective at labelling. With this in mind I tested three potential PupE lengths. The minimal C-terminal helix length, $PupE^{46-72,}$ the slightly longer $PupE^{41-72}$ with an additional 5 amino acids to act as a short linker between the “active” PupE and any additional tags, and the full length PupE, $PupE^{2-72}$.

Initially there appeared to be problems with the expression and detection of $PupE^{46-72}$ by Western blot with signal observed at ~15kDa, higher than the expected size of ~9KDa based on primary sequence data. All PupE constructs were designed with a ubiquitin leader to support expression of such a small peptide. The ubiquitin leader should be removed in vivo by the action of deubiquitinating enzymes (DUBs), with the expected size of un-cleaved Ubq-$PupE^{46-72}$ being ~17kDa it became a worry that ubiquitin was not being removed. To check this, N-Ethylmaleimide to inhibit ubiquitin cleavage by DUBs was used. The treated sample seemed to show increased levels of potentially un-cleaved Ubq-PupE with lower levels being observed in untreated samples indicating possible partial ubiquitin cleavage. This result seemed to directly contradict the literature as Ubq fusion has reportedly approached 100% efficiency. Due to this, the next step was to establish whether a standard Laemmli buffer system and 0.45µm pore PVDF membrane was simply unable to resolve the expected ~9KDa protein. To test this, I moved to a more sensitive Tris-Tricine SDS-PAGE system in order to better separate proteins of low

molecular weight along with 0.2 µm PVDF. Despite these efforts I was still unable to detect PupE at the expected ~9kDa mass. This prompted me to investigate the possibility of failed ubiquitin cleavage in another way. The Gly-Gly cleavage site between Ubiquitin and PupE was mutated to Ala-Ala, resulting in the prevention of cleavage by DUBS. However, after attempted transient expression this mutant was never detected, and I hypothesise that by inhibiting cleavage it was recognised as a ubiquitinated polypeptide and therefore degraded. Further work in the lab using non-ubiquitin fusions of PupE expressed in bacteria seemed to show that the true running size of PupE was indeed ~15kDa and not the expected ~9kDa; this suggests that Ubq-PupE was being processed correctly.

To test PupE as an effective scaffold for affinity tags all three lengths mentioned above in chapter 4 were made as N-terminus fusions with a cleaved Ubiquitin leader, a TwinStrep tag for affinity purification, as it is cheap, fast, and typically demonstrates low background/non-specific binding. This addition also allowed for a genetically encoded system in comparison to PUP-IT giving it the potential to work in a whole organism context with significantly improved physiological relevance. A 3xFLAG tag was also included for identification by Western blot but it should be noted that this could be easily changed if alternative tags would improve upon the effectiveness of the system. As well as the three PupE lengths negative control PupA versions were also made, with the same N-terminal TwinStrep and FLAG tags but with the substitution of the PupE C-terminal glutamate with a physiologically inactive alanine.

The three PupE lengths, alongside PupA negative controls, were successfully expressed in *N. benthamiana* and detected using the optimised Tris-Tricine SDS-PAGE system. Interestingly a mass shift was observed between the PupE

and their PupA counterparts with PupA appearing to run at a smaller size. I hypothesise that this is due to a change in the charge of the protein seen with the substitution of glutamate to alanine. The increased negative charge of PupE compared to PupA may repel the SDS that would normally coat peptides in SDS-PAGE, resulting in an overall reduction in negative charge compared to PupA, altering mobility in SDS-PAGE.

6.5 PafA: Active and promiscuous in plants

To see if PafA was active in plants I first needed to demonstrate and detect stable PafA expression. Initially I ran into some difficulties with vector stability in bacteria. It was possible that the PafA construct was toxic due to the fact that problems only occurred after the construct was transferred to a plant expression vector, with some protein expression driven by one of the prokaryotic or eukaryotic promoter elements in the vector resulting in PafA production. However, as the system had been successfully reconstituted in *E. coli* bacteria its more likely that these problems were a result of translation of an out of frame open reading frame, resulting in a toxic product, or an unstable DNA sequence that was prone to the formation of secondary structures. The problem was overcome by the addition of an intron, which either prevented spliced PafA expression in bacteria, interrupted toxic product expression or disrupted secondary structures.

After the introduction of the intron PafA was successfully expressed and detected by Western blot. I was also able to reconstitute the system by the transient co- expression of $PupE^{46\text{-}72}$ alongside PafA. This demonstrated good proteome labelling and detection within plants for the first time, with labelling only occurring in the presence of both PafA and PupE. This Suggests that that

system is highly active and a good potential tool for PPI technique for the study of plants.

6.6 GFP: Good for pupylation

One potential limitation of the system is its reliance of free lysines for labelling to occur. Despite apparent promiscuity and no defined primary sequence for pupylation, I had no idea how many available lysines would be required for effective labelling of target proteins. Green Fluorescent protein (GFP) was a good candidate for a testing as a model pupylation acceptor as it is easy to express, well characterised and is a commonly used protein tag with 20 free surface lysines.

Firstly GFP-PafA fusions were made, transiently expressed and detected by Western blot. It was also demonstrated that when co-expressed alongside PupE, GFP-PafA seems to remain active, based on the robust pupylation of the proteome. This confirms what has been reported in the literature that N and C-terminal PafA fusions remain capable of conjugating PupE. In my GFP-PafA construct I used a 5xGGSGG linker in order to give flexibility to allow PafA to access GFP and limit steric hindrance. It was a consideration that this linker length or composition may need optimisation in conjunction with an alternative orientation of PafA-GFP, however, given the observed functionality this was deemed unnecessary. Despite this, as it was made using a gateway expression vector, there is a chance that some inflexible amino acids remain between the GFP and PafA which has the potential to impair function in some way, although these were not sufficient to disrupt experimental needs and outcomes. I was also able to demonstrate that pupylated proteins could be captured and

enriched using the TwinStrep tag included in the PupE tag. This enrichment was used to explicitly demonstrate GFP pupylation by Western blot and subsequently confirmed by mass spectrometry.

After enrichment, proteins were sent to mass spectrometry, however, only two pupylated lysines were identified on GFP. This result is unexpected as when compared to the enriched GFP detected by Western blot it seems to demonstrate a large number of pupylated forms of GFP-PafA, indicative of a high degree and diversity of labelling of GFP. It was possible that overall coverage of GFP was poor, but when this was looked at it showed that unmodified peptides were present that covered almost all of the protein. So, what reasons could there be for the mass spectrometry detecting so few pupylated lysines? PupE itself contains two lysines, FLAG contains three lysines and TwinStrep has one. This information along with the literature reporting that PupE can self-pupylate indicates that it is highly likely there is some sort of branching effect occurring. With the highly promiscuous nature of PafA and this potential for self-pupylation of PupE I hypothesize that this branching effect is occurring and potentially inhibiting the ability of enzymes cutting for mass spectrometry, this access is likely further hindered by the fact that pupylation occurs on lysines and digestion was done using Trypsin/Lys-C likely resulting in multiple missed cleavages if a high number of lysines were pupylated.

This Possible branching would also explain what was detected by Western blot with GFP-PafA after enrichment seemingly visible even in the MDa range despite the size of GFP-PafA being only ~85kDa. This is far larger than the pupylation of one or two lysines. The suspected polypupylation would explain this as the more branching the higher the protein size but, additionally, it would

affect the movement through the gel, reducing mobility and indicating a much bigger protein size than present.

From this it is clear that some further optimisation of the system could improve the sensitivity of mass spectrometry analysis if detection of pupylated peptides is required; this would be a useful discrimination factor for determining whether a protein is a true or false positive by demonstrating the presence of the post-proteolysis PupE remnant attached to Lysine (e.g. GGE left from digestion with Trypsin).

In regard to the PupE construct its possible the number of lysines could be reduced. Minimising this polypupylation phenomenon. A first easy change would be to replace the FLAG tag with a tag containing no lysines such as a T7 tag, Glu-Glu tag, ALFA tag, E Tag, Rhol-D4 tag or Spot tag. By doing this is would reduce the number of lysines present by nine.

Next, it may be possible to change the lysines present in PupE; by changing them to and aspartic (D) or glutamic (E) acid not only would you stop polypupylation of PupE, but you could introduce sites allowing for the use of, for example, Asp-N or Glu-C for mass spectrometry digestion, possibly enabling better enzyme access, higher cutting efficiency and reducing the need to account for multiple missed cleavages during mass spectrometry analysis. Mutation of the most C-terminal lysine (Lys69) in PupE may be one of the most beneficial changes; blockage of trypsin cleavage here by polypupylation in the experiments described here would result in a peptide too large (and complex) for effective mass spectrometry analysis. Due to the location of the lysines in PupE in the and conserved C-terminal region it is possible that they are essential for function, therefore any modification here would need to be tested

for functionality. PupE Lys69 forms hydrogen bonds with PafA Asp34; a simple K<>D swap at these positions may maintain hydrogen bonding but allow for elimination of Lys69 (Barandun *et al*., 2013). Even if substitution of Lys69 did reduce efficiency it may be beneficial; an enzyme with a low turnover will require longer in proximity with a substrate and have the potential to increase signal to noise ratios. Alternatively, it may be possible to substitute Asp for one of the Gly residues in the C-terminal GGE motif after Lys69. However, there is still the likelihood that pupylation at Lys69 would block access for any protease used.

The last potential optimisation to investigate would be the TwinStrep tag as it is required for the enrichment of tagged proteins and is both effective and efficient. However, it may make sense to replace the one lysine it has but, like PupE, it is unclear if this lysine is critical to TwinStrep function or not. It is likely anyway that reductions of lysines in other parts of the PupE construct would be sufficient and that the one lysine present in TwinStrep would not result in such severe branching and there would be adequate space for protease digestion by changing the Western blot affinity tag and possibly PupE. Although these changes are intended to improve the sensitivity of mass spectrometry it may also help with improved resolution of pupylated proteins by Western blot. By limiting the polypupylation branching it should improve gel mobility resolving at a more accurate size giving more distinct and clear banding.

Comparable work in plants using other proximity labelling techniques, TurboID and min-TurboID, has demonstrated efficient and effective labelling in various plant tissues at different developmental stages. Similar to the work described here they have also shown labelling by Western blot and mass spectrometry successfully identifying known and novel interactors of proteins of interest.

However, unlike DogTag they require outside addition of biotin by vacuum infiltration. Other limitations include an inability to study PPIs in some abiotic stress conditions such as cold or drought stress. Due to DogTag's genetically encoded nature, and the vast array of Actinobacteria that natively use the PafA/ PupE system, it is likely possible to tailor the system to function in conditions and temperatures required to answer the user's question. Its functionality as a closed system also makes it a very attractive physiologically relevant technique for studying interactions in whole tissues and organisms in comparison to other close proximity labelling methods.

6.7 Application of the DogTag system to study Receptor like kinase interactions

After successful work to somewhat optimise the DogTag system in plants the next task was to attempt to apply it as a protein-protein interaction detection technique. With some expected benefits over current proximity labelling techniques such as BioID it may present an opportunity to better study PPIs that have traditionally proven difficult.

Receptor like kinases are a large family of integral membrane proteins that perceive and detect a diverse range of extracellular stimuli, including hormone signals, symbiotic interactions, cell wall status and pathogens. Two well studied members of this family are Flagellin Sensitive 2 (FLS2), involved in immune responses, and Brassinosteroid Insensitive 1 (BRI1), involved in plant development. As such, these proteins are involved in important plant processes and are highly likely to have a vast array of interacting proteins of various types. However, as integral membrane proteins likely contained within large integral membrane protein complexes, the study of this potential breath of interactions

has been difficult using established techniques. By being able to label interacting proteins without disruption to the membrane environment would be a huge advantage here, and something that should be possible with DogTag.

Along with these RLKs I am specifically interested in their interaction with protein S-acyl transferases (PATs). FLS2 and BRI1 are both subject to a dynamic fatty acid-based post translational modification called S-acylation. This modification seems crucial to function, but it is still unknown which PATs are involved in the S-acylation of BRI1 and FLS2. The DogTag proximity labelling technique should overcome the issues of transient PAT/RLK interaction, RLKs and PATs being integral membrane proteins, with PATs being poly-topic and redox sensitive and the RLK complexes requiring the lipid bilayer for stabilisation.

By the attachment of PafA to the RLK of interest and co-expression with PupE, proximal proteins would be pupylated allowing for capture and identification by Western blot or mass spectrometry. First, I successfully created and expressed FLS2- PafA and a BRI1-PafA fusions. Both were made using a 5xGGSGG linker as in the GFP-PafA fusions mentioned earlier. This should provide a good amount of flexibility for PafA, reducing any steric hinderance, and is a similar length to the 3xMYC linker that supports FLS2 function in FLS2-3xMYC-GFP fusions thereby increasing chance of retaiing functionof BRI1 and FLS2. This linker was also comparable to the linker used for BioID work in plants (Mair et al 2019). Like with the GFP-PafA fusions, the possibility of optimising the linker length and composition was considered but proved ultimately unnecessary for detectable activity. One possible novel use for varying linker lengths would be by increasing or decreasing the labelling radius of the system; by increasing or decreasing the linker it may be possible to use pupylation as a means to

measure distance between your protein of interest, in this case an RLK and a known interacting partner, for example BRI1 and its co-receptor BAK1. Its also possible by changing the linker length you would capture totally different proteins as you increased or decreased labelling radiuses allowing for inferential of distance of more novel potential interactors (Figure 6.1).

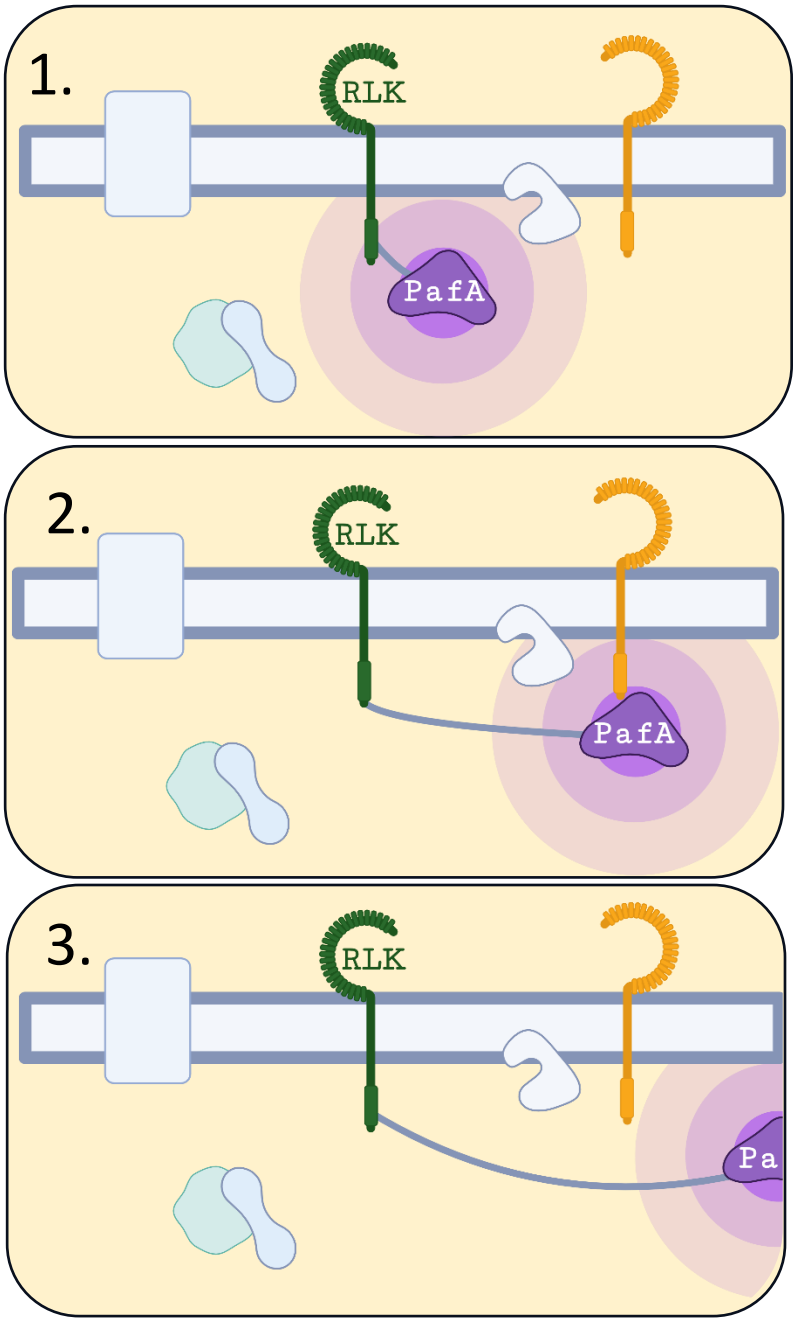


Figure 6.1. DogTag could present to opportunity to measure distances between proteins.

By changing the linker length between the protein of interest and PafA, it could be possible to use this system as a type of molecular ruler. 1. linker is not quite long enough to label a known co-receptor. 2. with a slightly longer linker length the known co-receptor should experience some labelling. 3. With an even longer linker length PafA is further away than with length 2 and the co-receptor should experience low levels or no labelling, allowing for a better estimate of protein distances.

6.8 PAT-RLK pairwise interactions revealed through DogTag

One benefit to using a proximity labelling technique is the ability to use it for pairwise analysis to validate potential interactions as well as a discovery tool for novel interactors. For the identification of the PAT or PATs that S-acylates BRI1 and FLS2 a pairwise approach was used. By expressing RLK-PafA fusions alongside PupE with each PAT it should be possible to identify labelled, and therefore proximal, PATs.

Being interested in PATs that interact with RLKs, and previous work shows that S-acylation occurs at the plasma membrane (Hurst et al. 2021), I focussed my efforts on PATs with known plasma membrane localisations. These PATs included 4,5,6,7,8,9,12,19, and 21 (Batistic 2012). As this pairwise experiment was being done by over-expression, and it was expected that pupylation would be very efficient, an initial control for false positive results was included. For this, PAT10 was selected as its reported to be present at the tonoplast membrane and therefore highly unlikely to be involved in the S-acylation of RLKs. To control for differing numbers and accessibilities of lysines on PATs, all PATs used were tagged at the C-terminus with GFP, a proven Pup acceptor.

Due to the need for activation of FLS2 with its ligand peptide flg22 a pilot pairwise study was only done once using previously identified PATs of interest; PATs 6, 9 and 12 (Charlotte Hurst, unpublished). PAT10 was not included in the FLS2 pairwise experiment. From this it is hard to draw any real conclusions, but it does seem to indicate that PAT9 is not proximal to FLS2 whilst PAT12 appears to be close enough for labelling to occur. Failure to detect PAT9 in either enriched sample despite good expression is especially interesting as it

has previously been identified as a PAT involved in the S-acylation of P2K1/DORN1, another receptor involved in plant immunity through detection of extracellular ATP levels (Chen *et al.* 2021). Work has shown that FLS2 is S-acylated constitutively as well as dynamically, if PAT12 did seem to be consistently identified it would be interesting to try and distinguish which form of acylation it was involved in. From this one study it is identified in both the non-active (no flg22) and active (with flg22) enriched samples; this could indicate a role in either type of S-acylation, but I hypothesize that it is more likely to be involved in dynamic S-acylation due to the stronger signal present after flg22 treatment. I think it is possible that labelling detected in the inactivated sample is due to the localisation of PAT12 and maybe not direct interaction but could indicate that PAT12 is present in the resting complex in a more distal position to FLS2 but comes into close proximity when FLS2 is activated. This would be a good example of where linker length could be optimised between PafA and FLS2 to reduce labelling radius to see if it was possible to maintain detection after activation but abolish detection before interaction indicating a likely physical interaction. The detection of PAT12 in the inactivated sample may also be due to over expression resulting in its abundance in the cell. Further repeats of this would have been beneficial in order to draw some more solid conclusions but due to the extra technical steps involved in activation of FLS2 for the pairwise studies, the focus was moved to BRI1 where an endogenous ligand (brassinosteroids) is present in plants.

For BRI1 multiple replicates were carried out, using a full range of plasma membrane PATs and tonoplast membrane PAT, PAT10 as a negative control. Due to the universal presence of brassinosteriods, BRI1’s activating ligand, in plants there should be a proportion of constitutive activation of BRI1 in *N.*

benthamiana so, unlike FLS2, should not require activation. The first point of note from these pairwise experiments is the identification of PAT10 in enriched samples. This was unexpected as its reported location is at the tonoplast membrane. Labelling may have occurred for a number of reasons; it is possible that PAT10 and BRI1 do actually encounter one another, possibly during protein synthesis and transport, or during BRI1 degradation at the vacuole. It is also possible that due to the over-expression of proteins in the system that there is some level of “leakage” however, if this was the case, it would be expected that every well-expressed PAT would be observed in the enriched samples, which is not the case. Finally, it is also impossible to rule out the possibility of free PafA in the cell as a result of BRI1-PafA turnover or proteolysis, which would result in the labelling of non-proximal proteins, however like with over-expression, this also seems unlikely as with this, again, you would expect all well-expressed PATs to be present in the enriched sample which is not what was observed. Due to this, the enrichment signal of PAT10 is considered to be the baseline signal above which a plasma membrane PAT must reach for consideration as a potential BRI1 interactor. Two replicates with the full range of plasma membrane PATs alongside PAT10 were carried out, along with a more focused replicate looking at some potentially positive PATs, along with a seemingly non-interacting PAT, PAT7. From the full replicates, PATs, 8, 9 and 12 came out as strong candidates for BRI1 interaction, whilst PAT 5 was borderline. PAT7, despite expressing well and being localised at the plasma membrane, did not appear in the enriched sample so was used as a more relevant negative control. The focused replicate confirmed PATs 8, 9 and 12 as strong candidates, but not PAT5. This is surprising as PAT 5 has been reported as a PAT involved in the S-acylation of P2K1, another receptor involved in plant

immunity. However, unlike with FLS2, PAT 9, which also has been reportedly involved in P2K1 S-acylation (Chen *et al.* 2021), was detected. The identification of PAT 8 alongside PAT9 is also not surprising as they make up a highly related protein pair with an amino acid similarity of 74% (Batistic 2012). The possibility of an overlap in function is therefore highly likely. PAT12 seems a good potential interactor of both BRI1 and FLS2 despite them being involved in regulation of different plant processes. However, a common PAT would correlate with the conserved nature of the dynamic S-acylation sites in BRI1 and FLS2. With all members of the RLK family containing the dynamic S-acylation sites it is possible that one or more PATs may be involved in the S-acylation of multiple RLKs.

In order to confirm these interactions next steps would perhaps include co-Immunoprecipitation (Co-Ips) experiments with the three strongly enriched PATs, 8, 9 and 12 alongside PAT 7 as a negative control. However, it is important to note that this may require a large amount of optimisation, as a likely transient interaction occurs between PATs and RLKs Co-IPs and so it may be difficult to detect an interaction. Another possible way to confirm a role for these PATS in RLK function would be with the use of PAT knockouts in Arabidopsis; by carrying out S-acylation assays it may be possible to demonstrate a change of BRI1 or FLS2 acylation in plant lines with these PATs of interest knocked out. Lastly, S-acylation boost assays may be able to indicate which PATs are able to S-acylate RLKs rather than just be in proximity; this would in turn differentiate which PATs affect RLKs from those that act downstream of PATs.

6.9 Use of DogTag for discovery of novel interactors

It is well established that BRI1 forms multi-protein complexes in the plasma membrane (Aker *et al.,* 2008) but there is a strong likelihood that BRI1 is also involved in a wide range of PPIs and protein-proximal interactions. Due to traditional methods not being well suited to integral membrane proteins like RLKs, a large portion of interactions have likely gone undetected by more established techniques. These issues could be overcome by the DogTag system, by maintaining the membrane environment during labelling many of these as yet unidentified, weak interactions and membrane stabilised interactions may be maintained and identifiable by mass spectrometry.

BRI1-PafA and PupE were individually or co-expressed transiently in *N. benthamiana.* Protein was extracted and after expression was confirmed sample was then enriched. After enrichment an aliquot was analysed by Western blot to look at the extent of labelling and capture. The remaining sample was in-gel trypsin digested and cleaned up, followed by analysis using nano-liquid chromatography mass spectrometry. This was repeated three times. Raw data was then analysed using MaxQuant with the parameters described previously. From all three replicates only 51 proteins were identified despite reanalysis to account for technical error. This was surprising given how effective DogTag seemed to be at labelling and did not seem comparable to Western blot levels of pupylation. Due to the small sample of proteins detected, alongside low numbers of peptides from each protein, full statistical analysis would not be meaningful, and as such a semi-quantitative analysis based on spectral counting was therefore performed. This allowed for the identification of a number of proteins that were seen solely or primarily in the BRI1-PafA/PupE samples but not seen in negative controls. These proteins were considered good candidates for BRI1 interactors. Due to the semi-quantitative nature of this

analysis, along with low numbers of proteins and peptides identified makes it difficult to draw major conclusions. However, 10 peptides were identified in all three repeats and not in controls mapping to HSP80/90 family members; this suggests a genuine association.

Heat shock proteins (HSPs) are a large family of highly conserved proteins in animals and plants. These chaperone proteins are involved in a multitude of processes including protein maturation, re-folding, and degradation. Although some HSPs are constitutively expressed others are expressed in response to various stress conditions, such as drought, salinity, and pathogens. This is especially vital in plants as they are unbale to change environment they must adapt in order to tolerate such stresses. HSP80/90 have been reportedly found in the cytosol, endoplasmic reticulum, and nucleus, with a proposed function in the regulation of steroid hormone receptors and protein translocation. The HSP90 class in particular seem to have an important role in signalling protein function and trafficking (Pratt *et al.,* 2003) with cytoplasmic HSP90 being described as having a role in pathogen resistance due to interactions with resistance proteins (Hubert *et al*.,2003). HSP80/90 family members were also identified co-factors in the folding and plasma membrane activity of the RLK CERK1 (Chen *et al.,* 2010). This all collectively makes HSP80/90 attractive likely interactors of BRI1, as a dynamically trafficked receptor at the plasma membrane involved in hormone signalling.

However, due to the large family size of HSPs, and described functional overlap (Hu *et al.,* 2009), this interaction would be challenging to validate. As chaperone proteins HSPs are involved in a vast array of processes making them likely critical to function. Knockout mutants have a high likelihood of either being embryo lethal, or having their effects masked by other HSPs that possess the

same functionality. This would make genetic validation and drawing any functional conclusions hard.

As a new technique it is beneficial to not only identify candidate novel interactors but also some known well established interactors for validation. Due to the extremely low number of proteins identified by the first discovery attempt some optimisations were made in order to increase sensitivity by mitigating potential masking of low abundance of true positives by abundant but non-specific false-positive proteins:

- **Protamine sulfate treatment**- to deplete highly abundant but likely irrelevant, proteins from input samples, potentially allowing for a clearer picture of other proteins present.
- **Addition of a BRI1 brassinosteroid binding mutant as a negative control**- allowing me to distinguish between proteins that interact with BRI1 as part of brassinosteroid induced activation, from those that are stable interactors or form part of BRI1 cycling.
- **On-bead digestion**- Other proteomic studies have shown that on bead digestion appears to give lower background and greater sensitivity in mass spectrometry-based analyses

One such abundant non-specific protein in plants is RuBisCo, to try and selectively precipitate RuBisCo and reduce abundance, samples were treated with protamine sulfate (Kim *et al.,* 2013). However, as this technique appears to work by precipitating out large proteins it seems protamine sulfate also removed BRI1 from samples, this suggests that protamine sulfate may not be as selective for RuBisCo as proposed. However other work involving FLS2-PafA using protamine sulfate has been successful to increasing mass spectrometry sensitivity, despite reduced signal seen by Western blot detection compared to

untreated samples (Sonica Chaudhry pers comm). This suggests that despite depletion of large proteins after protamine sulfate treatment significant proteins remain in the sample for analysis by mass spectrometry. A possible alternative solution to reduce the presence of these highly abundant non-specific proteins would be the use of tandem affinity purification, by attempting enrichment using both the FLAG tag and the TwinStrep tag the stringency for truly labelled proteins would increase, hopefully decreasing background and masking of proteins of interest. Additionally, for this it may be beneficial to change the FLAG tag for something with a high affinity purification strategy available, such as SPOT or ALFA tags, where nanobody agarose with low nM to pM affinity is available alongside antibodies for Western blotting.

As the DogTag system used here does not include inducible expression of the PupE label and BRI1 activation is happening in the plant all the time, it is useful to be able to distinguish between PPIs occurring as part of brassinosteroid induced activation, from those that are stable interactors or form part of BRI1 synthesis, maturation, and recycling. To do this I made a BRI1 brassinosteroid binding mutant as a negative control. This mutant copies the Gly644 to Asp mutation observed in the severe bri1-6 loss of function allele (Noguchi, Fujika *et al*. 1999), and is proposed to disrupt binding of brassinosteriods to the island domain of BRI1 within leucine-rich repeat 21, without disrupting protein folding (Hothorn, *et al.* 2011). This mutant will therefore not form a brassinosteroid induced dimer with BAK1 and will not initiate signalling downstream of BRI1. This will hopefully improve the relevance of proteins detected by mass spectrometry.

Lastly, to increase sensitivity further, on bead digestion was used due to its beneficial nature previously observed in the lab. This likely derives from

unbiased digestion of both non-specific and specifically bound proteins in comparison to elution with SDS-PAGE sample buffer, where the weak interaction between matrix and non-specifically bound proteins are more likely disrupted compared to the strong specific interactions seen between TwinStrep tagged proteins and magstrep type 3 XT beads (K_D in low pM range and very thermally stable). By choosing on bead digestion this should increase the detection ratio of specifically bound proteins to non-specific abundant proteins.

After optimisation there seems to have been an improvement in sensitivity. With the proteins detected increasing from 51 to 261. However as this optimised mass spectrometry was only done once it is impossible to draw conclusions beyond an indication towards possible increased sensitivity.

6.10 Future work

Despite significant steps towards the optimisation and application of the DogTag technique in plants for pairwise and discovery applications there are still a number of interesting areas that may be worth pursuing. As one of the most exciting applications for this technology is for discovery and my initial aims included the stable expression of the system in the model organism Arabidopsis thaliana. This would allow for the demonstration of the system in a whole organism context, one of its unique benefits. By stably expressing BRI1-PafA or FLS2-PafA from either native or overexpression promoters alongside PupE it would allow for the detection of potentially more physiologically relevant interacting partners by mass spectrometry, especially in the case of FLS2 which is not natively present in *Nicotiana benthamiana*, in which all the transient work has been carried out. Improvements to PupE in these experiments may include the addition of an inducible promoter, which would allow for control over the labelling window without having to disrupt the native cell environment. Alternatively, tissue, developmental stage or infection activated promoters could be used to drive PupE Expression and allow for identification of specific interactors from defined tissue, cell, developmental stage or pathogen challenged cells. Some work was done towards this stable work in *A. thaliana* with multiple rounds of transformations attempted via floral dipping. However, disappointingly, this work was lost due to the COVID lockdown in 2020 with not enough time remaining of the project to continue or reattempt.

As identified during the optimisation of the system with GFP-PafA, there are a number of changes that could be made to potentially improve the sensitivity of

mass spectrometry. By changing the PupE construct to minimise lysines present it could combat suspected poly-pupylation that likely blocks lysines, reducing the efficiency of sample digestion and deriving peptides too large or complex for identification. Ideally the C-terminal Lys69 present in PupE would be mutated and the FLAG tag changed to something without lysines such as T7 tag, Glu-Glu tag, ALFA tag, E Tag, Rhol-D4 tag or Spot tag as mentioned previously. There is also a lysine present in TwinStrep, but this may be crucial to function, and as the C-terminal lysine of PupE appears to be the main site for poly-pupylation, by changing this and the FLAG tag should be sufficient to vastly reduce poly-pupylation, reduce sample complexity/diversity and enhance protease accessibility. These changes should significantly improve the sensitivity of future discovery experiments using mass spectrometry.

The DogTag system shows promising potential as a proximity labelling technique in plants. Traditional approaches in plants as well as other organisms are limited for a number of reasons that include invasive and destructive techniques that do not truly maintain or represent the native environment of the cell where the interactions are occurring. This limits the ability to define components, particularly of mulyi-subunit protein complexes, and how they change in response to stimulus. DogTag has shown potential to overcome some of these challenges and to drive future research in a more physiologically relevant way, for both pairwise analysis and discovery of new interactions.

Bibliography

Aker J, de Vries SC. Plasma membrane receptor complexes. Plant Physiol. 2008 Aug;147(4):1560-4. doi: 10.1104/pp.108.120501. PMID: 18678747; PMCID: PMC2492642.

Alassimone J, Fujita S, Doblas VG, van Dop M, Barberon M, Kalmbach L, Vermeer JE, Rojas-Murcia N, Santuari L, Hardtke CS, Geldner N. Polarly localized kinase SGN1 is required for Casparian strip integrity and positioning. Nat Plants. 2016 Jul 25;2:16113. doi: 10.1038/nplants.2016.113. PMID: 27455051.

Ali GS, Prasad KV, Day I, Reddy AS. Ligand-dependent reduction in the membrane mobility of FLAGELLIN SENSITIVE2, an arabidopsis receptor-like kinase. Plant Cell Physiol. 2007 Nov;48(11):1601-11. doi: 10.1093/pcp/pcm132. Epub 2007 Oct 9. PMID: 17925310.

Barandun J, Delley CL, Ban N, Weber-Ban E. Crystal structure of the complex between prokaryotic ubiquitin-like protein and its ligase PafA. J Am Chem Soc. 2013 May 8;135(18):6794-7. doi: 10.1021/ja4024012. Epub 2013 Apr 23. PMID: 23601177.

Barandun J, Delley CL, Weber-Ban E. The pupylation pathway and its role in mycobacteria. BMC Biol. 2012 Nov 30;10:95. doi: 10.1186/1741-7007-10-95. PMID: 23198822; PMCID: PMC3511204.

Batistic O. Genomics and localization of the Arabidopsis DHHC-cysteine-rich domain S-acyltransferase protein family. Plant Physiol. 2012 Nov;160(3):1597-612. doi: 10.1104/pp.112.203968. Epub 2012 Sep 11. PMID: 22968831; PMCID: PMC3490592.

Bäumlein H, Wobus U, Pustell J, Kafatos FC. The legumin gene family: structure of a B type gene of Vicia faba and a possible legumin gene specific regulatory element. Nucleic Acids Res. 1986 Mar 25;14(6):2707-20. doi: 10.1093/nar/14.6.2707. PMID: 3960730; PMCID: PMC339693.

Beck DB, Narendra V, Drury WJ 3rd, Casey R, Jansen PW, Yuan ZF, Garcia BA, Vermeulen M, Bonasio R. In vivo proximity labeling for the detection of protein-protein and protein-RNA interactions. J Proteome Res. 2014 Dec 5;13(12):6135-43. doi: 10.1021/pr500196b. Epub 2014 Oct 29. PMID: 25311790; PMCID: PMC4261942.

Braun P, Gingras AC. History of protein-protein interactions: from egg-white to complex networks. Proteomics. 2012 May;12(10):1478-98. doi: 10.1002/pmic.201100563. PMID: 22711592.

Bücherl CA, Jarsch IK, Schudoma C, Segonzac C, Mbengue M, Robatzek S, MacLean D, Ott T, Zipfel C. Plant immune and growth receptors share common

signalling components but localise to distinct plasma membrane nanodomains. Elife. 2017 Mar 6;6:e25114. doi: 10.7554/eLife.25114. PMID: 28262094; PMCID: PMC5383397.

Burns KE, McAllister FE, Schwerdtfeger C, Mintseris J, Cerda-Maira F, Noens EE, Wilmanns M, Hubbard SR, Melandri F, Ovaa H, Gygi SP, Darwin KH. Mycobacterium tuberculosis prokaryotic ubiquitin-like protein-deconjugating enzyme is an unusual aspartate amidase. J Biol Chem. 2012 Oct 26;287(44):37522-9. doi: 10.1074/jbc.M112.384784. Epub 2012 Aug 31. PMID: 22942282; PMCID: PMC3481346.

Chen D, Hao F, Mu H, Ahsan N, Thelen JJ, Stacey G. S-acylation of P2K1 mediates extracellular ATP-induced immune signaling in Arabidopsis. Nat Commun. 2021 May 12;12(1):2750. doi: 10.1038/s41467-021-22854-1. PMID: 33980819; PMCID: PMC8115640.

Chen L, Hamada S, Fujiwara M, Zhu T, Thao NP, Wong HL, Krishna P, Ueda T, Kaku H, Shibuya N, Kawasaki T, Shimamoto K. The Hop/Sti1-Hsp90 chaperone complex facilitates the maturation and transport of a PAMP receptor in rice innate immunity. Cell Host Microbe. 2010 Mar 18;7(3):185-96. doi: 10.1016/j.chom.2010.02.008. PMID: 20227662.

Chinchilla D, Zipfel C, Robatzek S, Kemmerling B, Nürnberger T, Jones JD, Felix G, Boller T. A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. Nature. 2007 Jul 26;448(7152):497-500. doi: 10.1038/nature05999. Epub 2007 Jul 11. PMID: 17625569.

Drisdel RC, Green WN. Labeling and quantifying sites of protein palmitoylation. Biotechniques. 2004 Feb;36(2):276-85. doi: 10.2144/04362RR02. PMID: 14989092.

Gietz RD, Woods RA. Yeast transformation by the LiAc/SS Carrier DNA/PEG method. Methods Mol Biol. 2006;313:107-20. doi: 10.1385/1-59259-958-3:107. PMID: 16118429.

Gilchrist CA, Gray DA, Baker RT. A ubiquitin-specific protease that efficiently cleaves the ubiquitin-proline bond. J Biol Chem. 1997 Dec 19;272(51):32280-5. doi: 10.1074/jbc.272.51.32280. PMID: 9405433.

Gómez-Gómez L, Boller T. FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. Mol Cell. 2000 Jun;5(6):1003-11. doi: 10.1016/s1097-2765(00)80265-8. PMID: 10911994.

González Montoro A, Chumpen Ramirez S, Quiroga R, Valdez Taubas J. Specificity of transmembrane protein palmitoylation in yeast. PLoS One. 2011 Feb 24;6(2):e16969. doi: 10.1371/journal.pone.0016969. PMID: 21383992; PMCID: PMC3044718.

Greaves J, Gorleku OA, Salaun C, Chamberlain LH. Palmitoylation of the SNAP25 protein family: specificity and regulation by DHHC palmitoyl transferases. J Biol Chem. 2010 Aug 6;285(32):24629-38. doi: 10.1074/jbc.M110.119289. Epub 2010 Jun 2. PMID: 20519516; PMCID: PMC2915699.

Guth E, Thommen M, Weber-Ban E. Mycobacterial ubiquitin-like protein ligase PafA follows a two-step reaction pathway with a phosphorylated pup intermediate. J Biol Chem. 2011 Feb 11;286(6):4412-9. doi: 10.1074/jbc.M110.189282. Epub 2010 Nov 16. PMID: 21081505; PMCID: PMC3039397.

Hammond-Kosack KE, Jones JD. PLANT DISEASE RESISTANCE GENES. Annu Rev Plant Physiol Plant Mol Biol. 1997 Jun;48:575-607. doi: 10.1146/annurev.arplant.48.1.575. PMID: 15012275.

Hanlon CD, Andrew DJ. Outside-in signaling--a brief review of GPCR signaling with a focus on the Drosophila GPCR family. J Cell Sci. 2015 Oct 1;128(19):3533-42. doi: 10.1242/jcs.175158. Epub 2015 Sep 7. PMID: 26345366; PMCID: PMC4610211.

Häweker H, Rips S, Koiwa H, Salomon S, Saijo Y, Chinchilla D, Robatzek S, von Schaewen A. Pattern recognition receptors require N-glycosylation to mediate plant immunity. J Biol Chem. 2010 Feb 12;285(7):4629-36. doi: 10.1074/jbc.M109.063073. Epub 2009 Dec 11. PMID: 20007973; PMCID: PMC2836068.

Hemsley PA, Taylor L, Grierson CS. Assaying protein palmitoylation in plants. Plant Methods. 2008 Jan 11;4:2. doi: 10.1186/1746-4811-4-2. PMID: 18190696; PMCID: PMC2244627.

Hemsley PA, Weimar T, Lilley KS, Dupree P, Grierson CS. A proteomic approach identifies many novel palmitoylated proteins in Arabidopsis. New Phytol. 2013 Feb;197(3):805-814. doi: 10.1111/nph.12077. Epub 2012 Dec 17. PMID: 23252521.

Hondred D, Walker JM, Mathews DE, Vierstra RD. Use of ubiquitin fusions to augment protein expression in transgenic plants. Plant Physiol. 1999 Feb;119(2):713-24. doi: 10.1104/pp.119.2.713. PMID: 9952468; PMCID: PMC32149.

Hothorn M, Belkhadir Y, Dreux M, Dabi T, Noel JP, Wilson IA, Chory J. Structural basis of steroid hormone perception by the receptor kinase BRI1. Nature. 2011 Jun 12;474(7352):467-71. doi: 10.1038/nature10153. PMID: 21666665; PMCID: PMC3280218.

Howie J, Reilly L, Fraser NJ, Vlachaki Walker JM, Wypijewski KJ, Ashford ML, Calaghan SC, McClafferty H, Tian L, Shipston MJ, Boguslavskyi A, Shattock MJ, Fuller W. Substrate recognition by the cell surface palmitoyl transferase DHHC5. Proc Natl Acad Sci U S A. 2014 Dec 9;111(49):17534-9. doi: 10.1073/pnas.1413627111. Epub 2014 Nov 24. PMID: 25422474; PMCID: PMC4267385.

Hu W, Hu G, Han B. Genome-wide survey and expression profiling of heat shock proteins and heat shock factors revealed overlapped and stress specific response under abiotic stresses in rice. Plant Sci. 2009 Apr;176(4):583-90. doi: 10.1016/j.plantsci.2009.01.016. Epub 2009 Feb 5. PMID: 26493149.

Huang K, Sanders S, Singaraja R, Orban P, Cijsouw T, Arstikaitis P, Yanai A, Hayden MR, El-Husseini A. Neuronal palmitoyl acyl transferases exhibit distinct substrate specificity. FASEB J. 2009 Aug;23(8):2605-15. doi: 10.1096/fj.08-127399. Epub 2009 Mar 19. PMID: 19299482; PMCID: PMC2717768.

Hurst CH, Turnbull D, Hemsley PA. Determination of Protein S-Acylation State by Enhanced Acyl-Switch Methods. Methods Mol Biol. 2019;2009:3-11. doi: 10.1007/978-1-4939-9532-5_1. PMID: 31152391.

Hurst CH, Turnbull D, Gronnier J, Myles SM, Pflughaupt RL, Kopischke M, Davies P, Jones S, Robatzek S, Zipfel C, Hemsley PA. S-acylation stabilizes ligand-induced receptor kinase complex formation during plant pattern-triggered immune signalling bioRxiv 2021.08.30.457756; doi: https://doi.org/10.1101/2021.08.30.457756

Kang HJ, Voleti B, Hajszan T, Rajkowska G, Stockmeier CA, Licznerski P, Lepack A, Majik MS, Jeong LS, Banasr M, Son H, Duman RS. Decreased expression of synapse-related genes and loss of synapses in major depressive disorder. Nat Med. 2012 Sep;18(9):1413-7. doi: 10.1038/nm.2886. PMID: 22885997; PMCID: PMC3491115.

Kenny EF, O'Neill LA. Signalling adaptors used by Toll-like receptors: an update. Cytokine. 2008 Sep;43(3):342-9. doi: 10.1016/j.cyto.2008.07.010. Epub 2008 Aug 15. PMID: 18706831.

Kim DI, Jensen SC, Noble KA, Kc B, Roux KH, Motamedchaboki K, Roux KJ. An improved smaller biotin ligase for BioID proximity labeling. Mol Biol Cell. 2016 Apr 15;27(8):1188-96. doi: 10.1091/mbc.E15-12-0844. Epub 2016 Feb 24. PMID: 26912792; PMCID: PMC4831873.

Kim, Y.J., Lee, H.M., Wang, Y., Wu, J., Kim, S.G., Kang, K.Y., Park, K.H., Kim, Y.C., Choi, I.S., Agrawal, G.K., Rakwal, R. and Kim, S.T. (2013), Depletion of abundant plant RuBisCO protein using the protamine sulfate precipitation method. Proteomics, 13: 2176-2179. https://doi.org/10.1002/pmic.201200555

Kourelis, J., Kaschani, F., Grosse-Holz, F.M. et al. A homology-guided, genome-based proteome for improved proteomics in the alloploid Nicotiana benthamiana. BMC Genomics 20, 722 (2019). https://doi.org/10.1186/s12864-019-6058-6

Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 1970 Aug 15;227(5259):680-5. doi: 10.1038/227680a0. PMID: 5432063.

Lam SS, Martell JD, Kamer KJ, Deerinck TJ, Ellisman MH, Mootha VK, Ting AY. Directed evolution of APEX2 for electron microscopy and proximity labeling. Nat Methods. 2015 Jan;12(1):51-4. doi: 10.1038/nmeth.3179. Epub 2014 Nov 24. PMID: 25419960; PMCID: PMC4296904.

Li P, Li J, Wang L, Di LJ. Proximity Labeling of Interacting Proteins: Application of BioID as a Discovery Tool. Proteomics. 2017 Oct;17(20). doi: 10.1002/pmic.201700002. Epub 2017 Apr 10. PMID: 28271636.

Li QF, Lu J, Yu JW, Zhang CQ, He JX, Liu QQ. The brassinosteroid-regulated transcription factors BZR1/BES1 function as a coordinator in multisignal-regulated plant growth. Biochim Biophys Acta Gene Regul Mech. 2018 Jun;1861(6):561-571. doi: 10.1016/j.bbagrm.2018.04.003. Epub 2018 Apr 17. PMID: 29673687.

Liu Q, Zheng J, Sun W, Huo Y, Zhang L, Hao P, Wang H, Zhuang M. A proximity-tagging system to identify membrane protein-protein interactions. Nat Methods. 2018 Sep;15(9):715-722. doi: 10.1038/s41592-018-0100-5. Epub 2018 Aug 13. PMID: 30104635.

Li Y, Scott R, Doughty J, Grant M, Qi B. Protein S-Acyltransferase 14: A Specific Role for Palmitoylation in Leaf Senescence in Arabidopsis. Plant Physiol. 2016 Jan;170(1):415-28. doi: 10.1104/pp.15.00448. Epub 2015 Nov 4. PMID: 26537563; PMCID: PMC4704564.

Lozano-Elena F, Caño-Delgado AI. Emerging roles of vascular brassinosteroid receptors of the BRI1-like family. Curr Opin Plant Biol. 2019 Oct;51:105-113. doi: 10.1016/j.pbi.2019.06.006. Epub 2019 Jul 23. PMID: 31349107.

Martin DD, Beauchamp E, Berthiaume LG. Post-translational myristoylation: Fat matters in cellular life and death. Biochimie. 2011 Jan;93(1):18-31. doi: 10.1016/j.biochi.2010.10.018. Epub 2010 Nov 5. PMID: 21056615.

Martins S, Dohmann EM, Cayrel A, Johnson A, Fischer W, Pojer F, Satiat-Jeunemaître B, Jaillais Y, Chory J, Geldner N, Vert G. Internalization and vacuolar targeting of the brassinosteroid hormone receptor BRI1 are regulated by ubiquitination. Nat Commun. 2015 Jan 21;6:6151. doi: 10.1038/ncomms7151. Erratum in: Nat Commun. 2021 May 17;12(1):2982. PMID: 25608221; PMCID: PMC4713032.

May DG, Scott KL, Campos AR, Roux KJ. Comparative Application of BioID and TurboID for Protein-Proximity Biotinylation. Cells. 2020 Apr 25;9(5):1070. doi: 10.3390/cells9051070. PMID: 32344865; PMCID: PMC7290721.

Miyamoto T, DeRose R, Suarez A, Ueno T, Chen M, Sun TP, Wolfgang MJ, Mukherjee C, Meyers DJ, Inoue T. Rapid and orthogonal logic gating with a gibberellin-induced dimerization system. Nat Chem Biol. 2012 Mar 25;8(5):465-70. doi: 10.1038/nchembio.922. PMID: 22446836; PMCID: PMC3368803.

Nam KH, Li J. BRI1/BAK1, a receptor kinase pair mediating brassinosteroid signaling. Cell. 2002 Jul 26;110(2):203-12. doi: 10.1016/s0092-8674(02)00814-0. PMID: 12150928.

Nguyen TMT, Kim J, Doan TT, Lee MW, Lee M. APEX Proximity Labeling as a Versatile Tool for Biological Research. Biochemistry. 2020 Jan 28;59(3):260-269. doi: 10.1021/acs.biochem.9b00791. Epub 2019 Dec 3. PMID: 31718172.

Noguchi T, Fujioka S, Choe S, Takatsuto S, Yoshida S, Yuan H, Feldmann KA, Tax FE. Brassinosteroid-insensitive dwarf mutants of Arabidopsis accumulate brassinosteroids. Plant Physiol. 1999 Nov;121(3):743-52. doi: 10.1104/pp.121.3.743. PMID: 10557222; PMCID: PMC59436.

Özcelik D, Barandun J, Schmitz N, Sutter M, Guth E, Damberger FF, Allain FH, Ban N, Weber-Ban E. Structures of Pup ligase PafA and depupylase Dop from the prokaryotic ubiquitin-like modification pathway. Nat Commun. 2012;3:1014. doi: 10.1038/ncomms2009. PMID: 22910360; PMCID: PMC4351746.

Rana MS, Kumar P, Lee CJ, Verardi R, Rajashankar KR, Banerjee A. Fatty acyl recognition and transfer by an integral membrane S-acyltransferase. Science. 2018 Jan 12;359(6372):eaao6326. doi: 10.1126/science.aao6326. PMID: 29326245; PMCID: PMC6317078.

Röttig A, Steinbüchel A. Acyltransferases in bacteria. Microbiol Mol Biol Rev. 2013 Jun;77(2):277-321. doi: 10.1128/MMBR.00010-13. PMID: 23699259; PMCID: PMC3668668.

Roth AF, Feng Y, Chen L, Davis NG. The yeast DHHC cysteine-rich domain protein Akr1p is a palmitoyl transferase. J Cell Biol. 2002 Oct 14;159(1):23-8. doi: 10.1083/jcb.200206120. Epub 2002 Oct 7. PMID: 12370247; PMCID: PMC2173492.

Roth, G.E., Gierl, M., Radke, M., Meise, M., Lintermann, R., Korge, G. (2002). Drosophila Start is related to the vertebrate genes MLN64 and StAR, and is expressed in steroidogenic tissues. A. Dros. Res. Conf. 43 : 831C.

Roux KJ, Kim DI, Raida M, Burke B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J Cell Biol. 2012 Mar 19;196(6):801-10. doi: 10.1083/jcb.201112098. Epub 2012 Mar 12. PMID: 22412018; PMCID: PMC3308701.

Russinova E, Borst JW, Kwaaitaal M, Caño-Delgado A, Yin Y, Chory J, de Vries SC. Heterodimerization and endocytosis of Arabidopsis brassinosteroid receptors BRI1 and AtSERK3 (BAK1). Plant Cell. 2004 Dec;16(12):3216-29. doi: 10.1105/tpc.104.025387. Epub 2004 Nov 17. PMID: 15548744; PMCID: PMC535869.

Salaun C, Locatelli C, Zmuda F, Cabrera González J, Chamberlain LH. Accessory proteins of the zDHHC family of S-acylation enzymes. J Cell Sci. 2020 Nov 17;133(22):jcs251819. doi: 10.1242/jcs.251819. PMID: 33203738.

Santiago J, Henzler C, Hothorn M. Molecular mechanism for plant steroid receptor activation by somatic embryogenesis co-receptor kinases. Science. 2013 Aug 23;341(6148):889-92. doi: 10.1126/science.1242468. Epub 2013 Aug 8. PMID: 23929946.

Schmidt MF, Schlesinger MJ. Fatty acid binding to vesicular stomatitis virus glycoprotein: a new type of post-translational modification of the viral glycoprotein. Cell. 1979 Aug;17(4):813-9. doi: 10.1016/0092-8674(79)90321-0. PMID: 226266.

Shiu SH, Bleecker AB. Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. Proc Natl Acad

Sci U S A. 2001 Sep 11;98(19):10763-8. doi: 10.1073/pnas.181141598. Epub 2001 Aug 28. PMID: 11526204; PMCID: PMC58549.

Sikorski RS, Hieter P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics. 1989 May;122(1):19-27. doi: 10.1093/genetics/122.1.19. PMID: 2659436; PMCID: PMC1203683.

Sorek N, Poraty L, Sternberg H, Buriakovsky E, Bar E, Lewinsohn E, Yalovsky S. Corrected and Republished from: Activation Status-Coupled Transient S-Acylation Determines Membrane Partitioning of a Plant Rho-Related GTPase. Mol Cell Biol. 2017 Nov 13;37(23):e00333-17. doi: 10.1128/MCB.00333-17. PMID: 28894027; PMCID: PMC5686577.

Sutter M, Damberger FF, Imkamp F, Allain FH, Weber-Ban E. Prokaryotic ubiquitin-like protein (Pup) is coupled to substrates via the side chain of its C-terminal glutamate. J Am Chem Soc. 2010 Apr 28;132(16):5610-2. doi: 10.1021/ja910546x. PMID: 20355727

Tian X, Isamiddinova NS, Peroutka RJ, Goldenberg SJ, Mattern MR, Nicholson B, Leach C. Characterization of selective ubiquitin and ubiquitin-like protease inhibitors using a fluorescence-based multiplex assay format. Assay Drug Dev Technol. 2011 Apr;9(2):165-73. doi: 10.1089/adt.2010.0317. Epub 2010 Dec 6. PMID: 21133675; PMCID: PMC3065724.

Torii KU. Receptor kinase activation and signal transduction in plants: an emerging picture. Curr Opin Plant Biol. 2000 Oct;3(5):361-7. doi: 10.1016/s1369-5266(00)00097-2. PMID: 11019802.

Van Larebeke N, Engler G, Holsters M, Van den Elsacker S, Zaenen I, Schilperoort RA, Schell J. Large plasmid in Agrobacterium tumefaciens essential for crown gall-inducing ability. Nature. 1974 Nov 8;252(5479):169-70. doi: 10.1038/252169a0. PMID: 4419109.

Voinnet O, Rivas S, Mestre P, Baulcombe D. An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. Plant J. 2003 Mar;33(5):949-56. doi: 10.1046/j.1365-313x.2003.01676.x. Retraction in: Plant J. 2015 Nov;84(4):846. Erratum in: Plant J. 2015 Aug;83(4):752. PMID: 12609035.

Walker JC. Structure and function of the receptor-like protein kinases of higher plants. Plant Mol Biol. 1994 Dec;26(5):1599-609. doi: 10.1007/BF00016492. PMID: 7858206.

Wang X, Li X, Meisenhelder J, Hunter T, Yoshida S, Asami T, Chory J. Autoregulation and homodimerization are involved in the activation of the plant steroid receptor BRI1. Dev Cell. 2005 Jun;8(6):855-65. doi: 10.1016/j.devcel.2005.05.001. PMID: 15935775.

Xu Y, Fan X, Hu Y. In vivo interactome profiling by enzyme-catalyzed proximity labeling. Cell Biosci. 2021 Jan 29;11(1):27. doi: 10.1186/s13578-021-00542-3. PMID: 33514425; PMCID: PMC7847152.

Zhang J, Li W, Xiang T, Liu Z, Laluk K, Ding X, Zou Y, Gao M, Zhang X, Chen S, Mengiste T, Zhang Y, Zhou JM. Receptor-like cytoplasmic kinases integrate signaling from multiple plant immune receptors and are targeted by a Pseudomonas syringae effector. Cell Host Microbe. 2010 Apr 22;7(4):290-301. doi: 10.1016/j.chom.2010.03.007. PMID: 20413097.

Appendix : list of primers

	Name		Sequence
SM 1	UFSP FLAGdel F		GCGAGCGCTTGGAGCCAC
SM 2	UFSP FLAGdel R		TCCACCACGGAGACGGAG
SM 3	USP FLAGins F		tcatgacattgattataaggacgatgatgacaagGGCGACTCTGTCTTAGATG
SM 4	USP FLAGins R		tccttataatctccatcatggtccttataatcaccTGATTTCTCGAACTGCGG
SM 5	USFP LSDDVin s F		tgatgttGGCGACTCTGTCTTAGATG
SM 6	USFP LSDDVin s R		tcagaaagCTTGTCATCATCGTCCTTATAATC
SM 7	USFP FLPupE F		gttgaagaagctcctgaagctcaagtttctgaagatcttaaggaaagacaagaaaagCTTTCTGATGATG TTGGC
SM 8	USFP FLPupE R		cttcagtagatctagtagccttttgttgtcctcctccagtatccttagtagcCTTGTCATCATCGTCCTTATA ATC
SM 9	PupE>A F		GAAGGGGGGAgcgTGAAAGGGTG
SM	PupE>A		TGGACGAAGCTCCTAACG

10	R		
SM 11	PafA D57N F		ACTTTATTTGaatGTTGGTTCTCATC
SM 12	PafA D57N R		CTAGCTCCGTTTCTCAAG
SM 13	PafA Seq 400bp		Acggctaatacctacttcgc
SM 14	PafA Seq 884bp		GCACTTTTAGGGGTCAGAGG
SM 15	PafA Seq 1364bp		GGGAGGGTGTTTCTTCTGCTA
SM 16	PafA Seq 1879bp		CATCCAAGAGTTGCTCAGATTG
SM 17	PafA Seq 2377bp		gctgcagtgcttcttataagtga
SM 18	PafA Seq 2874bp		gatggttatgcgtggctaca
SM 19	PafA Seq 505bp		AACAATGATAACGTCTCAACAGC
SM 20	USFP FLPupE aa F		aaGTTGAAGAAGCTCCTGAAG
SM 21	USFP FLPupE aa R		CTTCAGTAGATCTAGTAGC
SM 22	PafA seq start		CTTCCTATTAGTGTATGTCAAG
SM 23	pRS-MCS-F		tgtaaaacgacggccagtga
SM 24	pRS-MCS-R		tgtgtggaattgtgagcgga

SM 25	ALD6pro F		ttgtaatacgactcactatagggcgaattgGCGGCCGCCACCGACCATGTGGGCAAATTC
SM 26	ALD6pro R		GATGATGATGATGATGATGATCTCTCTTGAGCTCCATTGTATTCTGATAGTATG TGTTTG
SM 27	GAIF		TACACAAACACATACTATCAGAATACAATGGAGCTCAAGAGAGATCATCATCAT CATCAT
SM 28	GAIR		AAAAGACTAATAATTCTTAGTTAAAAGCACTCTAGTCGACATTAAGGTCGGTGA GCATAG
SM 29	ENO2ter mF		GATTCTATGCTCACCGACCTTAATGTCGACTAGAGTGCTTTTAACTAAGAATTA TTAGTC
SM 30	ENO2ter mR		ttaaccctcactaaagggaacaaaagctgTTAATTAAATTTTTCAAACTGCAAATTCAAG
SM 31	RPL18Bp roF		gtaatacgactcactatagggcgaattgGGCGCGCCTTTAAGGAATAAGGATACTTCAAG
SM 32	RPL18Bp roR		ATAAGATTAACTTCATCGCTCGCAGCGAGCTCCATTTTGTTTTTTGTTTTCTTCT AATTG
SM 33	GID1F		GCTCGCTGCGAGCGATGAAGTTAATCTTATTGAGAGCAGAACAGTGGTTCCTC TCAATAC
SM 34	GID1R		AAAAATCATTATCCTCATCAAGATTGCTTTATTTAGTCGACACATTCCGCGTTTA CAAAC
SM 35	TDH1ter mF		TGAGATTTCGGCGTTTGTAAACGCGGAATGTGTCGACTAAATAAAGCAATCTTG ATGAGG
SM 36	TDH1ter mR		aaccctcactaaagggaacaaaagctgCCTGCAGGCGTTCAGGGTAATATATTTTAACCG
SM 37	STE2F		ATCAATTAGAAGAAAACAAAAAACAAAATGTCTGATGCGGCTCCTTCAT
SM 38	STE2R		TCCTCCTCGCCCTTGCTCACTGAGCTCCCCGTTTTTCTGCTTCTCGATG
SM 39	mEGFP-NF		TACACAAACACATACTATCAGAATACAATGGAGCTCGTGAGCAAGGGCGAGGA G
SM	mEGFP-		TTGATGATGATGATGATGATGATCTCTCTTTCCGGCCGACTTGTACAGCTCGTC

40	NR		CATGC
SM 41	mEGFP-CF		TGGCTTGATTCTATGCTCACCGACCTTAATTCGGCCGGAGTGAGCAAGGGCGAGGAG
SM 42	mEGFP-CR		GACTAATAATTCTTAGTTAAAAGCACTCTAGTCGACCTTGTACAGCTCGTCCATGC
SM 43	mChF		CATCGAGAAGCAGAAAAACGGGGAGCTCAGTGAGCAAGGGCGAGGAG
SM 44	mChR		CTCAATAAGATTAACTTCATCGCTCGCAGCTCCGGCCGACTTGTACAGCTCGTCCATGC
SM 45	HAPafA-STOP-F		aAAGGGTGGGCGCGCCGAC
SM 46	HAPafA-STOP-R		taCATTCCAGCAATCAACTTCTCCACTCTCTCATCAACAG
SM 47	PafA3HA-ATG-F		gGATAGGAGGATATTCGGTTTAGAGAACGAGTATGGAGTTACATGC
SM 48	PafA3HA-ATG-R		atGGTGAAGGGGGCGGCCGC
SM 49	2muURA 3F		TTCTGCTTAGTTTGATGCCTGGCAGTTTATGGCGGGCGTCTTCCTTTTTCAATGGGTAATAACTGA
SM 50	2muURA 3R		TTGAACGTTGTGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCATTATCAACCGGGGTGGAGC
SM 51	pFF PafA F		TTCATTTCGAGAAGATCGTAACGAGGATCGAGAAGTTGACGATAGGAGGATATTCGGTTTAG
SM 52	pFF PafA R		ATGCCAACTTTGTACAAGAAAGCTGGGTCGGCGCGCCCTAAGCGTAATCTGGAACGTC
SM 53	USMP FL F		cgagcagaagctgatcagcgaggaggacctgaacggcgccgtggacGGCGACTCTGTCTTAGATG
SM 54	USMP FL R		ccgttcaggtcctcctcgctgatcagctgctcgcccatgtggcccagACCTGATTTCTCGAACTG

SM 55	USMP LSDDV F		gagcagaagctgatcagcgaggaggacctgaacggcgaggccgtggacCTTTCTGATGATGTTGG C
SM 56	USMP LSDDV R		gccgttcaggtcctcctcgctgatcagctgctcgcccatgtggcccagACCTGATTTCTCGAACTG
SM 57	USMP F		gagcagaagctgatcagcgaggaggacctgaacggcgaggccgtggacGGCGACTCTGTCTTAGA TG
SM 58	USMP R		gccgttcaggtcctcctcgctgatcagctgctcgcccatgtggcccagACCTGATTTCTCGAACTG
SM 59	USMP FL F		gagcagaagctgatcagcgaggaggacctgaacggcgaggccgtggacGCTACTAAGGATACTGG AG
SM 60	pENTR-3HA-PafA-7xGGSG G In FRAME F		gGGCGCGCCGACCCAGCTT
SM 61	pENTR-3HA-PafA-7xGGSG G In FRAME R		ACCTCCAGACCCTCCGCCAC
SM 62	pENTR 7xGGSG G-PafA-3HA IN FRAME F		cGGAGGTTCTGGAGGTGGCGGATCAGG
SM 63	pENTR 7xGGSG		GCGGCCGCGGAGCCTGCT

	G-PafA-3HA IN FRAME R		
SM 64	UBQ GG>AA R		AGGACGAGATGAAGCGTC
SM 65	UBQ GG>AA F		CCGTCTCCGTGCCGCAGCGAGCGCTT
SM 66	pK7WG yeast adaptor f		cattcggttaaacaccacgcacgttgccatgcagcgtacg GGTCCTTTTCATCACGTGCT
SM 67	pK7WG yeast adaptor r		ataccgtcaccaggcggccgttcttggccttcttcgtacgGTTTTGCTGGCCGCATCTTC
SM 68	XVE f		atatcactagtgcggccgcctgcaggtcgaccatatgggaCTGATAGTTTAAACTGAAGG
SM 69	XVE r		ggcgggaaacgacaatctgatccaagctcaagctaagcttTGTTTGGGATGTTTTACTCC
SM 70	OlexA f		ctcgttcagctttttgtacaaacttgtgatatcactagtAGCTTGGGCTGCAGGTCGAG
SM 71	OlexA(PupE) r		ggtcttccccgttagggttttcacgaagatctgcatggtgCGACTAGCTTCAGCGTGTCC
SM 72	OlexA)PupE f		aagttcatttcatttggagaggacacgctgaagctagtcgCACCATGCAGATCTTCGTGA
SM 73	PupE(T3A) r		gaaaccgatgatacggacgaaagctgggaggcctggatcgTCACTCTCCCCCCTTCTGGA
SM 74	(PupE)T3A f		ggatttcgttaggagcttcgtccagaaggggggagagtgaCGATCCAGGCCTCCCAGCTT
SM 75	T3A r		atcagtcccatatggtcgacctgcaggcggccgcactagtAAGCCTATACTGTACTTAAC
SM	pFF		TTCATTTCGAGAAGATCGTAACGAGGATCGAGAAGTTGGTACCAAAGACAATA

76	TurboID F		CTGTGCCTCT
SM 77	pFF TurboID R		ATGCCAACTTTGTACAAGAAAGCTGGGTCGGCGCGCCCTATCAGTGGTGGTG GTGGT
SM 78	GGSGG Sense		AGGTggatctggtGG
SM 79	GGSGG Anti-sense		CTCCaccagatccAC
SM 80	3GGSGG Sense		AGGTggatctggtggaggcgggtcaggaggtggcggatccggtGG
SM 81	3GGSGG Anti-sense		CTCCaccggatccgccacctcctgacccgcctccaccagatccAC
SM 82	5GGSGG Sense		AGGTggatctggtggaggcgggtcaggaggtggcggatccggtggaggtggatcgggtggcggtggatctgga GG
SM 83	5GGSGG Anti-Sense		CTCCtccagatccaccgccacccgatccacctccaccggatccgccacctcctgacccgcctccaccagatcc AC
SM 84	5GGSGG SDM PafA-3HA F		tggaggtggaggttctggaggtggaggttctggaggtGATAGGAGGATATTCGGTTTAGAGAAC GAGTATGGAGTTACATGC
SM 85	5GGSGG SDM PafA-3HA R		gaacctccacctccagaacctccacctccagaacctccGGTGAAGGGGGCGGCCGC
SM 86	35s promoter with HpaI		GTTACCtgagacttttcaacaaagggt

	site F		
SM 87	35s Term with HpaI site R		CCATTGgatctagtaacatagatgac
SM 88	35s Term seq primer R		Gactgatgggctgcctgtat