Identification of a New Class of SUMO Proteases in Plants, and Investigation Into The Role of SUMOylation in Pathogen Perception

YATES, GARY

Identification of a New Class of SUMO Proteases in Plants, and Investigation Into The Role of SUMOylation in Pathogen Perception

Gary Yates

Thesis submitted for the Degree of Doctor of Philosophy Department of Biosciences

September 2017

Abstract

Plants, as primary produces of energy and the main source of food for many organisms, are both ecologically and economically important to sustaining life. Understanding the molecular mechanisms by which plants respond to stress is key to optimizing food production and maintaining biodiversity in an increasingly uniform global ecosystem. *Arabidopsis thaliana* is the major model plant species in which primary research provides insights into the workings of more important plant species. The research presented here advances our understanding of the role of SUMOylation and deSUMOylation in plants. The discovery of a previously undescribed group of SUMO proteases in plants, called deSUMOylating Isopeptidases (DeSis) has provided evidence for the first time of SUMOylation cycles outside the nucleus. Using proteomic comparisons from animal species, this group of SUMO proteases has been classified into 3 distinct sub-classes; DeSi1, DeSi2 and the plant specific DeSi3. Characterisation of the SUMO isopeptidase activity was achieved through biochemical assays in which SUMO1 conjugation chains were reduced in the presence of the DeSi proteins. Furthermore, investigation into the role of the DeSi3 protein At1g47740 (DeSi3a), has revealed key regulation of the immune response pathway. DeSi3a knockout plants were more resistant to pathogen attack, showed a higher response to the detection of pathogenic elicitor molecules and appeared to be primed for immune response. Together this provided evidence that DeSi3a is a major negative regulator of plant immunity. In addition to this, investigation into potential DeSi3a target substrates has revealed that the activation of the key receptor for the perception of bacterial pathogens, FLS2, is SUMOylation dependent. Further evidence suggests that the fungal pathogen receptor, CERK1, may also be regulated by SUMO. This thesis provides evidence that the regulation of pathogen perception is SUMOylation dependent and that DeSi3a is a key negative regulator of defence.

Contents

List of Publications

1. **The ubiquitin conjugating enzyme, TaU4 regulates wheat defence against the phytopathogen Zymoseptoria tritici**. Linda Millyard, Jack Lee, Cunjin Zhang, Gary Yates & Ari Sadanandom
   Scientific Reports 6:35683 · **2016** doi: 10.1038/srep35683

2. **SUMO proteases: uncovering the roles of deSUMOylation in plants.** Gary Yates, Anjil Kumar Srivastava and Ari Sadanandom
   Journal of Experimental Botany, 67, 9, pp. 2541–2548, **2016** doi**:** 10.1093/jxb/erw092

3. **SUMO is a critical regulator of salt stress responses in rice**. Anjil Kumar Srivastava; Cunjin Zhang; Gary Yates, Mark Bailey, Adrian Brown, and Ari Sadanandom
   Plant physiology, 170, 4, pp.01530.2015 **2016** doi: 10.1104/pp.15.01530

4. **Expression, Purification, and Enzymatic Analysis of Plant SUMO Proteases.** Gary Yates, Anjil Kumar Srivastava, Beatriz O'Rosa, Ari Sadanandom.
   Plant Proteostasis. Book Chapter. pp.125-133 **2016**. doi: 10.1007/978-1-4939-3759-2_10

5. **Ubiquitination In Plant Nutrient Utilisation**. Gary Yates & Ari Sadanandom. *Frontiers in Plant Science,* 4, 452 · **2013** doi: 10.3389/fpls.2013.00452 ·

6. **Identification of the domains of Cauliflower mosaic virus protein P6 responsible for suppression of RNA-silencing and salicylic acid-signalling.** Janet Laird, Carol McInally, Craig Carr, Sowjanya Doddiah, Gary Yates, Elina Chrysanthou, Ahmed Khattab, Andrew J. Love, Chiara Geri, Ari Sadanandom, Brian O. Smith, Kappei Kobayashi and Joel J. Milner
   Journal of General Virology, 94, 12, · **2013** doi: 10.1099/vir.0.057729-0

Under review, January 2018 - **SUMO conjugation to the pattern recognition receptor FLS2 is required to trigger intracellular signalling in plant innate immunity.** Beatriz O'rosa and Gary Yates, Vivek Verma, Jack Lee, Alberto Campanaro, Moumita Srivastava, Malcolm Bennett, Ari Sadanandom. *Nature Communications.*

List of figures

List of tables

Abbreviations

- Amp ampicillin
- BAK1 BRASSINOSTEROID INSENSITIVE1 ASSOCIATED KINASE1
- BIK1 BOYTRITIS INDUCED KINASE1
- BP base pairs
- BZEL BTB-ZINC FINGER EFFECTOR LYMPHOCYTES
- CERK1 CHITIN ELICITOR RECEPTOR KINASE1
- Chl chloramphenicol
- DeSi deSUMOylation isopeptidase
- DTT dithiothreitol
- DUB deubiquitinating enzymes
- EFR EF-Tu RECEPTOR
- ELS1 ULP1a/ESD4 LIKE SUMO PROTEASE
- ESD4 EARLY IN SHORT DAYS 4
- ETI effector-triggered immunity
- ETS effector-trigger susceptibility
- FLC FLOWERING LOCUS
- flg22 flagellin peptide fragment of 22 amino acids
- FLS2 FLAGELLIN SENSING 2
- FPLC fast protein liquid chromatograph
- GA gibberellin
- Gent gentamycin
- HR hypersensitive response
- HRP horseradish peroxidase
- HuDe1 *Homo sapiens* DeSi1
- HYP2 HIGHYPLIODY2
- IPTG isopropyl β-D-1-thiogalactopyranoside
- Kan kanamycin

- Kb kilo-base pairs
- kDa kilo daltons
- LB Luria Bertani (Media)
- LRR leucine rich repeat
- MAMP microbe associated molecular patterns
- MAPK mitogen activated protein kinase
- MAPKKKK MAPK kinase kinase kinase
- MKK MAPK kinase
- MS Murashige & Skoog (media)
- NLR nucleotide-binding, leucine-rich repeats
- OD optical density
- OTS1/2 OVERLY-TOLERENT-TO-SALT1/2
- PAGE Polyacrylamide gel electrophoresis
- PAMP pathogen associated molecular patterns
- PCD programm cell death
- PM posttranslational modifers
- PRR pattern recognition receptors
- Ps *Pseudomonas syringae*
- Pst *Pseudomonas syringae* DC3000
- PTI PAMP trigger immunity
- RBOHD Respiratory burst oxidase homolog protein D
- Rif rifampicin
- RK receptor kinase
- RLCK receptor-like cytosolic kinase
- RLK receptor like kinase
- ROS reactive oxygen species
- SA salicylic acid
- SAR systemic acquired resistance
- SDS sodium dodecyl sulfate
- SENP sentrin-specific proteases
- SIZ1 SAP & miz1
- SOC Super Optimal broth with Catabolite repression

- SUMO small ubiquitin-like modifier
- Ub ubiquitin
- USPL ubiquitin-specific protease-like

Declaration

This thesis is submitted to the University of Durham in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. The work (including data generated and data analysis) was carried out by the author except where explicitly states otherwise.

Gary Yates

Statement of Copyright

Acknowledgments

I would firstly like to thank my supervisor, Prof. Ari Sadanandom who provided critical input and knowledge throughout the duration of this degree. Secondly, Dr. Cunjin Zhang, and Dr. Beatriz Orosa for their teachings in the lab, both have outstanding lab skills and scientific understanding, which helped me become a competent researcher. Dr. Mark Bailey, Dr. Stuart Nelis, Helen Riordan and Dr. Charlotte Walsh for their welcoming friendship when I arrived in the lab, and their continued support throughout. A thank you also goes to all the members of the Sadanandom lab past and present for putting up with my strange ways.

Special appreciation goes to Prof. Andres Bachmair and Dr. Konstantin Tomanov for hosting me in Vienna and teaching me crucial biochemical skills. Also to Dr Jack Lee for allowing me to ware him down over the years until he eventually became a Scottish football fan (sort of!) and for *Infinity+1*. Extra special appreciation goes to Linda Millyard for proof-reading, refraining from killing me during my thesis write up and for the countless other contributions she made to make this happen.

A final thank you to all the people in the Bioscience department for there help over the years and to Prof Keith Lindsey and the department band, who gave me a chance to 'rock it' on bass.

Chapter 1.
Introduction

1.01 - *Arabidopsis thaliana* – The Model: A Brief History

Arabidopsis thaliana (Arabidopsis) is a well-established organism for studying all components of plant life and has been instrumental in deciphering mechanisms that have lead to a greater understanding of all life on planet Earth. However it is apparent that the research performed on this organism was only made possible by the early pioneers and champions of the past, which not only established Arabidopsis as a suitable botanic research tool, but also as a major source of innovation in developing molecular and genetics techniques (Lamesch *et al*., 2012). Arabidopsis has been studied scientifically since the turn of the 20th century, with the earliest paper describing its chromosomal arrangement published in 1907 by Fredrick Laibach. This first published article on Arabidopsis described the existence of 5 chromosomes that contain the Arabidopsis genome (Laibach 1907). Laibach also published further work in the early days of Arabidopsis research including flowering time and seed dormancy variation, and his research group created the first synthetic mutants in Arabidopsis using X-ray mutagenesis in the 1930s (Kroonneef & Meinke 2010). By 1955 auxotrophic mutants in Arabidopsis were being generated and an international community centred on Arabidopsis research started to grow from the late 1950s into the 1960s (Langridge 1955).

Before the genomic era, the plant was studied in various botany-based projects, however until 1980 there were only 65 publications recorded (Kroonneef & Meinke 2010). As the need for suitable genetic model organisms grew in the 1980s, the potential for Arabidopsis was highlighted several times within the literature. One of the more important articles published by Somerville and Ogren 1980, provided detailed insight into how effective Arabidopsis mutant studies were for understanding both physiological and molecular traits (Somerville and Ogren 1980). By 1985 the Arabidopsis genome size was calculated and this led to further validation for its use in genetic research, due to its relatively small sized genome (125 x 10^{6} base

pairs) (Leutwiler *et al*., 1984). In the late 1980s, possibly the most significant discovery at the time, was the stable introduction of transgenes into the germ-line. This was achieved by treating mature seeds with transgenic *Agrobacterium tumefaciens* (Agrobacteria) (Feldmann and Marks 1987). The significance of this proved to be crucial for all future Arabidopsis research as prior to this the main way to transform plants was through tissue culture based techniques, something that was not well established in Arabidopsis due to the genetic background incompatibility issues caused by natural variation. The research that followed this work led to the development of the well-established foral dip method, which used transgenic *Agrobacteria* to infect the flower bud creating stable transgenes in the developing embryos (Bechtold *et al*., 1993; Clough and Bent, 1998). This technology was taken a step further by the creation of T-DNA insertions, which exploited the floral dip methodology to create single gene knockout lines. (Haseloff and Amos, 1995). By the late 1990s *Arabidopsis thaliana* had become only the third multicellular organism, and the first plant, to have a fully sequenced genome. The resulting annotations and programmes like the 'Arabidopsis 2010 project' (an initiative aimed at adding functionality to all 25000+ Arabidopsis genes by the year 2010) (Chory e*t al*., 2000), has give rise to one of the most comprehensive genetic and molecular libraries in all of biology, collectively housed by the TAIR database/website (Kroonneef & Meinke 2010; Somerville and Koornneef, 2002; Lamesch *et al*., 2012).

1.02 – *Arabidopsis thaliana* - key features

The *Arabidopsis thaliana* genome consists of 125 x 10^6 base pairs of DNA within 5 chromosomes. The genome contains approximately 25,000 genes and knockout-lines for nearly all genes are available (Lamesch *et al*., 2012). It is a small plant with a rapid life cycle and it can produce new seeds in only 6 weeks after germination. Mutations are easily generated and can be induced by several well-established methods, but a major advantage is the fact that one plant can produce up to 10,000 seeds, making genetic screening much more effective. Furthermore the plant can self-pollinate and can therefore

generate homozygous mutants within 2 generations (Somerville & Meyerowitz 2002).

Another feature that is exploited for research is the interaction between the plant and invading pathogens. Infections can be traced in various ways and susceptibility of *Arabidopsis thaliana* ecotypes (or mutants) can be determined and scored for comparison. A well-studied example of this is the *Arabidopsis thaliana* interaction with the bacterial pathogen *Pseudomonas syringae* (Ps) (Katariri *et al.,* 2002). Ps is a Gram-negative biotrophic pathogen with various strains that can infect a wide range of hosts (Agrios 1997). The bacterium generally gains access to the host tissues via natural openings like stomata or by sites of wounding on the leaves, and once inside susceptible plants, can proliferate and overwhelm host defences. This leads to necrosis and chlorosis and the bacteria can spread quickly within the host (Hirano & Upper 2002). When mature *Arabidopsis thaliana* plants are inoculated with the strain *Pseudomonas syringae pv. tomato DC3000* (Pst), after 2 days dark patches appear on the leaf, making the leaf look water-soaked followed by the onset of chlorosis and necrosis by day 3. This is followed by a symptomatic speckling effect of necrotic spots around the infection sites as the pathogen spreads within the plant (Hirano & Upper 2002). In Arabidopsis seedlings, Pst infection can cause growth retardation in resistant plants, and measuring root growth of infected plants can give some insight into the virulence of the pathogen in a range of host ecotype or genotypes (Zipfel *et al*., 2004). The growth of the Pst within the plant can be measured by extracting samples of leaf tissue that are infected at different time points, and re-plating the bacteria in a serial dilution and counting the colonies that grow. This is useful as it gives an indication of how well the bacterium is growing therefore its virulence within a genotype (Katariri *et al.,* 2002). The plants response to pathogen perception can also be measured by detecting the production of Reactive Oxygen Species (ROS). These free radicals are produced in response to the detection of stress and act as early molecular signals that prepare the plant for a response to the stress (Thakur and Sohal, 2013). Measuring ROS can provide information about the plants ability to perceive pathogens, furthermore this response can be generated by applying synthetic molecules

such as flg22 (A 22 amino acid peptide which mimics a part of the bacterial motility organelle flagellum), which will activate the initial ROS burst without engaging a full infection response (Gómez-Vásquez et al., 2004).

1.03 - Plant immune system overview

As primary providers of energy to the eco-system, plants have to contend with biotic threats from nearly all families of life. Microbial threats, such as bacteria, fungi and oomycetes are a constant danger as most are found within the soil from which the plant grows. However, although plants are at risk from a plethora of pathogens, very few successfully infect the plant (Dickinson 2003). Plants are sessile organisms that have evolved sophisticated mechanisms of cellular regulation. Their immune system is multi-layered and appears more complex with each new detail reported in scientific articles. In order to defend against pathogenic threats, plants have robust cell walls and waxy coverings that act as a good physical barrier (Malinovsky *et al*., 2014). The physical barriers prevent the majority of infectious agents comprising the plant, however plants have naturally occurring entry/exit points such as stomata for gas exchange and transpiration, which some pathogens exploit to negate the natural barriers (Glazebrook 2005). Once breached however, the plant depends on a different style of defence in order to deal with the threat.

During biotic stress cell plasma membrane receptors act as an early trigger for the plants immune response. By recognition and interaction of a receptor, with a molecule of the infectious species, the plant is able to mount a defence response. Known as Pathogen/Microbe Associated Molecular Patterns (PAMPs/MAMPs), these molecules are conserved microbial elements which activate the plants PAMP-trigger Immunity (PTI) (Jones and Dangl*.,* 2006). Examples of PAMPs include bacterial flagellin, bacterial elongation factor (EF-Tu), and the fungal structural component chitin (Boller & Felix 2009). These PAMPs are detected by Pattern Recognition Receptors (PRRs) such as FLAGELLIN SENSING 2 (FLS2), EF-Tu RECEPTOR (EFR) and CHITIN ELICITOR RECEPTOR KINASE1 (CERK1) (Dunning *et al*., 2007; Kombrink *et al.,* 2011). The receptors are of a class of trans-membrane proteins called Receptor-like Kinases (RLK or RK) that are specific to plant species (Shiu and Bleecker,

2002). Within the RLK group there is a subclass of proteins that contain Lecuine-Rich Repeat (LRR) domains that are related to the toll-like receptors of animal systems (Trinchieri and Sher, 2007). Identification of RLK class is determined by the occurrence of the arginine-aspartic acid (RD) motif in the kinase domain, which is the phosphorylation activation site of the receptor of this class (Johnstone *et al*., 1996). LRR-RK proteins make up a family of PRRs that act as the detection molecules of stress signals. The receptors normally target highly conserved microbial elements for greater coverage of defence. FLS2 an non-RD kinase, is one such receptor for example, that can recognise a specific fragment of the bacterial motility apparatus, flagellin. The region of flagellin that FLS2 can recognise is a 22 amino acid long peptide (flg22) that is highly conserved in the bacterial kingdom (Naito *et al.,* 2008). Activation of PRRs such as FLS2, results in PTI as a first-line/early defence for the plant. During PTI downstream effects include but are not limited to; cell wall re-enforcement, increased reactive oxygen species production, MAP-kinase activation and altered gene expression (Zipfel 2008, Chisholm *et al.,* 2006). One way pathogens have evolved to cope with PTI is through production of effector molecules that supress the PTI response, usually resulting in successful infection. Some of the effector molecules from pathogens act upon the kinase domain of the plants PRRs such as FLS2, and can disrupt the signalling cascade that usually follows PAMP recognition (Ziang *et al.,* 2008). In cases where the effector interferes with the PTI response, this can cause increased virulence through Effector-Triggered-Susceptibility (ETS) (Jones and Dangl 2006). Some plants however, have co-evolved proteins which detect the effector proteins or changes in their target, these are known as Resistance (R) proteins. R proteins, when activated induce a secondary defence system called Effector Triggered Immunity (ETI), otherwise known as gene–for-gene resistance (Bent *et al.,* 2007). Effector proteins from different species have evolved to target a plethora of different substrates in the plant cell (Block *et al.,* 2008). R proteins evolve to recognise effector proteins and mount an ETI response, thus creating further selection pressure for the effector proteins to mutate to avoid detection. These interactions are mechanistically diverse, and it appears that over evolutionary timescales a so-

called molecular arms race between plant and pathogen has taken place. The high selection pressure caused by this molecular arms race has created what is known as the zig-zag model. This model suggests a way in which intense selection pressure, is a possible explanation as to why there is such a varied spectrum of targets and responses to effector protein targets and R proteins respectively (Jones & Dangl 2006). ETI can result in a Hypersensitive Response (HR) causing Programmed Cell Death (PCD) at a local level, whilst promoting Systemic Acquired Resistance (SAR) on the global level (Jones & Dangl 2006). SAR primes the plant for a second wave of infection meaning the defence responses are quicker and more effective essentially acting as plant 'memory' or 'experience' derived immunity (Durrant and Dong 2004). SAR signalling mechanisms remain elusive, salicylic acid (SA) signalling has been argued for and against in this role (Rasmussen 2002, Vernooij *et al*., 1994), but interestingly methyl salicylate has been proposed as the actual mobile signal. Being converted to and from SA, for recognition and signalling respectively, it is thought that the as SA is produced at a local level, the methyl salicylate can be synthesised and mobilised, then converted back to SA at systemic sites (Park *et al.,* 2007). SA has a major role in the SAR response and has been shown to be essential for establishing SAR in remote systemic tissues (Durrant and Dong 2004). Another proposed mechanism for SAR signalling is a lipid transfer based mechanism. Mutant Arabidopsis defective in the lipid-tranfer protein DIR1, mount a normal localised defence response but show no SAR response detected in systemic tissues, this suggests that the mobile signal maybe a lipid molecule (Maldonado *et al.,* 2002). The SAR response includes the production of the ROS. Production of ROS acts as a first defence signal and is essential for HR (Thakur and Sohal, 2013). Two ROS bursts have been measured during pathogen stress induction, the first is rapid and considerably larger, induced within minutes of elicitor detection and is triggered locally. The second, is a smaller burst and slower (approx. 50-70 minutes after detection) which is also observed systemically in the plant, both of which are required for SAR activation (Alverez 1998). Synthetic elicitors that mimic pathogen molecules can trigger induction of ROS production through the PAMP receptor pathway, which can lead to activation of cellular

defences (Gómez-Vásquez *et al*., 2004). The receptor for these types of elicitors can also be the target of effector molecules from the pathogen. The Arabidopsis LRR receptor FLS2 for example, is crucial for activation of immune responses, making FLS2 and FLS2 signalling a natural target for effectors (Wang & He 2010). Effector proteins AvrPto and AvrPtoB from the pathogen *Pseudomonas syringae* for example, bind to the LRK co-receptor BIK1, preventing interaction with FLS2, this is part of an essential assembly for immune response (Block & Alfano 2011).

1.04 - Plant pathogen interactions

Effector proteins can prevent proper activation of the plants PTI and thus enable successful infection of the pathogen (Feng and Zhou, 2012; Raffaele and Kamoun, 2012; Win *et al*., 2012). Effector molecules may target a specific virulence factor of the host, and effectively 'turn off' host immunity. However, in some cases, where plants are seemingly resistant to a pathogen, effector molecules are detected in the plant by Nucleotide-binding, Leucine-rich Repeats (NLR) proteins, NLR proteins are highly diverse and cover a multitude of targets. These proteins, as known as R proteins, can interact with the effector directly or indirectly (for example, upon activation of another immune component) (Chisholm *et al*., 2006; Dodds and Rathjen, 2010; Jones and Dangl, 2006). Previously known as the gene-for-gene hypothesis, R protein recognition of effector proteins is currently described as "the guard hypothesis". This model suggests R proteins have virulence factors that they 'guard' by detecting changes induced by effector molecules, which activate ETI (Rafiqi *et al.,* 2011). The mechanistic procedure by which this works is called a "bait and switch", where in the detected host protein, usually a virulence factor, is the bait and the R protein is the switch; activating when the 'bait' is modified by the effector (Collier & Moffet 2009). An extension of the guard hypothesis is "the decoy model". The decoy model implicates that the host protein, which is 'guarded' by the R protein, is a "decoy" i.e. not a virulence factor target, but rather one that emulates an effector target. This allows the plant to detect the pathogen before host defences are compromised and also

prevents the pathogen shutting down the plants immune system thus making the plant resistant.

Over 8,000 *Arabidopsis thaliana* proteins have been identified as having 'R protein' function and these proteins are involved in fungal, bacterial and oomycotal infections (Dreze *et al*., 2011; Mukhtar *et al.,* 2011; Wesseling *et al.,* 2014). An important observation in the field of plant-pathogen interactions is that in pathogenic species representing three different kingdoms, common host protein networks are targeted by the pathogens using unique effector proteins that have undergone convergent evolution to act on the same targets (Wessling *et al*., 2014). This interaction shows that the host target molecules are selectively found by different pathogens over evolutionary timescales (Dreze *et al*., 2011). Moreover it helps to identify various key components of the plant immune system.

1.05 - Activation of immune signalling cascades

PRRs are highly conserved PAMP receptors that cause activation of host defences. The LRR ecto-domain of PRRs is the recognition site for the PAMP molecules and triggers the initial signalling events for immune activation (Macho & Zipfel 2014). For example during recognition of flg22, FLS2 dimerises with and associates in a complex with the transmembrane kinase protein BRASSINOSTEROID INSENSITIVE1 ASSOCIATED KINASE1 (BAK1), propagating a phosphorylation cascade which results in PTI (Figure 1.1) (Macho & Zipfel 2014; Zipfel 2006). Similarly ERF associates with BAK1 upon recognition of EF-Tu, and similar defence response is mounted (Macho & Zipel 2014; Chinchilla *et al.,* 2007). CERK1 also dimerises after recognition of chitin oligomers and associates with other transmembrane proteins resulting in phosphorylation cascade (Liu *et al.,* 2012b; Cao *et al.,* 2014). The mechanistic pathway for PRR signals involves common components and interactions. Receptor-Like Cytosolic Kinases (RLCKs) interact with PRRs and in a conversed manner. CERK1, ERF and FLS2 all associate in complexes with the RLCK; BOYTRISTIS INDUCED KINASE1 (BIK1) (Zhang *et al.,* 2010a; Shi et al., 2013). BAK1 acts to phosphorylate BIK1 upon recognition of the PAMP, in the case of FLS2, PAMP detection results in BAK1 phosphorylating BIK1 and in

turn, RLCK phosphorylation of the FLS2-BAK1 complex prevents interaction with RLCK and resulting in downstream effects such as production of ROS (Figure 1.1) (Zhang *et al.,* 2010a, Macho & Zipfel 2014; Shi *et al.,* 2013). These phosphorylation events are part of a signalling cascade, which links the plants perception of PAMPs to its cellular responses. This is a highly conserved mechanism in eukaryotes and the family of proteins involved are known as Mitogen Activated Protein Kinases (MAPKs). MAPK cascades operate to amplify signals using three components that sequentially phosphorylate other kinase proteins resulting in downstream responses. This is initiated via stimulus of a MAPK-Kinase-Kinase-Kinase (MAPKKK) protein, which phosphorylates a MAPK-Kinase-Kinase (MKK) protein, which in turn phosphorylates a MAPK protein, which then trigger modifications in target proteins such as; transcription factors, effectors and various other substrates (Khokhlatchev *et al.,* 1998; Rodrigues *et al.,* 2011). In the *Arabidopsis thaliana* genome approximately 60 MAPKKKs have been identified using bioinformatics, along with 10 MKKs and over 20 MAPKs (Group *et al*., 2002). These mechanisms are involved in transducing many diverse signal-types in plants such as the hormone signalling, cell differentiation, transcriptional reprogramming (Wang *et al*., 2007; Zhang e*t al*., 1997; Xing 2008). The MAPK cascade is tightly related to the defence-signalling network in plants, SA treatment has been shown to induce MAPK activation in Arabidopsis roots (Mockaitis & Howell, 2000). Furthermore the specificity of one MAPK pathway over another seems to involve combinations of the MAPKKK, MKK and MAPK families. The proteins MPK4 and MPK6 have been shown to be targets of the pathogen effector HopAI1, which is able to de-phosphorylate these MAPK components and supress PTI when infecting plants (Li *et al*., 2007; Boller & Felix 2009).

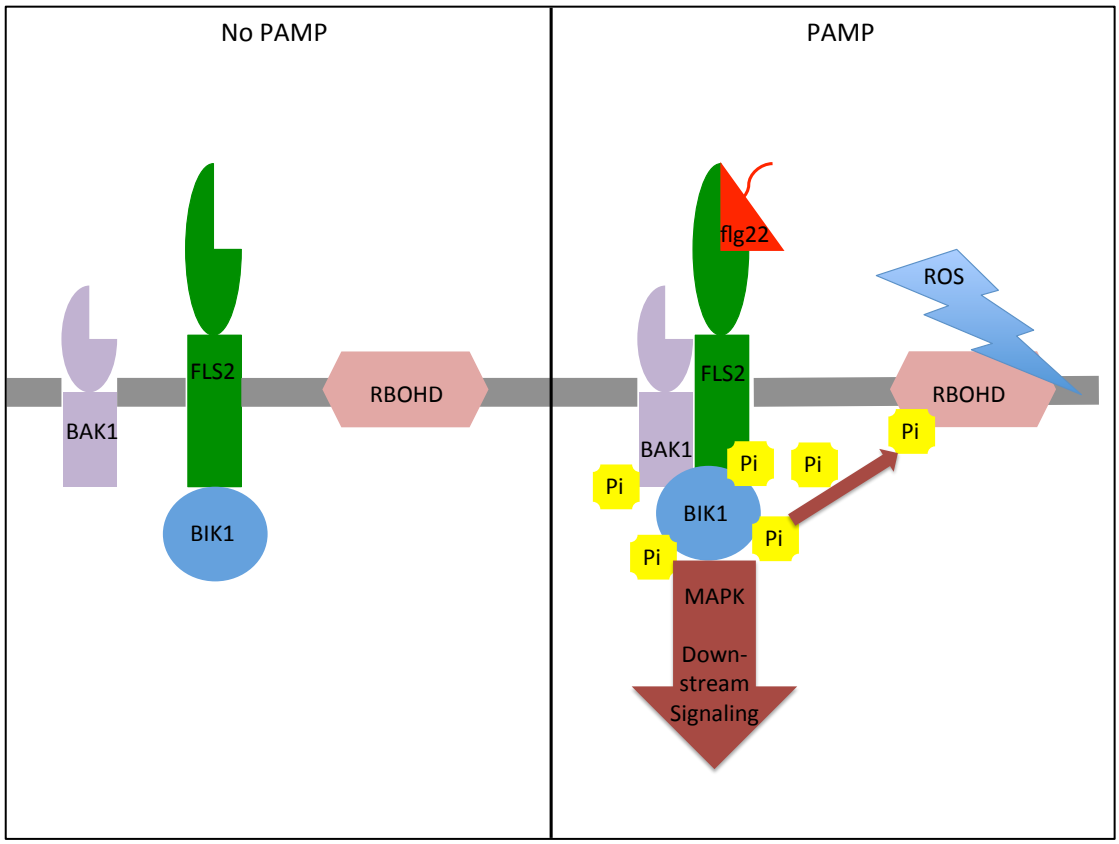

Figure 1.1. Activation of flg22 induced immune signalling. FLS2, BAK1, RBOHD and BIK1 are prior to PAMP recognition on the left panel. The right hand panel shows the re-configuration of the proteins into a complex under recognition of the PAMP flg22. FLS2 dimerises with BAK1 and this complex acts to phosphorylate BIK1. BIK1 induces phosphorylates of RBOHD, activating ROS production. Phosphorylation of BIK causes downstream signalling, through further phosphorylation events involving MAPKs.

1.06 - Post translational modification

Millions of years of evolution have allowed plants to develop many complex and sophisticated mechanisms that enable them to adapt to different environments across the world. These adaptations give them an advantage in certain adverse niches (Munns and Tester, 2008). This requires rapid and reversible physiological responses. Combinations of transcriptional, post-transcriptional, translational, and post-translational mechanisms allow plants to respond in this efficient manner (Edwards *et. al.,* 2012). Modification of protein activity is essential for quick adaptation to environmental stress (Kerscher *et al.,* 2006). This process can be achieved by the addition or removal of small molecules to target proteins, known as post-translational modifications/modifiers (PMs) (Seo and Lee, 2004). A plethora of molecules can bind to target proteins after translation, thereby modifying their activity, stability, and/or localization (Mann and Jenson, 2003). PMs are also involved in signalling mechanisms where not only signalling cascades but also the rapid removal of the signal are important (Vertegaal, 2011). There are many types of PMs in eukaryotic cells and arguably one of the most well-studied examples is the process of ubiquitination.

1.07 - Ubiquitin

Ubiquitin and ubiquitin-like modifiers are important in controlling cellular responses that alter the target protein stability and activity (Miura and Hasegawa, 2010). Highly conserved throughout the eukaryotic kingdom, ubiquitin is involved in most processes of plant life, including but not limited to stress, drought, and nutrient utilization (Sadanandom *et al.,* 2012; Dye and Schulman, 2007; Yates and Sadanandom, 2013). Ubiquitin is a short peptide of only 76 amino acids, which acts by forming bonds with target proteins by linkages via one of seven lysine residues and a C-terminal glycine (Mukhopadhyay and Riezman, 2007). In order for this linkage to occur, ubiquitination requires three specialized enzymes called E1, E2, and E3 (Hershko *et al.,* 1983). E1 is an ATP-dependant activation enzyme that charges the ubiquitin peptide at the aforementioned C-terminal glycine. The E2 is a conjugation enzyme, which binds the charged ubiquitin and can either allow

transfer to the E3 or directly ubiquitinate the substrate. Finally, the E3 is a ligase that gives substrate specificity by bringing ubiquitin and substrate together, facilitating the linkage to target lysines. *Arabidopsis thaliana* has two genes encoding E1 enzymes, up to 45 genes for E2 enzymes, and more than 1400 genes encoding E3 ligases (Kraft *et al.,* 2005). Ubiquitin is often attached to target proteins as chains, with the variation of chain length and the specific attachment sites altering the fate of the substrate (Walsh and Sadanandom, 2014). These chains can also be modified through the removal of ubiquitin molecules by DeUBiquitinating enzymes (DUBs); there are approximately 50 genes encoding DUBs in the Arabidopsis genome (Isono and Nagel, 2014). DUBs are also responsible for the processing of poly-ubiquitin chains into mono-ubiquitin and can also remove ubiquitin completely from its substrate (Callis *et al.,* 1990).

In the cell, the enzymes used for the ubiquitination process can be combined in various ways to give high specificity and control over a wide range of target substrates (Vierstra, 2009). Plant lines bred during the Green Revolution, in which yields of cereals such as *Oryza sativa* (rice) and *Zea mays* (maize) improved through the production of dwarf varieties, were in most cases a result of the failure to degrade DELLA proteins. In wheat lines this was caused by the deletion of important sequences in DELLA genes, resulting in truncated proteins. These mutations interfered with a highly conserved region containing a lysine residue, resulting in a failure to ubiquitinate DELLA due to the loss of the phosphorylation site, and hence repression of plant growth (Peng *et al.,* 1999; Pearce *et al.,* 2011). Gibberellin (GA)-mediated degradation of DELLA proteins is controlled by the E3 ligase SLEEPY1, which, with the addition of ubiquitin, causes DELLA proteins to be targeted to the 26S proteasome for degradation (McGinnis *et al.,* 2003). Removing the plants ability to degrade DELLA leads to constitutive growth repression in this same manner; this quality was exploited in crop plants to prevent yield losses caused by lodging, making plants shorter and sturdier and therefore more able to withstand wind and rain. E3 enzymes are emerging as the key

regulators for ubiquitination, providing the specificity to complex systems for fast responses.

1.08 - SUMO

The Small Ubiquitin-like MOdifier (SUMO) is a small polypeptide of approximately 100–115 amino acids. This protein was first identified in *Solanum lycopersicum* (tomato) plants in 1999, and was named due to its similarity to ubiquitin (Hanania *et al*., 1999). The SUMOylation process is described as the act of attaching a SUMO protein to a target substrate. It is a quick response mechanism to modify the behaviour of proteins under stress and is proving to be a major post-translational regulator in plants and other eukaryotic organisms (Gill, 2004; Hay, 2005; Downes and Vierstra, 2005; Vierstra, 2012). Like ubiquitination, SUMOylation is a dynamic process that is altered during biotic or abiotic stress and can change the stability of proteins or interfere in protein–protein interactions (Wilkinson and Henley, 2010). The functional consequences of SUMOylation are diverse and depend on the modified substrate protein. Adding and removing SUMO regulates various aspects of basic cellular processes in stress and other processes such as defence responses, DNA repair, nuclear transport, transcriptional regulation, chromosome segregation, ion channel activity, nitrogen metabolism, phosphate starvation, and regulation of flowering (Hanania *et al.,* 1999; Hoege *et al.,* 2002; Ross *et al.,* 2002; Stade *et al.,* 2002; Hotson *et al.,* 2003; Kurepa *et al.,* 2003; Lois *et al.,* 2003; Murtas *et al.,* 2003; Xia *et al.,* 2004; Gill, 2005; Miura *et al.,* 2005; Nacerddine *et al.,* 2005; Rajan *et al.,* 2005; Colby *et al.,* 2006; Yoo *et al.,* 2006; Catala *et al.,* 2007; Lee *et al.,* 2007; Lyst and Stancheva, 2007; Conti *et al.,* 2008; Park *et al.,* 2010). One example of the wider implications of SUMOylation comes from proteomic studies of Arabidopsis grown under stress conditions. There is evidence to suggest that SUMO conjugation during heat stress can result in controlling alternative splicing and other RNA-based DNA modifications although it hasn't been shown that SUMO binds RNA or DNA directly (Miller *et al.,* 2013). This implies that the

action of SUMOylation can go far beyond simply modifying individual proteins for protection against stress and can change the entire transcriptome, allowing major changes in physiological responses.

Within the Arabidopsis genome, eight SUMO proteins (AtSUMO) were found through bioinformatics approaches. It is believed that some of the AtSUMO isoforms are expressed in specific conditions or at specific times. In addition to Arabidopsis, SUMO families have been identified in many different crop plants such as rice, maize, *Triticum aestivum* (wheat), *Sorghum bicolor* (sorghum), and *Poplulus spp*. (poplar) (Kurepa *et al.,* 2003; Reed *et al.,* 2010; van den Burg *et al.,* 2010). SUMOylation occurs in a series of enzymatic reactions very similar to ubiquitination, which include activation, conjugation, and ligation. The SUMO E1 has two small subunits called SUMO activation enzymes 1a and 1b (SAE1a and SAE1b) and a large subunit (SAE2); this complex requires ATP, and the E2 conjugation enzyme—SUMO conjugation enzyme 1 (SCE1)—acts to bring the SUMO and E3 together (Figure 1.2). E3 enzymes act to attach SUMO to substrates; however, only two SUMO E3 ligases, HIGHPLOIDY2 (HPY2) and SAP & Miz1 (SIZ1), are found in the *Arabidopsis thaliana* genome (Kurepa *et al.,* 2003; Miura *et al.,* 2005; Miura and Hasegawa, 2010). Interestingly, despite the obvious similarity between SUMOylation and ubiquitination, the SUMO system has a strikingly low number of E3s compared to ubiquitination.

The reversible covalent bond formed between the C-terminal glycine residue of SUMO and an amine side chain of the lysine residue of the target protein is catalysed by E3 SUMO ligases. However, conjugation of SUMO to target proteins can also occur without the help of an E3 SUMO ligase, meaning E2 enzymes can also directly facilitate the SUMO-substrate linkage (Wilkinson and Henley, 2010).

The Arabidopsis SUMO E3 SIZ1 has been shown to mediate freezing tolerance, playing a crucial role in phosphate deficiency and functions in defence responses (Miura *et al.,* 2005; Yoo *et al.,* 2006; Miura *et al.,* 2007a; Miura *et al.,* 2007b). The AtSIZ1 homologs in rice (OsSIZ1 and OsSIZ2) are equally as

versatile in their role in stress responses and adaptations. OsSIZ1/2 have a role in the control of heat shock, with an increase in sensitivity to abscisic acid shown in knockout studies. The OsSIZ1 protein can in fact partially complement Arabidopsis siz1 knockout plants, showing great conservation across the monocot-dicot kingdom (Park *et al.,* 2010). The wide range of processes that the SUMO E3s are involved in compared to ubiquitin E3s, coupled with vast difference in numbers of each, implies that there is something other than E3s in the SUMO system driving specificity.

1.09 – De-conjugation by SUMO proteases

The process of removing SUMO from its substrate is called DeSUMOylation. Like SUMO conjugation, DeSUMOylation is a rapid process that makes the SUMOylation cycle reversible. This is achieved through isopeptidase activity of the SUMO protease enzymes such as is seen in the class SENtrin specific Proteases (SENPs). SENPs are cysteine proteases that are response for the activation of pre-SUMO into mature SUMO by cleavage of the di-GLY residue of the SUMO protein (Gareau *et al.,* 2010). The most extensive studies on these proteases are in mammalian systems. In *Homo sapiens* for example, there are 6 SENPs; SENP1-3 and SENP5-7. These proteases are all found in various sub compartments of the nucleus and show some cross over in activity. For example, SENP 1 and 2 can target SUMOs 1, 2 and 3, and are capable of both SUMO maturation and deconjugation of mono-SUMO from its substrate (Figure 1.2) (Kolli *et al.,* 2010). Where as SENP 3 and 5 specifically target SUMOs 2 and 3 but have little activity on SUMO 1, they are also involved in deconjugation of mono-SUMO from the target substrate (Hickey *et al*., 2012). SENP 6 and 7 seem to be specifically targeted to poly-SUMO2 and poly-SUMO3 chain deconjugation, and no activity has been shown in respect to SUMO maturation (Shen *et al.,* 2009 & Bekes *et al.,* 2011).

In addition to this, a separate class of SUMO proteases called DeSUMOylation Isopeptidases (DeSi) have been found to have extremely high specificity in relation to target substrates (Shin *et al.,* 2012). This class of proteases has 2 member proteins in *Homo sapiens* DeSi1 and DeSi2, so far only functionality has been identified for DeSi1 and it appears these enzymes have

deconjugation activity but lack the SUMO maturation capabilities (Figure 1.2) (Shin *et al.,* 2012, Suh *et al.,* 2012). Furthermore a class related the DeSis known as the Ubiquitin-Specific Protease-Like (USPL) protease has been identified in *Homo sapiens*. USPL1 is critical in the cell proliferation stage of the cell cycle and seems to have higher affinity for SUMO2 and SUMO3 then for SUMO1 (Kolli *et al.,* 2012; Schulz *et al.,* 2012). The role of SUMO proteases in plants is discussed in the forthcoming chapter (chapter 3), with an overview of the current literature and investigation of putative DeSi proteases provided.

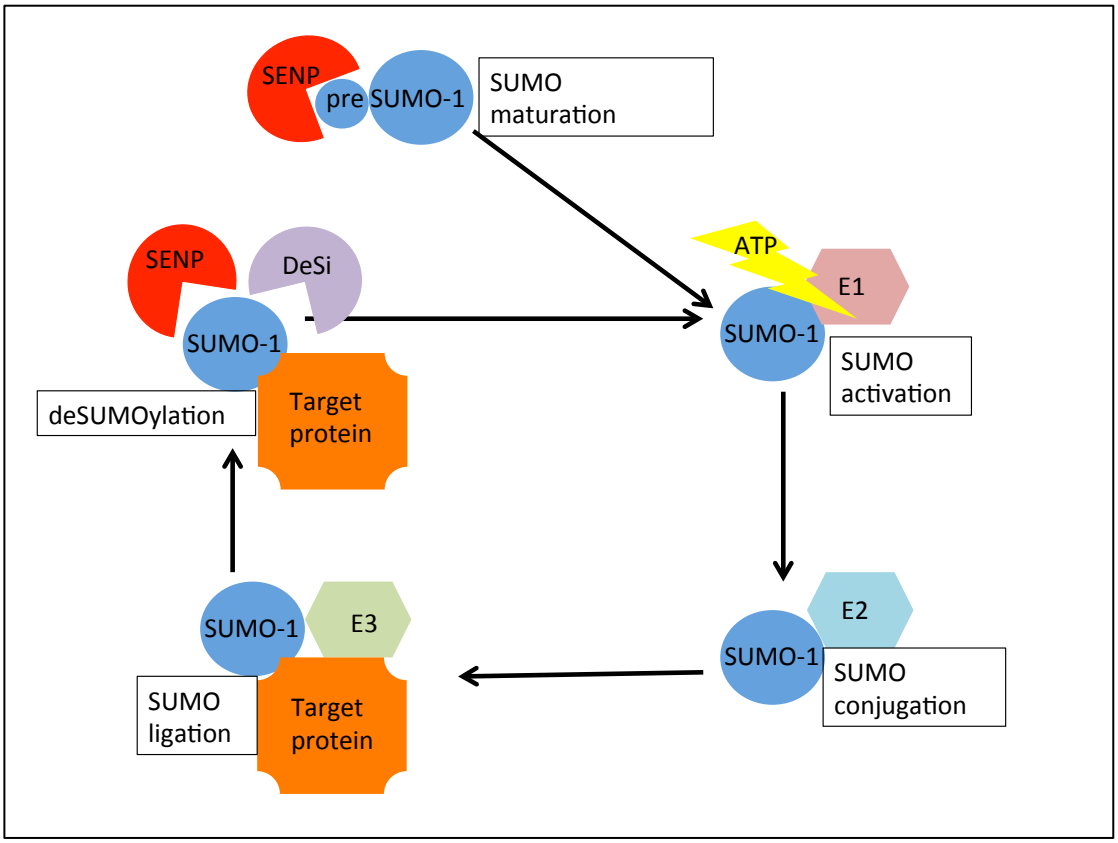

Figure 1.2. Overview of SUMOylation & deSUMOylation system. The SUMO1 protein is produced as an inactive precursor form preSUMO1. Maturation of the preSUMO1 occurs through the cleaving preSUMO1 by SUMO proteases of the class SENP. The mature SUMO then undergoes activation by E1 (a process that requires ATP), conjugation by the E2 and then substrate ligation via interaction with the E3. This covalent attachment of SUMO to the substrate protein can be reversed by SUMO proteases of either SENP or DeSi class. SUMO1 is then recycled for subsequent rounds of SUMO attachment.

1.10 – Thesis objectives

The complexity of the molecular interactions that are at play in eukaryotic cells is enormous. Advancing our understanding of the cellular interplay has caused changes in paradigms and this consistent challenging of prior understanding has lead to new trends being established. This thesis aims to reveal a new subclass of SUMO proteases in plants and underpin the significant roles SUMOylation and deSUMOylation have in biotic stress responses. Furthermore, it exposes greater detail on a relatively new level of molecular regulation: deSUMOylation.

The underlying working hypothesis for this thesis is as follows;

In response to stress, SUMOyaltion attachment is increased on a global scale. DeSUMOylation is a critical process that provides specificity by targeting proteins depending on the type of stress perceived.

If this hypothesis were true, the plants identification of a specific stress would cause changes in the levels of protein and/or mRNA of the corresponding SUMO proteases. If an individual SUMO protease is down regulated in response to a specific stress, for example stress 'A', whilst not to other stresses, it is more likely that the SUMOylated target is involved in the pathway to deal with stress 'A'. Likewise if the same SUMO protease gene is knockout, it may show enhanced response to stress 'A', or display the "stress 'A' " phenotype under normal conditions.

Working under this hypothesis the objectives of this project are described in 3 parts as follows:

The aim of the work was to first identify proteins with DeSi domain regions. Secondly, to order the identified proteins using phylogenetic analysis, this enables better understanding of the DeSi family structure. Thirdly, after bioinformatic and phylogenetic analysis, cloning of the identified DeSis for genetic and protein analysis leading to the validation of SUMO deconjugation activity. This would be the first description of this class of SUMO proteases in plants. The next aim was to ascertain the cellular localisation of the DeSi proteins and compare that with the localisation of previously described SUMO proteases. The cellular localisation information was used to identify a subset

of DeSi proteins for further investigation, these were then used to identify and/or generate genetic knockouts and overexpressing plants for the chosen DeSi genes. In order to analyse any phenotypic changes and to test for resistance or susceptibility to specific stresses, these mutant plants were assayed establishing a stress-response correlation. Further analysis was to reveal the mechanistic actions, providing insight to possible DeSi targets and function. Using the previous results, the next aim was to identify possible DeSi target proteins. Then to use a bioinformatics approach to identify SUMOylation sites on the potential DeSi targets. Test if mutation of the predicted SUMOylation site on the putative DeSi target proteins affects the phenotype and or response to stress and develop further genetic tools to aid analysis of the role of SUMOylation within the identified target(s).

Chapter 2
Materials and Methods

2.1 Materials

The chemicals used for the practical work in this thesis were bought from Sigma-Aldrich, Fisher Scientific, VWR or Melford.

2.1.1 Vectors

pENTR D-TOPO Kan Gateway entry vector (Life Technologies)
pDEST 15 Amp N-terminal GST tag E. coli expression (Life Technologies)
pDEST 17 Amp N-terminal HIS tag E. coli expression Life Technologies
pEarleygate 104 Kan N-terminal YFP tag plant expression (Earley *et al*., 2006)
pEarleygate 201 Kan N-terminal HA tag plant expression (Earley *et al*., 2006)
pET9d (c2corr) Kan N-terminal His-tag protein expression vector (Budhiraja *et al*., 2009; Tomanov *et al.,* 2011)
pET9d kan No tag protein expression vector (Budhiraja *et al*., 2009; Tomanov *et al*., 2011)
pMAL Amp N-Terminal MBP-tag bacterial expression vector (Budhiraja *et al.,* 2009; Tomanov *et al.,* 2011)
pET-Tag3 Kan Strep-HA-His-tag bacterial expression vectpr (Budhiraja *et al.,* 2009; Tomanov *et al.,* 2013)
pQE70 – Amp His protein expression vector (Canonne *et al*., 2011; Colby *et al.,* 2006)

2.1.2 Bacterial strains

Organism	Strain	Resistance
Agrobacterium tumefaciens	GV3101pMP90	Rifampicin and Gentamicin
Escherichia coli	DH5a	n/a
Eschericha coli	BL21-ROSETTA(De3)	Chloramphenicol
Pseudomonas syringae pv.	tomato DC3000	Rifampicin and Kanamycin

2.1.3 Antibodies

Antibody; Host; Working concentration (TBST); Supplier.

anti-HA Rat 1:10,000 Roche

anti-HIS Mouse 1:10,000 Roche

anti-SUMO1 Rabbit 1:5,000 Manufactured inhouse

anti-GFP Rabbit 1:8000 Abcam

anti-GST Rat 1:5000 Abcam

anti-RAT-Hrp 1:20,000 Sigma

anti-Rabbit-Hrp 1:20,000 Sigma

anti-Mouse-Hrp 1:15,000 Sigma

2.1.4 Enzymes

Polymerases MyTaq™ Red Mix - Bioline

Q5® Hot Start High-Fidelity DNA Polymerase - New England BioLabs

SYBR® Green JumpStart™ Taq ReadyMix™ - Sigma-Aldrich

Gateway Life Technologies pENTR D-TOPO - Fisher

Life Technologies Gateway cassette LR clonase II - Fisher

Reverse Transcription Invitrogen SuperScript® II Reverse Transcriptase - Fisher

Invitrogen RNaseOUT™ Recombinant Ribonuclease Inhibitor - Fisher

Restriction Enzymes MluI - New England BioLabs

2.1.5 Antibiotics – working concentration

Ampicillin 100mg/l
Chloramphenicol 34mg/l
Kanamycin 50mg/ml
Rifampicin 12.5mg/ml
Gentamicin 25mg/ml

E. coli and *Argobacterium* were grown in Luria Bertani (LB) medium containing 10g/l bacto-tryptone, 5g/l yeast extract and 10g/l NaCl (pH7.0) with 1.5% v/w agar.

2.2 Methods

2.2.1 Plant Growth

Arabidopsis thaliana seeds were sown in wet Levington F2 plus sand compost or media (see below). Seeds were stratified for three days at 4°C before transferring to Panasonic MLR Plant Growth Chambers. Chamber conditions for long day: 16 hours light at 22°C and 8 hours dark at 20°C. For short day: 10 hours light at 22°C and 14 hours dark at 20°C with a constant humidity of 70%.

Media - 2.15g Murashige & Skoog medium (1/2MS) and 10g of Sigma agar (excluded for liquid media) was added to 1 litre of purified water and autoclaved for 15 minutes at 121°C. Media was then cooled to 50°C and supplements added if needed, before pouring into petri-dishes.

2.2.2 Seed Sterilisation

10-15mg of seeds were place in 1.5ml eppendorf tubes and placed in a air-tight sealable box within a fume hood cabinet. Three millilitres of concentrated Hydrochloric acid was added to a beaker inside the box containing 97ml of hypochlorite. The box was sealed and left overnight (approx. 16 hours). Upon opening the box the beaker was removed and the

box resealed and transported to a sterile laminar flow cabinet for airing (minimum of 2 hours).

2.2.3 Arabidopsis Phenotyping

Root length:

Seeds were sterilised as described above and sown onto plates containing 1/2MS media and placed in dark conditions at a constant temperature of 4°C. After 3 days, the plates containing the seeds were removed and place in long day conditions. After 3 days of growth seedlings were moved to plates containing 1/2MS media with or without an elicitor (flg22 or chitosan). The seedlings were allowed to grow for a further 5, 6 or 7 days in long day conditions before they were photographed with a ruler to normalise the scale. The pictures were then analysed using ImageJ software, which allowed accurate measuring of the roots whilst equilibrating the scale of each picture to the ruler. The measurements were then analysed statistically using a 2-tailed T-TEST, and averages calculated and plotted on a bar chart. Additionally after photographs were taken, several seedlings that represented the mean of each genotype were placed together on a single plate for comparison.

Fresh weight:

For fresh weight measurements the protocol was followed as above (Root length). After the initial photographs were taken, the seedlings were removed and dried on a paper towel before being weighed on a electronic balance. The seedlings were counted and weighed in groups of 10 or 20 individuals and averages calculated afterwards. Statistical analysis was performed using a 2-tailed T-TEST, standard error was calculated and results plotted on a bar graph.

2.3 - DNA/RNA analysis

2.3.1 Taq Polymerase PCR

For 20µl PCR reactions, a final concentration of 1x ReddyMix (containing 0.625 units ThermoPrime Taq DNA polymerase) (Thermo Scientific), was mixed with 0.5µM forward primer, 0.5µM reverse primer and 5ng of template cDNA and run in a heat cycler (TC-3000G Techne machine) under the following standard conditions:

95°C for 5 minutes then 30 cycles consisting of 3 steps:

Step 1. 95°C 15 seconds

Step 2. 30 seconds 45-65°C (annealing temperatures depend on primer Tm)

Step 3. 30 seconds at 72°C (1 minute per 1 kb of expected product)

Finally heated for 5 minutes at 72°C.

2.3.2 Q5 Polymerase Proof-Reading PCR

Fifty microliter PCR reactions were made containing a final concentration of 1x Q5 reaction buffer (NEB, Ipswich, USA), 200µM of dNTPs, 0.5µM of forward primer, 0.5µM of reverse primer, 5ng template cDNA, 1 unit Q5 high-fidelity DNA polymerase (NEB) in sterile distilled water. PCR was performed in a heat cycler (TC-3000G Techne machine) using the following standard conditions:

98°C for 2 minutes, then 30 cycles consisting of 3 steps:

Step 1. 98°C 10 seconds,

Step 2. 45-65°C 15 seconds,

Step 3. 72°C (20 seconds for 1 kb of the expect DNA product)

Followed by 5 minutes at 72°C.

5 µl 10x DNA loading dye was added to PCR products, which were run on an agarose gel.

2.3.3 Colony PCR

Colony PCRs were based on the conditions of Taq polymerase PCR (see above) with one amendment; the template DNA was derived from an individual

colony of bacteria suspended in 20µl sterile distilled water. 1µl of this was used in as the DNA template in each reaction.

2.3.4 Site Directed Mutagenesis

For a 40µl PCR reaction a final concentration of 1X Q5 reaction buffer, 200µM of dNTPs, 0.5µM of forward primer, 0.5µM of reverse primer, 2ng template cDNA Q5 high-fidelity DNA polymerase and 25.8µl of sterile distilled water.
98 °C for 2 minutes, then 20 cycles consisting of 3 steps:
Step 1. 98°C 10 seconds,
Step 2. 45-65°C 15 seconds,
Step 3. 72°C (20 seconds for 1 kb)

For degradation of methlyated DNA template, 1µl of DpnI (NEB) was added directly to the 40µl PCR reaction, the solution was mixed, spun down and incubated at 37°C for 2 hours. The reaction was deactivated by heating at 80°C for 20 minutes and the mutated plasmid was transformed into competent DH5α *E. coli* cells.

2.3.5 RNA Extraction

Frozen leaf tissue was ground extensively with a chilled pestle and mortar under liquid nitrogen. Plant Total RNA Kit (Sigma-Aldrich Spectrum™) was used to extract RNA with no changes from the protocol provided in the kit. RNA concentration was calculated by measuring absorbance at wavelengths of 260 and 280 nm using a NanoDrop™ 1000 Spectrophotometer (Thermo Scientific).

2.3.6 cDNA Synthesis

1-2µg of RNA (normalised to the lowest) was mixed with sterile distilled water to make a final volume of 10µl. 1µl of oligo dT (10mM) (VWR, Radnor, USA) was added to the RNA mixture, which was then heated for 5 minutes at 65°C. 4µl of 5x strand buffer (Invitrogen, Grand Island, USA), 2µl of DTT (Invitrogen), 10mM of dNTPs(VWR) and 1µl of RNase OUT (Invitrogen) was added to the RNA mixture, it was then heated to 42°C for 2 minutes before

adding 1µl of Superscript II (Invitrogen). The mixture was then heated at 42^{0}C for 50 minutes then 70°C for 15 minutes.

2.3.7 RT-PCR

Twenty microliter reactions at a final concentration of 1x SYBR Green JumpStart Taq ReadyMix (Sigma), 0.5µM forward primer, 0.5µM reverse primer, 1µg of template cDNA, 2.5% ROX reference dye in sterile distilled water were run in a StepOnePlus Real-Time PCR System (Applied Biosystems) under the following conditions:

95°C for 20 seconds then 40 cycles consisting of 2 steps:

Step 1. 95°C 3 seconds

Step 2. 30 seconds 55-65°C (annealing temperatures depend on primer Tm)

Then 2 minutes at 72°C

Technical repeats were done in triplicate for each sample and were normalised against a reference gene. Comparisons were performed using the ΔΔCT method (http://www6.appliedbiosystems.com/support/tutorials/pdf/performing_rq_gene_exp_rtpcr.pdf).

2.3.8 Agarose Gel Electrophoresis

Gels were set with between 0.8-1.2% agarose, with higher concentrations needed for better separation of the smaller fragments. 1x TAE buffer (Biorad, West Berkeley, USA) was added to the 0.8-1.2% of agarose (Melford, Ipswich, England), this was then heated in a microwave pending melting of the agarose. For a 100ml gel solution, 0.75µl of ethidium bromide (Fischer Scientific, Waltham, USA) was then added for DNA visualisation. After 30 minutes the gel set and was then submerged in a gel tank containing 1x TAE buffer. Hyperladder (either 50bp or 1Kb depending on size of fragment) (Bioline) was pipetted into the wells as a marker of fragment size and the samples were run at ~100 volts. The gel was then visualized under a UV using a Gene Flash machine and Quantity One program on the computer, highlighting DNA.

2.3.9 Gel extraction/DNA Recovery

DNA fragments were extracted by excising the appropriate band using a scalpel blade. The removed gel slice containing the band of interest was placed inside a pre weighed 1.5ml eppendorf tube. The weight of the gel was then calculated, this was done by weighing the tube plus the gel and subtracting the weight of the tube. The gel extraction was done as per the protocol provided in the QIAquick gel extraction kit (Qiagen, Limburg, Netherlands). In the final step of the elution was done using 30µl of water.

2.3.10 Gateway Cloning

pENTR™ Directional TOPO® Cloning Kits (Invitrogen) was used for inserting DNA fragment into entry plasmid. The pENTR/D-TOPO entry vector (1ug) was mixed with 5µl of gel extracted PCR product (as described above), and 1µl of Salt Solution buffer. The mixture was made up to a volume of 6µl using DNAase free water and then incubated at room temperature for 5 minutes before being moved to an ice bucket. For transformation into competent DH5α *E. coli* cells, 2µl of the mixture was use as described below (2.4.2).

2.3.11 DNA Mini-prep

Ten millilitres cultures of liquid LB media were prepared with appropriate antibiotics: (50µg/ml of kanamycin for DTOPO, pEarleyGate104/201 or 50µg/ml of carbenicillin for pDEST15) and inoculated with *E.coli*. The culture was grown overnight then spun down for 10 minutes, at 5,000rpm and 4°C. The plasmid was purified from the pellet using QIAprepR spin miniprep kit (Qiagen) following manufacturers instructions. The concentration of plasmid was measured using a Nanodrop ND-1000 spectrophotometer.

2.3.12 LR Reaction (Destination vector insertion)

In order to transfer genes of interest into the appropriate destination vector, 50-150ng of pENTR/D-TOPO vector and 0.5µl of destination vector (150ng) were mixed with 1µl of TE buffer (pH 8.0) and 0.5µl of LR clonase II enzyme (Invitrogen), the solution was then spun down by centrifugation. The sample was incubated at room temperature for 1.5 hours. To inhibit the reaction,

Proteinase K solution (Invitrogen) was added and incubated at 37°C for 10 minutes. The solution was then ready for transformation into *E.coli*.

2.4 - Microbiology

2.4.1 Bacterial Growth

E.coli was grown at 37°C overnight on LB media as described above. *Argobacteria* were grown at 28°C overnight or 36 hours (plated culture). *Pseudomonas syringae* were grown at 30°C on King B media for 24 hours.

2.4.2 Bacterial Transformation - Generation of competent cells

Bacteria were streaked onto a LB plate with or with appropriate antibiotics at 28°C for *Argobacteria* and 37°C for *E.coli* for 48 and 24 hours respectively. 10ml of liquid medium was then inoculated with a single colony and incubated at 28°C and 37°C (for *Agrobacteria* and *E.coli* respectively) with shaking for 20 hours. The 10 ml culture was used to inoculate 250ml of LB, which was grown at 18°C and 22°C (for *Agrobacteria* and *E.coli* respectively) with shaking for 18 hours. The liquid culture was cooled on ice for 10 mins and centrifuged at 2000g and 4°C for 12 minutes. The pellet was resuspended in 20ml of precooled TE buffer (10mM Tris pH 8, 1mM EDTA) and incubated for 10 minutes in on ice. The cells were centrifuged at 2000g for 12 minutes and resuspended in 20ml or pre-chilled LB. The cells were measured into 100 µl aliquots and flash frozen in liquid nitrogen.at -80°C for later use. For one transformation experiment, 1µl of the DNA product (plasmid from D-TOPO, LR reactions) was added to 400µl of competent *E.coli* or *Agrobacteria* cells. The cells were then heated for 30 seconds at 42°C (*E.coli*) or 37°C (*Agrobacteria*) before being transferred ice. Cells were incubated at 37°C (*E.coli*) and 28°C (*Agrobacteria*) with 500µl of SOC medium (Super Optimal broth with Catabolite repression). After 1 hour the cells were then spun down and re-suspended in 200µl of fresh LB media, 50µl of the cell solution was spread onto LB agar plates containing the appropriate antibiotic for the selection and grown at 37°C for 16 hours (*E.coli*) or 28°C for 40 hours (*Agrobacteria*).

2.5 – Protein Analysis

2.5.1 SDS PAGE

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was used to analyse protein content. The Min-Protean Tetra Cell system (Bio-Rad, Hercules, USA) was used in all cases with the gels made as follows:

Stacking gels: 5% acrylamide, 0.125M Tris pH 6.8, 0.1% SDS, 0.1% ammonium persulphate and 0.01% TEMED.

Resolving gels: ranged from 10-15% acrylamide, 0.375 M Tris pH 8.8, 0.1% SDS, 0.1% ammonium persulphate and 0.04% TEMED.

4x SDS PAGE loading buffer was used to denature and dilute sample, the buffer contained 40% glycerol, 8% SDS, 200mM Tris pH6.8, 0.05% bromophenol blue, 1% betamercaptoethanol. Electrophoresis was ran at 80-100V with the cassette submerged in 1x Running Buffer: 25mM Tris, 192mM glycine, 0.1% SDS.

2.5.2 Coomassie Blue Staining

After removing the stacking gel, the resolving gel was stained for 40 minutes with shaking in Coomassie Blue Stain (0.25% Brilliant Blue, 50% methanol, 10% glacial acetic acid). After this, the gel was rinsed with water and soaked in Coomassie-destain (10% methanol, 10% glacial acetic acid) for 2-4 hours or until protein bands were clearly visible.

2.5.3 Western/Immuno-Blotting

After SDS-PAGE, the stacking gel was removed and the resolving gel was transferred to a PVDF membrane. Submerging the membrane in 100% methanol for 1 minute ensured activation, the membrane was then transferred to 1x transfer buffer (25mM Tris, 192mM glycine, 10% methanol) for 5 minutes before the transfer apparatus was put together. Sponge and blotting paper were used to sandwich the gel and membrane within a clamp-ready cassette. The cassette was placed in a conductor unit within a gel tank and filled with chilled 1x transfer buffer and run at 30 volts for 16 hours at 4°C. Membranes were removed and blocked with a 5% semi-skimmed milk solution in TBST (50mM Tris pH 7.4, 150mM NaCl, 0.1% TWEEN 20) for one

hour at room temperature. Membranes were then rinsed twice with 1x TBST before adding primary antibodies, primary anti-bodies were suspended in 1x TBST (See Table 2.2 for conditions). After the pre-determined incubation time, the antibody was removed and membrane rinsed with 1x TBST, before undergoing 5x 5 minute washes in TBST on a rocking platform. Once this wash cycle was complete the membranes were incubated with appropriate secondary antibodies for one hour (See table 2.3). Membranes were then subjected to another wash cycle, using the same steps and wash buffer as mentioned above. Not allowing the membrane to dry, ECL solution 1 (2.5mM luminol, 0.4mM p-coumaric acid, 100mM Tris pH 8.5) and solution 2 (0.02% hydrogen peroxide, 100mM Tris pH 8.5) were mixed 1:1 and applied to the membrane by pippetting. The ECL solution was left on the membrane for 1 minute before sealing in a light-proof cassette. In a dark room, the cassette was opened and a Fujifilm X-ray film (Fisher) placed on the membrane. The chemical reaction of the ECL on the HRP antibody caused light, the film was removed after various time periods and exposed film developed with an Xograph Compact 4x Automated Processor (Xograph Imaging Systems).

2.5.4 Protein Expression

Transgenic *E. coli* strains BL21 and Rosetta were grown in 10ml LB cultures containing the appropriate antibiotic (50µg/ml of kanamycin and 25µg/ml of carbenicillin for pDEST15) for 16 hours at 37°C. Expression profiling of recombinant proteins was achieved by testing the optimum conditions using the following procedure;

1ml of the overnight culture was added to a 2-litre flask containing 100ml of LB and appropriate antibiotics. This 100ml culture was placed on a shaker and grown at 37^{0}C until the optical density (O.D.) at 600nm of the culture was between 0.6-0.8. Two 1ml samples were then taken spun down and kept onto ice. 1µM of IPTG (Fischer Scientific) was added to the 100ml culture for induction of transgene expression. At 1, 2 and 3 hours after addition of IPTG, further samples were taken, the volume of which depended on the $O.D._{600}$ measured at the time, so that all samples contained the same the number of bacterial cells as the 1ml pre-induced sample. All samples were then spun

down, pellets collected and supernatant discarded. One of the samples from each time point was processed as total protein extract, whilst the other sample of the same time point was split into the soluble and insoluble fractions. For the total protein extract the pellet was mixed in 60µl of sterile distilled water and 20µl of 4x SDS PAGE loading buffer before heating at 98°C for 3 mins and put back on ice. The insoluble/soluble pellet was mixed in 60µl of BugBuster (Novagen, Billerica, USA)]with 1 tablet of cOmpleteTM Mini EDTA-Free Protease Inhibitor Tablets (Roche, Indianapolis, USA) (BugBuster mix). The re-suspended pellet was spun down for 1 minute at 12,000rpm. The supernatant (soluble fraction) was collected into a new eppendorf tube and 20µl of 4x SDS loading buffer before heating for 3 minutes at 98°C. The insoluble fraction was mixed in 60µl of sterile water and 20µl of 4x SDS loading buffer was added before heating for 3mins at 98°C. For analyse, 10µl of each fraction was loaded onto a SDS-PAGE to analyse the protein content, the proteins were then visualised by comassie staining or immuno-blotting. Once analysis was compete, the optimal conditions were recorded and repeated with a 500ml culture, and the entire culture was harvested at the appropriate time and pellet used for purification.

2.5.5 Purification of Untagged Proteins

The SUMOylation machinery protein SCE was untagged and therefore the HiTrap column for ion exchange chromatography was used for purification. Transgenic bacteria containing SCE cDNA were grown as described above and cells spun down at 5,000rpm for 15 mins. The pellet was then resuspended in 5ml of phosphate buffer, along with 1µg/ml aprotinin , 1µg leupeptin and 10mM DTT, by gently agitating on a rocker at 4°C for 20mins. The resuspended bacteria was then ultra-centrifuged at 75,000 xg for 30mins at 4°C (Beckman Optima, Beckman Coulter, Indianapolis, United States). The supernatant was then subjected to Fast Protein Liquid Chromatography (FPLC) using a HiTrap column (see below).

2.5.6 Purification of Tagged Proteins

All purifications from bacterial cultures were generated from the above protein expression procedure and processed as follows:

The pellet was weighed and BugBuster mix was added at 2.5ml of mix, per 1 gram of pellet. The pellet with the BugBuster mix was incubated at room temperature on a rocking platform for 20 minutes or until the pellet was fully solubilised. The sample was then centrifuged for 15 minutes at 10,000rpm. Discarding the pellet, the supernatant was then removed using a needle and syringe to avoid uptake of debris. The supernatant was then filtered by pressure filtration using a 0.45µM filter, and stored on ice until being applied to one of the following purification protocols.

2.5.7 FPLC

Purification using GSTrap, HISTrap, MBPTrap and HiTrap columns (GE healthcare, Little Chalfont, USA) followed the same protocol but used different buffers. The following describes GSTrap purification, the buffers used for HIStrap and MBPtrap and HiTrap purifications are listed after the protocol.

Proteins with a GST-tag were purified using 1ml GSTrap 4B columns attached to an ÄKTA Avant machine (GE healthcare) and FLPC was performed as follows;

The column was washed with 5ml of purified water at a flow rate of 2ml/min before being equilibrated with 10ml of Binding buffer (140nM NaCl, 2.7mM KCl, 10mM Na2HPO4, 1.8mM KH2PO4, pH 7.4) The extracted cellular supernatant was added to the column at a flow rate of 0.2ml/min and the flow through collected. The column was then washed with 10ml of binding buffer at a flow rate of 1ml/min or until a the flow through became clear (i.e. UV absorbance returns to the basel level as is monitored by the AKTA machine, a sample of the flow through from the wash was also collected. A gradient elution was performed to release the recombinant protein, with an increase of 0-100% in elution buffer (50mM Tris base, 20mM Reduced glutathione, pH 8.0) over 10 mins at a flow rate of 1ml/min. The elution was collected in 10 ml fractions that were then tested for protein content.

<u>HIStrap Buffers:</u>

Binding buffer: 20mM sodium phosphate, 0.5M NaCl, 20–40mM imidazole, pH 7.4.
Eultion buffer: 20mM sodium phosphate, 0.5M NaCl, 1M imidazole, pH 7.4

MBPtrap Buffers:
Binding buffer: 20mM Tris-HCl, 200mM NaCl, 1mM EDTA, 1mM DTT pH 7.4
Elution buffer: 10mM maltose in binding buffer

HiTrap Buffers:
Binding Buffer: 10mM Na-phosphate, 10mM NaCl, 1mM DTT and 20% Glycerol. Final pH 6.5
Elution Buffer: 10mM Na-phosphate, 1M NaCl, 1mM DTT and 20% Glycerol. Final pH 6.5

2.5.8 His-Small Scale Batch

SUMO conjugation chain purification used was performed on His-Bind Resin (Novagen) using the following procedure:
400µl His-Bind slurry (resin in ethanol) was centrifuged at 500g for 1 minute and supernatant was discarded. The resin was then washed twice with 400µl sterile water, each time removing the supernatant between spins. Next the resin was charged by 3 washes with 400µl Charge Buffer (50mM NiSO4) and then equilibrated with 2 washes with 400µl Binding Buffer (5mM imidazole, 0.5M NaCl, 20mM Tris-HCl pH 7.9). 2ml of the processed cellular extract was added and incubated at room temperature with gentle agitation, for 15 minutes. The sample was then spun down as described above and supernatant removed and discarded. The remaining resin was then washed 3 times with 600µl Binding Buffer and 2 times with 600µl Wash Buffer (60mM imidazole, 0.5M NaCl, 20mM Tris-HCl pH 7.9). For protein elution, 500µl of elution buffer (1M imidazole, 0.5M NaCl, 20mM Tris-HCl pH 7.9) was added before centrifuging and collecting the supernatant. This elution step was repeated resulting in 2 samples.

2.5.9 *In Vitro* SUMOylation Assay

In vitro SUMOylation assays were set up in the following proportions, for 250µl reactions; 4µg of SAE, 1µg of SCE, 10µg of SUMO, 5µg of PIAL2M and 5mM ATP (pH 7.5) in 1x SUMO buffer (20mM Tris, 5mM MgCl2 pH 7.4) were mixed in 1.5ml eppendorf tubes. The samples were then kept at 30°C for 1-4 hours. The reaction was stopped by addition of 1x SDS sample buffer or by purification by His small-scale batch method.

2.5.10 *In Vitro* DeSUMOylation Assay

To test SUMO protease activity against SUMO chains, deSUMOylation assays were set up as follows: 10µM purified SUMO isoforms, 5µM of the tested protease and 1mM DTT and 100mM NaCl in 1x SUMO buffer. SUMO conjugation chains were made as previously described until the addition of SDS buffer. To enable deSUMOylation assay to work SUMO chains had to be free of SDS buffer. Therefore experiments were set up to allow, first the SUMO chain conjugation, and then immediately used in the deSUMOylation assays. Working samples were diluted to 10µl of SUMO chains, and 2µl of tested protease in 12µl of SUMO buffer giving a final volume of 24µl. After mixing samples were then kept at 30°C for 3, 6 or 16 hours. The reactions were stopped by addition of 4x SDS sample buffer and heating to 98°C for 3mins.

2.5.11 Confocal Microscopy

Leaf section were cut (approximately 0.5cm^2) and placed on a microscope slide (Fischer Scientific), a droplet of water added before covering the leaf section with a 22x22mm cover slip (Menzel-Glaser, Waltham, USA). The slide was place on the stage of a Leica SP5 confocal laser-scanning microscope (Leica, Berlin, Germany). A 64x objective oil lens was used for viewing and an Ar-Ion gas laser was used to excite YFP at 514nm and GFP at 488nm.

2.6 - Pathogen response

2.6.1 Detection of ROS

Detection of reactive oxygen species was performed on 4 week old plants, where 1cm^2 discs were cut for leaf tissue under sterile conditions. The discs were immediately transferred to a 48 well plate-containing 0.5ml of sterile water. The plate containing the leaf discs were put under gentle agitation and left for 22 hours at room temperature. After this time, water was removed and replaced with Photon Detection Buffer: 20µg/ml of HorseRadish Peroxidase (HRP) (Sigma-Aldrich), 34µg/ml of Luminol (Sigma-Aldrich), or Photon Detection Buffer plus an elicitor (flg22 or chitinosan). The plate was immediately transferred to a dark box which contained a photon counting camera, and light emissions record over 40-90 mins.

2.6.2 Pathotesting

Pseudomonas syringae DC3000 (Pst) were grown on plates containing King's B agar and incubated in the dark for two days at 28°C. After this time, King's B liquid media was inoculated from the plate and placed on a shaker and grown overnight at 28°C. The bacteria was then centrifuged at 4000 rpm for 10mins at room temperature. The pellet was resuspended in sterile water and spun again under the same conditions previously described. An OD_{600} of 0.002 (1X106cfu/ml) was reach through serial dilution of the sample. From each 4 week-old plant, 4 leaves were pressure infiltrated and place in the back in the same growth conditions are before (short day). After 2, 3, 4 or 5 days, 1cm2 leaf discs were cut from infected leafs and crushed with 240ml of sterile water. The sample was then serial diluted along the microtitre plate, taking 40ml of the initial/pervious sample and diluting 1 in 5 each new well. Taking 10µl of each dilution, the samples were droplet spotted on a King's B agar plates and left to air dry for 10mins. They were then transferred to a 25°C room for 36 hours before pictures were taken for colonies to be counted. For each genotype this was repeated on 3 plates for each time point. (Katagiri *et al*., 2002).

Chapter 3

The discovery and validation of a new class of SUMO proteases

3.1 Introduction

In order to understand SUMO protease phylogeny and function, it is necessary to investigate the better-studied SUMO proteases. Ubiquitin-Like-Proteases (ULP) and ULP-like proteins belong to the cysteine protease family of proteolytic enzymes (Mukhopadhyay and Dasso, 2007). This superfamily of cysteine proteases has a highly conserved catalytic motif whilst the functional aspects have diversified greatly. These often-specialized proteases are widespread in eukaryotes and in plants, with many present within a species (Gilles and Hochstrasser, 2012). SUMO proteases act by reversing the SUMOylation process, resulting in the removal of SUMO from its target (Mukhopadhyay and Dasso, 2007). Generally, proteases that are responsible for the removal of SUMO are said to act through a process called deSUMOylation, as opposed to SUMO proteases that cause SUMO maturation, ULPs can do both. DeSUMOylating proteases cleave the isopeptide bond precisely between the terminal glycine of SUMO conjugates and the lysine residue of its cognate substrate, releasing free SUMO from the target protein ready for further conjugation cycles. These proteases play critical roles in maintaining the equilibrium in SUMO signalling (Muller *et al.*, 2001). Analyses of SUMO proteases using gain-of-function and loss-of-function studies have shown involvement in various cellular processes such as hormone signalling, plant defence, abiotic stress, nuclear transport, enzyme activity, cell cycle progression, and plant development, mainly through the regulation of gene expression (Melchior, 2000; Lois *et al.*, 2003; Murtas *et al.*, 2003; Conti *et al.*, 2008, 2014; Budhiraja *et al.*, 2009; Hickey *et al.*, 2012; Bailey *et al.*, 2015; Nelis *et al.*, 2015; Sadanandom *et al.*, 2015). SUMO proteases remain largely understudied, especially in crop plants. ULPs that specifically remove SUMO from target proteins were initially identified in *Saccharomyces cerevisiae* and studies showed they were necessary for cell cycle progression (Li and Hochstrasser, 1999). Since first discovered, various putative

deSUMOylating enzymes have been identified from human and yeast studies (Gong *et al.*, 2000; Li and Hochstrasser, 2000; Johnson, 2004). Recent investigations into the activities of SUMO proteases in plants have contributed greatly towards our understanding of the role of SUMO in plant stress. The seven identified SUMO-specific proteases in *Arabidopsis* are EARLY IN SHORT DAYS 4 (ESD4), ULP1a/ESD4 LIKE SUMO PROTEASE (ELS1), ULP1b, ULP1c/OVERLY TOLERANT TO SALT 2 (OTS2), ULP1d/OTS1, ULP2a, and ULP2b (Kurepa *et al.*, 2003; Colby *et al.*, 2006; Miura *et al.*, 2007a; Miura and Hasegawa, 2010; Hermkes *et al.*, 2011). So far only a few *bona fide* SUMO proteases like ESD4 and OTS1/OTS2 have been characterized using genetic, physiological, and biochemical approaches in *Arabidopsis*, and none have been characterized from crop plants (Reeves *et al.*, 2002; Murtas *et al.*, 2003; Conti *et al.*, 2008, 2014).

A separate class of SUMO protease identified in animal species are characterised by the DeSUMOylating Isopeptidase (DeSi) motif (Gilles and Hochstrasser 2012). This class of SUMO protease is in the cysteine protease superfamily of the clan known as C97, the first and only SUMO protease identified in this clan (Mukhopadhyay and Dasso, 2007). Unlike ULPs, DeSi proteins act only on the removal of SUMO conjugates, and have not been show to catalyse SUMO maturation (Shin *et al.,* 2012, Suh *et al.,* 2012). To date there is limited data available on the function of the DeSi1 and DeSi2 proteins and only 1 substrate has so far been describe, BTB-ZINC FINGER EFFECTOR LYMPHOCYTES (BZEL) (Shin *et al.,* 2012). In mice, BZEL is a transcriptional repressor which is active when SUMOylated, it is also a substrate for DeSi1, making DeSi1 a key regulator of the repressor function (Shin *et al.,* 2012). The aim of this chapter is to identify and biochemically test a group of putative DeSi proteases from *Arabidopsis thaliana.*

3.2. Identification of putative DeSi-type SUMO proteases in plants.

Using bioinformatic software; BLAST (Geer *et al*., 2010), CUSTAL OMEGA (Sievers *et al.,* 2011), MUSCLE (Edgar, 2004) and MEGA6 (Tamura *et al*., 2013), analysis of the *Arabidopsis thaliana* proteome was performed to identify putative new SUMO proteases. Previous publications have shown that within the *Arabidopsis thaliana* proteome there are 7 SUMO proteases of the class described as Ubiquitin-Like Proteases (ULPs) (Kurepa *et al.*, 2003; Colby *et al.*, 2006; Miura *et al.*, 2007a; Miura and Hasegawa, 2010; Hermkes *et al.*, 2011). The ULP SUMO proteases have a characteristic SENP-type catalytic triad; HXnDXnC (Li and Hochstrasser, 2003). The SENP-type motif consists of three amino acids that are structurally in close proximity to each other when the protein is properly folded. However the primary structure shows the sequence contains an unspecific amino acid sequence of undetermined length, which linearly separates each active amino acid of the SENP-type catalytic triad (Figure 3.01). This means that the SENP-type motif is more difficult to identify from the primary structure because the catalytic triad only comes together at the tertiary structure level.

In contrast to the ULPs, the catalytic motif of the DeSi-type protease consists of a Histidine (H), then a indeterminate sequence of amino acids followed by a characteristic NCN triad (Gilles and Hochstrasser 2012). This class of SUMO proteases was first identified in animal systems, and the catalytic motif that was identified was used here to search the *Arabidopsis thaliana* proteome This search revealed 8 proteins containing the HXnNCN motif, which is characteristic of DeSi-type SUMO proteases. Analysis of the primary structure of the 8 proteins was carried out by protein alignment software; MUSCLE. The output from this analysis revealed that these 8 proteins have conserved regions that align with not only with the human DeSi1, but also to each other (Figure 3.02).

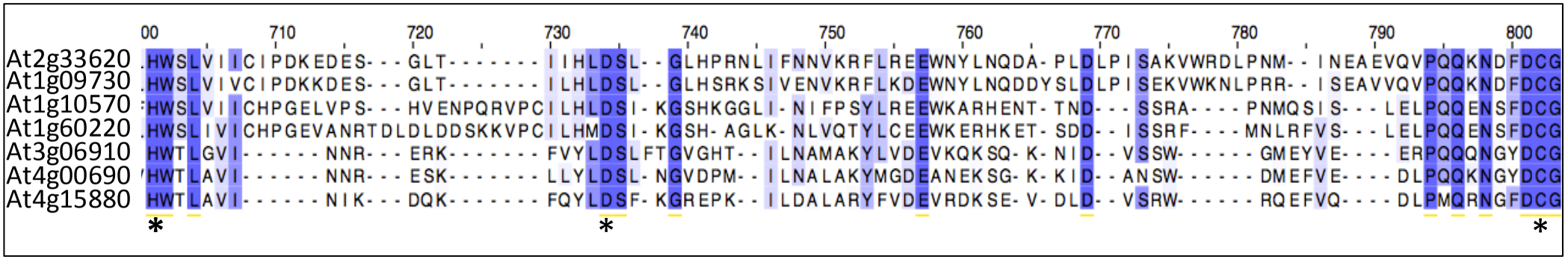

Figure 3.01. Seven *Arabidopsis thaliana* ULP proteins alignment to reveal conserved catalytic motif.
Partial protein alignment showing the amino acid sequence of 7 *Arabidopsis thaliana* proteins ULP proteases. Similarity in sequence is highlighted by the darkened vertical stripes behind the amino acids. Asterisks highlight the 3 amino acids of the catalytic motif. Protein sequences were extracted from the TAIR website and used as the input for the MUSCLE alignment software.

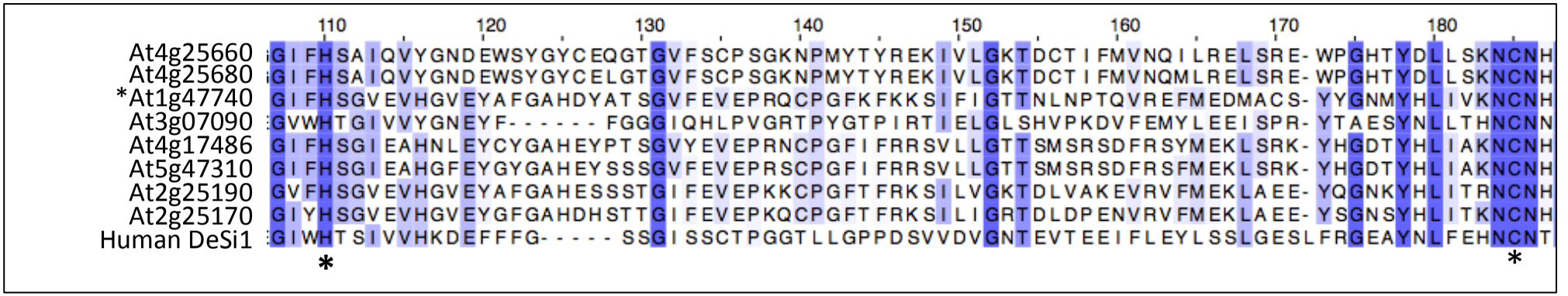

Figure 3.02. Eight *Arabidopsis thaliana* DeSi-like proteins alignment to the Human DeSi 1.

Partial protein alignment showing the amino acid sequence of 8 *Arabidopsis thaliana* proteins and the *Homo sapiens* Desi 1 protease. Similarity in sequence is highlighted by the darken vertical stripes behind the amino acids. Asterisks highlight the Desi catalytic motif; H-NCN. Protein sequences were extracted from the TAIR website and used as the input for the MUSCLE alignment software

To compare the difference within the catalytic regions of ULP and DeSi proteins, a schematic block cartoon of the protein structures is shown in figure 3.03. The primary structure of the proteins is presented diagrammatically highlighting the layout of the key amino acids. From this figure, it is clear that these protein groups are of a different class to each other, the primary function of them may crossover but there is certainly reason to believe they function distinctly based on the catalytic structure

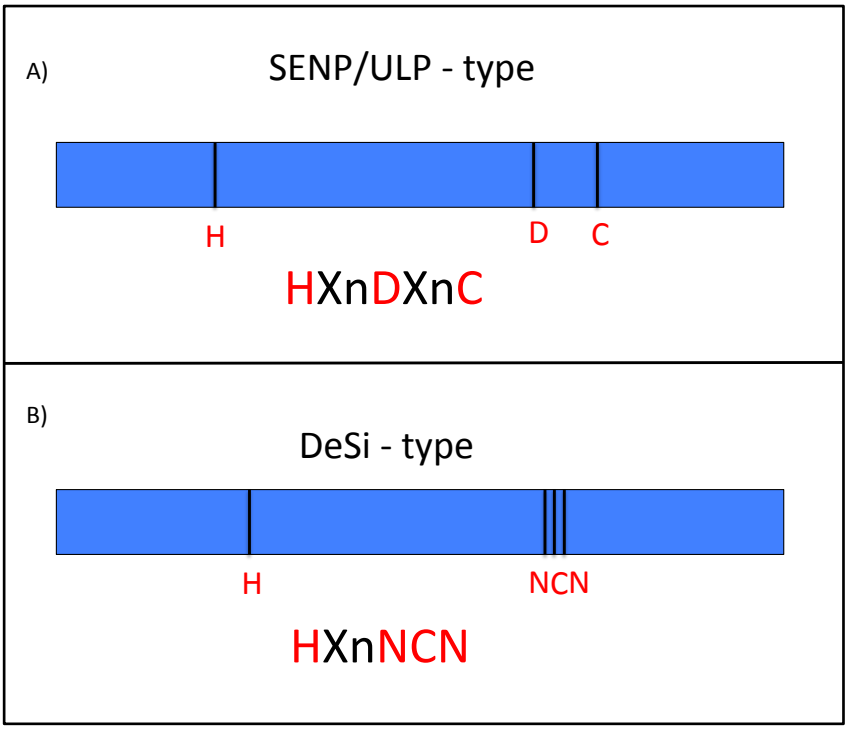

Figure 3.03. Schematic diagram shows two types of SUMO proteases and their characteristic catalytic motif. A) Shows the catalytic triad of the SENP-type SUMO protease. These are represented by a characteristic Histidine - Aspartic Acid - Cysteine (H-D-C) triad, with any amino acids (X) of an undetermined length (n) between each.
B) The DeSi catalytic motif is characterised by Histidine-Asparagine-Cysteine-Asparagine (H-NCN). Here the triad NCN is sequential whereas there is an undetermined amino acid sequence between this and the H.

From the alignment of the 8 proteins identified as having the DeSi-type motif, it is not fully clear what phylogenetic relationship exist between each protein. In order to address this, phylogenetic analysis of the 8 Arabidopsis DeSi proteins was performed using the MEGA6 software package. Based on alignment shown in figure 3.02 a phylogenetic tree was constructed and revealed that these DeSi proteins could be subdivided into 3 groups (Figure 3.04). The phylogenetic tree shows that of the 3 sub-groups, one contains a larger set of proteins than the other two. Sub-group 1 and 2 contain only 1 and 2 proteins respectively compared to 6 proteins in sub-group 3. The tree shows a common ancestral protein branches into 2 lineages, one that led to the DeSi1, and one that led to DeSi 2 and DeSi3. At the 2nd node, DeSi2 and DeSi3 proteins also separate resulting in a group of 5 DeSi3 proteins and 2 DeSi2 proteins. Based on genomic proximity and sequence similarity it is highly likely that both DeSi2 proteins, and *At5g25170* & *At5g25190* are examples of gene duplication.

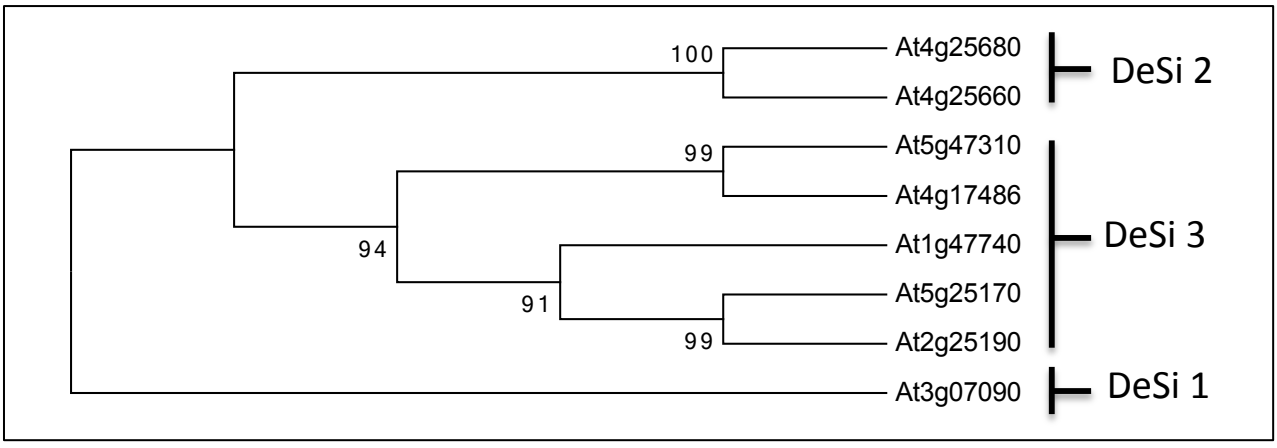

Figure 3.04. Phylogenetic analysis of *Arabidposis thaliana* DeSi proteins reveals 3 distinct groups. DeSi protein sequences were aligned using the software MUSCLE, and phylogeny inferred by neighbour-joining tree without distance corrections using MEGA6. Numbers shown at nodes represent boot strap values based on 1000 replicates. Three distinct groups are revealed by the analysis, with the largest number of proteins classed in the 3rd group, labelled DeSi3. The evolutionary history was inferred by using the Maximum Likelihood method based on the Jones et al. w/freq. model [1]. The tree with the highest log likelihood (-2206.1596) is shown. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 2.5899)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 11.2509% sites). The analysis involved 8 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 33 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 [2]

Previous publications showed that animal systems have only 2 distinct DeSi sub-groups (Nayak & Müller., 2014)., therefore to better understand which animal DeSi sub-group corresponds to the which sub-group from Arabidopsis, phylogenetic analysis was performed using MEGA6 software (Figure 3.05). In order to make the analysis robust, multiple putative DeSis from various plant species, along with the known DeSis from three animal species were used in the input sequence. The putative DeSi homologues from rice, wheat, barley, potato and other plant species (listed in Figure 3.05) were found using BLAST software with the input query being each of the Arabidopsis DeSis in turn. DeSi1 and DeSi2 from human, zebra fish and dog were used as an out-group control. Figure 3.05 shows that 3 distinct groups form from the multi-species input, but only 2 of these groups contain animal DeSi proteins. The proteins that form the groups DeSi1 and DeSI2, are common in both animal and plant kingdoms, but a 3rd group appear to be plant specific, which will be from here on called DeSi3.

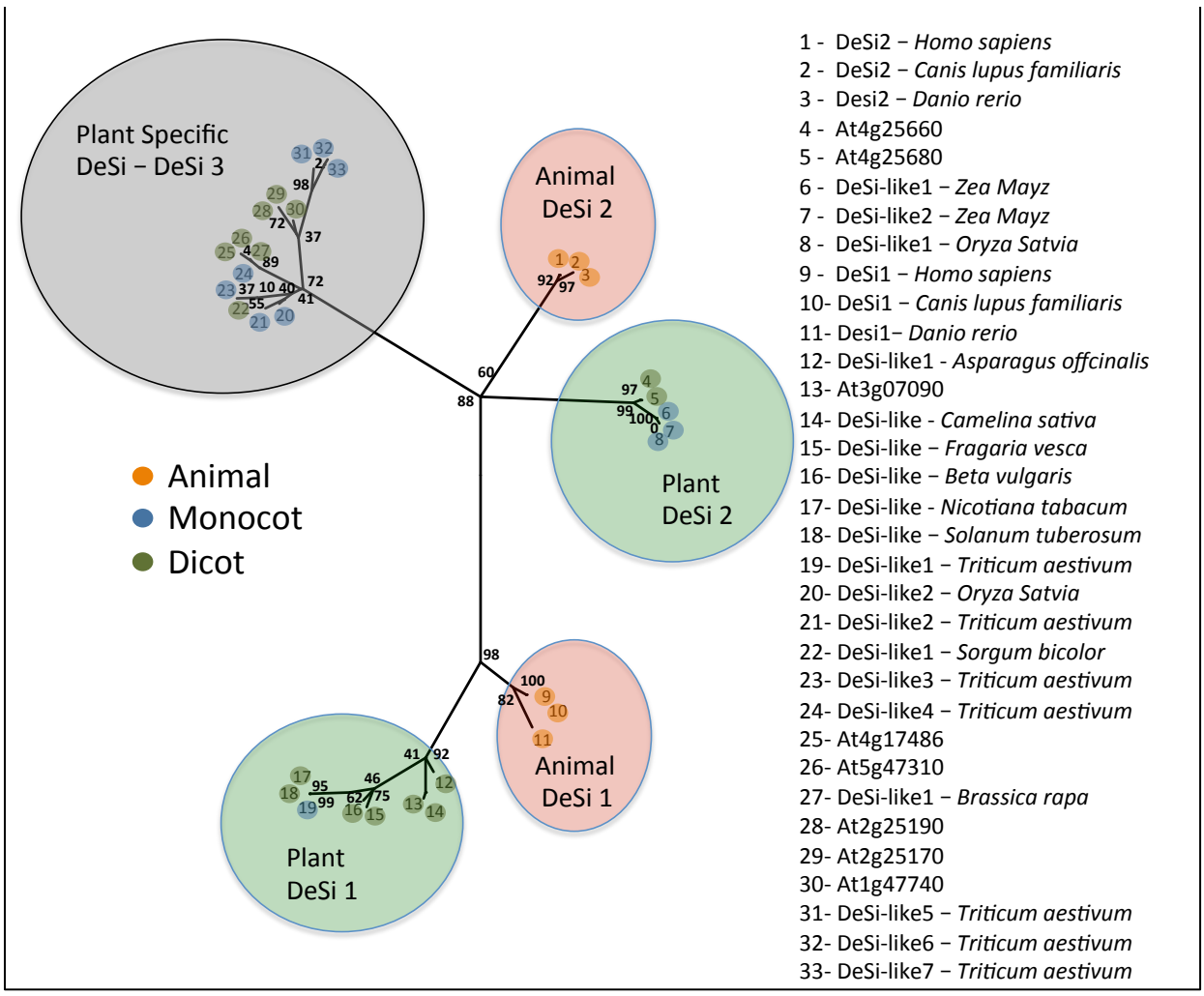

Figure 3.05. Phylogenetic analysis reveals a plant specific group of DeSi proteins. Molecular Phylogenetic analysis by Maximum Likelihood method reveals 3 distinct groups within the plant species. Two of the groups, plant DeSi1 and plant DeSi2 independently pair with animal DeSi1 and DeSi2. The third group represents plant specific DeSis or DeSi3. Coloured numbers at the end of the branches correspond to DeSi proteins and the species explained in the key on the right. The DeSi proteins have been categorised as either animal species (orange circles), monocots (blue circles) or dicots (green circles). Small bold numbers represent boot-strapping values of the tree based on 1000 replicates.
The evolutionary history was inferred by using the Maximum Likelihood method based on the Jones et al. w/freq. model [1]. The tree with the highest log likelihood (-2206.1596) is shown. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 2.5899)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 11.2509% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 33 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 87 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 [2].

To further validate the likelihood that the plant DeSi function like the human DeSi1 (HuDe1), computational 3D modelling using the PYMOL protein structure software, was carried out on the *Arabidopsis thaliana* At1g47740 protein. This was initially achieved using the Phyre2 web portal for protein modelling, prediction and analysis (Kelley et al 2015), and then compared with the crystal structure of HuDe1 in PYMOL (Figure 3.06). Both the proteins have a specific shape to the beta-sheets, which when viewed from the angle shown in figure 5, create a cross over at the centre point with the sheets spiralling in both directions. Interestingly the predicted structure is very similar to HuDe1 in the beta-sheet region but not at the rest of the protein, which equates to the N-terminal. Due to the position of the catalytic triad in relation to the beta-sheet, there is more evidence to support these proteins as being orthologous.

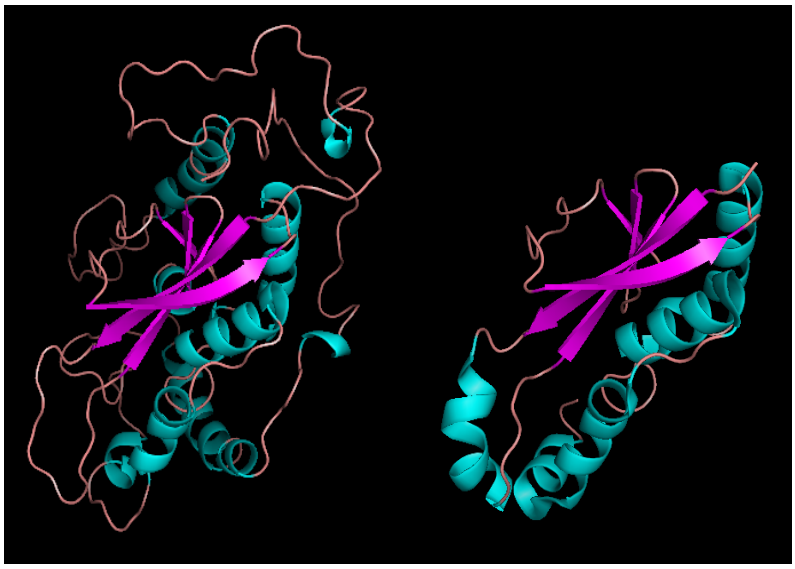

Figure 3.06. Predicted 3D structure of At1g47740 shows striking similarity to Homo sapiens DeSi1 protein. Prediction of 3D structure of At1g47740 was orientated to the angle as HuDe1 using the software PYMOL (Delano 2002). The beta sheet arrangement in the two proteins is nearly identical, with the backbone structure of the alpha helices also really similar until the extended N-terminal that is not present in HuDe1. Predicted structure of At1g47740 was performed using the Phyre2 web portal for protein modelling, prediction and analysis (Kelley et al 2015). Input sequence was acquired from the TAIR website.

3.3. Cloning of genes encoding the *Arabidopsis thaliana* DeSI proteins

For investigation into the function of the 8 Arabidopsis DeSi proteins, one gene from each sub-group was chosen cloned using gene specific primers (table one). Total RNA was extracted from the Arabidopsis Columbia-0 ecotype and cDNA generated using the SuperScript™ II Reverse Transcriptase kit from Invitrogen. This cDNA was used as the template for cloning the open reading frames of the genes, the resulting fragment would contain only the protein-coding region i.e. excluding introns and UTR regions. Table one shows the expected fragment length and primers used for amplification. Once the desired coding DNA fragment was amplified, it was analysed on an agarose gel by size separation electrophoresis for confirmation of the predicted amplicon (Figure 3.07a). Figures 3.07b and 3.07c show a diagram of the full length and the truncated At1g47740 protein respectively, with the truncated version created to contain only the active site motif. This was made as an attempt to increase protease activity by potentially increasing activity against non-specific SUMO substrates. In addition to this, the active site cysteine residue of At1g47740 (Figure 3.0.3B) was mutated to serine as a measure to cease protease activity and validate that this truly is the active cysteine. The mutated At1g47740-C/S protein was used as a comparison throughout the rest of the work.

Gene name	Forward-DNA primer Sequence	Reverse-DNA primer Sequence	Expected fragment Size (bp)
AT4G25680	caccATGACGGAGG TTGTTCTGCA	CTACTGAAACAGTAGAACAT	759
AT3G07090	caccATGGCTGAGG AAGCGCATAA	CTAAGCGTTTACATTGAG	798
AT4G47740	caccATGTTGAACGGAAAAGAAGAGC	CCTTTCTTTCAAGGAGCTGCT	840
AT4G47740-FRAGMENT	caccATGGGTATATTTCACTCTGGTGT	TTGGGATCTTTTTCCCGGTA	293
AT4G47740-C/S	caccATGTTGAACGGAAAAGAAGAGC	CCTTTCTTTCAAGGAGCTGCT	840
AT4G47740-C/S-FRAGMENT	caccATGGGTATATTTCACTCTGGTGT	TTGGGATCTTTTTCCCGGTA	293

Table 1. List of primer sequences used to amplify cDNA gene fragments. The gene names correspond to the gene of interest coding DNA. Forward sequences include the 'cacc' sequenced for pENTRE/DTOPO directional cloning vector kit. Each primer set was homologous to the start and end sequence of the coding regions of each gene, with the exception of At1g47740-Fragment and At1g47740-C/S-Fragment, which had sequences corresponding to the cDNA sequence flanking the DeSi motif.

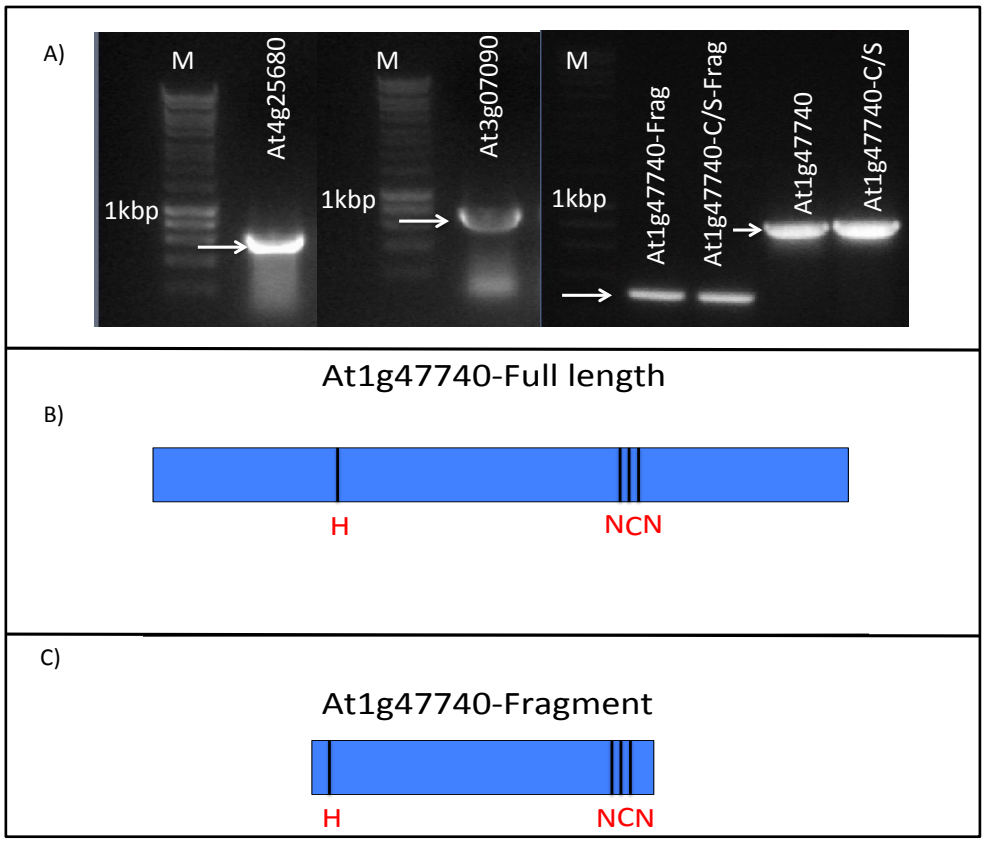

Figure 3.07. Cloning *Arabidopsis thaliana* DeSi cDNA. RNA was extracted from Arabidopsis seedlings and cDNA synthesised using the SuperScript™ II Reverse Transcriptase kit obtained from Invitrogen. The cDNA then formed the template for the amplification of the DeSi coding DNA. After amplification the cDNA fragment was used in a pENTR™/D-TOPO™ Cloning Kit reaction and transformed into *E.coli* cells. The plasmid containing the At1g47740-C/S and At1g47740-C/S-Fragment, first underwent a site directed mutagenesis procedure to create the C-to-S mutant form of the catalytic triad, before being transformed to *E.coli*.
A) shows the resulting cDNA fragment of each gene analysed on an agrose gel by electrophoresis size separation. Arrows point to the predicted cDNA fragment size, the gene names a written above the corresponding band. DNA maker labelled (M), has been labelled at the 1000 base pair band (1kbp). B) and C) show a diagram of the protein structure with the difference being that the fragment contains a truncated version of the protein which is 'trimmed' to a few amino acids either side of the catalytic motif.

After verifying the clones by sequencing and having the desired cDNA fragments now in the entry vector, the transfer of the fragments into suitable destination vectors was enabled via the process of Gateway compatible recombination. This reaction, which was performed using the Gateway™ LR Clonase™ Enzyme mix from Invitrogen, enabled transfer of the cDNA fragment into the preferred destination vector, in this instance, pDEST15. The pDEST15 destination vector has an N terminal GST tag and a selection cassette for ampicillin allowing positive selection on antibiotic containing media. The LR reaction was transformed into DH5alpha cells and grown on media containing ampicillin. The positive colonies were checked by PCR, using various primer sets (Figure 3.08) (Table 2).

Gene name	Forward-DNA primer Sequence	Reverse-DNA primer Sequence	Expected fragment Size (bp)
AT4G25680	ACAAGTTTGTACAAAAAAGCAGGCT	CTACTGAAACAGTAGAACAT	1294
AT3G07090	ACAAGTTTGTACAAAAAAGCAGGCT	CTAAGCGTTTACATTGAG	1321
AT4G47740	ATGTTGAACGGAAAAGAAGAGC	ACCACTTTGTACAAGAAAGCTGGGT	921
AT4G47740-FRAGMENT	ATGGGTATATTTCACTCTGGTGT	ACCACTTTGTACAAGAAAGCTGGGT	395
AT4G47740-C/S	ATGTTGAACGGAAAAGAAGAGC	ACCACTTTGTACAAGAAAGCTGGGT	921
AT4G47740-C/S-FRAGMENT	ATGGGTATATTTCACTCTGGTGT	ACCACTTTGTACAAGAAAGCTGGGT	395

Table 1.2. List of primers used for colony PCR. The gene names correspond to the gene of interest coding DNA. Primer sets consisted of one gene primer and one vector sequence. For At4g25680 and At3g07090, the forward primer sequence was taken from part of the GST sequence of pDEST15 and the reverse primer was homologous to the reverse compliment of the end sequence of each gene. From the other genes, the forward primer sequence was homologous to the start sequence of each gene and the reverse primer was the reverse compliment of the 'AttB1' sequence of the pDEST 15 vector.

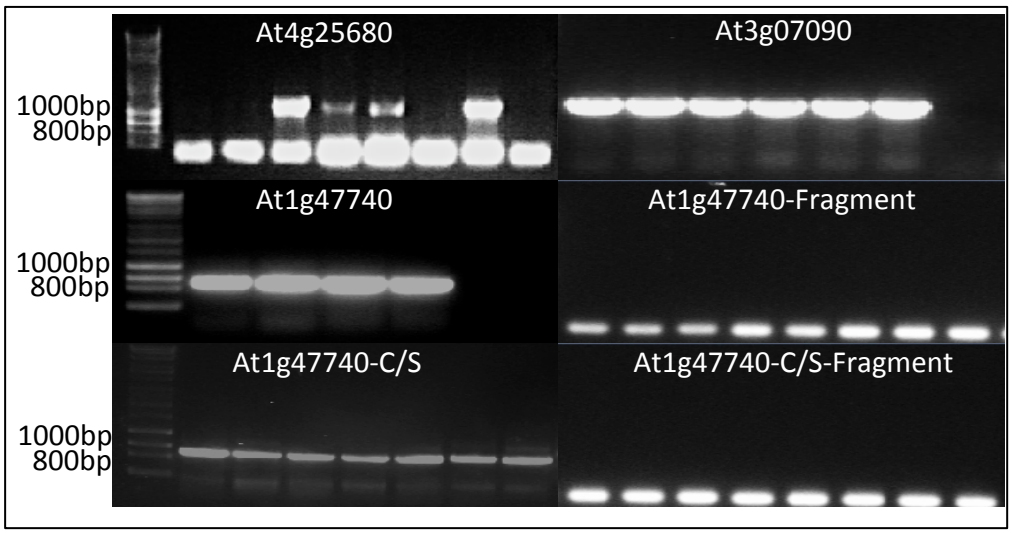

Figure 3.08. *Arabidopsis thaliana* DeSi genes transformed into the bacterial expression vector pDEST15. The figure shows the amplification of DeSi genes from *E.coli* colonies that grow on selection media. In At4g25680 (top left panel) 4 of the 8 colonies tested positive for the desired fragment size. For At3g07090 (top right) all 6 colonies tested positive. The middle left panel shows At1g47740, where 4 of the 6 colonies tested positives. In the other 3 panels shown (At1g47740-Fragment, At1g47740-C/S and At1g47740), all colonies that were sampled tested positive for the desired band size.

Once confirmed, the colonies were grown overnight, spun-down, lysed and plasmid DNA extracted using ZR Plasmid Miniprep™ - Classic from obtained from Zymo. The purified pDEST15 vectors containing the desired gene fragments were then transformed into another *E.coli* strain, Rosetta. This strain contains a vector that has tDNA codons from eukaryotic systems optimised for enhancing expression of eukaryotic genes in bacterial organisms (Novy et al 2001). The Rosetta strain of *E.coli* has a resistance cassette for the antibiotic chloramphenicol, therefore for positive selection, the transformed bacteria was grown on media containing both ampicillin (resistance cassette of pDEST15) and chloramphenicol (resistance cassette of Rosetta). The desired gene fragments were then ready to be tested for optimising expression. Each recombinant *E.coli* containing the genes were expressed in the following conditions:

Grown overnight at 37, a 1:1000 dilution used in a subculture that was grown until O.D.600 reached a reading of 0.6. 1ml samples were taken and processed

as pre-induced controls, IPTG was added at 1uM and culture grown at 28 degrees. 1ml samples were then taken every hour for 3 hours. The samples were then processed into 3 categories; total extract, soluble extract and insoluble extract. These extracts were then prepared for SDS-PAGE analysis, after which the gels were coomassie-stained and protein profiles could be observed on top of a light-box. This process was carried out for the following proteins, At1g47740 At4g25680, At3g07090, At1g47740-C/S, At1g47740-fragment and At1g47740-C/S fragment. As can be seen in Figure 3.09 the accumulation of recombinant protein increased after induction using IPTG, and increased further over time. At1g47740 was found to be present more prominently in the insoluble fractions. However after 3 hours there was an accumulation in the soluble fraction, making optimally harvestable after 3 hours of induction (Figure 3.08 top). At4g25680 was expressed optimally in the soluble fraction of hour 2, as more clearly indicated in the "Tot" lane. By the 3rd hour the protein was being removed from he soluble fraction and accumulating more in the insoluble fraction (Figure 3.08 bottom). The Conditions under which the other recombinant DeSi proteins were best expressed was recorded and can be seen in Table 3.

Gene Name	Optimal Induction Time	Optimal Expression Temperature	IPTG (uM)
AT4G25680	2 Hours	28°	1
AT3G07090	3 hours	28°	1
AT4G47740	3 hours	28°	1
AT4G47740-FRAGMENT	1 hour	28°	1
AT4G47740-C/S	3 hours	28°	1
AT4G47740-C/S-FRAGMENT	1 hour	28°	1

Table 1.3. Optimal expression conditions for recombinant Arabidopsis DeSi proteins. Gene names correspond to the gene of interest protein product. Conditions were tested over 3 time points and results analysed by SDSPAGE. Temperature and IPTG concentrations for optimal expression were constant for all proteins tested.

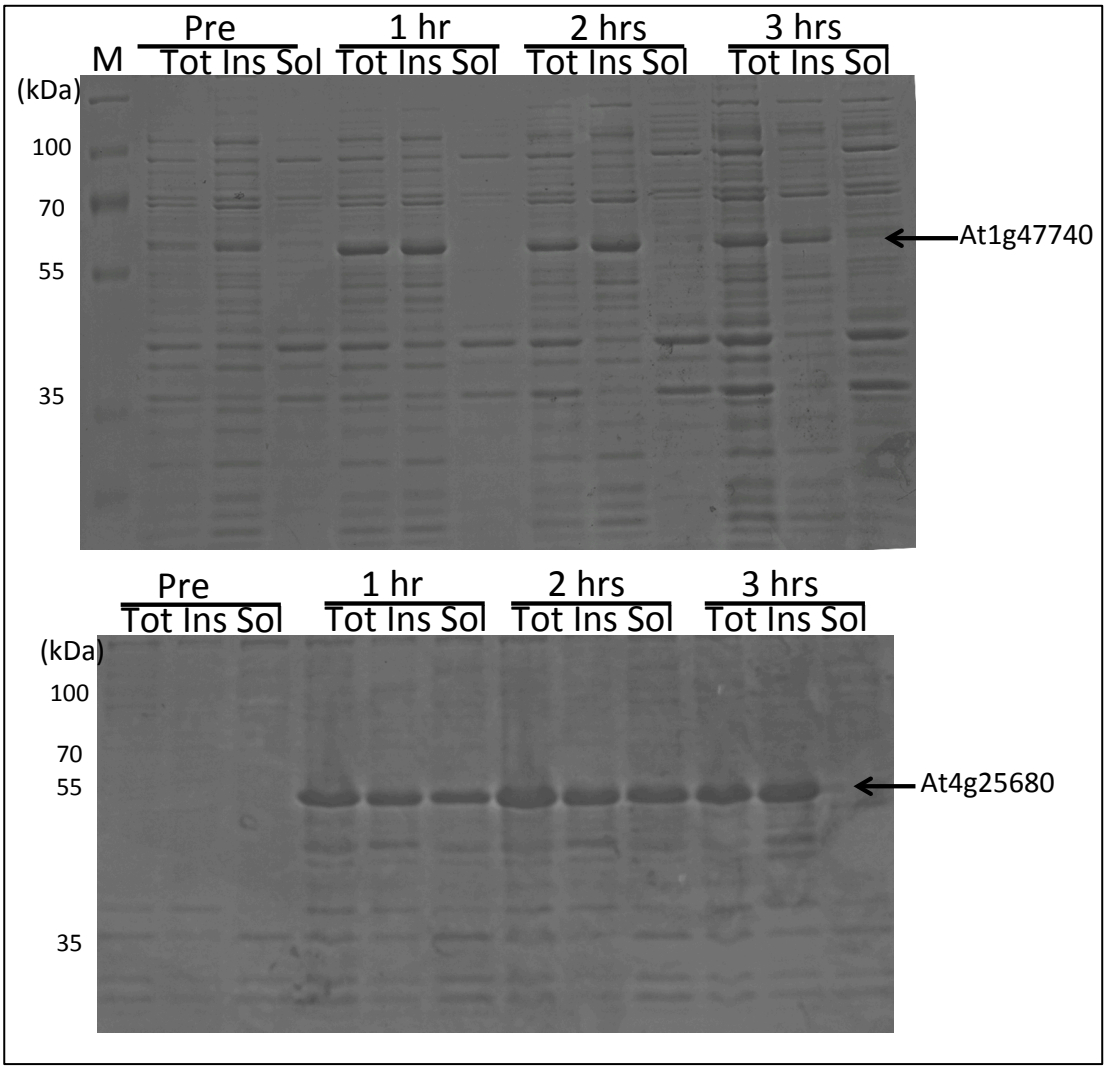

Figure 3.09. Expression of recombinant Arabidopsis DeSi proteins in *E.coli.*
Transgenic bacteria were grown until an OD600 of 0.6, where upon samples were taken and processed as follows: Total extract (Tot), Insoluble fraction (Ins) and Soluble fraction (Sol). Samples taken before IPTG was added are used as pre-induced (Pre) controls. After addition of 1uM of IPTG samples were taken every hour for 3 hours, the samples were then ran on a SDS-page gel. The top panel shows At1g47740 expression, protein indicated by arrow, the pre-induced sample contained some expression, however the protein quantity in the soluble fraction was low. It increased over time and after 3 hours of induction the protein was expressed well enough in the soluble fraction. Bottom panel shows the expression of At4g25680, indicated by arrow. Here there is no expression of the gene in the non-induced phase, after IPTG however, the protein expressed to a high level, increasing over the first 2 hours, and eventually disappearing from the soluble fraction in the 3rd hour.

3.4. Purification of *E.coli* expressed recombinant DeSi proteins

With the expression optima of each protease now known, the proteins were expressed and prepared for purification by lysing spun down cells in a bugbuster/protease inhibitor mix. This allowed the 'burst' cells to be spun-down and the soluble fraction extracted as supernatant. The supernatant was then syringe filtered and flow-through used as the starting material for the purification.

The cell extract was run through a GSTrap column (GE healthcare) attached to an AKTA machine. The column was enriched with the recombinant proteins and then eluted, by a gradating elution-to-binding buffer mixture, resulting in fractions of different protein content. These fractions were then analysed on SDS-PAGE and coomassie stained for confirmation of the purified protein (Figure 3.10 and Figure 3.11). The appropriate fraction was collected, the buffer was exchanged and protein concentrated using concentration columns from Amicon.

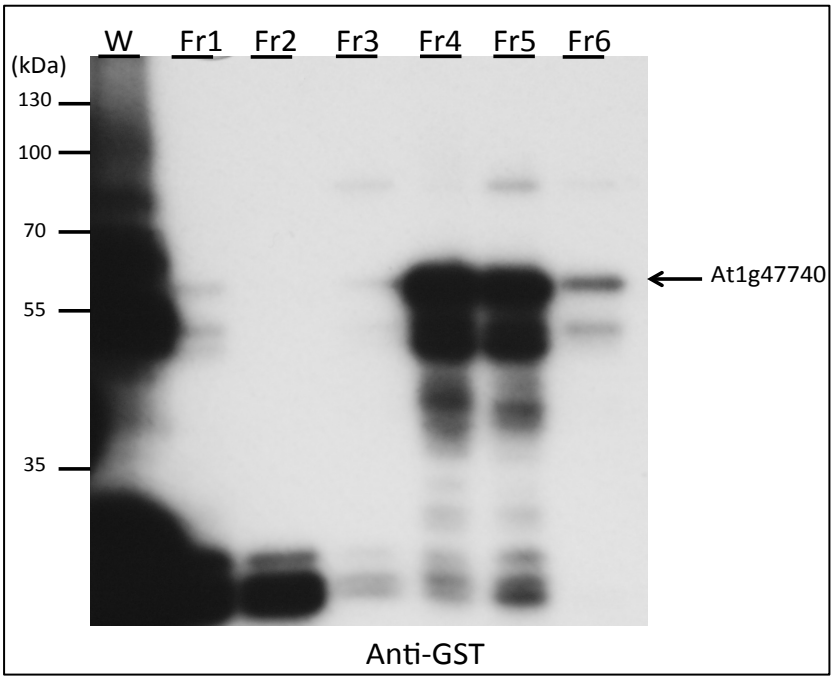

Figure 3.10. Purification of recombinant At1g47740 protein. Transgenic bacteria containing the At1g47740 cDNA fragment was grown at the optimal conditions for protein expression then cells were harvested. The supernatant was extracted and loaded onto a GE healthcare GST trap column. The column was attached to the AKTA machine and washed before fractions were eluted. Samples from all 6 fractions were added to 4x SDS loading buffer and proteins separated by SDS PAGE. Proteins were then transferred to a nitro membrane and probed with Anti-GST. In the first lane of the blot (marked W) the wash was loaded followed by the 6 fractions (Fr1-Fr6). The arrow shows the At1g47740 protein, which can be seen in Fr4, Fr5 and Fr6. Just below the asterisk is break down product from the protein. Fr6 showed the most pure yet abundant protein and so was used for further processing.

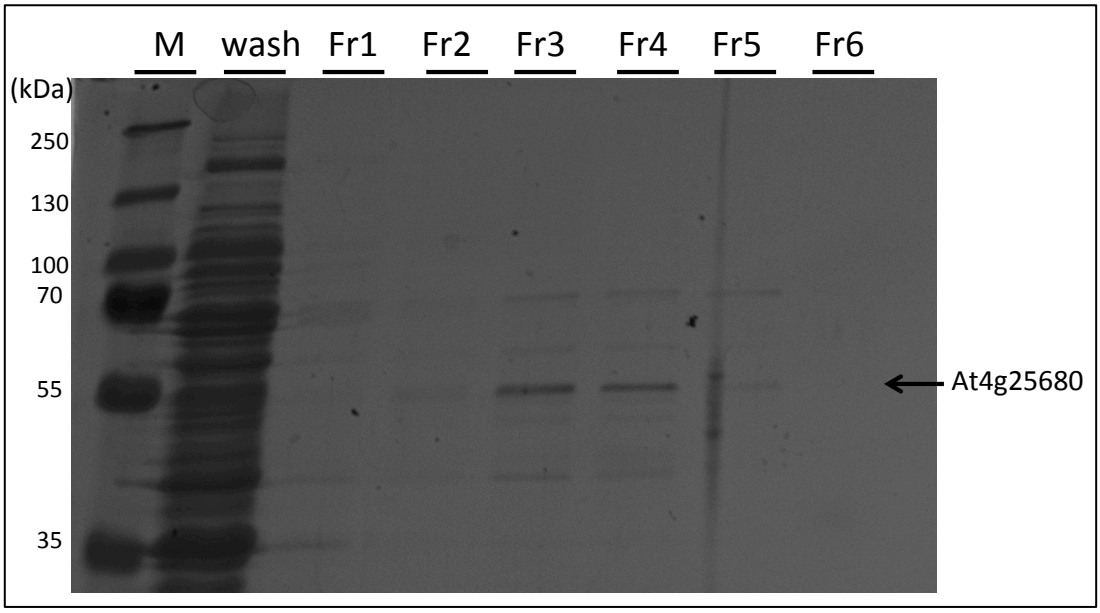

Figure 3.11. Purification recombinant At4g25680 protein. Transgenic bacteria containing the At4g25680 gene was grown at the optimal conditions for protein expression then cells were harvested. The supernatant was extracted and loaded onto a GE healthcare GST trap column. The column was attached to the AKTA machine and washed before fractions were eluted. Samples from all 6 fractions were added to 4x SDS loading buffer and ran on SDS PAGE. Protein was then stain with the protein identifier coomassie. The first lane of the blot (marked M) shows the proteins marker, in the second lane (marked W) the wash was loaded followed by the 6 fractions (Fr1-Fr6). The arrow points to the expected size of the At4g25680 protein, which can be seen in Fr3, Fr4 and Fr5. Although the protein did not come off the column cleanly, as can be seen by the upper and lower bands, Fr4 was used for further processing.

3.5. Purification of *E.coli* expressed *Arabidopsis thaliana* SUMO conjugation machinery

In order to test the activity of the DeSi proteins, poly-SUMO1 substrates were constructed using the *Arabidopsis thaliana* SUMO1 conjugation enzymes; SAE1 (SUMO E1), SCE (SUMO E2), PIAL2 (SUMO E4) and SUM1. Clones were obtained from, and are previously described by Tomanov *et al* 2014.

SUMO E1 is a heterodimeric protein made of 2 subunits SAE1 and SAE2. SAE2 is a small peptide of around 18kDa, it has a 6xHis tag on the C-terminus. SAE1 is imade of 2 smaller subunits, 1a and 1b (35kDa & 36kDa respectively), which form a dimer subunit of around 71kDa. SAE was purified using a HIStrap column (GE healthcare) attached to an AKTA machine. Samples of the fractions were processed and separated by SDSPAGE, the proteins then transferred to a blotting membrane and probed with antibodies raised against His (Figure 3.12).

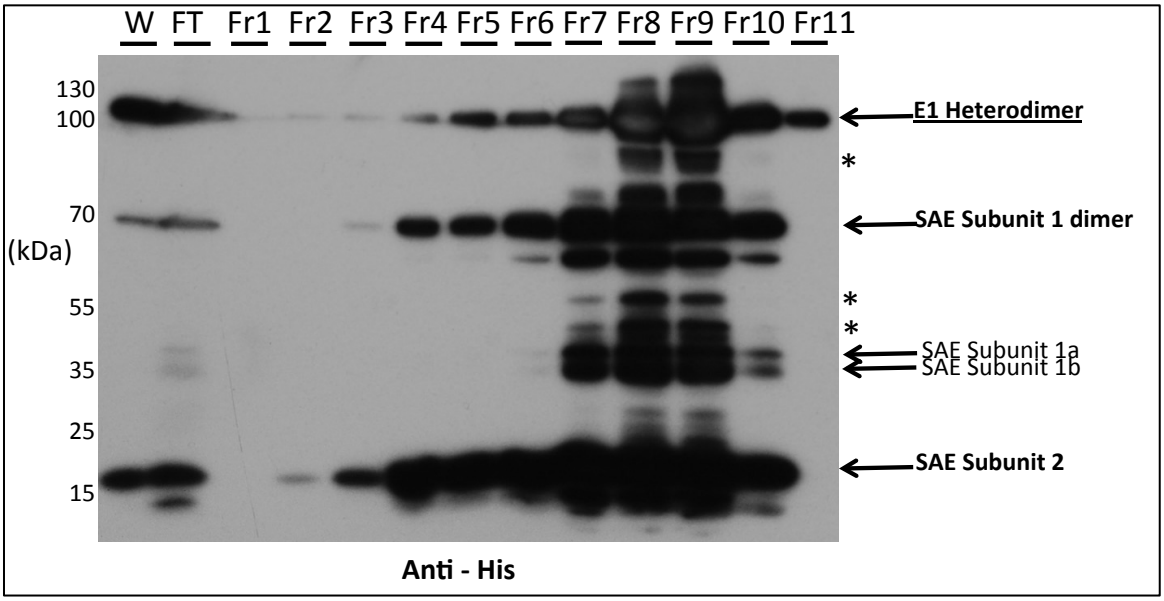

Figure 3.12. Purification of bacterially expressed SAE (E1). Transgenic bacteria containing the SAE subunits was grown and protein expressed, cells were then harvested and subject to FLPC purification. Samples from all 11 fractions, were added to 4x SDS loading buffer and subjected to SDS PAGE and immunoblot analysis using antibodies raised against His. The first lane of the blot (marked W) the wash was loaded, in the second lane the flow through from the extract (marked FT) was added. The 11 fractions were then loaded sequentially. The expected size of each of the components of SAE are indicted with arrows and labelled with names. The complete SAE heterodimer is together around 100kDa, the asterisks on the figure mark partial protein assembly. Fr11 is the only fraction that contains the complete combined subunits as all the rest have other bands or don't have the 100kDa band. Fr11 was used for further processing.

SCE was previously shown to have limited activity when fused with a protein tag. Therefore in order to purify the untagged SCE, ion exchange chromatography was used. A HiTrap column obtained from GE Healthcare was attached to an AKTA machine. Lysate was added to the column and then washed with binding buffer. The elution of was performed using binding buffer with added NaCl in increasing concentrations resulting in a gradient-elution. The resulting eluted fractions were analysed by SDS PAGE and coomassie stained, the results then viewed on top of a light box (Figure 3.13)

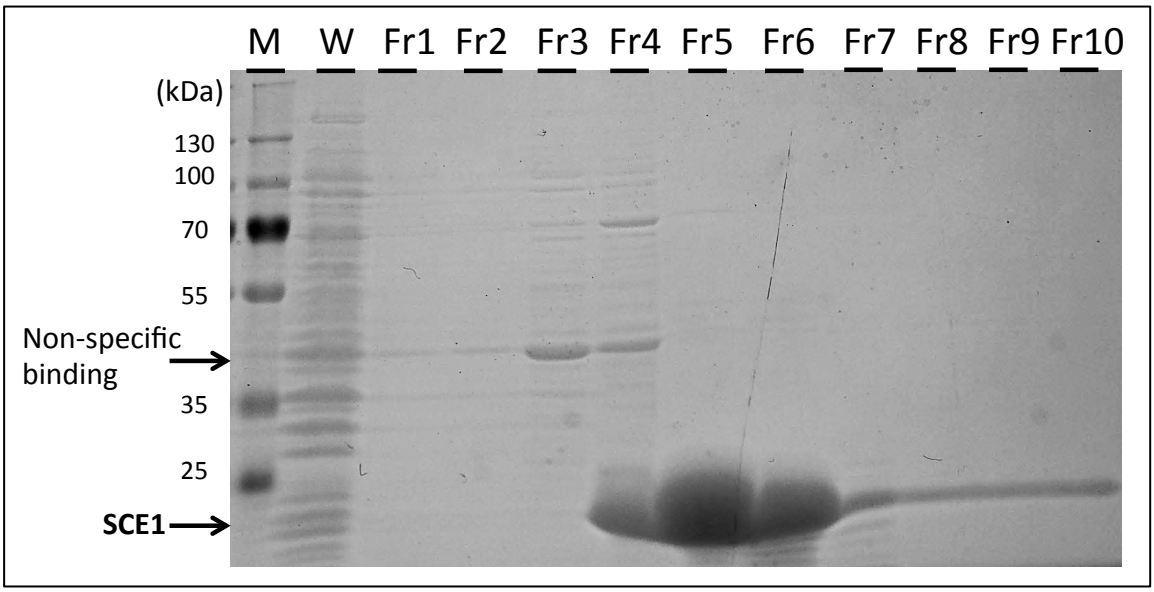

Figure 3.13. Purification of bacterially expressed SCE1. Protein was expressed in *E.coli* cells, spun down and soluble extract harvested. Extract was ran through ion exchange column attached to a AKTA machine and elutes were gathered in 1ml fractions. Samples of the fractions were ran on an SDS PAGE gel and coomassie stained. The expected size of the untagged protein was 17kDa marked by labelled arrow, some non-specific binding can be seen also in Fr3 and Fr5. Fractions 5, 8, 9 and 10 showed clean purifications, where as the other fractions have productions above the arrow (Fractions 1-4) or below (Fractions 6 and 7). Fr5 was used for further processing.

PIAL 2 fragment was expressed with an MBP tag in order to aid protein folding and then purified on a MBP column (GE healthcare). The purified fractions were processed as described above and immunoblot analysed using an antibody raised against MBP. Fraction number 10 produced a clean single band at the expected size of the recombinant protein (Figure 3.14).

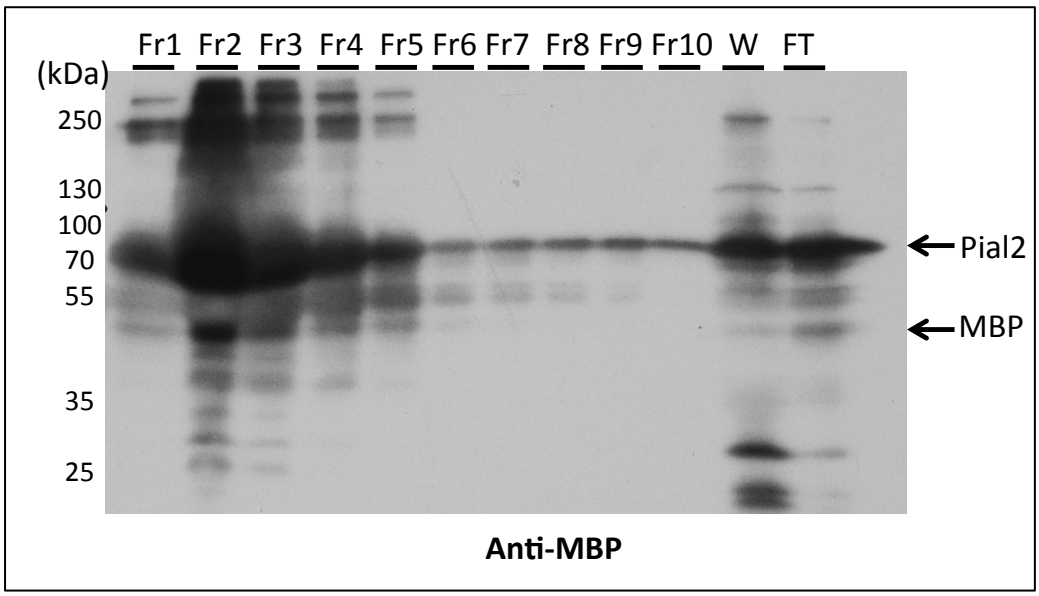

Figure 3.14. Purification of bacterially expressed PIAL2 fragment. Transgenic bacteria containing the Pial2 fragment was grown and protein expressed, cells were then harvested and subject to FLPC purification. Samples from all 10 fractions were added to 4x SDS loading buffer and subjected to SDS PAGE and immunoblot analysis using antibodies against MBP. The 10 fractions were then loaded sequentially left to right, then the wash (marked W) loaded, and finally the flow through from the extract (marked FT) was added. The expected size of the around Pial2-MBP fragment is around 100kDa, the arrow labelled 'Pial2' marks the expected size on the recombinant protein. The arrow labelled 'MBP' show expected size of free MBP protein (42kDa). Fr10 is the only fraction that contains the combined subunits as all the rest have other bands or don't have the 100kDa band. Fr10 was used for further processing.

SUMO1 was expressed in *E.coli* and cell lysate extracted. The recombinant SUMO1 with an attached His tag, was purified using a HIStrap Column (GE healthcare) attached to an AKTA (FLPC). The fractions were then analysed by SDSPAGE and coomassie stained (Figure 3.15). Clean purification of SUMO1 was seen in fractions 9, 10 and 11.

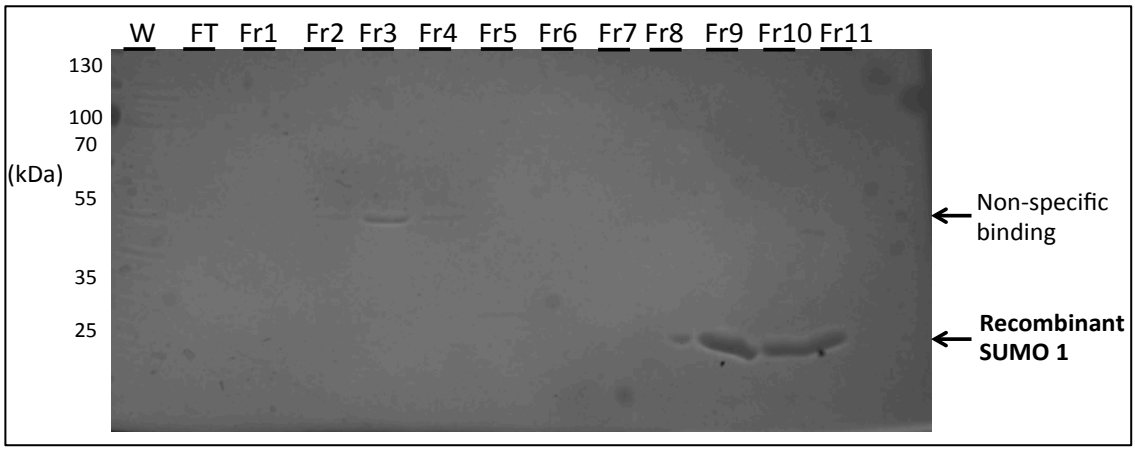

Figure 3.15. Purification of bacterially expressed SUMO1. Transgenic bacteria containing the recombinant SUMO1 transgene was grown and protein was expressed, cells were then harvested and subject to FLPC purification. Samples from all 11 fractions were added to 4x SDS loading buffer and subjected to SDS PAGE and coomassie stained. The first lane of the blot (marked W) the wash was loaded, in the second lane the flow through from the extract (marked FT) was added. The 11 fractions were then loaded sequentially. The expected size of the recombinant SUMO1 is 22kDa, the labelled arrow marks this protein. Other protein bands can be seen in Fr3 and Fr4 indicted by arrow labelled "Non-specific binding'. Fr9 and Fr10 contain the purified recombinant SUMO 1 and were used for processing.

3.6. *In vitro* SUMO conjugation assay

This assay was previously described by Bhudiraja et al. (2009) and was modified in the following ways. 20 µl reactions containing 4 µg SAE, 1 µg SCE, 100 µg SUMO proteins, 5 µg PIAL2M and 5 mM ATP (pH 7.5) were incubated in SUMO buffer at 30 degrees for 1,2 or 4 hours. The reactions were terminated by addition of 4x SDS Laemmli buffer and heated at 98 degrees for 3 minutes. The chains were subjected to SDSPAGE and prepared for immunoblotting using antibodies raised against His (Figure 3.16). The duration of the incubation period had an affect on the amount of conjugates seen. For optimal harvesting of poly-SUMO1 conjugation chains, reactions were incubated for a minimum of 4 hours. After this the reaction were halted as previously described or by addition directly to His-Bind resin (Novogen) for purification. This allowed the poly-SUMO1 chains to be used in further assays without interference from the conjugation enzymes. After purification using the HIS-batch method, the SUMO chains were used as the substrate in the deSUMOylation assays.

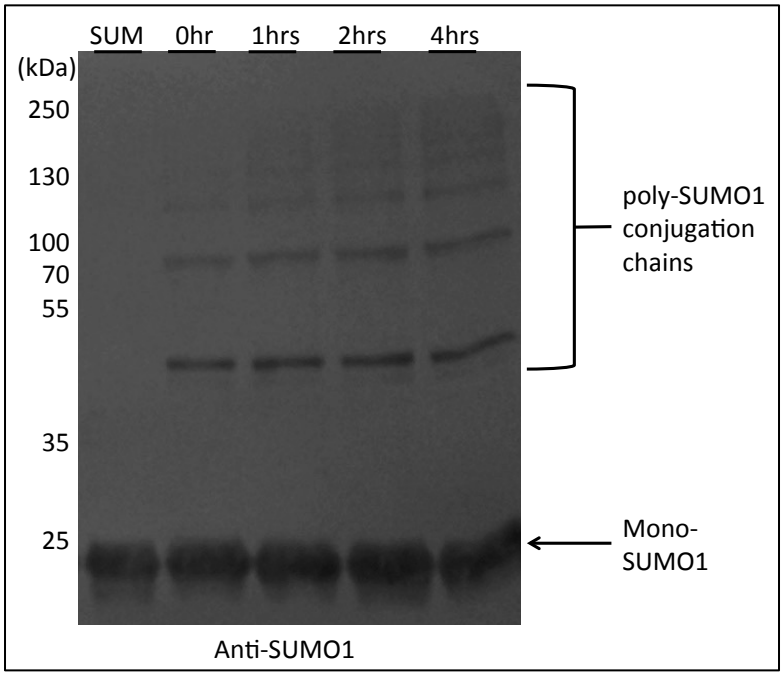

Figure 3.16. *In vitro* poly-SUMO1 conjugation from bacterially expressed Arabidopsis SUMO components. SAE1, SCE1, SUMO1 and PIAL2 purified proteins were mixed with SUMO buffer containing ATP. Samples were then taken at the point of mixing (0hr), 1 hour later (1hr) then 2 and 4 hours later (2hrs, 4hrs respectively). Samples were added to 4x SDS loading buffer and ran on SDS PAGE. The proteins were the transferred to a blotting membrane and probed with anti-Strep. The first lane on the left (SUM) shows mono-SUMO1 only, marked with labelled arrow. The next 4 lanes show an increase in poly-SUMO1 conjugation as a result of time.

3.7. *In vitro* DeSUMOylation.

For testing the SUMO protease enzymatic activity, putative DeSi proteins were incubated with poly-SUMO1 chains. The reaction mixes were set up as follows: 10 µg free SUMO isoforms, 5 µg of the putative protease and 1mM DTT in 1x SUMO buffer. Reactions were incubated for 16 hours at 30 degrees and terminated by the addition of 4x Laemmeli buffer and heated at 98 degrees for 3 minutes. 10ul of the reaction was then analysed by SDSPAGE and immunoblotting using antibodies raised against *Arabidopsis thaliana* SUMO1. The blot showed that there is reduction in the intensity of the poly-SUMO1 chains when DeSi proteins are added (Figure 3.17). The activity of the proteases against the poly-SUMO1 chains was quantified using the ImageJ software which counted the pixels in each lane. The measurement from the lane marked 'No protease' was used to normalise the scale on which the other lanes would be measured. The measurements for each lane were then calculated as a percentage of intensity set against the no protease control

(Figure 3.17C). The effectiveness of the proteolytic activity can be inferred by the percentage of pixel intensity on each lane of the immunoblot. Each of the proteases tested had some level of reduction in pixel intensity, with the exception of At1g477740-C/S, which looked very similar to the no protease control (Figure 3.17A). After quantification however, At1g47740-C/S showed a 9% reduction in pixel intensity (Figure 3.17C). XopD was used as a positive control and completely abolished poly-SUMO1 chains, with an almost 100% reduction in intensity. All 3 full length Arabidopsis DeSi proteins reduced poly-SUMO1 chains by 29-49%. Interestingly both truncated versions of the At1g47740 gene showed increased amount of reduction, the At1g47740-fragment reduced poly-SUMO1 chains by 34% more than the full-length version. At1g47740-C/S-fragment also showed signs of increased activity compare to the full-length At1g47740-C/S with a 38% reduction in intensity. Figure 3.17B shows a repeat of the experiment concentrating on only proteins At1g47740 and At1g47740-C/S, here the purification of the proteases was better and more activity is seen. For transparency in experimental validation the graph shown in Figure 3.17C is made against assays that included all of the tested proteases in one experiment. However, the repeated experiments in which At1g47740 was primarily the focus, more protease activity was observed (Figure 3.17B).

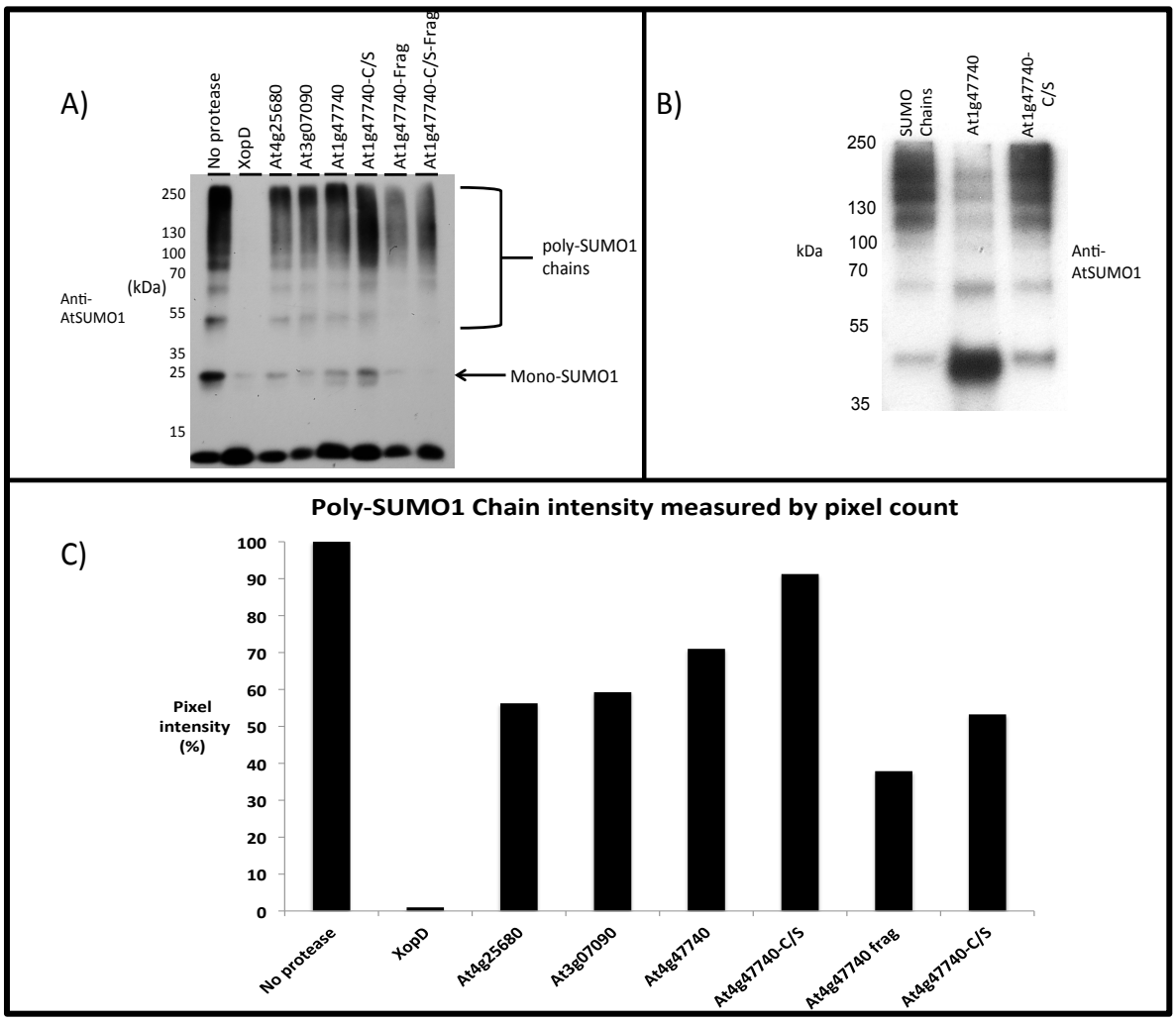

Figure 3.17. In vitro DeSUMOylation of SUMO conjugation chains by Arabidopsis DeSi proteins. Equal amounts of poly-SUMO1 conjugation chains were added to individual 1.5ml tubes, then SUMO protease buffer with or without DeSi proteins added also in equal amounts. The SUMO protease buffer contained (in order left to right); No protease, XopD, At4g25680, At3g07090, At4g47740, At4g47740-C/S, At4g47740 fragment and At4g47740-C/S fragment. A) The first lane to the left (labelled 'No protease') shows unaffected poly-SUMO1 chains, in the second (labelled XopD) poly-SUMO1 chains are completely abolished by the present of the XopD protease. In the next 6 lane poly-SUMO1 chain reduction can be seen in all cases expect At4g47740-C/S, were very little reduction is observed. The last 2 lanes on the right, show the activity of At4g47740 fragments which appear to be more active then the full length counterparts, even in the case of the C/S mutant. B) shows the activity of the At1g47740 and At1g47740-fragment when tested with against SUMO chains in a later repeat of the assay. C) Quantification of poly-SUMO1 chain intensity by way of pixel counting using the image analysis software ImageJ. Each individual lane of the immunoblot in (A) was measured for Grey Scale pixel intensity and compared for calculating the amount of poly-SUMO1 chain present. The lane labelled 'No protease' was used to normalise the scale, and then intensity of the other lanes calculated as a percentage set against this measurement. At4g25680, At3g07090 and At4g47740 had a reduced intensity of 44%, 49% and 29% respectively, whereas the At4g47740-C/S had reduction of 8%. The fragmented DeSis had the furthest reductions with At4g47740-frag causing 63% reduction and At1g47740-C/S-frag causing 47% reduction. The lowest band on the gel shows a SUMO1 breakdown product identified by mass spectrometry (Tomanov *et al.*, 2015).

3.8. *In vitro* removal of SUMO from linked substrate.

In order to check if the SUMO protease activity of the plant DeSi's can cleave the gly-gly motif of the preSUM1 protein, DeSUMOylation assays were setup using the plant protein FLC fused to the SUMO1 protein. The DeSi3 protease At1g47740 was used along with OTS1 as a positive control to check for SUMO maturation activity. The reaction was setup as follows:

10ug FLC-preSUM1 and 5ug protease incubated in SUMO buffer plus 1mMDTT at 30degrees and incubated for 12 hours (Figure 3.18). In the reaction containing OTS1, there was a reduction in the band at the size SUMO-FLC is expected. In contrast to this, the reaction incubated with At1g47740 appeared to show no reduction as did the no protease control (Figure 3.18).

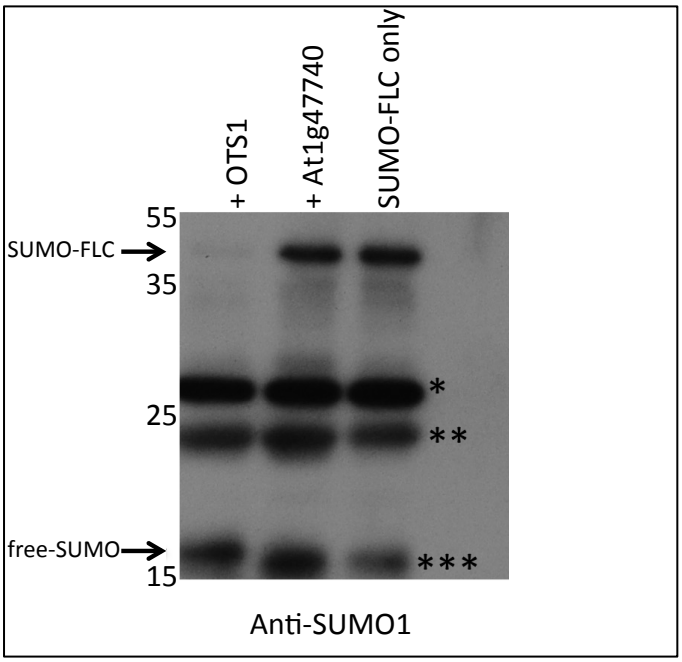

Figure 3.18. DeSi protease At1g47740 does not cleave SUMO from SUMO-FLC substrate. Equal amounts of SUMO-FLC were added to 1.5 ml tubes and incubated for 12 hours with either OTS1, At1g47740 or without a protease. Samples were then analysed by SDSPAGE and immunoblotting using anitbodies raised against SUMO1. The arrow labelled SUMO-FLC indicates the expected size of the substrate, the arrow labelled free-SUMO points to the expected size of the cleaved SUMO. The SUMO-FLC substrate is reduced in the presence of AtOTS1 but not in the presence of the At1g47740. The regions indicated by '*' and '**' show non-specific bands appearing in the lanes with the substrate due to improper purification and/or degradation of SUMO-FLC. The region marked '***' shows untagged SUMO degradation.

3.9. Discussion

There is a fundamental difference in the catalytic motif structure of ULP and DeSi proteins. Although both act by removing SUMO, there is also a functional difference in the way this is catalysed, ULPs are capable of the removal of SUMO1 from its 'preSUMO1' immature form, whereas DeSi proteases are unable to catalyse this reaction. Both can act by removing SUMO from substrates but appears they have different targets (Shin *et al.,* 2012). In the *Arabidopsis thaliana* proteome there are seven proteins identified as having the ULP motif and 8 containing the DeSi motif. Phylogenetic analysis of the 8 DeSi proteases revealed 3 distinct sub-groups with various numbers of proteins in each sub-group (Figure 3.04). Within these sub-groups; One DeSi protease (At4g07090) is classed in the sub-group 'DeSi1', based on its phylogenetic relationship with the DeSi1 paralogues in animal species. Two DeSi proteases (At3g25660 & At3g25680) are in the sub-group 'DeSi2" because of their similarity to DeSi2 paralogues in animal species. Five DeSi proteases (At1g47740, At4g17486, At5g47310, At2g23170 and At2g23190) are classed in the subgroup now known as 'DeSi3'. The DeSi3 sub-group is evolutionarily further removed from the animal DeSis than are the DeSi1 and DeSi2 subgroups. The DeSi3 sub-group has undergone the greatest gene expansion of all 3 sub-groups in Arabidopsis and the implication of this would be that the function of this sub-group had diversified at a greater rate than the DeSi1 or DeSi2 sub-groups. Though not certain, the larger set of proteins in the DeSi3 class could also imply that the target substrates are also of a highly diverse nature, for example, as explored in the next two chapters, the LRR-RK group of proteins could be an example of a family of related but diversified proteins which are targeted by individual DeSi3 proteins.

Cloning of the coding regions of a DeSi gene from each sub-class was performed to provide molecular tools for investigating the function of their protein products. Generation of a Cys-to-Ser mutant of At1g47740 allowed investigation of the catalytic function of the DeSi motif, as did the generation of the At1g47740 fragments (At14g47740-Fragment and At1g47740-C/S-

Fragment), which contained only the DeSi catalytic motif region of the proteins. After expressing these recombinant DeSi genes, *in vitro* deSUMOylation assays were set up to ascertain if the putative DeSi proteins could remove SUMO1 from poly-SUMO1 conjugation chains (Figure 3.17). The assay showed that each DeSi tested was capable of reducing poly-SUMO1 chains by a significant amount compared to the no protease control. The only exception is the At1g47740-C/S mutant protein, which showed only a 9% decrease in poly-SUMO1 chains. This was expected as the catalytic cysteine may contribute more to the activity if the protein than in other DeSi type proteases. Interestingly none of the tested DeSi proteases were able to completely abolish the poly-SUMO1 chains, unlike the positive control XopD. The At1g47740 fragments showed a higher reduction in poly-SUMO1 chains then any of the full length DeSis, but still did not match the efficiency of XopD. When repeating the experiments and focusing only on the protease activity of At1g47740, it was possible to enhance SUMO chain reduction. This was possibly due to the fact that the expression and purification steps were more stringent when less proteins were being produced (Figure 3.17B), as opposed to simultaneous testing of all proteases as is shown in Figure 3.17A.

It was previously shown that the DeSi proteins from mice could not cleave immature SUMO1, unlike the ULPs. Therefore in order to test for this type of catalytic activity in Arabidopsis DeSis, At1g47740 was tested for its ability to remove SUMO1 from an SUMO-FLC substrate. The ULP protease OTS1 was used as a control and incubated with the SUMO-FLC substrate, as was At1g47740. After a 12 hour incubation there was no notable difference in the SUMO-FLC amount when incubated with At1g47740, as opposed to OTS1, which completely cleaved SUMO1 off the substrate.

Collectively this data confirms the presence of a family of DeSi proteins in *Arabidopsis thaliana*, and shows that like the animal paralogues, these are capable of deSUMOylation but not maturation of SUMO1.

Chapter 4
Desi3a is a Negative Regulator of Plant Defence

4.1. Introduction

The results described in chapter 3 indicate that the *Arabidopsis thaliana* proteome contains a family of previously undiscovered SUMO proteases (DeSi proteins). Furthermore, one sub-group of this family of DeSi proteins is specific to plants. Based on this data, it was decided that a protein from each DeSi subgroup was to be investigated for potential differences in cellular localisation (Figure 4.01). The results of this experiment lead to the conclusion that the At1g47740 protein was interesting for several reasons and worth further investigation. This was due to its apparent dual membranous localisation (Figure 4.01) and the fact it is a plant specific DeSi (DeSi3) with no obvious/direct paralogs within the *Arabidopsis thaliana* proteome. As the At1g47740 protein is the only protein found at the end of its branch shown on the phylogenetic tree in Figure 3.04, the confidence in a single gene knockout revealing a phenotype was high. This is due to the reduced chance of redundancy observed in non-duplicated genes compared to genes that are obviously the result of duplication events, as is the case for the exocyst complex component gene *SEC10* (*At5g12370*) (Vukašinović, *et al.,* 2014) and most likely also the case for the Arabidopsis DeSi2 genes *At4g25660* and *At4g25680*.

The pathogen-host paradigm, which is often described as an evolutionary arms race, has revealed that over relatively short time scales both pathogen and host can adapt to overcome their respective adversary (Anderson *et al.,* 2010). This strong driver for genetic change lies on top of an already very complex and evolutionary old immunity vs pathogen interplay. The role of SUMOylation in this interplay however is only just beginning to be uncovered (Castaño-Miquel, *et al.,* 2017). This chapter contains evidence that the SUMOylation process is integral to the plants defence mechanisms and in particular, in the plants perception of pathogens. At1g47740 appears to play a

role in the negative regulation of plant defence, and is involved in the signalling resulting from pathogen detection.

4.2.1 DeSi cellular localisation

For further investigation of the role of the DeSi proteases YFP fusion proteins were made for tracking cellular localisation. A protein from the sub-group DeSi1 and DeSi3 along with both DeSi2 proteins (due to the predicted gene duplication event resulting in both DeSi2 proteins) were chosen and the cDNA gene fragment was cloned into the YFP tagged plant expression vector series pEARLYGATE (pEG) (Earley *et al.,* 2006). The pEG vectors control gene of interest expression via the CaMV35S, promoter and has an N-terminal YFP tag. These vectors were used to transform the bacterium *Agrobacterium tumefaciens* GV3101 cells; this strain of bacteria is used to transiently express proteins of interest when infiltrated into *Nicotiana benthamiana* leaves (Vijn & Govers 2003). 3 days after syringe infiltration, 1cm squared sections of the leaves were carefully cut and placed under a microscope slide for viewing in a fluorescence microscope (Zeiss SP5). An empty vector expressing YFP only was used as a control to show the tag was not guiding the sub-cellular localisation of the proteases.

The foremost observation was that unlike all currently known SUMO proteases, the DeSis do not localise only to the nucleus of the cell. As can be seen in Figure 4.01 the ULP2a control is concentrated to the nucleus whereas the At4g25680 and At1g47740 are absent from the nucleus. Proteases At4g25660 and At3g07090 can see found in both the nucleus and in speckles along the plasma membrane. Furthermore At1g47740 seems to be bound to the plasma membranes. This experiment lacked the necessary cellular markers to validate the compartmentalisation of the proteases, but it gives some indication of the approximant areas within the cell that these proteins are found.

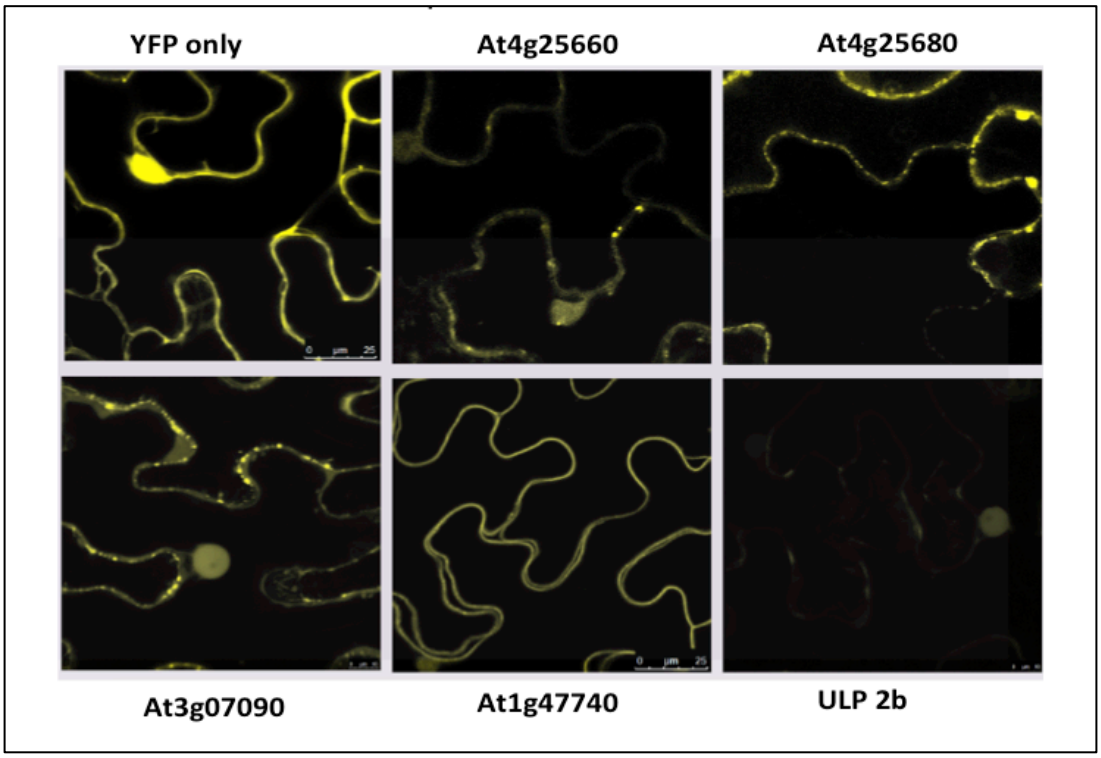

Figure 4.01. YFP tagged DeSi proteins localise to non-nuclear compartments when expressed in *Nicotiana benthanmiana* leaves.
Agrobacteria containing the coding regions of the protein of interest was syringe infiltrated in *N. benthanmiana* leaves. After 3 days 1cm squared sections of leaf were dissected, mounted on microscope slides and viewed by confocal microscope. The panels in the figure are labelled with gene names corresponding to the protein of interest. The 'YPF only' panel (top left) shows the localisation of the YFP protein from untransformed pEG104 vector. YFP-ULP2b represents the localisation of the ULP class of SUMO proteases and is found in the nucleus. All the DeSi proteins tested can be seen in abundance outside the nucleus, with YFP-At1g47740 and YFP-At4g25680 showing no nuclear localisation, where as YFP-At3g07090 an YFP-At4g25660 could be seen in both the nucleus and in foci near the cellular membrane.

4.2.2 Identification of *At1g47740* mutant plants

In order to ascertain the relationship between the Arabidopsis DeSi proteins and their function in the plant, transgenic seeds, harbouring T-DNA inserts for *At1g47740* gene knockouts (Figure 4.02), were attained from the Arabidopsis Seed Bank (Scholl *et al.,* 2000). These seeds were grown and genotyped for confirmation of gene interruption.

Primer sets for the T-DNA and for the gene were used in combination in PCR reactions using the genomic DNA extracted from the seedling as the template DNA (table1). Testing for both the T-DNA and the gene enables analysis of whether the tested plant is a homozygous or heterozygous for the T-DNA insert.

The diagram shown in figure 4.02 represents the genomic location of the T-DNA insert of the transgenic seeds. Using PCR reactions, heterozygote and wild-type plants could be distinguished from the homozygote plants, which are the preferred genotype for further experiments (Figure 4.02b).

In addition to isolating plants lines with *At1g47740* loss-of-function, plants were generated that over-expressed the DeSi protein via the 35S promoter in the col-0 background. This was done using the floral dip method and transgenes were prepared in the pEG vector series with the protein of interest fused with a GFP tag and transformed into *Agrobacteria*. The plants were then grown to maturation and seeds collected. The transgenic seeds were resistant to the antibiotic BASTA and therefore positive selection was achieved by sowing seeds on soil watered with 1:1000 BASTA to water mix. Seedlings which grew normally on this selection were allowed to mature and seeds collected for segregation assays. Plants that produced transgenic seedlings at a 1:3 (1:2:1) ratio were assayed by growth on BASTA media and the positive seedlings allowed to grow until maturity. After selection, 2 over-expressing lines, GFP tagged *At1g47740*-OX1 and GFP tagged *At1g47740*-OX2, were used with col-0 and *At1g47740*-KO, for checking the transcript levels of *At1g47740*. After RNA extraction and cDNA synthesis as described in 3.2.1, PCR reactions were carried out using primers matching the *At1g47740* coding region, shown in table 1. The results shown in Figure 4.03, show that along with col-0,

At1g47740-OX1 and *At1g47740*-OX2 produce a DNA fragment at the estimated size of the desired gene fragment. In addition, it also shows that no transcript is present in the knockout line (Figure 4.03).

Gene Name	Foward primer	Reverse Primer	Expected fragment size
At1g47740	ATGTTGAACGG AAAAGAAGAGC	CCTTTCTTTCAA GAGCTGCT	840
T-DNA	GTTCCGAAATC GGCAAAAT	ATTTTGCCGATT TCGGAAC	290

Table 4.1. Primer sequences for genotyping At1g47740 knockout plants.
The sequences of the DNA primers used for identification of the At1g47740 gene and the TDNA insertion cassette. The expected DNA fragment size is indicated on the right hand column.

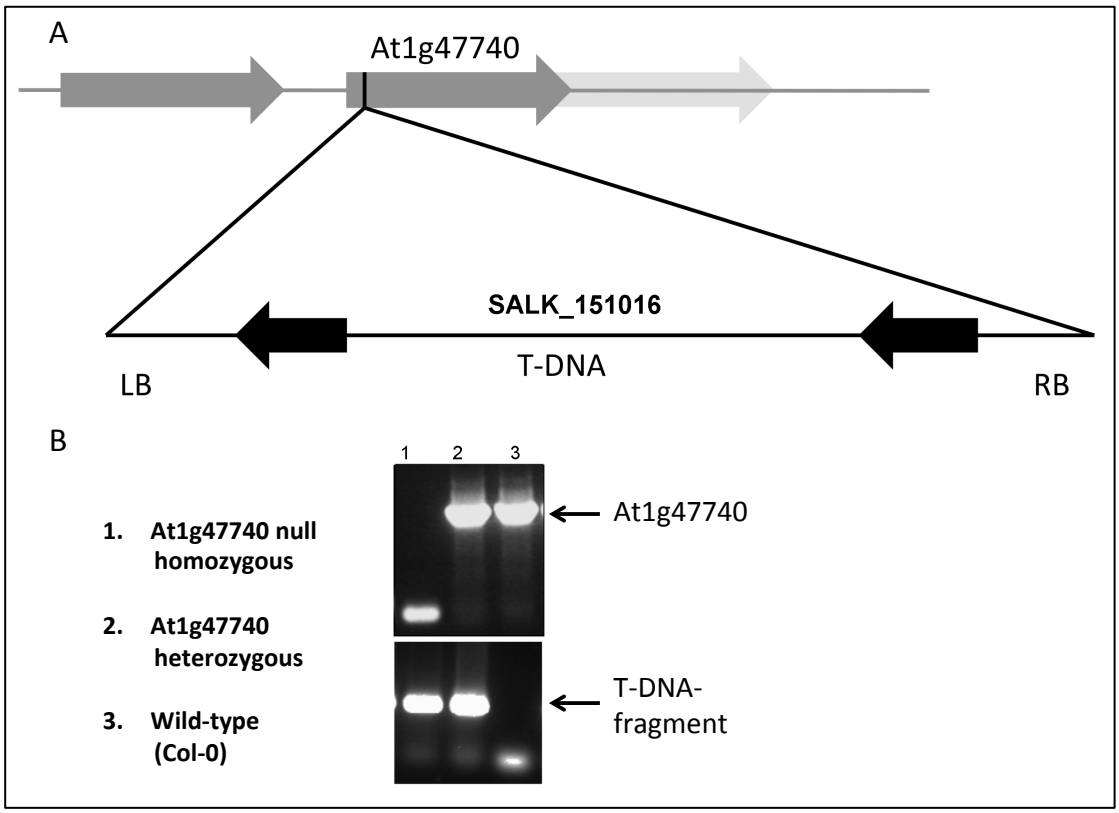

Figure 4.02. T-DNA insertion results in At1g47740 gene knock-out.
A schematic diagram of the position of the T-DNA is shown in (A). Black arrows indicate the T-DNA orientation and grey arrows indicate the genes in the genome with *At1g47740* labelled. LB and RB indicate the left and right boarder of the T-DNA respectively, SALK_151016 is the seed line acquired from the NASC seed bank. Black line shows the position in the *At1g47740* gene where the T-DNA insertion is located. PCR of genomic DNA extracted from wild-type (col-0) and At1g47740 mutant lines was performed and results shown in (B). Two PRC reactions where performed; one to amplify the *At1g47740* gene (upper panel) and one to amplify the T-DNA insertion (lower panel). Numbers correspond to genotypes labelled on the left, the PRC shows that the *At1g47740* gene is amplified in col-0 (3) and one of the *At1g47740*-T-DNA lines (2), whereas the T-DNA is amplified in both the *At1g47740*-T-DNA lines, resulting in identification of a homozygote T-DNA insertion (1). Labelled arrows in (B) show the expected DNA fragment as indicated.

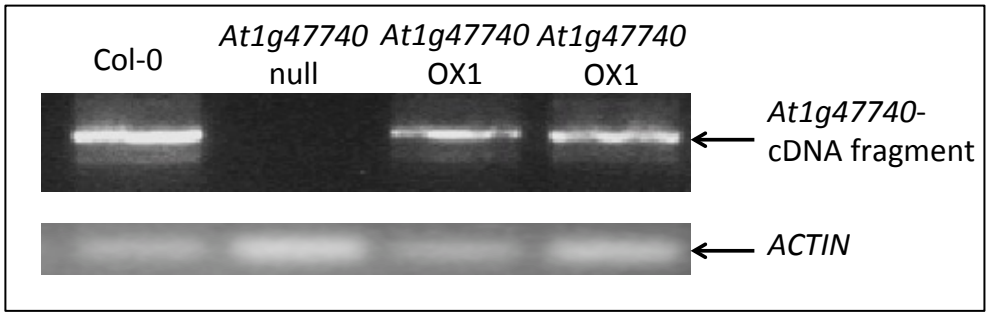

Figure 4.03. PCR analysis of cDNA shows the presence of *At1g47740* transcripts in wild-type and over-expressing lines, but not the knockout line.
Seedlings were grown for 10 days then RNA extracted and cDNA synthesised. PCR reactions were carried out using primers matching the coding region of the *At1g47740* cDNA. The presence of a DNA band at the expected size confirms transcription in the corresponding genotype. Actin was used as a loading control for the normalisation of the cDNA template used in each reation.

For conformation that the transgene is expressed in the over-expressing lines, seedlings were grown for 10 days and then frozen in liquid nitrogen, total protein was then extracted and analysed by SDS-PAGE and immunoblotting. Using antibodies raised against GFP, the immunoblot was probed, the GFP tagged At1g47740 recombinant protein had an expected molecular weight of 62kDa. The presence of a band at this expected size confirmed the recombinant protein was expressed, however, it is clear that the protein is either more stable or expressed to a higher level in the *At1g47740*-OX2 line than *At1g47740*-OX1 line (Figure 4.04). No protein band was seen in the col-0 samples, this was to be expected, as the native protein is not fused to GFP.

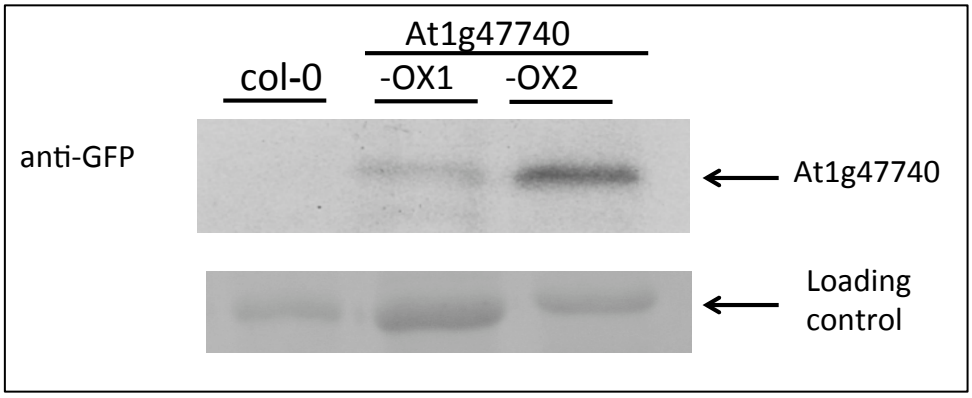

Figure 4.04. In 10 day old seedlings expression recombinant At1g47740 protein is presence in both over-expressing lines.
10 day old seedlings were grown in liquid culture for 24 hours. Sample consisting of 20 seedlings were flash frozen, then total protein content was extracted. Samples were then western blot analysed using antibodies raised against GFP. The top panel shows the bands at the expected size of the recombinant GFP-At1g47740 as indicated by the arrow, the bottom panel shows the total extract as a loading control for the normalisation of the samples. No protein is present at the expected size in the col-0 sample, however both *At1g47740*-OX1 and –OX2 display a protein band with expression at a higher level in the *At1g47740*-OX2 line.

4.2.4 Assays to test for phenotypic differences in KO and Overexpressing plants versus wildtype

Wild-type (col-0), *At1g47740* Knockout (*At1g47740*-KO) and GFP-*At1g47740* overexpression lines (*At1g47740*-OX1 and OX2) were tested in stress response assays using biotic stress inducers in the growth medium. The pathogen response elicitor flg22 or the pathogen PstDC3000 (Underwood *et al*., 2013) was used as a biotic stress inducer in the media and seedling response was observed. Plants were first grown on normal 1/2MS media for 3 days and then transferred to normal 1/2MS media, or 1/2MS media containing flg22. Root length was measured 5 days after plants were transferred onto flg22 media and plants grown on normal 1/2MS media were compared to those grown on 1/2MS media containing flg22. All the plants grown on normal 1/2MS media show the similar root elongation, however when grown on the flg22 containing media, *At1g47740*-KO plants have shorter roots when compared to col-0 and *At1g47740*-OX lines (Figure 4.05). Furthermore plants overexpressing the *At1g47740* gene appear to be unaffected by the flg22 in the media and root length is comparable to the plants grown on normal media. Quantification of the root lengths was calculated using averages of 30 plants per genotype for each experiment. The experiment was done 3 times and results were consistent. Figure 4.06 shows a bar graph of the average root growth calculated for each genotype on both the aforementioned media conditions. Interestingly, the *At1g47740* overexpressing lines showed no significant difference in average root length on both treatments, whereas the *At1g47740*-KO plants showed a reduction greater than that seen in the wild-type (col-0). Statistical significance was calculated using two-tailed T-test and p-values shown in table 2.

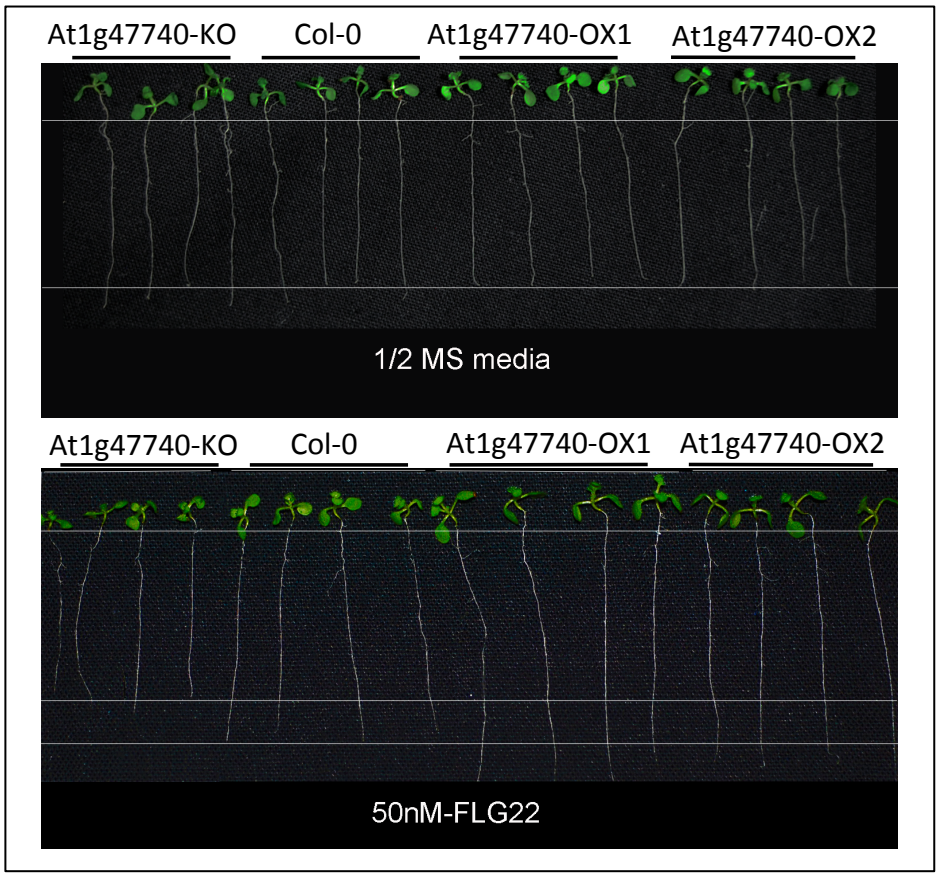

Figure 4.05. *At14g47740*-KO plants show increased root growth inhibition compared to col-0 when grown on media containing flg22.
30 plants of each genotype were used on both 1/2MS media with 50nMflg22 and 1/2MS media only, and root length measured digitally using imageJ software. All plants were germinated and grown for 3 days on 1/2MS media before being transferred to plates containing 1/2MS or 1/2MS plus flg22. All genotypes transferred onto 1/2MS media showed consistent root elongation after 8 days (top panel), however only 2 genotypes (*At1g47740*-KO and col-0) showed reduced root growth on media containing flg22 (bottom panel). Further more *At1g47740*-KO plants were smaller than col-0 plants when grown on media containing flg22. White horizontal lines boarder the roots of the plants highlighting the difference in root length.

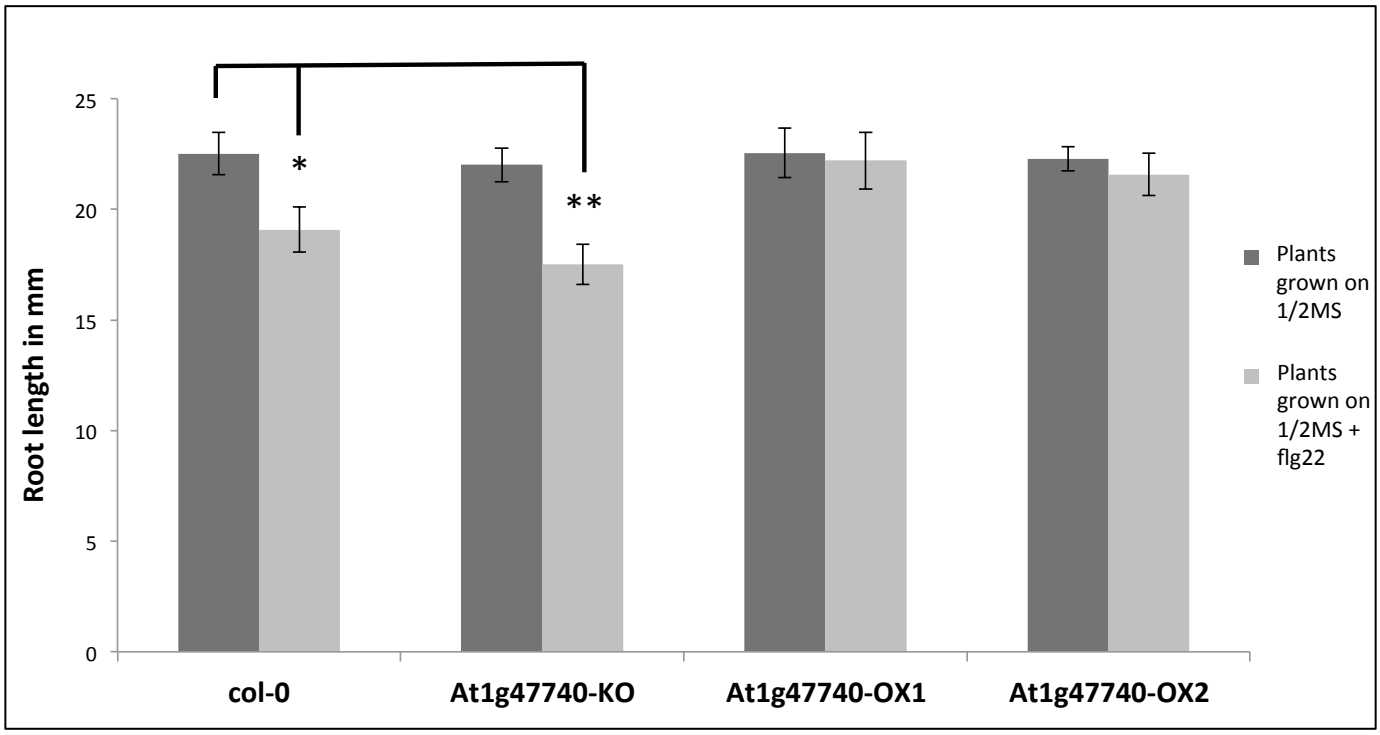

Figure 4.06. The average root length of *At1g47740*-KO plants grown on media containing flg22 was less than that of col-0.
The root length measurements of 30 plants of each genotype was used to calculate an average for each genotype for each treatment, the results were then plotted on a bar graph. Both the *At1g47740* overexpressing lines (*At1g47740*-OX1 and *At1g47740*-OX2) showed an insignificant difference when comparing average root length on 1/2MS media and 1/2MS media containing flg22. Contrary to this, both col-0 and *At1g47740*-KO showed a reduced average root length when grown on media containing flg22. The difference between the average root length of col-0 and *At1g47740*-KO was significant when grown on flg22 containing media. (*) represents p-value of <0.01, (**) represents p-value of <0.001. Error bars show the ±SE as calculated from the SD.

Genotype	1/2MS vs FLG22 p-value
col-0	0.00437
At1g47740-KO	0.00038
At1g47740-OX1	0.5435
At1g47740-OX2	0.2699

Table 4.2. Statistical analysis of root length using two-tailed T-test reveals significant p-values for *At1g47740*-KO but not *At1g47740*-OX lines. Root length measurements for each genotype were taken on both 12MS media and 1/2MS media plus flg22. The root length from both treatments were compared and statistically analysed using a two-tailed T-test. The results show that both col-0 and *At1g47740*-KO had a significant difference in root length when comparing the two treatments. The p-value of *At1g47740*-KO is over a factor of ten lower than that of col-0, showing that the difference is statistically greater in *At1g47740*-KO plants than the col-0 wild-type. Contrary to this, the *At1g47740*-OX lines showed no statistically difference in root length between the treatments.

For quantification of the amount of root growth inhibition the stress media caused, the averages from the plants grown on flg22 containing media was subtracted from the average root length of the same genotype grown on 1/2MS only media. This assay further analyses the data of figure 4.06 and determines if the reduction in root length is due to overall slower growth of a genotype or whether the presence of the flg22 peptide caused the plants to restrict root growth. Figure 4.07 shows that of the 4 genotypes tested, *At1g47740*-KO plants were most affected by the flg22 stress as already indicated in figure 4.06. After 7 days on media containing flg22, *At1g47740*-KO plants showed an average of 4.5mm inhibition of root length compared to plants grown on normal media, whereas col-0 only showed a 3.4mm inhibition, an on average reduction of around 25% more inhibition in *At1g47740*-KO plants, than col-0 plants. In contrast, the *At1g47740* overexpressing plants are not obviously affected by the flg22 media, and display a minimal to no reduction in root length when treated. The p-values of the data showed that the inhibition in root growth of *At1g47740*-KO plants was 10 fold less than that of col-0, meaning the results are significantly more robust for the *At1g47740*-KO and verifies an increase in response to flg22 over the wild-type.

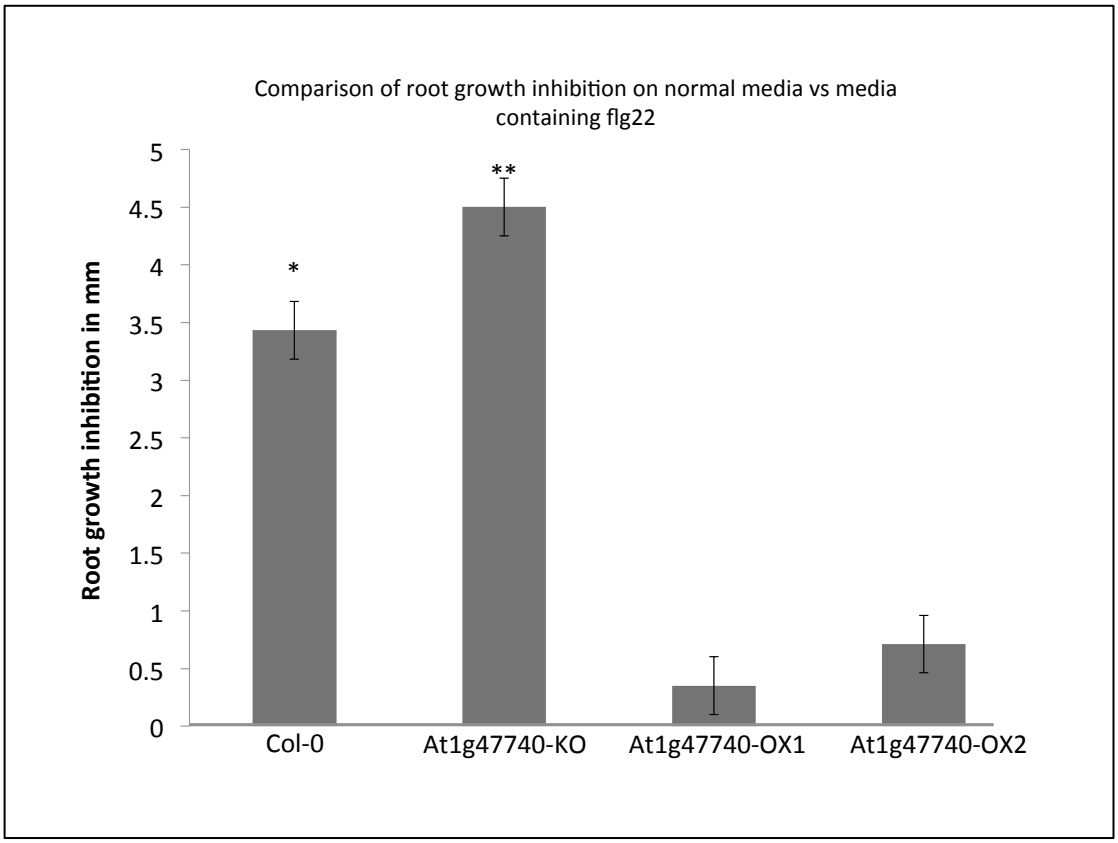

Figure 4.07. Quantification of root growth inhibition shows flg22 caused increased inhibition in *At1g47740*-KO plants compared to col-0.
The average root length of each genotype grown on flg22 containing 1/2MS media was subtracted from the average root length of those grown on normal 1/2MS media. The results were plotted on a graph and show the biggest inhibition of all genotypes occurred in the *At1g47740*-KO line, with a 4.5mm reduction in root length. Col-0 plants grown on flg22 media showed an average reduction of 3.4mm compared to those grown on normal 1/2MS media. (*) Indicates a statistical p-value of <0.01, (**) indicates a p-value of <0.001, comparing growth of each genotype on both medium. Error bars represents ±SD across all replicates.

To further understand the phenotypic differences between the genotypes, fresh weight data was collected from 10 day old seedlings. The purpose of this assay is to determine if the root growth inhibition shown in the previous figures is a consequence of the disruption of the root elongation mechanisms or a global growth arrest due to switch from growth to defence activated by the presence of the flg22 peptide. 30 plants per genotype were carefully removed from their media and dried on a paper towel before being weighed in groups of ten and average calculated. This experiment was repeated 3 times and results were consistent throughout. The fresh weight of all plants grown on normal 1/2MS media was consistent between all the genotypes. Some plants grown on media containing flg22 showed a reduced fresh weight compared to that on normal media. However no significant difference in the fresh weight due to flg22 was observed in the *At1g47740* overexpressing lines (Figure 4.08). In contrast, the col-0 plants and the *At1g47740*-OK plants showed a reduction in fresh weight when grown on media containing flg22, in a similar fashion to the root length assays where *At1g47740*-KO shows a greater reduction in fresh weight compared to col-0. The reduction in fresh weight on flg22 containing media compared to normal 1/2MS media was approximately 10% and 20% for col-0 and *At1g47740*-KO respectively.

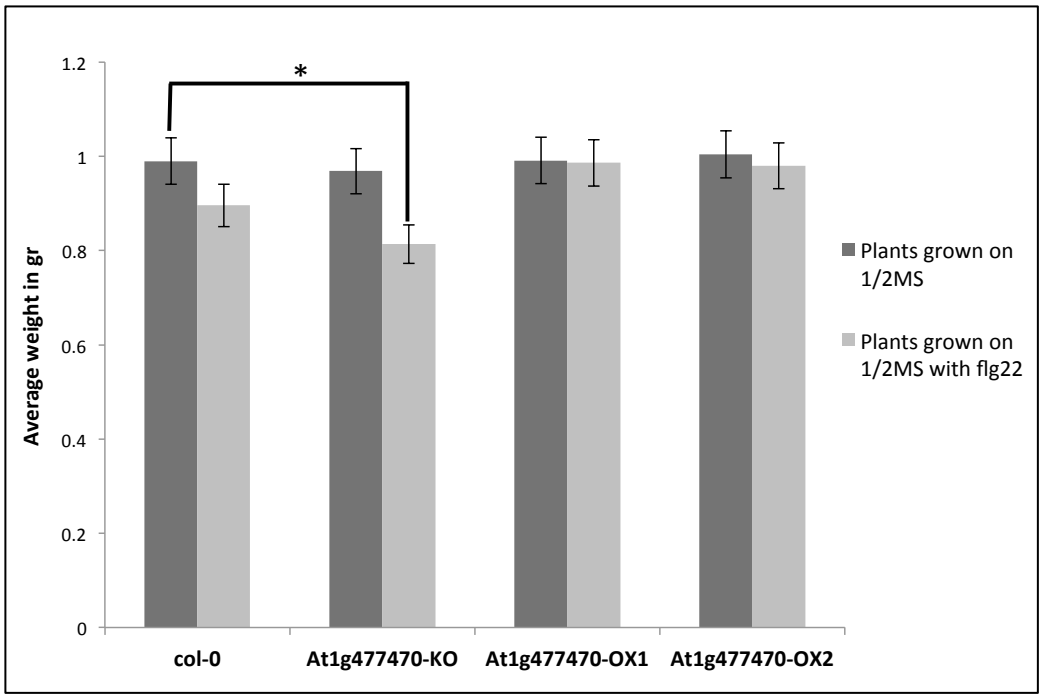

Figure 4.08. *At1g47740*-KO plants grown on media containing flg22 show a greater reduction in fresh weight compared to col-0. Plants were grown for 3 days on 1/2MS media then transferred to plates with 1/2MS media with and without flg22. After 7 days of further growth the plants were carefully removed, towel dried and weighed. 30 plants for each genotype was used and the experiment was repeated 3 times. The average weight of each genotype for both treatments was calculated and plotted on the graph above. Only col-0 and *At1g47740*-KO plants had a significant reduction in fresh weight, with col-0 showing a 10% reduction when grown on media with flg22, and *At1g47740*-KO showing a 20% reduction. The *At1g47740* overexpressing lines showed no significant difference when comparing growth on normal media to that on media containing flg22. * represents p-value <0.01

4.2.5 Response to pathogen PST DC3000

To understand the extent to which the *At1g47740* gene is involved in response to pathogen attack, *Pseudomonas syringae* (Pst) infection assays were carried out on 4-week-old plants. Pst was prepared as a liquid culture at a concentration of 10^6 cfu mL^{-1} and infiltrated into *Arabidopsis thaliana* leaves using a 1ml syringe. Plants were allowed to grow normally and at 2 dpi and 4 dpi, leaf disks were cut from the infiltrated leaves. The Pst was extracted and a serial dilution was done for quantification by plating 10ul of each dilution on plates containing kings-B media. A tray containing 12 plants of each genotype was used for the infiltration of the Pst and second tray was used with 12 plants of each genotype for 10mM $MgCl_2$ infiltration as a control, the experiment was done in triplicate and repeated. At 2dpi there was a marginal difference in the amount of bacteria present in all genotypes, but the *At1g47740*-KO plants showed fewer Pst then the other genotypes (Figure 4.09). At 4 dpi a greater difference was observed, the col-0 plants showed an increase in Pst cfu by over 1log as did the *At1g47740* overexpressing lines, which showed a growth increase in planta closer to 1.5log. In *At1g47740*-KO plants however, the Pst growth was significantly reduced, with the titre of Pst only increasing by 0.5log, a significant reduction from that shown in the other genotypes (Figure 4.09).

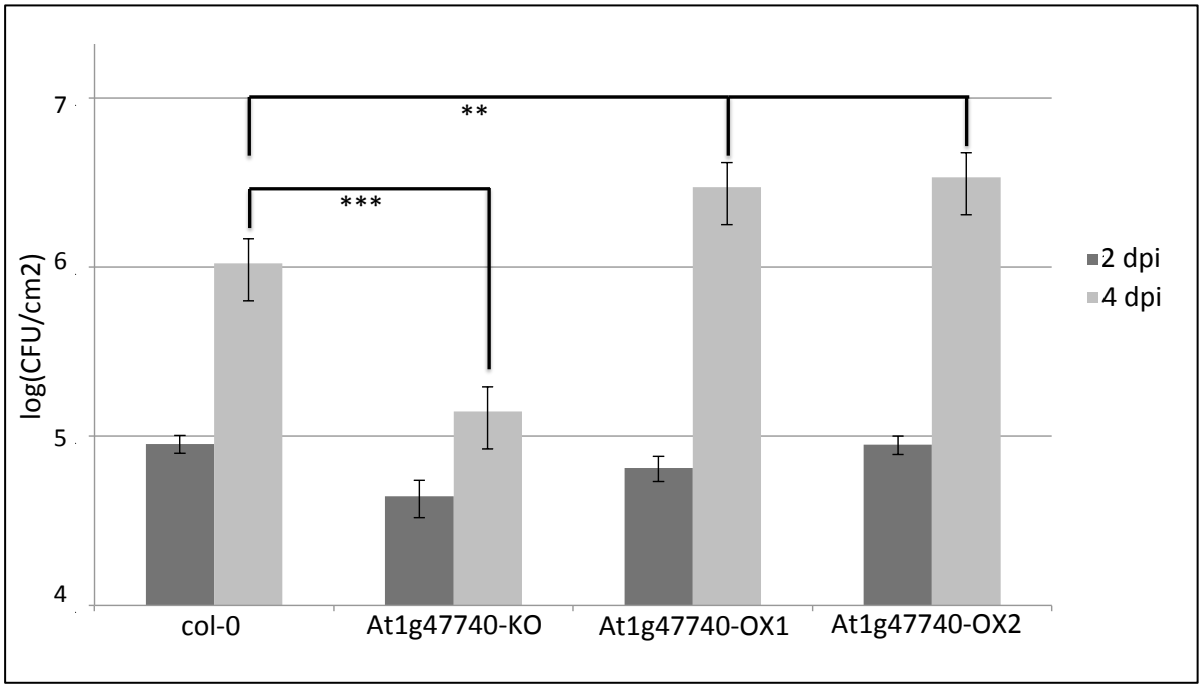

Figure 4.09. *At1g47740*-KO plants are more resistant than col-0 to the plant pathogen PST.

PST was pressure infiltrated into leaf surface at a concentration of 10^6 cfu mL^{-1}. At 2 and 4 days post infection (dpi), Pst levels from infiltrated leaves were quantified, each data set consists of 4 leafs from different plants, the experiment was done in triplicate and repeated. Error bars represent ±SD of 3 replicates. *** represents p-values of <0.00001 and ** represents p-values of <0.0001

In addition, leaves representing the symptoms of each genotype were chosen for imaging. Figure 4.10 shows the hypersensitive response of each genotype visualised by the area of necrosis observed on the leaf surface. This assay provided evidence to support the role of the *At1g47740* gene in pathogen response. When the gene is knocked out the effectiveness of PST pathogenicity is restricted, whereas the plants over-expressing the *At1g47740* gene appear to be more, or at least equally, susceptible to the pathogen as col-0. The spread of necrosis observed on the infiltrated leaves differs between genotypes with the least amount observed in the *At1g47740*-KO line. The data in Figure 4.09 supports this visual representation shown in Figure 4.10, collectively it appears that the *At1g47740* gene is involved in the negative regulation of plant defence against bacterial pathogens.

Figure 4.10. Pathogenic symptoms of Pst infiltration reveal resistance phenotype for *At1g47740*-KO plants.
PST was infiltrated into plant leaves as described in Figure 4.07. 5 dpi, leaves were chosen by blind selection and photographed to represent the respective genotype. *At1g47740*-KO plants showed less necrosis than either col-0 or the *At1g47740* overexpressing lines. The experiment was repeated and results were consistent.

4.2.6 Transient ROS production under elicitor induction

A fundamental function of plant stress response is the generation of reactive oxygen species (ROS) (Thakur and Sohal, 2013). To check if the ROS levels are altered under bacterial attack, leaf discs were cut from 4 week old plants and incubated in water for 24hours on a 24 well plate. The water was then removed and a hydrogen-peroxidase and luminal solution added prior to 250nM of flg22. The plate was then put in a dark box with a photon counting camera for 90 minutes, all experiments were repeated at least 3 times and results were consistent. FLS2 knockout plants, that do not product ROS under flg22 treatment, were used as a negative control.

In plants where the *At1g47740* gene was knockedout, a larger ROS production was seen compared to wild-type although the rate at which the ROS burst occurs appears not to differ (Figure 4.11). In the *At1g47740*-OX1 line the amount of luminance was more comparable with col-0 than of *At1g47740*-KO. The *At1g47740* overexpressing plants also showed a bigger degree of variability than the other genotypes as indicated by the error bars, but surprisingly did not show an overall reduction in ROS burst compared to col-0.

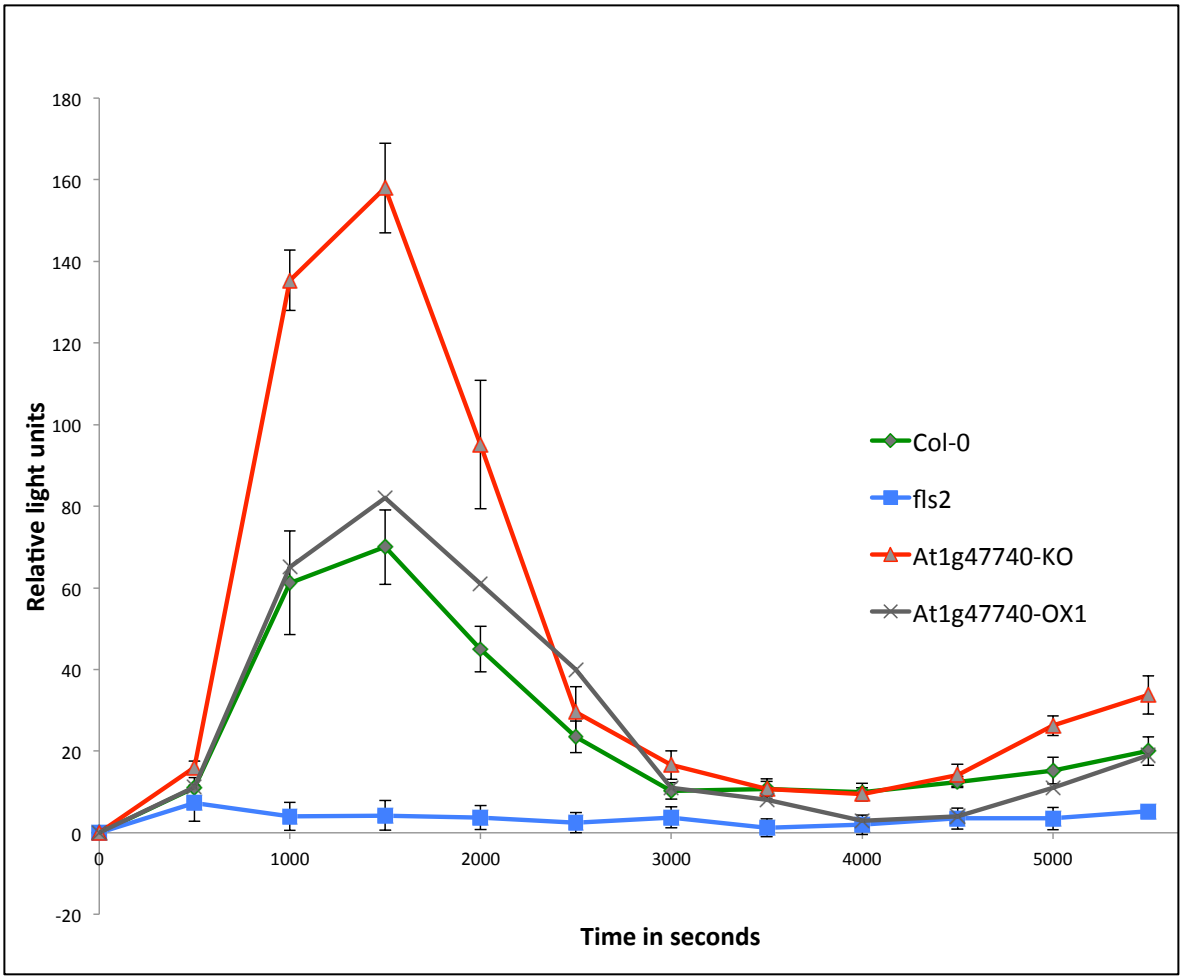

Figure 4.11. ROS generation in *At1g47740*-KO plants is greater than that of col-0 when induced by flg22.
Transient ROS production in response to the bacterial elicitor flg22 in col-0 (green-diamond), *fls2* (blue-square), *At1g47740*-KO (red-triangle) and *At1g47740*-OX1 (grey-cross) was measured using a luminal, hydrogen-peroxidase based solution and a photon counting camera. 9 leaf discs were cut from 6 plants of each genotype and placed in a 24 well plate in triplicate. The relative light units (RLUs) are shown on the x-axis whilst time in seconds is displayed on the y-axis. Both col-0 and *At1g47740*-KO showed a similar induction time with the ROS production peak around 23 minutes. Comparison of the peaks between the three genotypes reveal *At1g47740*-KO plants produce a much higher level of luminance then that of col-0 and *At1g47740*-OX1, *fls2* is used as a negative control and produces no change/peak under flg22 elicitation. Error bars show plus and minus SE as calculated by deviation of 3 different wells for each genotype. Experiments were repeated a minimum of 3 times and results were consistent.

At1g47740-KO showed an increased ROS burst when treated with flg22, this led to the hypothesis that the transcript levels of the *At1g47740* gene would be reduced when the same treatment was applied in the wild-type plants. Col-0 plants were grown for 10 days and then transferred to liquid media 24 hours before treatment with 250nM flg22. Prior to treatment, samples were taken and frozen and set as time point 0 hours. Flg22 was added and after 3 hours samples were taken and frozen for processing. Total RNA was extracted from frozen samples and cDNA generated using the SuperScript™ II Reverse Transcriptase kit from Invitrogen. qRT-PCR was then performed using the primers shown in (Table 3). The qRT-PCR was normalised using primers for the housekeeping gene actin and FKR1 was used as a positive control as its expression is induced under flg22 treatment. As can be seen in Figure 4.12, the levels of the *At1g47740* transcript are significantly reduced after treatment with flg22, with a reduction in transcript levels of over 12 fold. FKR1 transcript levels showed an increase over the same time scale of around 10 fold. Interestingly the basal level of *At1g47740* transcript seems to be fairly high, the reduction observed after flg22 treatment indicates that the At1g47740 protein product or the transcript itself may act as a negative regulator of plant defence against bacterial pathogens.

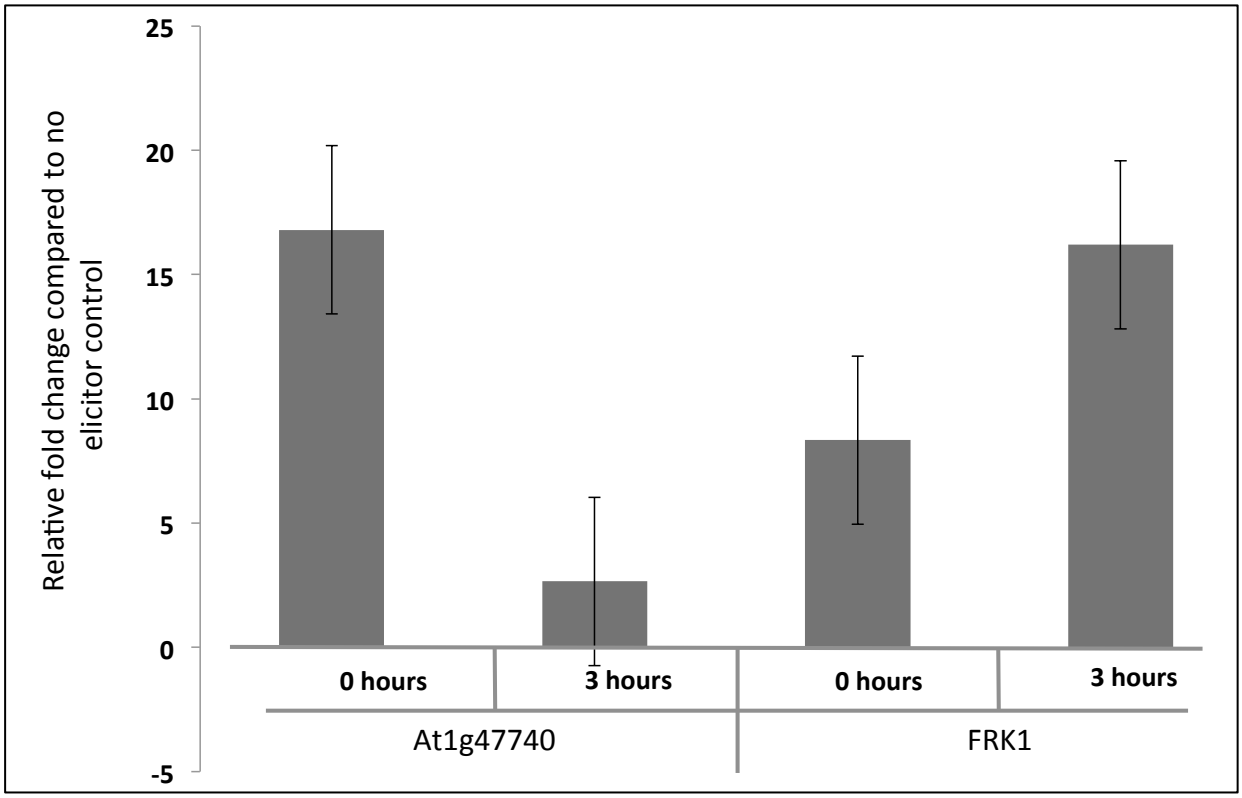

Figure 4.12. Detection of the elicitor peptide flg22 causes a reduction in *At1g47740* transcript levels.
qRT-PCR was performed on 11 day old plants after treatment with or without flg22. Primers were designed to identify the transcripts (mRNA) levels of *At1g47740* and FRK1. Two time points were measured, the 1st prior to addition of flg22 and the 2nd 3 hours later. The graph shows a reduction in *At1g47740* transcript levels by well over ten fold during the 2 time points, where as the control gene (FKR1) is up-regulated in the same time frame. Results were normalised against the actin housekeeping gene. Experiment was repeated in triplicate and results were consistent. Error bars represent ±SE of 3 repeats.

4.2.6 Protein levels change under elicitor treatment

To check if the At1g47740 protein levels change under elicitor detection a western blot analysis was carried out. 10 day old *At1g47740*-OX1 seedling were transferred to liquid culture 24 hours before flg22 treatment. 20 seedlings were removed and frozen prior to flg22 treatment then 250nM of flg22 was added and 20 plants harvested at time points of 30mins, 60mins and 90mins post-treatment. The frozen samples underwent protein extraction and the resulting extract was used for immunoprecipitation on beads enriched for GFP. The elute from the immunoprecipitation was then analysed by western blot and immunoblot using antibodies raised against GFP. The GFP tagged At1g47740 recombinant protein had an expected molecular weight of 62kDa, Figure 4.13 shows the immunoblot with the 4 time points; pre-treatment (0), 30mins post-treatment (30), 60mins post-treatment (60) and 90mins post-treatment (90). At time point '0', the abundance of recombinant protein detected by the western blot was greater then any of the preceding time points. At 30mins post-treatment the immunoblot showed a reduction in the intensity of the band compared to '0' indicating a reduction in protein levels. This was the largest change seen between 2 time points with time points '60' and '90' producing bands similar to '30'. The membrane was then stained with ponceau to show equal loading of the samples (bottom panel Figure 4.13)

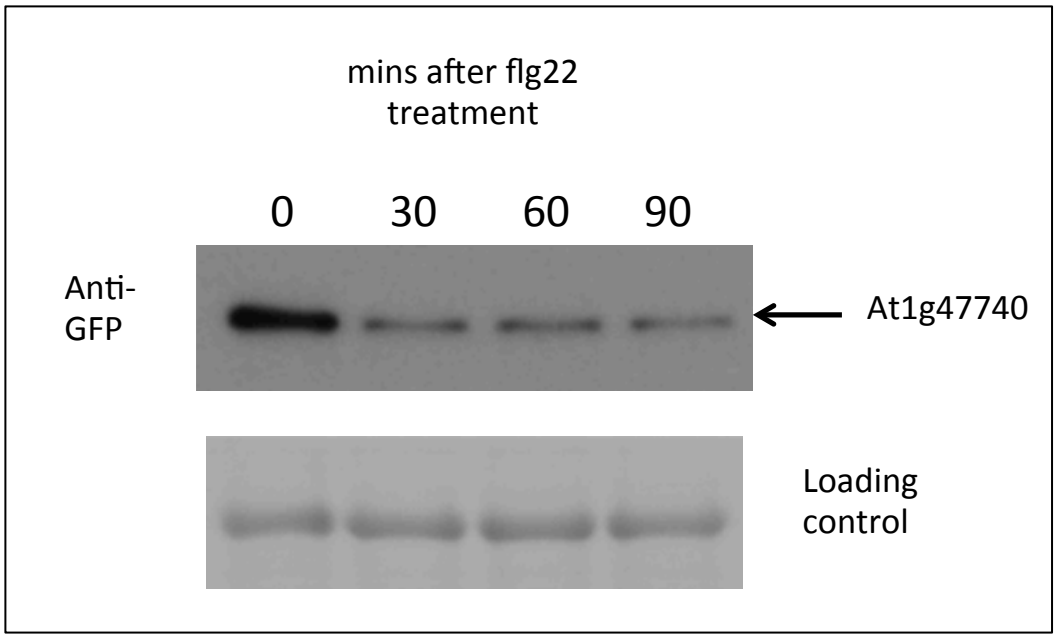

Figure 4.13. In 10 day old seedlings recombinant *At1g47740* protein levels are reduced by the elicitor flg22.
10 day old seedlings were grown in liquid culture for 24 hours. 4 samples consisting of 20 seedlings were flash frozen at 4 different time points, the total protein content was extracted and immunoprecipitation carried out. Samples were then western blot and immunoblot analysed using antibodies raised against GFP. '0' represents samples harvested prior to addition of flg22, '30', '60' and '90' represent the time in minutes that samples were harvested after addition of flg22. The top panel shows the bands at the expected size of the recombinant At1g47740 as indicated by the arrow, the bottom panel shows the total extract as a loading control for the normalisation of the samples.

As the evidence was growing that At1g47740 was involved in the bacterial/biotrophic pathogen defence via the FLS2 receptor, it was decided to investigate if there was a similar relationship between the *At1g47740* gene and the fungal/necrotrophic pathogen defence via the CERK1 receptor. 10 day old *At1g47740*-OX1 seedling were transferred to liquid culture 24 hours before chitosan treatment. 20 seedlings were removed and frozen prior to treatment then 1uM of chitosan was added and 20 plants harvested at time points of 15mins, 30mins and 60mins post-treatment. The frozen samples underwent protein extraction. The elution from the extraction was then analysis by western blot and immunoblot using antibodies raised against GFP. The GFP tagged At1g47740 recombinant protein had an expected molecular weight of 62kDa, Figure 4.14 shows the immunoblot with the 4 time points; pre-treatment (0), 15mins post-treatment (15), 30mins post-treatment (30) and 60 mins post-treatment (60). At time point '0', the abundance of recombinant protein detected by the western blot was greater then that of time point '15', however unlike with the flg22 treatment, the the difference was not obvious. Furthermore return to basal state was seemed to occur by the next time point '30' minutes. The membrane was stained with ponceau to show equal loading of the proteins.

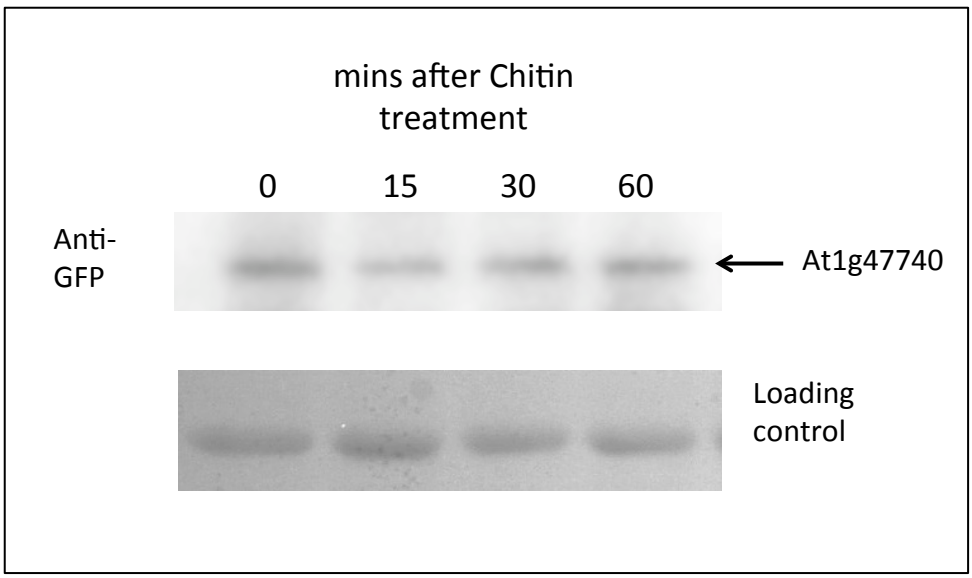

Figure 4.14. In 10 day old seedlings recombinant At1g47740 protein levels are reduced by the elicitor chitosan.
10 day old seedlings were grown in liquid culture for 24 hours. 4 samples consisting of 20 seedlings were flash frozen at 4 different time points, the total protein content was extracted. Samples were then western blot and immunoblot analysed using antibodies raised against GFP. '0' represents samples harvested prior to addition of chitosan, '15', '30' and '60' represent the time in minutes that samples were harvested after addition of chitosan. The top panel shows the bands at the expected size of the recombinant At1g47740 as indicated by the arrow, the bottom panel shows the total extract as a loading control for the normalisation of the samples.

4.2.7 ROS production under chitin treatment

To further investigate the fungal defence pathway, *At1g47740* knockout plants and complimenting overexpressing lines were used with wild-type and *cerk*1 knockout plants (deficient in the perception of chitin/chitosan), in ROS detection assays (Cao *et al*., 2014). This was to determine if there was a difference in the plants response to the elicitor. Leaf discs were cut from 4 week old plants and incubated in water for 24hours on a 24 well plate. The water was then removed and a hydrogen-peroxidase and luminal solution added prior to 1μM of chitosan. The plate was then put in a dark box with a photon counting camera for 90 min, all experiments were repeated at least 3 times and results were consistent.

As with the response to flg22, *At1g47740*-KO plants show an increase in initial ROS burst (Figure 4.15), the over-expressing lines are more comparable to col-0. The *cerk1-1* knockout line showed a small/weak response opposed to the expected no response, but this was probably due to the activation of other receptors due to the non-uniform nature of the elicitor. The *At1g47740*-KO lines also appear to have a prolonged duration of ROS production compared to col-0 and *At1g47740*-OX1, with the return to basal levels taking much longer. The peak of ROS production in col-0 when induced by chitin/chitosan was only 75% of that measured at the peak of *At1g47740*-KO (Figure 4.15).

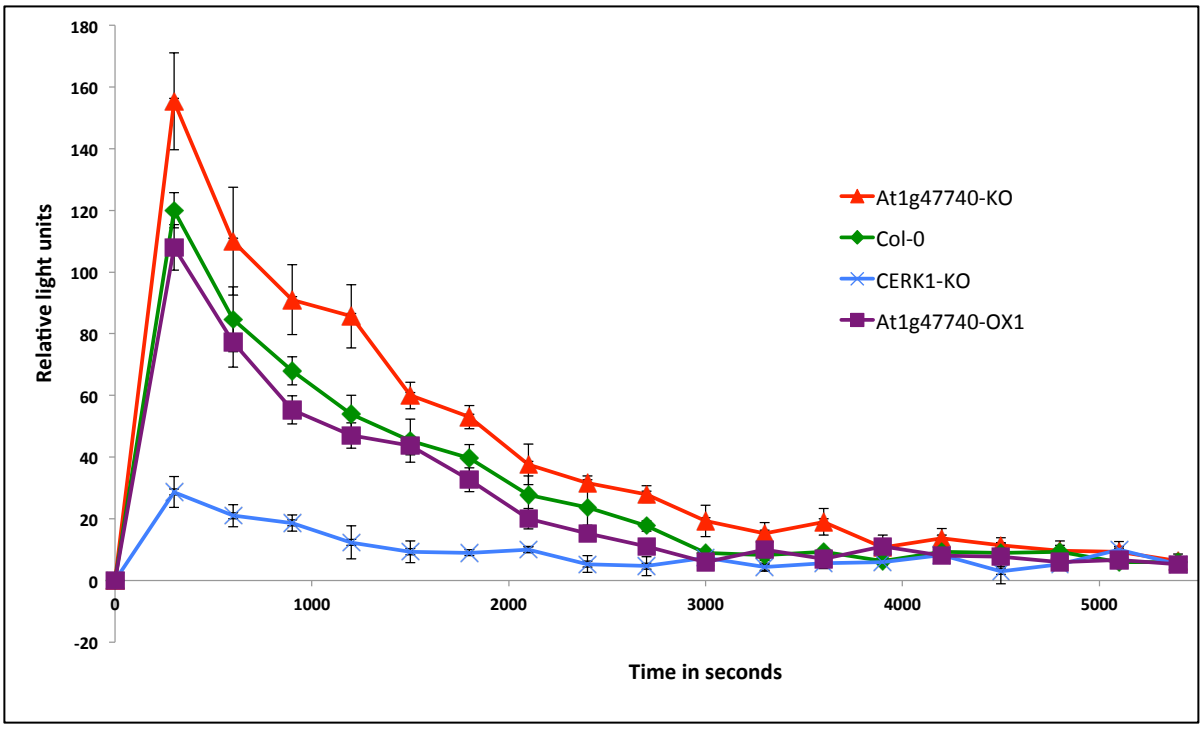

Figure 4.15. The elicitor chitinosan induces a larger ROS burst in *At1g47740*-KO plants compared to col-0.
Transient ROS production in response to the bacterial elicitor chitosan in col-0 (triangle), *crek1-1* (cross), *At1g47740*-KO (circle) and *At1g47740*-OX1 (squared) was measured using a luminal, hydrogen-peroxidase based solution and a photon counting camera. 9 leaf discs were cut from 6 plants of each genotype and placed in a 24 well plate in triplicate. The relative light units (RLUs) are shown on the x-axis whilst time in seconds is displayed on the y-axis. Comparison of the peaks between the three genotypes reveal *At1g47740*-KO plants produce a much higher level of luminance then that of col-0 and *At1g47740*-OX1, *cerk1-1* is used as an negative control. Further more, *At1g47740*-KO plants appear to have a longer 'lag' on the graph, indicating higher ROS production throughout the time period. Error bars show the plus and minus SE as calculated by deviation of 3 different wells for each genotype. Experiments were repeated a minimum of 3 times and results were consistent.

4.2.8 Transcript levels of defence genes when treated with chitin

As a way of investigating the results of the effects of chitosan in plants that are *At1g47740* null, quantitative RT-PCR was performed on col-0 and *At1g47740*-KO, with and without chitosan treatment. The aim of this experiment was to determine if the *At1g47740* gene had a global affect on down stream response to chitosan detection. Plants were grown for 10 days on plates containing 1/2MS media and then transferred to 1/2MS liquid media 24 hours before treatment with 1uM chitosan. Chitosan was added and after 1 hour samples were taken and frozen for processing. Total RNA was extracted from frozen samples and cDNA generated using the SuperScript™ II Reverse Transcriptase kit, qRT-PCR was then performed using the primers shown in (Table 4). The qRT-PCR was normalised using primers for the actin housekeeping transcript. Several pathogen response genes were chosen to measure transcript levels to ascertain any differences between the col-0 and the *At1g47740*-KO plants with and without chitosan treatment (Figure 4.16). A general trend observed from this experiment was that the transcript levels for several genes were up-regulated in the *At1g47740*-KO plants compared to col-0 independently of any elicitor treatment. This was true for *TRX*, *JAZ5*, *FRK1* and *WRKY33*; and to a lesser extent *PDF* and *ORA1* (For details of gene roles please refer to introduction section – 1.07 Responses to pathogen). After chitosan treatment however, several genes showed an increase in transcript levels by more than 2 fold, including; *MYB51, TRX, JAZ5, WRKY29, PAD3* and *FRK1*. Of these genes, *MYB51, TRX, WRKY29, PAD3* and *FRK1* transcript levels were hugely increased in the *At147740*-KO plants compared to col-0. This experiment was repeated only once due to adverse circumstances, and although the results were not then fully robust, it provided enough evidence that the investigation was going in the right direction. Down stream targets and affects are not in the remit of this thesis, but further investigation into this would be highly insightful.

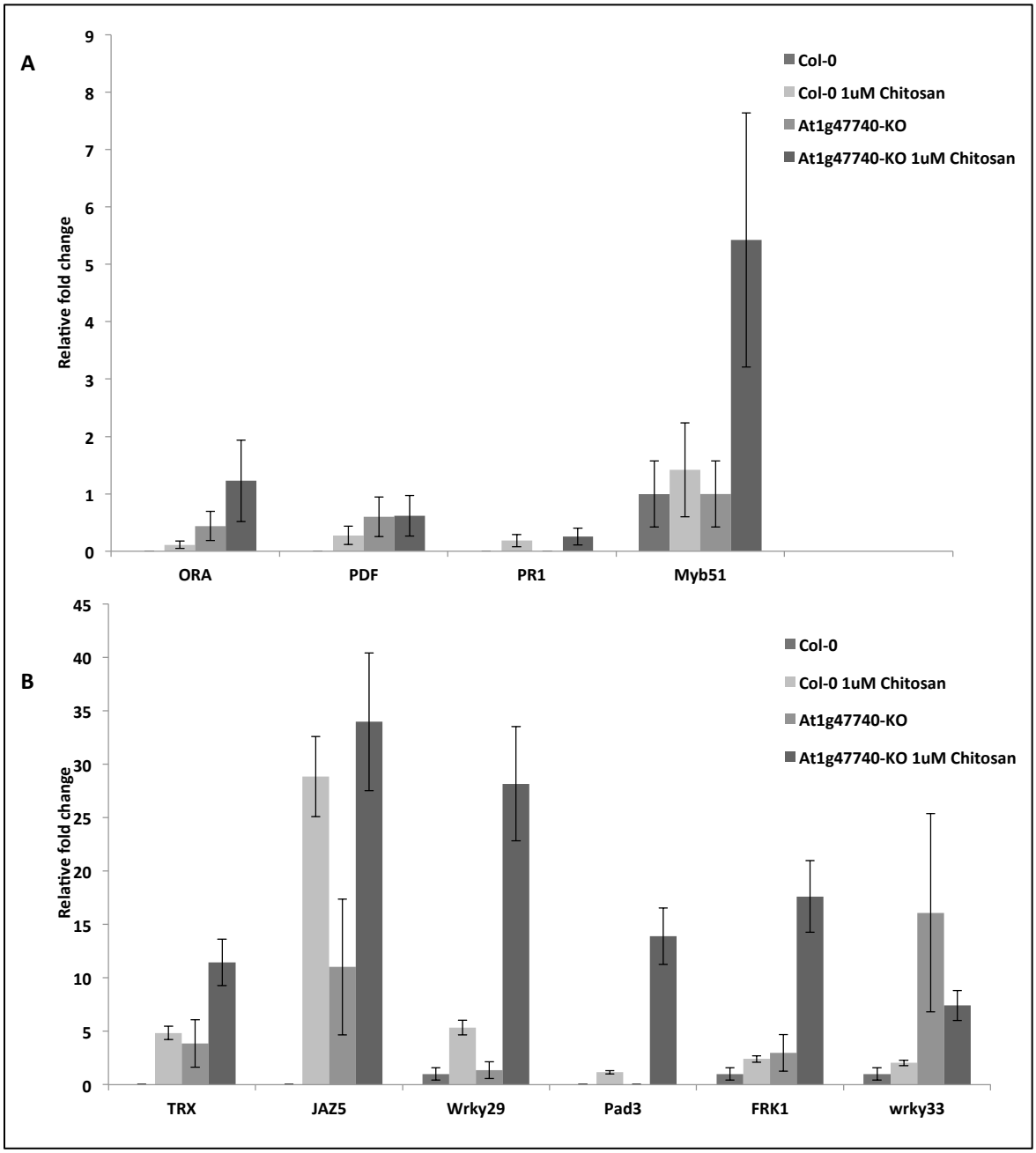

Figure 4.16. *At1g47740*-KO plants show an increase in pathogen response genes compared to col-0.

qRT-PCR was performed on 11 day old plants after treatment with or without chitosan. Primers were designed to identify the transcripts (mRNA) levels of various pathogen response genes listed in table 4. Two conditions were tested with two genotypes, the 1st condition 1 hour after the addition of water, and the 2nd 1 hour after the addition of chitosan. The graph shows the fold change on the y-axis compared to col-0 plants prior to treatment. Experiment was repeated only once. Error bars represent ±SE of 3 technical repeats.

4.3 Discussion

This chapter has explored the effects in *Arabidopsis thaliana* when knocking out the gene encoding the DeSi protease; At1g47740. It appears that the protein is located within a large membrane compartment in or around the cell periphery, making the likely location the plasma membrane, the tonoplast or both (Figure 4.01). Further investigation into the localisation, using appropriate cellular compartment markers, would allow full verification of the exact compartment. The interesting discovery for the scope of this project was that the SUMO protease was seen outside the nucleus; this is the first time this observation has been made in plants. In order to ascertain the function of the *At1g47740* gene *in planta,* transgenic plants were grown which contained the *At1g47740* gene knockout out via a TDNA insertion (Figure 4.02) or plants that overexpressed the *At1g47740* gene via 35S promoter. These genotypes, along with the wild-type ectotype col-0, were used in root growth assays to determine the affects on root growth inhibition when exposed to the pathogen elicitor peptide flg22. *At1g47740*-KO plants showed a greater reduction in root growth when grown on media containing flg22, compared to col-0 (Figure 4.05). This trend was also observed by calculating the amount of root growth inhibition the plants experienced when treated with flg22 (Figure 4.07). Whilst the *At1g47740* overexpressing plants showed no significant difference in root growth when grown on the media containing flg22, the *At1g47740*-KO plants had an increased statistically significant reduction in root length compared to col-0 (table 2). This is further reflected by the results of the fresh weight assays when plants grown in these same conditions, *At1g47740*-KO plants weigh less than col-0 when grown on media containing flg22, where as *At1g47740* overexpressing plants show little or no difference. Collectively this implicated that *At1g47740*-KO plants are further comprised in growth compared to col-0 when exposed to the elicitor flg22, something that could be explained by an increase in immune response. This evidence was enhanced by the fact that *At1g47740* overexpressing plants grew on flg22 containing media as well as they did on media without flg22.

To test if this data translated to differences in response to actual pathogen attack, bacterial infiltrations were carried out on the leaves of all genotypes using the plant pathogen PST DC3000. The results of this pathotest would determine if the *At1g47740* gene had a role in the regulation of the plants response to the bacterial pathogen by tracking the growth of the pathogen *in planta* and observing any physical symptoms on the plant leaf. In line with the previous findings the *At1g47740*-KO plants showed a reduced titre of bacteria compared to col-0 and *At1g47740* over-expressing lines after 4 days of infection, *At1g47740*-KO plants also had reduced leaf symptoms on day 5 of the treatment (Figures 4.09 & 4.10).

As the At1g47740 protein product was located in the periphery of the cell, and it has been shown to have a role in pathogen defence, it was hypothesized that perhaps it functioned in the early part of the plant defence; pathogen detection. ROS are produced by the plant in response to the detection of the elicitor peptide flg22 via the LRR-RK protein FLS2 (Underwood *et al*., 2013). *At1g47740*-KO plants showed a higher production of ROS compared to col-0 and indicated that the At1g47740 protein product acts upstream of any transcriptional changes brought on by pathogen detection (Figure 4.11). Furthermore *At1g47740* overexpressing lines showed large variability in ROS production, and although the measurements never reached the increased levels of the *At1g47740*-KO plants, it was observed that *At1g47740* overexpressing plants produced ROS higher and lower than col-0 with no obvious reason for the inconsistency. Col-0 and *At1g47740*-KO plants showed highly consistent ROS levels in comparison.

As the evidence grew that the *At1g47740* gene is in some way negatively regulating plant immunity under pathogen attack, it was logical to assume transcription levels for this gene would drop under pathogen detection. Therefore quantitative RT-PCR was performed on col-0 plants before and after flg22 treatment and revealed that the mRNA levels for the gene were indeed reduced after flg22 treatment (Figure 4.12). In line with this observation, protein levels of plants overexpressing *At1g47740* were checked before and after flg22 treatment. This assay revealed a similar trend to the

At1g47740 transcript levels, where proteins levels also reduce after flg22 treatment (Figure 4.13).

Interestingly, the same trend is seen in response to the fungal pathogen elicitor chitin (Chitosan). At1g47740 proteins levels are reduced, possibly to a lesser extent than seen under flg22 treatment, when treated with chitosan (Figure 4.14). This was slightly unexpected but required further investigation as a more universal role for At1g47740 could be uncovered. ROS production assays using the elicitor chitosan showed that *At1g47740*-KO plants had an increased ROS production compared to col-0 (Figure 4.15), much like the trend observed in the flg22 induced ROS assays.

To gauge if the *At1g47740*-KO plants would have transcriptional differences to col-0 under chitosan treatment, quantitative RT-PCR was performed on 10 different pathogen response transcripts. This experiment was a way to look down stream at the affects of chitosan in the *At1g47740*-KO background without really having planned or anticipated taking the research in this direction. The results showed that in the *At1g47740*-KO background, some pathogen response genes are constitutively up-regulated independently of elicitor treatment, whereas some are highly up-regulated in response to chitosan treatment. This experiment was only repeat once due to the time restraints, but gave a possible insight into the importance of the *At1g47740* gene in pathogen defence.

Taken together the work in this chapter showed the protease At1g47740 has a role in the response to pathogen attack, and is likely to be negatively regulating plant immune responses. The mechanism by which this works, assuming the deSUMOylating function is involved, could be as follows:

When the plant perceives pathogenic elicitors, SUMOylation of an unknown component in the periphery of the cell is increased and results in activation of the plants immune response. In the default state, the At1g47740 protein levels are high and constant deSUMOylation of the unknown target occurs freely. However under pathogen attack, the At1g47740 protein levels drop, allowing the SUMOylation levels of the unknown target to increase, this would mostly likely coincide with an increased rate of SUMOylation of the unknown target, resulting in activation of cellular responses.

If this theory were correct, in *At1g47740*-KO plants, defence mechanisms would be primed even before perception of the pathogen due to the cells inability to deSUMOylate the unknown target. In this sense At1g47740 could be regarded as a negative regulator of plant defence, and what makes that interesting is that it appears to do so without directly interacting with nuclear components, a first in terms of SUMOylation.

Chapter 5
Investigation of the potential DeSi3a targets

5.1 Introduction

The previous chapter provided strong evidence that the DeSi protease At1g47740 is involved in plant immunity and that phenotypic differences are significant when the *At1g47740* gene is knocked out. However, due to only having one *At1g47740*-KO line it was important to create complimenting lines with the recombinant *At1g47740* gene to check if the phenotypic changes seen in the *At1g47740*-KO line could be rescued. Due to time restrictions and issues with cloning procedures, the *At1g47740* promoter region was not cloned, and instead the cDNA was inserted into a vector harbouring the 35S promoter.
For further investigation of the mechanistic nature of the At1g47740 protein, the identification of putative target substrates was achieved using the following information:

1) The target substrate must be located in the cell periphery i.e. the plasma membrane
2) The target substrate must act early in pathogen detection/response
3) The target substrate must have known or predicted SUMOylation sites with a high confidence interval
4) The target substrate must be involved regulation of pathogen defence

Taking these details into account, together with the known pathogen signalling pathways, it was possible to make an initial potential target list. FLS2 and CERK1 are LRR-RKs, which are the main PRRs involved in the perception of bacterial and fungal pathogens respectively. The fact that knockout plants of *FLS2* and *CERK1* show no ROS response and the *At1g47740*-KO shows the opposite, made these 2 proteins first on the initial list of potential targets. However, based on the high specificity of the *Homo sapiens* DeSi1, the question that over shadowed this prediction was; Is it likely that the At1g47740 protein could target more than one substrate? i.e. both

FLS2 and CERK1. Based on the fact that the number of SUMO proteases within the *Arabidopsis thaliana* proteome is overwhelmingly low compared the number of SUMO targets, it was assumed that the SUMO proteases could have multiple targets and that this 'system' could still allow for very high specificity, especially within a protein family i.e. LRR-RKs. However, this line of thought led to the obvious conclusion that perhaps a single protein that interacted with both FLS2 and CERK1 was the main target of the At1g47740 protease. Since upon activation by a PAMP, both FLS2 and CERK1 interact with BIK1, and in both cases BIK1 is phosphorylated by BAK1, BAK1 and BIK1 were added to the short list of potential candidates (Zhang *et al.,* 2010a, Macho & Zipfel 2014; Shi *et al.,* 2013). The over stimulation of ROS production in the *At1g47740*-KO background led to the conclusion that a potential target could also be RBOHD. Due to the availability of FLS2 mutants within the research group, and the limited time remaining to complete the research, it was decided that FLS2 would be further investigated as an At1g47740 protease target. Since the previously discussed parallels between FLS2 and CERK1 were uncovered during the research of chapter 4, CERK1 was also chosen for further investigating.

5.2.1 Identification of SUMO Sites on Potential At1g47740 Target Substrates

Using the short list deduced as described above, At1g47740 target proteins were checked for SUMOylation sites. The preliminary step to ascertain SUMO interactions for each of the potential target proteins was to inspect the amino acid sequence for identification of SUMOylation sites. This was achieved using a custom made in-house plant SUMOylation prediction software (Nelis 2014, unpublished thesis). The results from the SUMO prediction software provided not only identification of the lysine to which SUMO attachment was likely, but also a confidence interval based on known SUMO-substrate interactions in

plants. The predictions made by the software were based on typical SUMOylation motifs comprising 6 amino acids that contain a central SUMO 'accepting' lysine (as described in section 1.1.1). The Arabidopsis FLS2 protein sequence was processed by the SUMO prediction software and showed that there are several potential SUMO sites within the protein. However due to the fact only 1 of these SUMO sites was in the kinase-domain of the protein, it was hypothesised that this was the most likely candidate for significant SUMO interaction inside the cell (Figure 5.01). In the FLS2 kinase domain the SUMO site K1120 is predicted with the highest possible confidence (99%), and the motif SLKQEE is common in other known plant protein SUMO sites (Impen *et al.,* 2014). The CERK1 protein sequence was also analysed for SUMO sites and unlike FLS2 only 3 sites were identified (Figure 5.02). An interesting observation for this experiment was that both FLS2 and CERK1 had in common a single predicted SUMO site in their respective kinase domain. The CERK1 SUMO site K495, was predicted at a confidence of 93%, and is represented by the amino acid sequence SAKVDV.

FLS2 – Protein sequence

MKLLSKTFLILTLTFFFFGIALAKQSFEPEIEALKSFKNGISNDPLGVLSDWTIIGSLRHCNWTGITCDSTGHVVSV
SLLEKQLEGVLSPAIANLTYLQVLDLTSNSFTGKIPAEIGKLTELNQLILYLNYFSGSIPSGIWELKNIFYLDLRNNLL
SGDVPEEICKTSSLVLIGFDYNNLTGKIPECLGDLVHLQMFVAAGNHLTGSIPVSIGTLANLTDLDLSGNQLTGK
IPRDFGNLLNLQSLVLTENLLEGDIPAEIGNCSSLVQLELYDNQLTGKIPAELGNLVQLQALRIYKNKLTSSIPSSL
FRLTQLTHLGLSENHLVGPISEEIGFLESLEVLTLHSNNFTGEFPQSITNLRNLTVLTVGFNNISGELPADLGLLTN
LRNLSAHDNLLTGPIPSSISNCTGLKLLDLSHNQMTGEIPRGFGRMNLTFISIGRNHFTGEIPDDIFNCSNLETLS
VADNNLTGTLKPLIGKLQKLRILQVSYNSLTGPIPREIGNLKDLNILYLHSNGFTGRIPREMSNLTLLQGLRMYSN
DLEGPIPEEMFDMKLLSVLDLSNNKFSGQIPALFSKLESLTYLSLQGNKFNGSIPASLKSLSLLNTFDISDNLLTGT
IPGELLASLKNMQLYLNFSNNLLTGTIPKELGKLEMVQEIDLSNNLFSGSIPRSLQACKNVFTLDFSQNNLSGHI
PDEVFQGMDMIISLNLSRNSFSGEIPQSFGNMTHLVSLDLSSNNLTGEIPESLANLSTLKHLKLASNNLKGHVP
ESGVFKNINASDLMGNTDLCGSKKPLKPCTIKQKSSHFSKRTRVILIILGSAAALLLVLLLVLILTCCKKKEKKIENS
SESSLPDLDSALKLKRFEPKELEQATDSFNSANIIGSSSLSTVYKGQLEDGTVIAVKVLNLKEFSAESDKWFYTEA
KTLSQLKHRNLVKILGFAWESGKTKALVLPFMENGNLEDTIHGSAAPIGSLLEKIDLCVHIASGIDYLHSGYGFPI
VHCDLKPANILLDSDRVAHVSDFGTARILGFREDGSTTASTSAFEGTIGYLAPEFAYMRKVTTKADVFSFGIIM
MELMTKQRPTSLNDEDSQDMTLRQLVEKSIGNGRKGMVRVLDMELGDSIVSLKQEEAIEDFLKLCLFCTSSRP
EDRPDMNEILTHLMKLRGKANSFREDRNEDREV

Position	Type	Confidence	Sequence
K573	II	98%	FSKLES
K832	I	94%	CCKKKE
K854	I	95%	ALKLKR
K971	II	92%	LEKIDL
K1120	I	99%	SLKQEE

Figure 5.01 The Arabidopsis FLS2 protein sequence contains predicted SUMO sites.
The amino acid sequence for the *Arabidopsis thaliana* protein FLS2 was extracted from the TAIR website, was used as the query sequence in the plant SUMO site prediction software. Multiple SUMOylation sites were predicted by the software, but only 5 sites had a >90% confidence. Further more only one SUMO site was predicted in the kinase domain of the protein (K1120, shown in red text), the other site (shown in various colours) are within the intramembranous regions of the protein.

CERK1 – Protein sequence

MKLKISLIAPILLLFSFFFAVESKCRTSCPLALASYYLENGTTLSVINQNLNSSIAPYDQINFDPILRYNSNIKDKDRI
QMGSRVLVPFPCECQPGDFLGHNFSYSVRQEDTYERVAISNYANLTTMESLQARNPFPATNIPLSATLNVLVN
CSCGDESVSKDFGLFVTYPLRPEDSLSSIARSSGVSADILQRYNPGVNFNSGNGIVYVPGRDPNGAFPPFKSSK
QDGVGAGVIAGIVIGVIVALLLILFIVYYAYRKNKSKGDSFSSSIPLSTKADHASSTSLQSGGLGGAGVSPGIAAIS
VDKSVEFSLEELAKATDNFNLSFKIGQGGFGAVYYAELRGEKAAIKKMDMEASKQFLAELKVLTRVHHVNLVR
LIGYCVEGSLFLVYEYVENGNLGQHLHGSGREPLPWTKRVQIALDSARGLEYIHEHTVPVYVHRDIKSANILIDQ
KFRAKVADFGLTKLTEVGGSATRGAMGTFGYMAPETVYGEVSAKVDVYAFGVVLYELISAKGAVVKMTEAV
GEFRGLVGVFEESFKETDKEEALRKIIDPRLGDSYPFDSVYKMAELGKACTQENAQLRPSMRYIVVALSTLFSST
GNWDVGNFQNEDLVSLMSGR

Position	Type	Confidence	Sequence
K74	I	98%	NIKDKD
K276	II	92%	STKADH
K495	II	93%	SAKVDV

Figure 5.02 The Arabidopsis CERK1 protein sequence contains predicted SUMO sites.
The amino acid sequence for the *Arabidopsis thaliana* protein CERK1 was extracted from the TAIR website, was used as the query sequence in the plant SUMO site prediction software. Three SUMO sites were predicted and all sites had a >90% confidence. Only one SUMO site was predicted in the kinase domain of the protein (K495, shown in red text), the other site (shown in various colours) are within the intramembranous regions of the protein.

5.2.2 Generation of Transgenic Plants

For further investigation of the putative SUMO sites, *FLS2* mutant seeds were generated by another lab member; Dr Beatriz Orosa, and donated as part of a collaboration of projects. These seeds included: Two *fls2* knockout lines with complementation of the wild-type *FLS2* & the mutated SUMOylation site *FLS2*K/R cDNA under the FLS native promoter pFLS:FLS2wt and pFLS:FLS2K/R. The FLS2K/R is mutated to prevent SUMO attachment at the predicted lysine residue. These transgenic plants were checked and verified, using multiple lines of each genotype by Dr O'rosa (data not provided).

The *cerk1-1* transgenic seeds, harbouring Ds transposon inserts for the *cerk1-1* knockouts (Figure 5.03), were obtained from the Arabidopsis Seed Bank (Scholl et al., 2000). These seeds were homozygote knockouts for *cerk1* as confirmed by PCR analysis (Figure 5.03). No DNA fragment was amplified from the PCR reaction that used cDNA from the *cerk1-1* plants, however the col-0 cDNA produced a band at the expected size for the *CERK1* coding region fragment. The Ds transposon inserted into the region that codes for the kinase domain of the DNA of the *CERK1* gene (Figure 5.03).

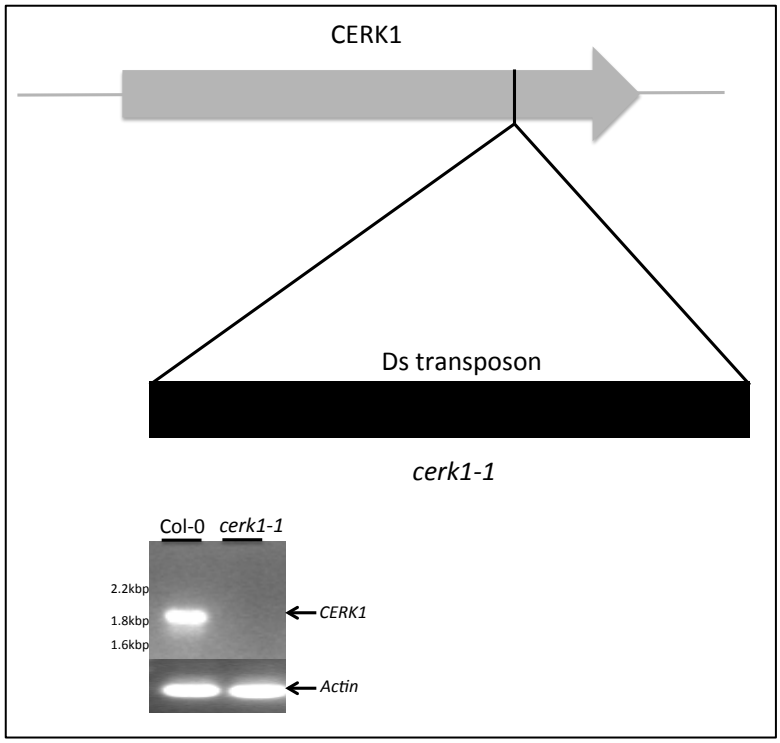

Figure 5.03. Confirmation of insertational knockout mutation of *cerk1-1.*
A diagrammatic representation of the *CERK1* gene showing the position of the Ds transposon insertion in the region that encodes for the kinase domain. Confirmation of the knockout was achieve by extracting the RNA from col-0 and cerk1-1 plants, synthesising cDNA and performing a PCR reaction using primers for the *CERK1* coding region. The figure shows that a section of DNA is amplified in the 'Col-0' lane at the size expected for the *CERK1* coding region. This is not seen in the lane '*cerk1-1*', providing evidence that the *CERK1* gene is not transcribed in *cerk1-1* plants. Actin primers were used as a control for the cDNA and for the PCR reaction.

The *CERK1* gene-coding region was cloned from *Arabidopsis thaliana* cDNA using the method as previously described (3.3.), using the primers shown in table 5.1. The cloned fragment was then gel extracted and used in a pENTR/DTOPO reaction to generate the pENTR plasmid carrying the *CERK1* cDNA fragment. The plasmid was transformed into *E.coli* DH5-alpha cells and plasmid DNA collected as described in (3.3.). The generation of complementation mutants in the *cerk1-1* knockout background was attempted using the foral dip method. Both wild-type *CERK1* (*CERK1*) and *CERK1*-K74R, K276K, K495R (*CERK1*-3K/R) were cloned into a vector that over-expressed the respective protein via the 35S promoter. This was done using the pEarleyGate104 vector and transformed into *Agrobacteria*. The

plants were then grown to maturation and seeds collected. The transgenic seeds were resistant to the antibiotic BASTA and therefore positive selection was achieved by sowing seeds on soil watered with 1:1000 BASTA to water mix. Seedlings which grew normally on this selection were allowed to mature and seeds collected for segregation assays. Plants that produced transgenic seedlings at a 1:3 (1:2:1) ratio, were assayed by growth on BASTA media and the positive seedlings grown to maturity. Only *CERK1*-3K/R complimenting plants were successfully selected in the screening process and no *CERK1* complimenting plants survived the BASTA treatment (Figure 5.04). Due to a time restraint, this procedure was not attempted again instead the *CERK1*-3K/R seedlings were used along side col-0 and *cerk1-1* knockout plants for further experiments.

Gene Name	Foward primer	Reverse Primer	Expected fragment size
At1g47740 gene	ATGTTGAACGG AAAAGAAGAGC	CCTTTCTTTCAA GAGCTGCT	840
At1g47740 cDNA	GTTCCGAAATC GGCAAAAT	ATTTTGCCGATT TCGGAAC	279
CERK1 gene	caccATGAAGCTAAA GATTTCTCT	AGAGAAATCTTTAG CTTAT	1854
CERK1 cDNA	GTTCCGAAATCGGC AAAAT	ATTTTGCCGATTTC GGAAC	253

Table 5.1. Primer sequences for *At1g47740* and *CERK1* plants.
The sequences of the DNA primers used for identification of the At1g47740 gene and the TDNA insertion cassette. The expected DNA fragment size is indicated on the right hand column.

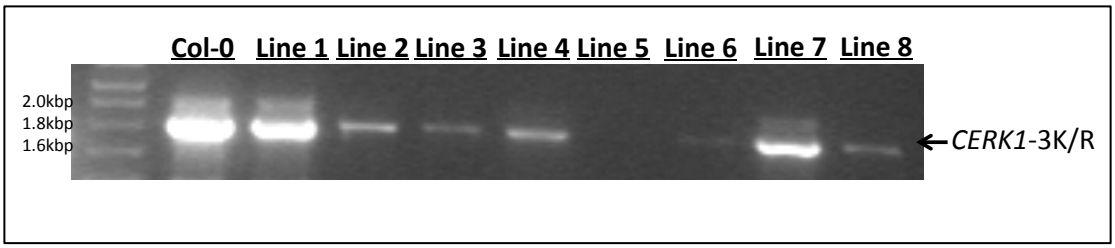

Figure 5.04. Genotyping of *cerk1-1* plants dipped with *CERK1*-3K/R construct.
After positive selection on resistance media, seedlings were grown for 10 day and then DNA was extracted from leaf discs. The genomic DNA extracted from each plant was normalised to that extracted from col-0 DNA. PCR was performed on the extracted DNA using primers for the *CERK1* coding region. The figure shows the results of col-0 and 8 transgenic lines. Lines 2, 3, 4, 6 and 8 appear to have a less intense band than col-0, line 5 has no band, and lines 1 and 7 have bands with intensity more similar to col-0. The intensity of the band correlates with gene copy number, in this case, it would appear lines 1 and 7 are likely carrying 2 copies and are therefore homozygous, where as lines 2, 3, 4, 6 and 8 have a single gene and are there for heterozygous. The *cerk1-1* lines were not used in this experiment the as non-transformed line 5, served as a negative control.

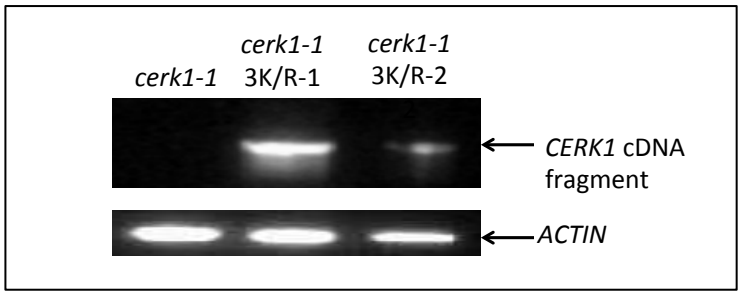

Figure 5.05. Transcript of *CERK1*-3K/R is detected by PCR in the *cerk1-1* background.
RNA from *cerk1-1, cerk1-1*-3K/R lines 1 and 7 was extracted and used to make cDNA. The cDNA was used as a template and in a PCR reaction using CERK1 primers. No transcript was detected in the cerk1-1 as expected, line 1 of the CERK1-3K/R dipped seedlings displayed a bright band, and line 7 displayed a duller band.

In addition to the above, *At1g47740*-KO plants were used to generate complementation lines harbouring the *At1g47740* gene under a 35S promoter. This was to see if the phenotypic changes in the *At1g47740*-KO plants could be rescued by the insertion of the *At1g47740* cDNA. The procedure was carried out in the same way as described above and 7 lines of *At1g47740*-complementing genotype were generated and verified with the expected gene fragment present in all lines tested (Figure 5.06). The band generated in col-0 is larger than that from the transgenic seedlings, this was expected as the primers used for DNA amplification spanned an intronic region. The intron region is not included in the cloning of the *At1g47740* gene fragment from cDNA, thus the col-0 DNA fragment is larger than that of the transgene. From this screen, 2 lines were carried forward for further experiments, from now referred to as *At1g47740*-COMP1 and *At1g47740*-COMP2.

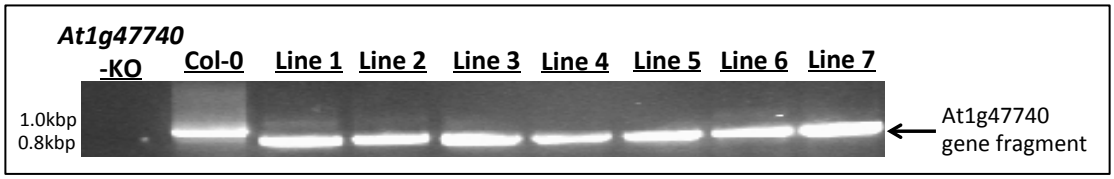

Figure 5.06. Genotyping of *At1g47740*-KO seeds dipped with a vector containing the *At1g47740* gene coding region. After positive selection on resistance media, seedlings were grown for 10 day and then DNA was extracted from leaf discs. The genomic DNA extracted from each plant was normalised to that extracted from col-0 DNA. PCR was performed on the extracted DNA using primers for the *At1g47740* coding region. All lines tested positive for the At1g47740 gene fragment. The gene fragment amplified from col-0 DNA was larger on the gel than in the transgenic plants, this is due to the intronic regions within the genomic DNA that is not present in the transgenic lines.

The *At1g47740*-COMP1 and COMP2 lines were used with *At1g47740*-KO and col-0 for RNA extraction and cDNA synthesis as previously described (1.5). This allowed verification of transcription of the transgene and for comparison of transcription levels in normal non-stressed state. Figure 5.07 shows the results of the PCR reactions in which transcription of the transgene was verified along with col-0, *At1g47740*-KO was used to show no transcription was present in the genotype prior to floral dipping.

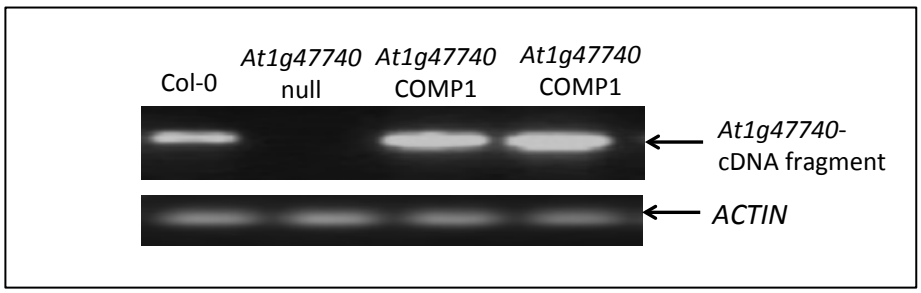

Figure 5.07. Transcript of *At1g47740* is detected by PCR in the *At1g47740*-COMP1 & -COMP2 background.
RNA from col-0, *At1g47740*-KO *At1g47740*-COMP1 and COMP2 was extracted and used to make cDNA. The cDNA was used as a template and in a PCR reaction using *At1g47740* primers and results analysed on an agarose gel. No transcript was detected in the *At1g47740*-KO as expected, whereas col-0, *At1g47740*-COMP1 and COMP2 all displayed bands at the expected size for the *At1g47740* transcript.

5.2.3 Expression of Recombinant Protein

For verification of the stability of the protein produced by the transgenic seedlings, western blot analyses were performed. Seedlings containing the transgenes for pFLS:FLS2wt and pFLS:FLS2-K/R in the *fls2* background, were grown on plates containing 1/2MS, agar media for 9 days and then transferred to a 6 well-plate containing liquid 1/2MS media for 24 hours. Seedlings were then collected, dried on a paper towel, weighed and flash frozen in liquid nitrogen. Samples were then ground to a fine powder using a mortar and pestle, and protein extraction buffer added. The sample was then spun in a centrifuge, supernatant collected and diluted with 4x SDS buffer. The samples were then heated at 98 degrees for 3 minutes and analysed by SDS-PAGE. Figure 5.08 shows the results for the immunoblot in which the FLS2 protein was attached to a GFP tag. In both pFLS:FLS2wt and pFLS:FLS2-K/R extracts, protein was detected at the expected size of the FLS2 protein plus GFP. This confirmed that both transgenic lines produced equally stable protein and that the level of protein produced in each, was approximately the same (Figure 5.08A). In order to determine if the mutated FLS2 was able to undergo SUMOylation co-immunoprecipitation assays were performed by the project collaborator Beatriz O'rosa, and showed that the pFLS:FLS2-wt can be pulled down with SUMO1 attached where as the K to R form showed no SUMO present when equal amounts of protein were used in the assay (Figure 5.08B) (Data kindly donated by Beatriz O'rosa).

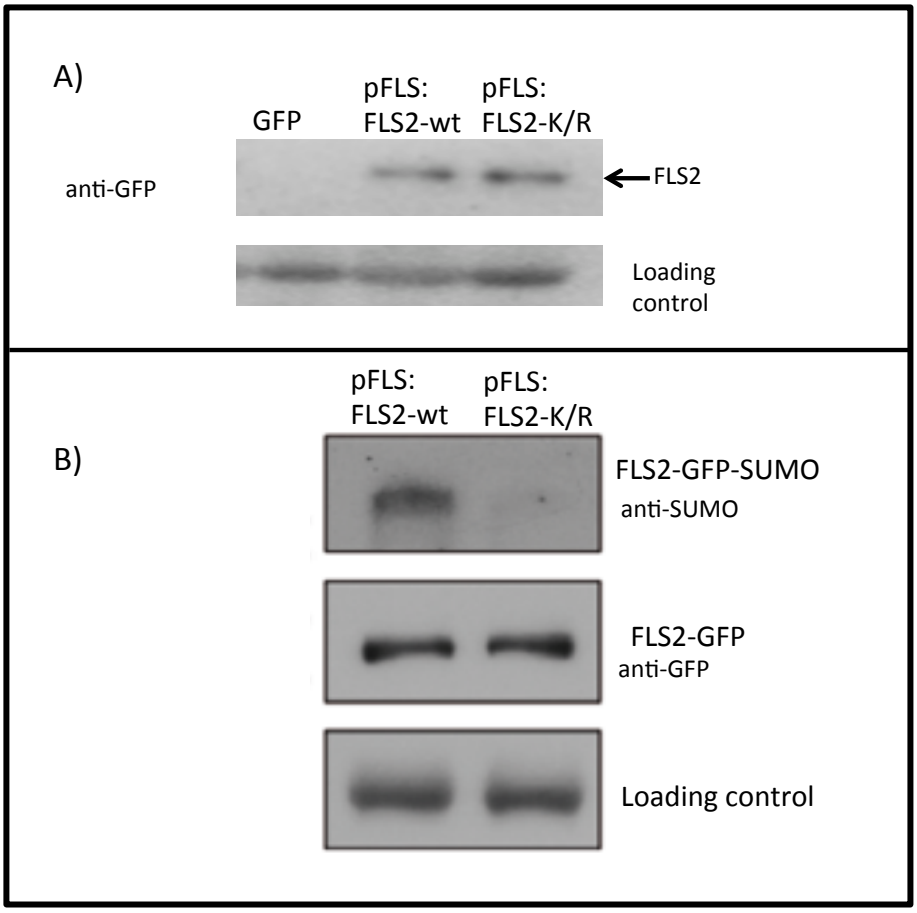

Figure 5.08. Expression of transgenic FLS2 is equal in both pFLS:FLS2wt and pFLS:FLS2-K/R, however SUMOylation is only seen in the pFLS:FLS2-wt.
Forty seedlings from each genotype, empty vector (GFP) control, pFLS:FLS2wt and pFLS:FLS-K/R were grown on 1/2MS media. After 10 days, seedlings were collected and flash frozen readied for protein extraction. The total protein content was extracted and then separated by SDS-PAGE and immunoblot analysed using antibodies raised against GFP. A) Protein bands appeared at the expected size of FLS2 plus GFP in both pFLS:FLS2wt and pFLS:FLS2-K/R samples, but not in the GPF empty vector plants. Furthermore both bands appear at approximately the same intensity showing protein in both wild-type and the K/R mutant are expressed equally and are equally stable. B) pFLS:FLS2-K/R shows no SUMOylation (upper panel) compared to the pFLS2:FLS2-wt, the middle panel shows equal amounts of protein were pulled down in this assay.

To determine if the CERK1 protein was equally stable in both wild-type and 3K/R form, transient expression assays were set-up using *Nicotiana tobaccum*. Using the pEarlyGate104 expression vector, which carries an N-terminal YFP tag, the coding region for the wild-type *CERK1* (CERK1-wt), and the mutant version that contains the triple K to R mutation (CERK1-3K/R)

were transformed into *Agrobacterium* along with the empty vector control (YFP). The transgenic bacteria were then grown overnight, spun down and diluted in $MgCl_2$ to an O.D.600 of 0.4. The re-suspended bacteria were then syringe infiltrated into 4 week old *Nicotiana benthamiana* leaves and the plants were allowed to grow for 3 more days. After 3 days, the leaf samples were removed weighed and flash frozen in liquid nitrogen. Leaf samples were ground to a fine powder and total protein was extracted. Samples were then analysed by SDS-PAGE and western blot. Probing with GFP antibody, protein at the correct estimated size was found in both CERK1-wt and CERK1-3K/R samples (Figure 5.09). The protein levels in both samples appeared in relatively equal proportion, although further experimentation would allow this to be verified.

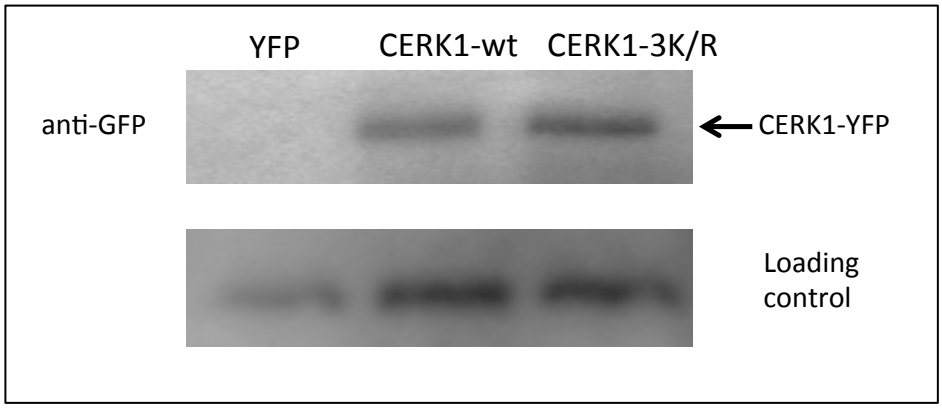

Figure 5.09. Transient expression of CERK1-YFP and CERK1-3K/R-YFP reveals consistent stability between both recombinant proteins.
Agrobacteria which contained either, no transgene (YFP), unaltered coding DNA for CERK1 (CERK1-wt) or the triple K to R mutated CERK1 coding DNA (CERK1-3K/R), were syringe infiltrated into *Nicotiana tabaccum* leafs. After 3 days the leaves were harvested and total protein extracted. The samples were analysed by western blot and probed using anti-GFP antibodies. In both CERK1-wt and CERK1-3K/R the protein was seen at the correct estimated size of the recombinant CERK1 plus YFP. The protein in both samples appears to be equally stable.

The stability of At1g47740 protein in both the *At1g47740*-COMP1 and COMP2 lines was determined in both the absence and presence of the elicitor peptide flg22. Seedlings were grown on plates containing 1/2MS, agar media for 9 days and then transferred to a 6 well-plate containing liquid 1/2MS media for 24 hours. Each genotype was then treated with water or flg22 30 minutes prior to harvesting samples. Seedlings were collected, dried on a paper towel, weighed and flash frozen in liquid nitrogen. Samples were then ground to a fine powder using a mortar and pestle, and protein extraction buffer added. The sample was spun in a centrifuge, supernatant collected and diluted with 4x SDS buffer. The samples were then heated at 98oC for 3 minutes and analysed by SDS-PAGE and immunoblotting using antibodies raised against GFP. In both *At1g47740*-COMP1 and COMP2 lines treated with water the recombinant protein was seen, however it appeared that *At1g47740*-COMP1 expressed the protein in more abundance than *At1g47740*-COMP2. In the samples treated with flg22, the recombinant At1g47740 protein was less stable in both lines and is only just visible (Figure 5.10).

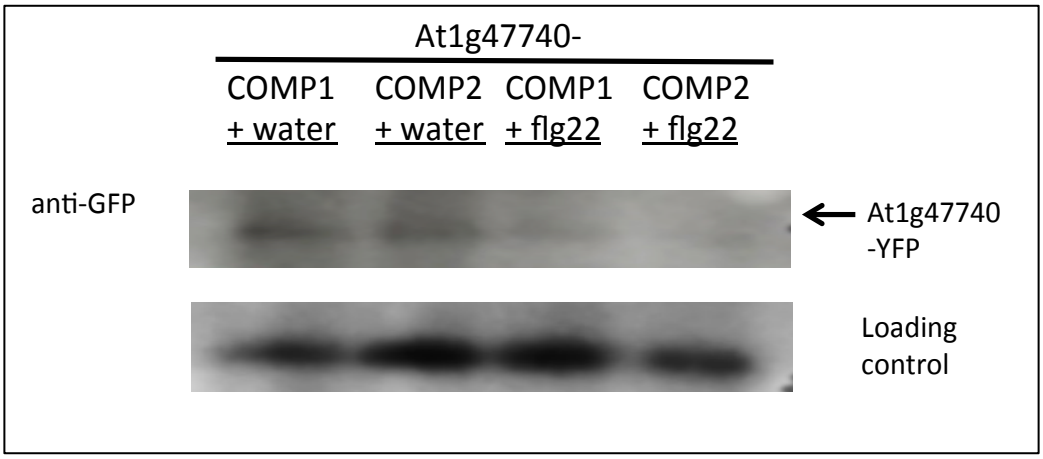

Figure 5.10. The stability of recombinant At1g47740 protein in transgenic plants is affected by the addition flg22 peptide.
40 transgenic plants of the genotypes *At1g47740*-COMP1 and *At1g47740*-COMP2, were grown on 1/2MS for 10 days and treated with water (+water) or 250nM flg22 (+flg22). Thirty minutes after treatment, plants were flash frozen and total protein was extracted. The samples were mixed with 4x SDS buffer and analysed by SDS-PAGE, immunoblotting was then carried out using antibodies raised against GFP. The samples that were extracted from plants treated with water showed more recombinant At1g47740 protein than samples treated with flg22. Given that the loading control shows less total protein in the 'At1g47740-COMP1 +water' sample, it appears that this line expresses the recombinant At1g47740 at a higher level than *At1g47740*-COMP2.

5.2.4 Root Growth Assay For Comparison of pFLS:FLS2wt And pFLS:FLS2-K/R on Media Containing flg22

The check if the plants response to the elicitor flg22 is affected in the transgenic lines, *fls2*, pFLS:FLS2wt and pFLS:FLS2-K/R seedlings were grown along with col-0. After 3 days growth on 1/2MS media, the seedlings were transferred onto media containing 250nM of flg22 peptide or fresh 1/2MS and grown for 7 days. Plants complimented with wild-type FLS2 (pFLS:FLSwt) showed a similar root growth as the wild-type plants (Figure 5.11). Plants complemented with the pFLS:FLS2-K/R construct showed no reduction in root growth compared to wild-type and were more similar to *fls2* knockout plants (Figure 5.11). Thirty seedlings from each genotype were photographed and roots measured digitally using imageJ software. An average of each genotype was calculated and plotted on a bar graph (Figure 5.12). The results show that the reduction of root growth in col-0 plants when treated with flg22 compared to no treatment, was almost identical to that seen in pFLS:FLS2wt plants. Furthermore, *fls* plants showed no reduction in growth on flg22 media, and pFLS:FLS2-K/R displayed the same with both genotypes seemingly unaffected by the presence of flg22 (Figure 5.12). Statistical significance was calculated using a two-tailed student T-Test and the resulting p-values revealed col-0 and pFLS:FLS2wt had a p-value of <0.001, whereas *fls2* and pFLS;FLS2-K/R showed significant differences with p-values >0.05 (Table2).

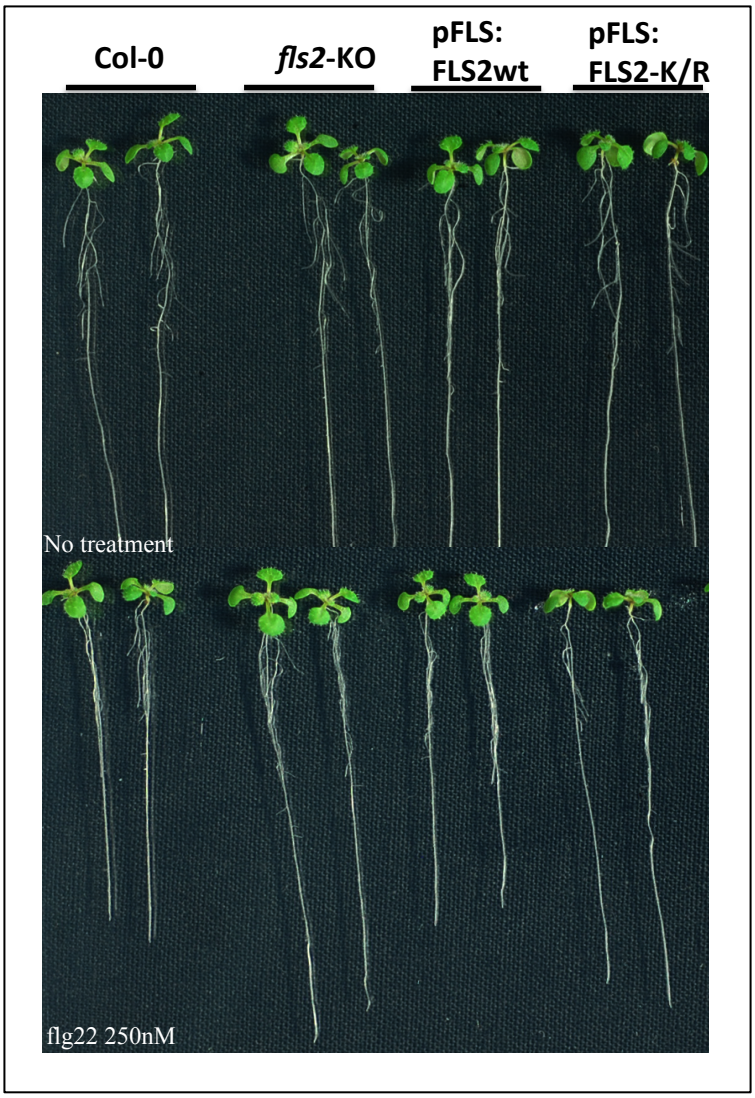

Figure 5.11. Seedlings carrying mutation of *FLS2* predicted SUMO site respond to flg22 treatment in a similar fashion to *fls2* knockout plants
Seedlings were grown for 3 days on 1/2MS media and then transferred to either 1/2MS media containing 250nM flg22 or 1/2MS media only. After 7 days on the new media seedlings were photographed and roots measure. The picture represents the seedlings in the mean for each genotype measured. All plants look similar in 1/2MS media conditions (top panel), however the seedlings treated with flg22 showed differences in root length (bottom panel). Col-0 and pFLS:FLS2wt plants displayed similar reduced growth in root length, where as *fls2* and pFLS:FLS2-K/R appeared to be insensitive to the treatment.

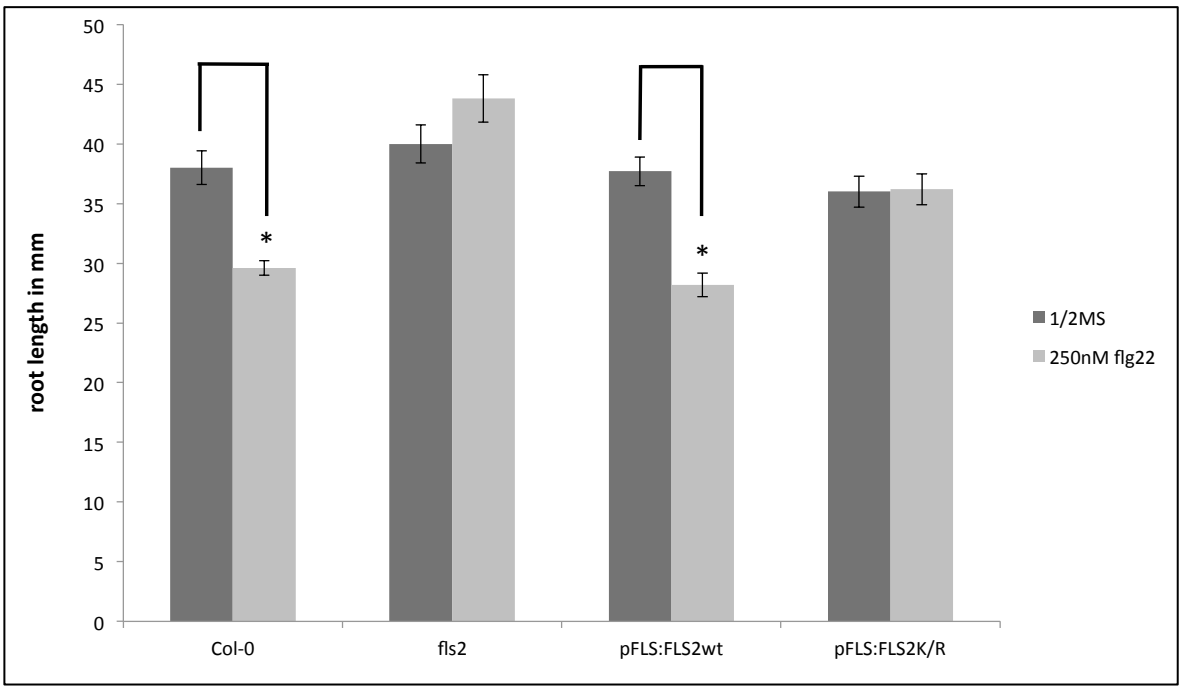

Figure 5.12. Quantification of the average root length of, plants grown on 1/2 MS media versus 1/2MS media containing flg22.
Thirty seedlings of each genotype were grown for 3 days on 1/2MS media and were transferred to either 1/2MS media containing 250nM flg22 or 1/2MS media only. After 7 days on the new media seedlings were photographed and roots were measured using imageJ software. The average root length of each genotype on both media were calculated and plotted on a bar graph. A comparison of each genotype on both media types was calculated for statistical significance using a two-tailed student T-Test, p-value of <0.001 (*).

Genotype	1/2MS vs FLG22 p-value
col-0	$2.15x10^{-5}$
fls2	$1.16x10^{-1}$
pFLS:FLS2wt	$9.94x10^{-4}$
pFLS:FLS2-K/R	$8.07x10^{-1}$

Table 5.2. Statistical analysis of root length using two-tailed T-test reveals significant p-values for pFLS:FLSwt but not pFLS:FLS2-K/R lines. Root length measurements for each genotype were taken on both 12MS media and 1/2MS media plus flg22. The root length from both treatments were compared and statistically analysed using a two-tailed T-test. The results show that both col-0 and pFLS:FLS2wt had a significant difference in root length when comparing the two treatments. Contrary to this, the pFLS:FLS2-K/R and fls2 lines showed no statistical difference in root length between the treatments.

5.2.5. Assay For Root Growth Comparison of *At1g47740* Complimenting Lines When Treated With Flg22

To determine the ability of the *At1g47740* gene coding region to rescue the phenotype of *At1g47740*-KO plants, the procedure for ascertaining differences in root length as previously described (5.2.4) was repeated using the *At1g47740*-COMP1 and COMP2 lines, along with col-0 and *At1g47740*-KO. After 3 days growth on 1/2MS media, the seedlings were transferred onto media containing 250nM of flg22 peptide or fresh 1/2MS and grown for 7 days. These results were then averaged for each genotype on both 1/2MS media and 1/2MS media containing flg22, and plants representing the mean were photographed together for comparison (Figure 5.13). As previously shown, *At1g47740*-KO plants displayed a more severe root growth inhibition on media containing flg22 than col-0. However, in the *At1g47740*-COMP1 and COMP2 lines, this growth inhibition was abolished and like in the *At1g47740*-OX1 and OX2 lines, no significant difference in root length could be seen (Figure 5.13). The average root length of each genotype is shown in Figure 5.14, and clearly shows that flg22 has little or no affect on *At1g47740*-COMP1 and COMP2 plants. In addition, two-tailed T-Test analysis of each genotype was calculated comparing flg22 treatment with non treated plants (Table3). This analysis confirmed that the difference in root length in *At1g47740*-KO plants was greater than the difference in col-0 plants when comparing the 2 treatments (<0.0001 and <0.001 respectively), and there was no difference out side normal distribution for the complimenting lines (Table 3).

Figure 5.13. The At1g47740 transgene can rescue the phenotype of At1g47740-KO plants treated with flg22.
Seedlings were grown for 3 days on 1/2MS media and then transferred to either 1/2MS media containing 250nM flg22 or 1/2MS media only. After 7 days on the new media seedlings were photographed and roots measure. The picture represents the seedlings in the mean for each genotype measured. All plants look similar in 1/2MS media conditions (top panel), however the seedlings treated with flg22 showed differences in root length (bottom panel). *At1g47740*-KO plants displayed reduced growth in root length greater than col-0, where as *At1g47740*-COMP1 and COMP2 appeared to be insensitive to the treatment.

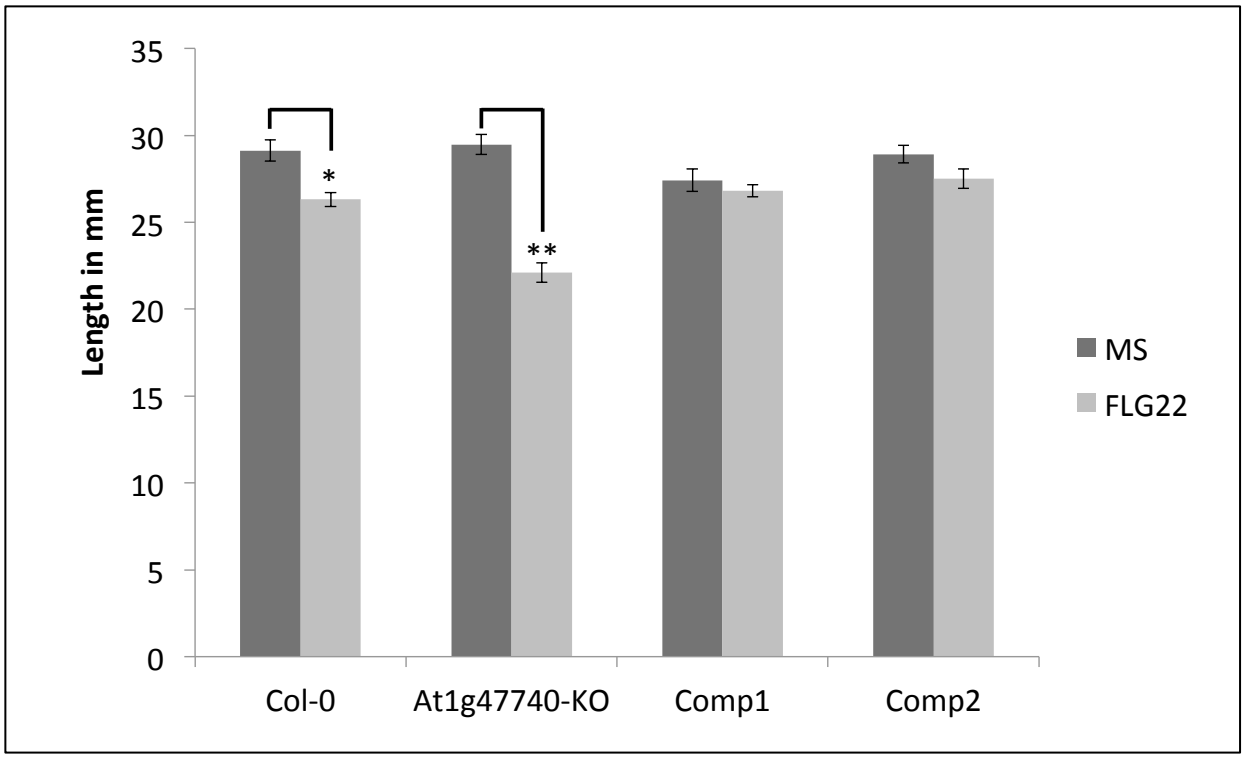

Figure 5.14. Quantification of the average root length of, plants grown on 1/2 MS media versus 1/2MS media containing flg22.
Thirty seedlings of each genotype were grown for 3 days on 1/2MS media and were transferred to either 1/2MS media containing 250nM flg22 or 1/2MS media only. After 7 days on the new media seedlings were photographed and roots were measured using imageJ software. The average root length of each genotype on both media were calculated and plotted on a bar graph. A comparison of each genotype on both media types was calculated for statistical significance using a two-tailed student T-Test, p-value of <0.001 (*) and <0.0001.

Genotype	1/2MS vs FLG22 p-value
col-0	0.0004782
At1g47740-KO	9.54×10^{-9}
At1g47740-COMP1	0.433815896
At1g47740-COMP2	0.07611571

Table 5.3. Statistical analysis of root length using two-tailed T-test reveals no significant p-values *At1g47740* complimenting lines. Root length measurements for each genotype were taken on both 12MS media and 1/2MS media plus flg22. The root length from both treatments were compared and statistically analysed using a two-tailed T-test. The results show that both col-0 and pFLS:FLS2wt had a significant difference in root length when comparing the two treatments. Contrary to this, the pFLS:FLS2-K/R and fls2 lines showed no statistical difference in root length between the treatments.

5.2.6 Effects Of SUMO-Site Mutations on Transient ROS Production Under Pathogenic Elicitor Induction

In order to ascertain whether the ROS production caused by the elicitors flg22 and chitin is influenced by the SUMO site mutations in FLS2 and CERK1, bioassays for detection of ROS were setup as follows;

$1cm^2$ leaf discs were cut from 4 week old plants and incubated in water for 24 hours in a 24 well plate. The water was then removed and a hydrogen-peroxidase and luminal solution added prior to 250nM of flg22 or 1uM of chitin. The plate was then put in a dark box with a photon counting camera for 80 minutes, all experiments were repeated at least 3 times and results were consistent.

To determine the affects of the *FLS2* mutations in response to flg22, pFLS:FLS2wt, pFLS:FLS2-K/R and *fls2* leaf disc were treated by flg22 and ROS production was recorded (Figure 5.15). Plants complemented with wild-type *FLS2* coding DNA produced ROS in an expected fashion with a peak of production around 22 minutes after flg22 treatment. The typical 'bell curve' seen in col-0 plants (Figure 4.09) was produced by the pFLS;FLS2wt plants, however the pFLS:FLS2-K/R plants did not respond in a same way (Figure 5.15). The pFLS:FLS2-K/R plants produced ROS in a very low abundance and this was more comparable to *fls2* than col-0 plants. However, the pFLS:FLS2-K/R leaf discs did show a slight elevation in ROS production over the same time period as the pFLS:FLS2wt plants, unlike *fls2*, which showed no elevation over the same period. These results indicate that flg22 induced production of ROS in *Arabidopsis thaliana* is dependant on the SUMOylation of the FLS2 protein.

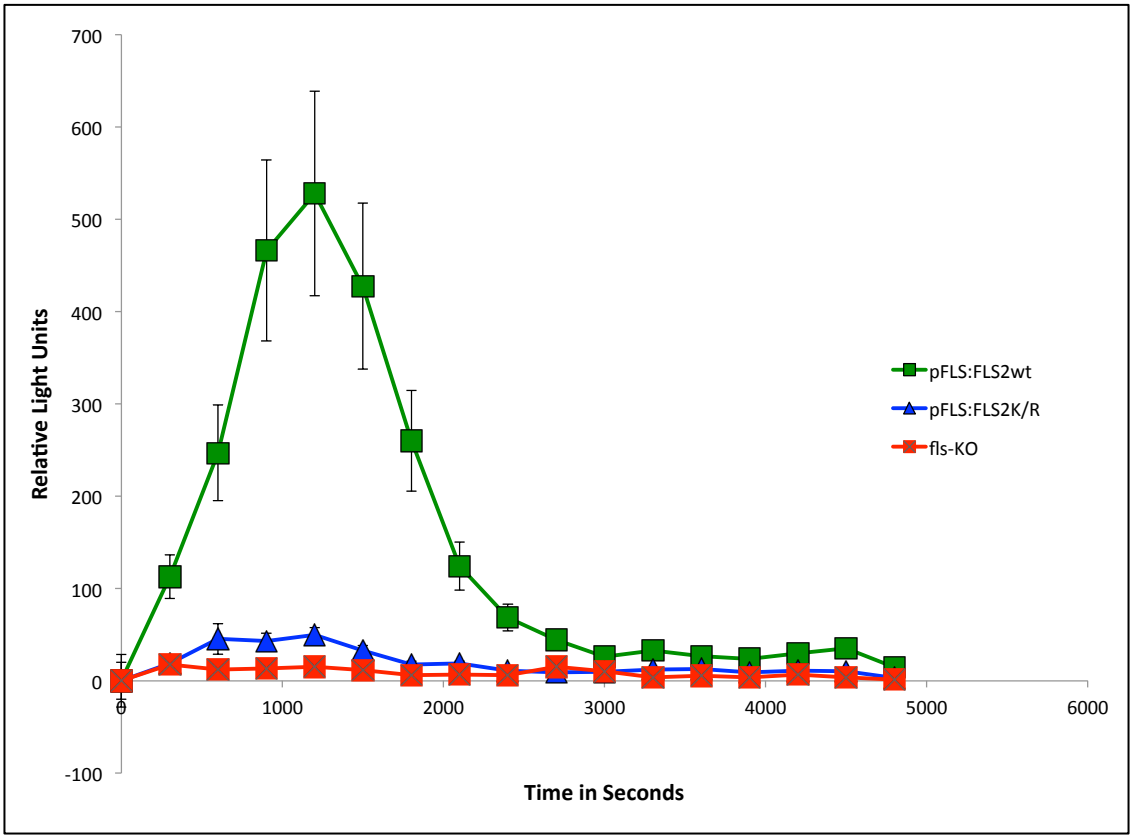

Figure 5.15. Flg22 induced ROS generation via the FLS2 receptor is SUMOylation dependent.
Transient ROS production in response to the bacterial elicitor flg22 in pFLS:FLS2wt (green-square), pFLS:FLS2-K/R (blue-triangle) and *fls2* (red-cross) was measured using a luminal, hydrogen-peroxidase based solution and a photon counting camera. 9 leaf discs were cut from 6 plants of each genotype and placed in a 24 well plate in triplicate. The relative light units (RLUs) are shown on the x-axis whilst time in seconds is displayed on the y-axis. The induction time at which the ROS production peaked was around 22 minutes after flg22 addition. pFLS:FLS2wt produced a relatively high level of ROS compared to pFLS:FLS2-K/R and *fls2*, however a slight elevation on ROS levels is seen in pFLS:FLS2-K/R compared to *fls2*. Error bars show plus and minus SE as calculated by deviation of 3 different wells for each genotype. Experiments were repeated a minimum of 3 times and results were consistent.

Transient ROS production assay were set up to test if the elevated levels of ROS observed in *At1g47740*-KO plants, can be offset by the addition of the *At1g47740* transgene. The transgenic lines *At1g4770*-KO, *At1g47740*-COMP1 and COMP2, were tested along with col-0 and *fls2* plants, is the same way as described previously (5.2.6). In the *At1g47740*-KO line, ROS production was observed at over 3 fold that of col-0, however, the plants with the complimenting *At1g47740* transgene did not produce ROS at the same level

(Figure 5.16). Furthermore *At1g47740*-COMP1 showed higher levels of ROS than that of *At1g47740*-COMP2, with a 0.5 fold increase and no significant increase seen over col-0 respectively. Due to the decreased levels of ROS produced in these two complimenting lines, it is further evidence that the At1g47740 protein plays a major role in the negative regulation of the production of ROS during the plants response to the elicitor flg22.

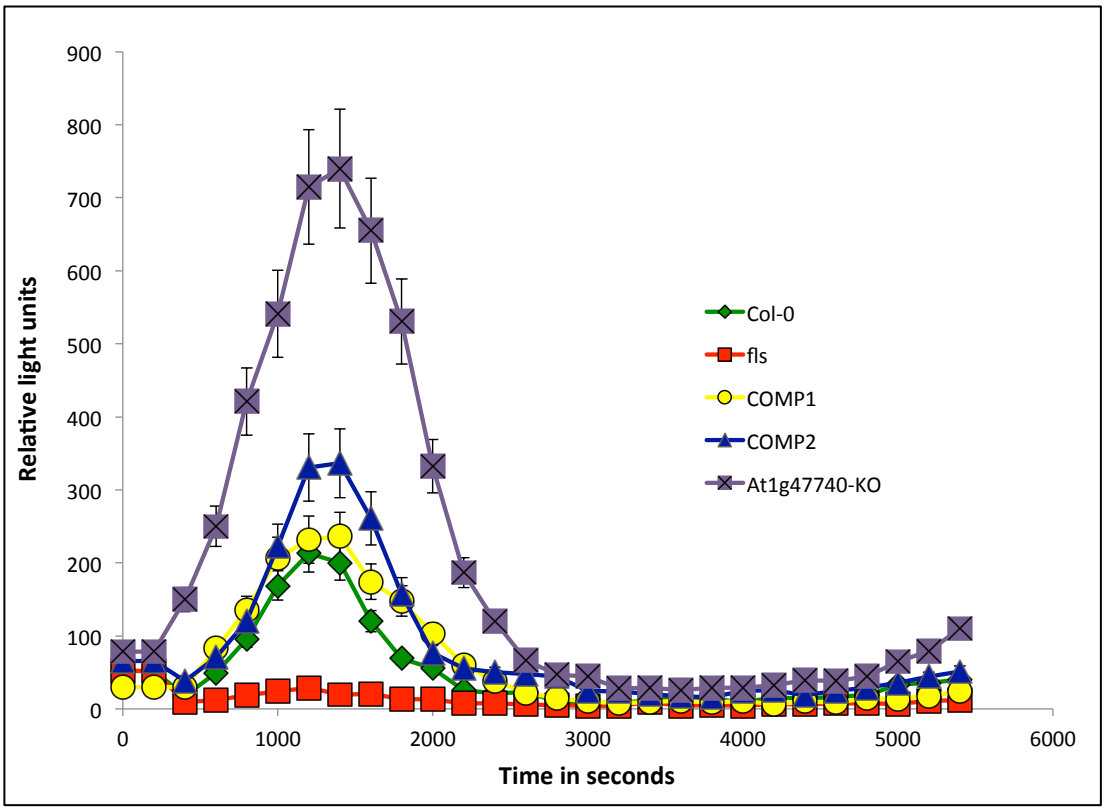

Figure 5.16. The elevated levels of ROS generation in *At1g47740*-KO plants is partially rescued by the *At1g47740* transgene
Transient ROS production in response to the bacterial elicitor flg22 in col-0 (green-diamond), *fls2* (red-square), *At1g47740*-KO (purple-cross), *At1g47740*-COMP1 (yellow-circle) and *At1g47740*-COMP2 (blue-triangle) was measured using a luminal, hydrogen-peroxidase based solution and a photon counting camera. 9 leaf discs were cut from 6 plants of each genotype and placed in a 24 well plate in triplicate. The relative light units (RLUs) are shown on the x-axis whilst time in seconds is displayed on the y-axis. The data shows that *At1g47740*-KO plants have a significantly higher level of ROS production compared to all other genotypes. The *At1g4770*-COMP1 and COMP2 lines showed a decrease in ROS production and are more comparable with col-0 than *At1g47740*-KO. Error bars show plus and minus SE as calculated by deviation of 3 different wells for each genotype. Experiments were repeated a minimum of 3 times and results were consistent.

To test the role SUMOylation plays in the CERK1 function, ROS assays were setup using leaf discs from the following plants; Col-0, *CERK1*-3K/R-1, *CERK1*-3K/R-2 and *cerk1-1.* Chitosan was used as the elicitor in this assay and the results showed a typical response from the col-0 plants with a sharp raise in ROS production after about 20 minutes and a prolonged return to basal levels, as also shown in Figure 4.13 (Figure 5.17). The *cerk1-1* leaf discs produced a much lower amount of ROS compared to col-0, and the onset of the initial burst was later than that seen in col-0 (Figure 5.17). Leaf discs from the *CERK1*-3K/R-1 and *CERK1*-3K/R -2 plants, showed a very similar ROS production profile to that seen in cerk1-1, however levels were very slightly higher in both these lines. Without a reconstituted wild-type *CERK1* line, it is difficult to ascertain the affect the mutated SUMO sites had on the response to chitosan. However, the *CERK1*-3K/R lines did not produce ROS of a significant enhanced amount over that of the *cerk1-1* lines, and if the same trend from the FLS2 SUMO mutants was applied to the *CERK1*-3K/R lines, the expected result would be what is observed in Figure 5.16 (i.e the K/R mutation reduces the plants ability to produce ROS derived from the elicitor chitosan). This does not mean the results would follow the same trend, but it is very likely that the reconstituted wild-type CERK1 would return ROS levels closer to that observed in col-0, based on the results of the *FLS2* mutants (Figure 5.15).

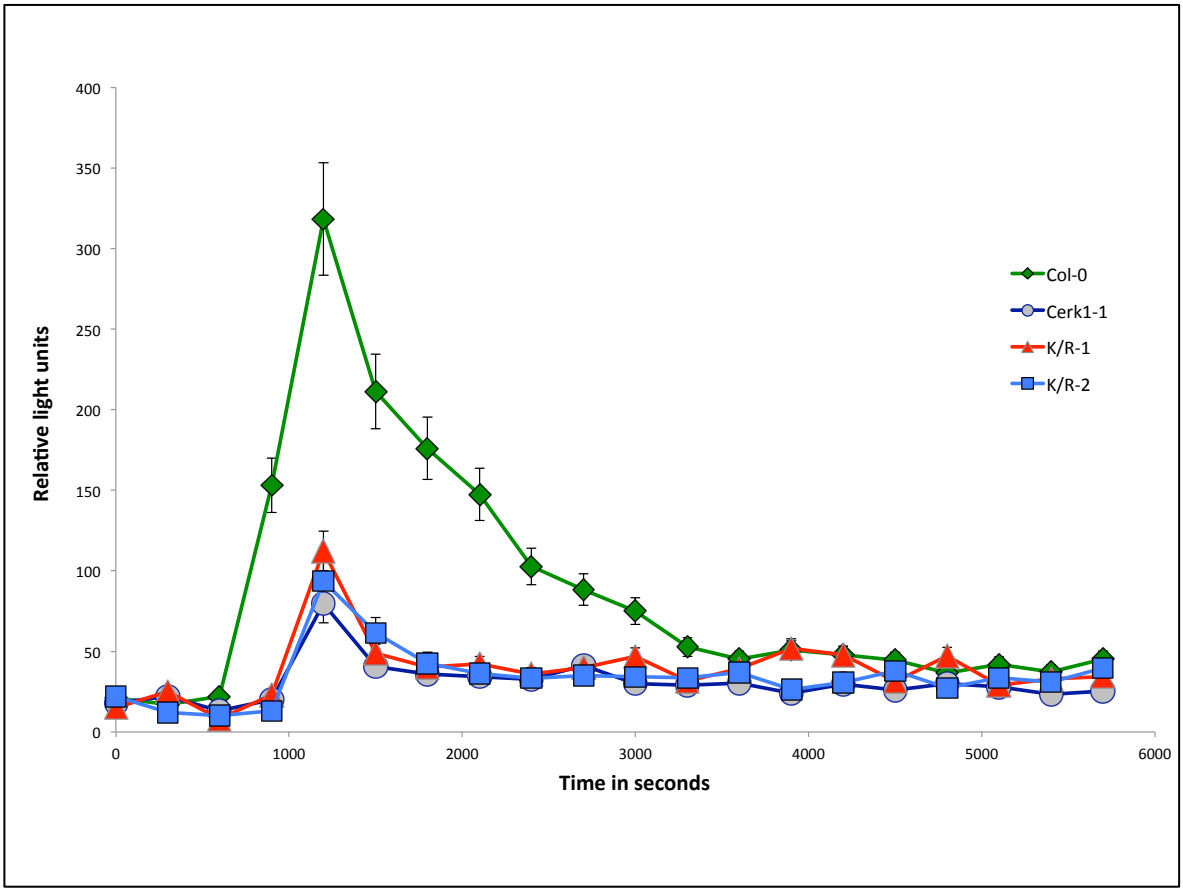

Figure 5.17. CERK1-3K/R does not recuse the ROS induction in *cerk1-1* plants.
Transient ROS production in response to the bacterial elicitor chitosan in col-0 (green-diamond), *cerk1-1* (grey-circle), CERK1-3K/R-1 (red-triangle) and CERK1-3K/R-2 (blue-square) was measured using a luminal, hydrogen-peroxidase based solution and a photon counting camera. 9 leaf discs were cut from 6 plants of each genotype and placed in a 24 well plate in triplicate. The relative light units (RLUs) are shown on the x-axis whilst time in seconds is displayed on the y-axis. Complementation of the mutated *CERK1*-3K/R construct failed to rescue the *cerk1-1* phenotype. ROS production from the 3K/R lines was comparable to *cerk1-1*. Error bars show the plus and minus SE as calculated by deviation of 3 different wells for each genotype. Experiments were repeated a minimum of 3 times and results were consistent.

5.3. Discussion

Collectively this chapter has discussed the possibility of identifying 2 target substrates for the DeSi3 protease At1g47740. The first step in verifying the putative target substrates was to screen the potential protein sequences for SUMOylation attachment sites. This step was crucial in the identification process, as it allowed almost instantaneous recognition of SUMO attachment motifs and gave credibility to the likelihood that the potential target protein was SUMOylated. Further validation using laboratory methods is of course necessary, however this method made sure that the identified target substrate was at least likely to be SUMOylated whether or not a target of the Atg47740 protease remained to be determined. Due to time limitation, and the bioinformatics supporting the SUMO attachment sites, only FLS2 and CERK1 were investigated as potential targets. To add biological data to support the theory that FLS2 and CERK1 are SUMO targets, it was necessary to generate or acquire the appropriate mutant plants. Dr Beatriz Orosa, a postdoctoral researcher from within the research group, also worked on FLS2, and generated the mutant *Arabidopsis* plants used in this research. The data verifying the genetics of these mutant plants was not included in this thesis, as the author did not generate it. However, it is provided in the supplement data on the collaborated forthcoming publication; Orosa and Yates *et al*., 2017 (*under review*). The *FLS2* and *CERK1* mutant plants used in this research allowed for greater understanding of the role SUMOylation plays in the function of these proteins. Mutating the identified SUMO site, K to R, has been shown to make proteins unable to be SUMOylated (Conti *et al*., 2014). This important genetic tool provided a strong lead that the SUMOylation of FLS2 and CERK1 was key to activating the cellular defence responses upon elicitor detection. An interesting result observed in this research was that both wild-type and K-to-R mutants of FLS2 and CERK1 were equally stable when expressed in *Arabidopsis thaliana* and *Nicotiana benthamiana* respectively (Figures 5.08 and 5.09). This led to the conclusion that SUMOylation of both of these proteins did not affect the stability of the protein and therefore SUMOylation had a more mechanistic role than controlling protein fate. The phenotypic responses of the mutant plants provided further clues of the

mechanistic function SUMO has on these receptors. Due to the similarities in the phenotypic response to flg22 in both root growth and ROS production, it could be concluded that mutation of FLS2 SUMO attachment site mutant, i.e. pFLS:FLS2-K/R, makes the plants respond effectively like *fls2* knockouts. Due to the lack of the reconstituted *CERK1* wild-type gene in the *cerk1-1* background, it is not possible to determine if the same is true for CERK1 K-to-R mutants, although it remains to be seen if it works in parallel to the FLS2.

Further to this, an important part of this research was to provide evidence that the *At14g47740* gene or its protein product is directly linked to the phenotype shown in previous chapters. This question was raised due to only having one knockout line for this genotype, and because no articles have been published using this line. It was therefore imperative to show that complimenting this genotype with the *At1g47740* coding DNA, would rescue the phenotype of the *At1g47740*-KO plants. It would have been beneficial to have cloned the *At1g47740* promoter region and have the transgene expressing under its native promoter instead of the constitutive, 35S promoter. However, the cloning of the identified promoter region was attempted several times using different primer sets, but the DNA region would not amplify. This may be due to tight binding of DNA-interacting proteins in the genomic area spanning the promoter region, or tightly wound DNA-histone stretches in this region. Due to the unsuccessful cloning of the promoter, it was decided that expressing the transgene under the 35S promoter in the *At1g47740*-KO background, should at least partially elevated the phenotypic difference, if the *At1g47740* gene was responsible for the phenotype.

After making the transgenic lines; *At1g47740*-COMP1 and COMP2, it was important to check that the transgene protein product was being expressed, like was shown for the *At1g47740*-OX1 and OX2 lines (Figure 4.03). It was more difficult to see the protein band in the At1g47740-COMP1 and COMP2 line than the OX1 and OX2 lines, this could have been due to the plants gene silencing mechanisms coming into play, or simply the region in which the transgene inserted into the genome (Figure 5.11). The transgene appeared to be stable despite the difficulties in seeing the protein, and the phenotype

testing administered herein supported strongly that the *At1g47740*-COMP1 and COMP2 lines were phenotypically more like col-0 than *At1g47740*-KO plants (Figures 5.13, 5.14 and 5.16). In conclusion this chapter has shown that the At1g47740 protease is responsible for the negative regulation of defence, particularly in the early stages of infection and in pathogen perception. Mutation of the predicted SUMO sites in of the LKRs, FLS2 and CERK1, prevents the proteins ability to signal properly upon PAMP recognition. No evidence provided here confirms interaction between At1g47740 and FLS2 or At1g47740 and CERK1, due to this further analysis is required for validation that FLS2 and/or CERK1 is a target of the DeSi3 protein. However, strong evidence is provided that FLS2 and CERK1 are SUMO substrates and SUMOyation plays a crital role in their respective functions. All three proteins appear in physical proximity, and *At1g47740*-KO plants showed opposing phenotypes in response to flg22 than that of *fls2* plants.

Chapter 6
Discussion

6.1 Summary

The role that SUMOylation plays in regulating cellular responses in plants is only partially understood, this is because more research has focused on SUMO attachment, leaving SUMO removal a relatively understudied discipline. The aim of this thesis was to provide insight in to the field of deSUMOylation and provide evidence that this is a fundamental process critical for regulating cellular functions. Over the course of this research, several key advancements to the field of SUMOylation/deSUMOylation have been shown;

- Discovery of a new class of SUMO proteases in plants
- The first SUMO proteases localising outside of the nucleus
- The SUMO protease At1g47740 (DeSi3a) is a negative regulator of defence against pathogens.
- The 2 major LRR-RK's, FLS2 an CERK1, are SUMO substrates
- PAMP activated ROS production via the FLS2 receptor is SUMOylation dependant
- Mutation of the SUMO attachment site impacts FLS2 and CERK1 normal function

The implication of the results are further discussed below, by collating these outcomes and combining them in part with the work done by Dr Beatriz Orosa on FLS2 SUMOylation, a more complete analysis can be presented. The data provided by Dr Orosa will not be shown but can be found referenced in, the as yet unpublished article, '*The Immune response receptor FLS2 is SUMOylated in response to flg22'* (Orosa *et al*., 2017).

6.2 Discovery of DeSi proteases in plants

The results in chapter 3 showed that 8 proteins, within the *Arabidopsis thaliana* proteome, contained the signature DeSi catalytic domain previously identified in animal species (Shin *et al*.,2012). Due to the fact there was such a high number of proteins containing the catalytic domain in Arabidopsis, compared to the number found in animal species, phylogenetic analyse of the proteins was deemed necessary to understand more about this new class of proteases. The most striking result from the phylogenetic analysis was the fact that three sub-classes of DeSi could be characterised. Using human, dog and zerbrafish DeSi1 and DeSi2 as an out-group control, the analysis showed that one of these sub-classes appears to be unique to plants (Figures 3.04 and 3.05). Of the 3 sub-classes, only one of the DeSi protein, At3g07090, was of the subclass DeSi1, 2 in the sub-class DeSi2 and 5 classed DeSi3. This could be significant as it implies that if both animals and plants shared common DeSi ancestors (DeSi1 and DeSi2), the function they serve could be conserved. The resulting divergence from the original DeSi-type to create DeSi3 would likely be driven by changes in target substrates. As plants also have a larger expansion of the SUMO protein family than is seen in animal species, it is likely that plant species also have an expansion in the de-SUMO protein families (Hay 2005). If true, this implies that members of the DeSi3 subclass are likely to target other SUMO proteins as well as or instead of SUMO1. Beyond the scope of this research, a peptide-screening array or protein interaction screen for each protease may reveal if they interact more with certain classes of SUMO than others.

The validation of the enzymatic activity of the DeSi proteases was achieved using the substrate SUMO1 conjugation chains. The activity shown was fairly low compared to that of the control protease XopD (Figure 3.16) (van den Burg *et al*., 2010). The reason for this could be due to the lack of a

SUMOylated substrate providing specificity for the protease. Without a target substrate to allow the protease to dock, only 'ambient' enzymatic activity would be likely. However due to the fact that each protease may only target a few SUMOylated substrate proteins, finding the appropriate targets before confirming deSUMOylation activity did not seem logical. Instead the creation of a truncated version of the DeSi3 protein At1g47740, which contained the catalytic domain only, was used to determine if deSUMOylation activity would increase. This was in an attempt to remove the part of the protein that provided the specificity and monitor if this increased the enzymatic activity (Figure 3.17a) (Colby *et al*., 2006; Hotson *et al*., 2003). As was discussed in section 3.9, the truncated protease did show an increase in deSUMOylation activity, however so did the truncated version with the mutated C-to-S. Due to the similarities in the chemistry between cysteine and serine as discussed in 3.9, and with the advantage of hindsight, mutating the cysteine to other amino acids such as alanine could have provided more insight to the enzymatic activity (Carrea *et al.,* 1993; Lee *et al.,* 2015). Perhaps producing several different mutations and testing them all against the wild-type and record differences in activity.

The At1g47740 protease was unable to cleave SUMO from the preSUMO1 FLC attached (preSUM-FLC) substrate, unlike the ULP: - OTS1 (Figure 3.17). This provided evidence that the DeSi proteases found in plants, share similar function to those found in animal species; i.e. in both species DeSi proteins can cleave SUMO1 from SUMO1 conjugated target, but cannot cleave the gly-gly terminal found in the preSUMO1 form. Taken together it has been shown that the identified and tested DeSi proteases described in this thesis are bon-fide SUMO proteases. Target recognition for each of these DeSi protease would further valid this, and would also provided insight to the function. As previously discussed, it is likely that each DeSi has more than one target substrate, however it is also likely that if this was true, the targets of the same protease would be involved in similar pathways, i.e. in the same pathway, or a related pathway.

6.3 At1g47740 Acts as a Negative Regular of Defence Through Desumoylation of a Target Substrate

During the testing of the DeSi proteases identified in chapter 3, the DeSi3 protein At1g47740 was chosen for further investigation for 3 main reasons,

1) It represents a plant specific sub-class DeSi3:

This was important because it provided insight into the types of processes this sub-class of protease could be involved in, thus underpinning the importance of the sub-class.

2) The protein is compartmentalised to the cell periphery and/or the plasma membrane with no evidence of nuclear localisation:

Due to this being the first time in plants that SUMO proteases have been seen outside the nucleus, curiosity was a partial driver in the decision to use At1g47740, as this gave mechanistic clues to the processes it was involved in, and also potentially identifying a SUMO regulated system, outside of the nucleus.

3) No other protein in the sub-class is on the same branch after phylogenetic analysis:

It was important to exclude proteins that may have homologues due to the possibility of functional redundancy, as is seen in the partial redundancy between OTS1 and OTS2 (Conti *et al*., 2008).

In an initial screening process (data not shown), several stress inducers were used to check for obvious phenotypic differences between col-0 and the *At1g47740*-KO plants. During this screen a phenotypic difference in the response to the elicitor peptide flg22 was identified in the *At1g47740*-KO plants. Due to only having one knockout line, col-0 plants were transformed to over-express the *At1g47740* gene to see if the phenotypic changes were reversed. With the advantage of hindsight, it would have been more useful to generate these in the *At1g47740*-KO background at this stage, rather than later in the research. However the lines generated from transforming col-0

with the *At1g47740* constructed, (OX1 and OX2) did provide important information in the preliminary stages of the research. One of these important factors was that *At1g47740*-OX1 and -OX2 appeared to be insensitive to the presence of flg22 in the growth media. This was almost the opposite of what was seen in the *At1g47740*-KO plants grown on the same media that seem to be hypersensitive the presence of flg22, compared to col-0. Furthermore the *At1g47740*-OX1 and OX2 lines perform similarly to *fls2* knockout lines when grown on media with flg22 present. However this trend is not seen when monitoring the ROS production of the various genotypes. Although *At1g47740*-KO plants produce a higher level of ROS than col-0, the opposite is not true for the over-expression lines. Due to the similarities in the root growth assays between *fls2* and *At1g47740*-0X1 and –0X2, it would be expected that the ROS production would also be similar. However, the over-expression lines produced ROS with a greater variation than seen in the other genotypes, although it was never recorded at the high values seen in the *At1g47740*-KO plants. There are various explanations to account for this behaviour. Firstly due to the presence of the native *At1g47740* gene in the –OX1 and -OX2 background, the plant could be partially silencing in the transgene (Matzek *et al*., 2002). Some evidence for this is provided in Figure 3.04, where the protein levels, at least in –OX1 are low. Secondly, if FLS2 is the target of At1g47740, then over expression of the protease would not necessary prevent the SUMOylation of FLS2, meaning that signalling could still occur. In this scenario, the FLS2 deSUMOylation would most likely be heightened, however unable to prevent the initiation of SUMO attachment to FLS2. Thirdly plants over-expressing a transgene, even if they are from the same line, have been shown to carry individual variation in the expression/silencing of the transgene (Matzke *et al*., 2008). In the research presented here, this is relevant because the *At1g47740*-OX1 and –OX2 lines are effectively tetraploid (4n) for the *At1g47740* gene and studies have shown that this can result in more endogenous silencing than in the same plants who are diploid (2n) for the same gene (Finn *et al*., 2011).

If any of these scenarios applied to this research, it could provide an explanation as to why the over-expressing plants showed variable results in

the ROS production assay. This was one of the reasons it was necessary to generate partial complementation by over-expressing *At1g47740* in the *At1g47740*-KO background.

The generation of *At1g47740*-COMP1 and –COMP2 lines allowed further determination of the influence the *At1g47740* gene had in the response to PAMPs. The At1g47740 cDNA fragment, under the 35S promoter, was reconstituted in the At1g47740 background but was considered only a partial complementation due to the lack of a native promoter. This however did not prevent the ability of these lines to reverse the effects observed in the knockout line. Due to the over-expression of the At1g47740 transgene, it is unclear why the levels of the ROS production in the COMP1 and COMP2 lines are not below the levels observed in col-0, as this would be the opposite of the high levels seen in the knockout plants. However this assay looks only at the early part of the defence signal cascade, when the -COMP1 and -COMP2 plants were grown on media with flg22, they seemed to respond like *fls2* or pFLS2:FLS2-K/R plants, and were apparently insensitive to the flg22. The correlation between the At1g47740 protein and FLS2 is speculative, however the evidence presented here shows that further investigation would likely uncover a direct or indirect link between the components. In accordance with the reduction in At1g47740 protein levels when plants were treated with flg22 (Figure 5.10), and the reduced transcript levels in the same conditions (Figure 4.12), the target of At1g47740 must require increased SUMOylation for activation of defences. This means that in –COMP1 and –COMP2 lines the reduction in At1g47740 protein and/or transcript levels during flg22 recognition is not enough to prevent deSUMOylation of the target, leading to the apparent insensitivity witnessed in the –COMP1 and COMP2 plants during the root length assays. Whereas in col-0 plants, as shown in Figure 4.12, the transcript levels of *At1g47740* are reduced by 12 fold, upon flg22 treatment. This reduction, along with any At1g47740 protein degradation, would inhibit deSUMOylation of its target substrate and thus allow the activation of defences. In the *At1g47740*-KO plants the over stimulation of the defence responses are a direct result of failure to deSUMOylate the target substrate. This proposed model allows for the At1g47740 target substrate to be the

receptor FLS2, however, it does not exclude the possibility that the target substrate could also be another component of the complex formed upon PAMP recognition.

The majority of this study is focused on the early stages of pathogen attack, however the infection cycle within the plants can lasts for days or even weeks. To see if *At1g47740* had any affect on the longer-term infection, the bacterial pathogen PST was used to infect the *At1g47740* genotypes for comparison against col-0. This assay revealed that *At1g47740*-KO plants showed increased resistance to PST, compared to col-0 and OX1 and OX2. Moreover, the opposite was true for the over-expressing lines, -OX1 and -OX2, as they were hypersensitive to PST compared to col-0 and *At1g47740*-KO (Figures 4.09 and 4.10). Although there is not much scope within this study to understand the longer-term effects of the At1g47740 role in pathogen attack, it does not seem likely from the results shown, that *At1g47740*-KO plants merely delay the infection onset. The *At1g47740*-KO plants displayed signs of true resistance and were not simply delaying the infection at the point of perception but also slowing/preventing the pathogens progression (Figure 4.09). The implications of this could lead to identification of possible homologues in crops species, where further research could be conducted with an aim to increase disease resistance in the species of interest.

6.4 PAMP Induction ROS Activation Via The FLS2 Receptor Is Sumoylation Dependent

Using FLS2wt, FLS2-K/R and *fls2* transgenic plants, this study provided strong evidence that the predicted SUMO attachment site K1174 was necessary for pathogen perception. When grown on media containing flg22, the pFLS2:FLS2-K/R plants responded in a similar fashion to *fls2* plants (Figure 5.11). As FLS2 has been shown to be crucial in the early stages of bacterial pathogen attack, primarily because it is responsible for pathogen perception, it could argued that the lysine residue K1120, is responsible for proper signal propagation also (Zipfel 2008, Chisholm et al 2006). Due to the

mutation not affecting the extracellular domain of the FLS2 protein, it is likely that FLS2-K/R protein is able to bind flg22 as normal, unlike in *fls2* plants where there is no receptor at all. If this is true then one of three things could be likely upon binding flg22:

A) FLS2-K/R is unable to bind with itself and/or its interaction partners e.g. BAK1, thus preventing proper defence activation.

B) FLS2-K/R interacts as normal with its binding partners however is unable to activate downstream signalling possibly inhibiting the phosphorylation cascade. Thus prevents proper defence.

C) FLS2-K/R interacts as normal with binding partners, but is not able to signal for auto-degradation. Resulting in improper defence signalling.

Since showing that FLS2 mediated ROS production is dependent on the predicted SUMO attachment K1120, further research has been conducted that proves FLS2 is SUMOylated in response to flg22 (Orosa *et al.,* 2017 unpublished). Using co-immunoprecipitation assays, this work also showed that mutation of the FLS2 K1120 does not affect its interaction of BAK1. Put together this rules out the previous suggested theory 'A' (FLS2-K/R is unable to bind with itself and/or its interaction partners e.g. BAK1, thus preventing proper defence activation), and proves that upon flg22 perception FLS2-K/R interact with BAK1 as normal, however something prevents both the activation of ROS and proper activation of downstream defence signals. To help understand this, and tie together 2 elements of this research, a speculative look at the metaphorical *other side of the coin* is necessary.

If the high levels of ROS seen in the *At1g47740*-KO plants are a direct result of failure to deSUMOylate FLS2, then FLS2 must be constitutively SUMOylated in this line, resulting in the over activation of defence responses (this could act as a priming mechanism for defence in the knockout line). This means that the levels of SUMOylation of FLS2 are correlated to the strength of the immune response, i.e. in the At1g47740-KO line, more SUMOylated FLS2 results in a larger defence response). If this is true, then the previously suggested theory 'C', (FLS2-K/R interacts as normal, but is not able to signal for auto-degradation after activation resulting improper defence signalling)

cannot be true as the opposite would be expected in *At1g47740*-KO background. Thus if the SUMO site mutation (K1121R) prevented auto-degradation of FLS2, then in the hyper-SUMO form predicted in the At1g47740-KO background, FLS2 would be oppositely affected resulting in a high level of FLS2 degradation. This is unlikely as it would result in a delayed and/or prolonged ROS signal in the knockout background compared to col-0. The argument against this is supported by the earlier onset of ROS seen in *At1g47740*-KO plants compared to col-0 (Figure 4.11 and 5.16). Therefore, by elimination, the only theory left is suggestion 'B', (FLS2-K/R interacts as normal with its binding partners however is unable to activate downstream signalling possibly inhibiting the phosphorylation cascade. Thus prevents proper defence). This means SUMOylation could be involved in activating the signalling mechanisms resulting in or involving the MAPK-cascade, or the even activation of RBOHD by BAK1 (Figure 6.1) (Zhang *et al.,* 2010a, Macho & Zipfel 2014; Shi *et al.,* 2013; Henry *et al*., 2013). This would explain the inability of FLS2-K/R to respond normally upon flg22 detection, as well as the *At1g47740*-KO plants being hyper sensitive to flg22.

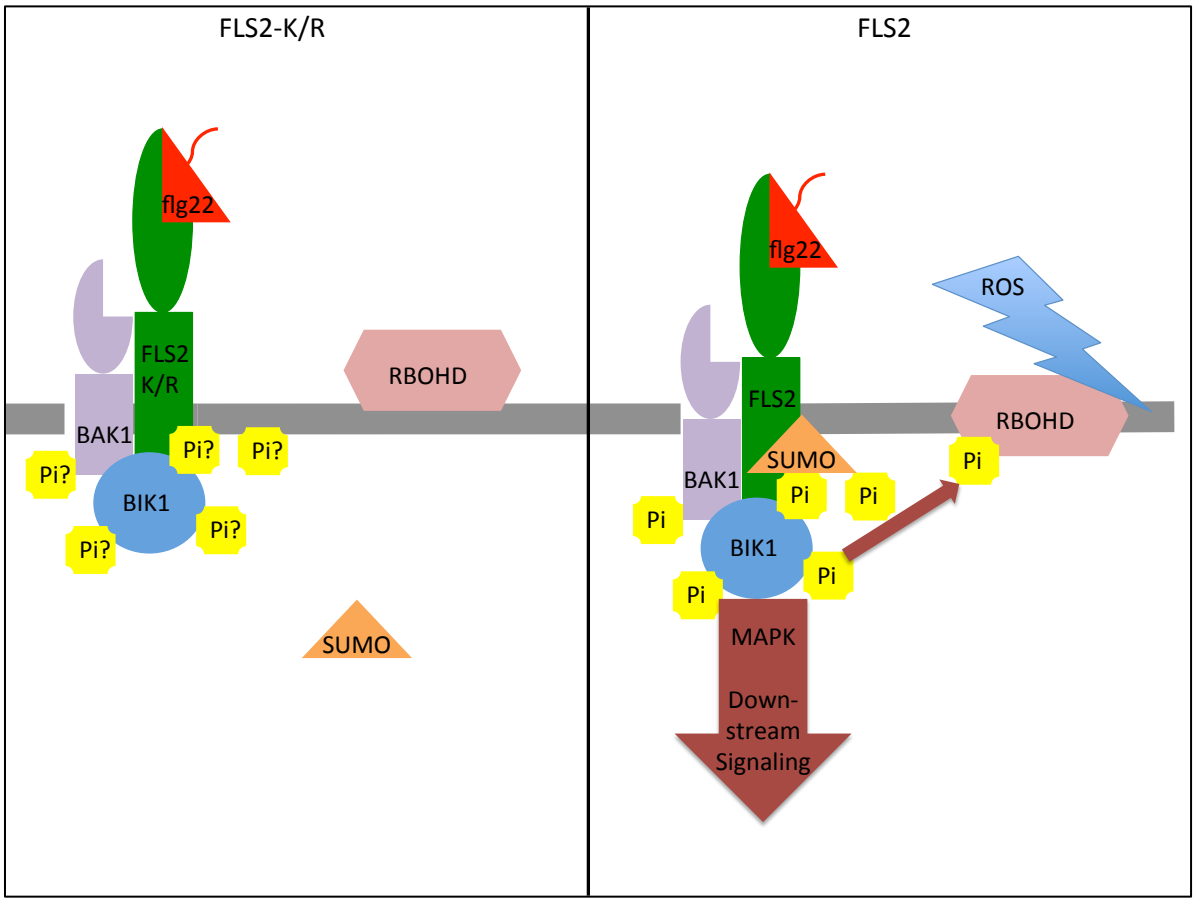

Figure 6.1. Proposed role of SUMOylation in the flg22 induced activation of FLS2 mediated signalling.
Right hand panel shows the existing FLS2 activation complex as shown in Figure 1.1 with the inclusion of SUMO attached FLS2. Left hand panel shows the equivalent with the FLS2-K1120R mutation. Here flg22 is detected and the complex required for signalling is formed, however in the absence of SUMO, proper signalling does not occur. The phosphorylation state of the complex in the FLS-K/R paradigm is unknown.

6.5 CERK1 as a Potential At1g47740 Target

In chapter 4, evidence is provided that the elicitor chitosan affects col-0 and *At1g47740*-KO plants differently (Figure 4.15). Like with flg22 treatment, chitosan caused an increased burst of ROS production in *At1g47740*-KO plants compared to col-0. However unlike the flg22 treatment, that in *At1g47740*-KO plants produced a ROS activity peak of >3 fold greater than observed in col-0 (Figure 4.11), when treated with chitosan the increase compared to col-0 was consistently closer to 0.3 fold at the peak (Figure 4.15). The comparative difference between the 2 genotypes suggests that in *At1g47740*-KO plants the hyper-production of ROS is 10 fold greater when treated with flg22 compared to chitosan. In accordance with this, the levels of At1g47740 protein in samples treated with chitosan did not reduce in the same manner as the samples that were treated with flg22 (Figures 4.13 and 4.14). Although a slight reduction can be seen in protein levels after 15 minutes, it was very subtle each time the experiment was repeated, and is questionable whether it was significant. A more thorough examination using immunoprecipitation for protein enrichment or using pathogens to replace the elicitor might determine better if the protein levels were dropping. Together this suggests that perhaps the At1g47740 protein plays a slightly different role in response to chitosan. This is partially supported with the read-out from the qRT-PCR results shown in Figure 4.16. The data from these experiments showed that in the *At1g47740*-KO background, not only were some pathogen response-associated genes up-regulated after chitosan treatment, but were also up-regulated prior to treatment. Interestingly most of these genes were in the 'necrotrophic' defence pathway, with only very little differences seen in the 'biotrophic' pathway even prior to treatment. Although further experiments on the *At1g47740* mutant plants, including use of fungal pathogens such as *Botrytis,* are needed to validate these results, this preliminary data supports the idea that in the response to chitosan the At1g47740 plays a different long-term role than in response to flg22.

However in the early stages of the response to chitin, if CERK1 was a target of At1g47740, the hyper SUMOylation state of this receptor would be the reason

for the increased response to the elicitor, as was suggested for FLS2. This is supported by the ROS production assays, as *At1g4774*0-OX1, produced lower amounts of ROS than col-0, even although this was marginal and not close to the basal levels produced by *cerk1-1* plants. Whereas the same plants when treated with flg22, did not produce ROS levels consistently below that of col-0. This could indicate a more direct correlation between At1g47740 levels and CERK1 SUMOylation.

The failure to produce *cerk1-1* complementation lines expressing the wild-type form of CERK1 meant that proper determination of the effects of mutating the predicted SUMOylation site could not be drawn, but only speculated upon. However chapter 5 showed that when transiently expressed both CERK1 and CERK1-3K/R were fairly equally stable in the total protein extract. Due to time restrictions, a planned transient expression and co-immunoprecipitation assay using tagged SUMO1 and tagged CERK1 or CERK1-3K/R, was not completed. If CERK1 was immunoprecipitated with SUMO1 attached and CERK1-3K/R was not, this would have validated the SUMO prediction site and proved CERK1 was SUMOylated. However as this data is missing, only speculation can be provided to fill in the blanks. The use of the *CERK1*-3K/R lines in the ROS production assay provided some insight to the function of the SUMO site on the CERK1 protein. The results cannot be validated due to a lack of reconstituted wild-type CERK1 in the *cerk1-1* background. However assuming that normal function would have been restored in these lines, as was seen in the pFLS:FLS2wt plants, the CERK1-3K/R lines did not rescue the phenotype of the cerk1-1 plants. Again, due to the lack of a transgenic wild-type, it is unclear if the inserted CERK1-3K/R transgene was viable, but assuming it was, it meant that the SUMO sites predicted in the CERK1, are necessary for chitosan induced CERK1 mediated ROS production. Therefore as is the case with FLS2, PAMP induced ROS production via the CERK1 receptor is SUMOylation dependent.

6.6 Future Prospects

Most of the logical suggestions regarding the interactions of DeSi3a and FLS2 have been carried out, and proper validation of the interaction has been made via co-immunoprecipitation and co-localisation assays (Orosa *et al*., 2017 unpublished). An interesting transgenic tool that would gave better understanding of the interplay of FLS2 and DeSi3a *in vivo*, would come from over-expressing the DeSi3a protein in the pFLS2:FLSwt and pFLS2:FLS-K/R. This would allow comparisons of phenotypic responses to flg22 and provide a link to the interactions through phenotype.

Much more work is needed to generate data showing that CERK1 and DeSi3a interact. Firstly SUMO-CERK1 co-immunoprecipitation for validation of SUMO attachment, secondly CERK1-DeSI3a pull-down assays or co-localisation assays to show direct interaction. Thirdly, as previously mentioned, CERK1 wild-type reconstituted in *cerk1-1* background. This along with full infection assays would provide a more complete story than is presented here.

6.7 Concluding Remarks

In order to address the original aims of this thesis the following statement is presented;

Using a bioinformatics approach the identification of a new class of SUMO proteases, of the type DeSi, was uncovered in *Arabidopsis thaliana.* A protein from each identified sub-class was shown by biochemical analysis to cleave SUMO1, thus validating the predicted enzymatic activity. These DeSi proteins are the first SUMO proteases visualised localising outside of the nucleus. The SUMO protease At1g47740 (DeSi3a) appears to act as a negative regulator of pathogen defence, and displayed pathogen resistance phenotypes when the corresponding gene is knockout. The 2 major pathogen receptors, FLS2 and CERK1, are SUMOylation substrates and are likely targeted by DeSi3a for deSUMOylation. The PAMP activated ROS produced via the FLS2 receptor is SUMOylation dependent, and the same is likely true for the CERK1 receptor. Mutations of the SUMO attachment sites, from K-to-R, makes FLS2 and CERK1

incapable of normal function, effectively operating as knockouts of their respective genes

Proposed model:

In normal conditions SUMOylation of FLS2 is kept to a minimum by the presence of DeSi3a. Under elicitor detection FLS2 is SUMOylated, the levels of DeSi3a protein are decreased to prevent FLS2 deSUMOylation. The reduction in DeSi3a levels may be a consequence of, or the cause of, increased FLS2 SUMOylation. FLS2 conducts signalling mechanisms to produce effective downstream immune responses, and transcription and protein levels of DeSi3a are further reduced.

References:

Alvarez, M.E., Pennel, R.I., Meijer. P-J., Ishikawa, A., Dixon, R.A., Lamb, C. (1998) Reactive oxygen intermediates mediate a systemic signal network in the establishment of plant immunity. Cell, 92:773-784

Agrios, G.N. (1997) Plant Pathology. San Diego: Academic Press

Anderson, J.P., Gleason, C.A., Foley, R.C., Thrall, P.H., Burdon, J.B., Singh, K.B. (2010) Plants versus pathogens: an evolutionary arms race. Functional Plant Biology, 37(6), 499–512.

Bailey, M., Srivastava, A., Conti, L., Nelis, S., Zhang, C., Florance, H., Sadanandom, A. (2016) Stability of small ubiquitin-like modifier (SUMO) proteases OVERLY TOLERANT TO SALT1 and -2 modulates salicylic acid signalling and SUMO1/2 conjugation in *Arabidopsis thaliana*. Journal of Experimental Botany, *67*(1), 353–363.

Bekes, M., Prudden, J., Srikumar, T., Raught, B., Boddy, M. N. et al. (2011) The dynamics and mechanism of SUMO chain deconjugation by SUMO-specific proteases. Journal of Biological. Chemistry, 286, 10238–10247.

Bent, A.F., Mackey, D. (2007) Elicitors, effectors, and R genes: the new paradigm and a lifetime supply of questions. Annual Reviews in Phytopathology, 45:399-436.

Block, A., Li, G., Fu, Z.Q., Alfano, J.R. (2008) Phytopathogen type III effector weaponry and their plant targets. Current Opinion in Plant Biology, 11(4):396-403.

Block, A., & Alfano, J. R. (2011) Plant targets for *Pseudomonas syringae* type III effectors: Virulence targets or guarded decoys? Current Opinion in Microbiology, 14(1), 39–46.

Boller, T. & Felix, G. (2009) A Renaissance of Elicitors: Perception of MicrobeAssociated Molecular Patterns and Danger Signals by Pattern-Recognition Receptors. Annual Review of Plant Biology, 60(1):379–406

Budhiraja, R., Hermkes, R., Muller, S., Schmidt, J., Colby, T., Panigrahi, K., Coupland, G., Bachmair, A. (2009) Substrates related to chromatin and to RNA-dependent processes are modified by Arabidopsis SUMO isoforms that differ in a conserved residue with influence on desumoylation. Plant Physiology, 149, 1529–1540.

Callis, J., Raasch, J.A., Vierstra. R.D. (1990) Ubiquitin extension proteins of Arabidopsis thaliana. Structure, localization, and expression of their promoters in transgenic tobacco. The Journal of Biological Chemistry, 265, 12486–12493

Canonne, J., Marino, D., Jauneau, A., Pouzet, C., Brière, C., Roby, D., Rivas, S. (2011) The Xanthomonas type III effector XopD targets the Arabidopsis transcription factor MYB30 to suppress plant defense. Plant Cell, 23, 3498-511.

Cao, Y., Liang, Y., Tanaka, K., Nguyen, C. T., Jedrzejczak, R. P., Joachimiak, A., and Stacey, G. (2014) The kinase LYK5 is a major chitin receptor in Arabidopsis and forms a chitin-induced complex with related kinase CERK1. eLife, 3.

Carrera, A.C., Alexandrov, K., Roberts, T.M. (1993) The Proceedings of the National Academy of Science, U.S.A., 90, 442.

Castaño-Miquel, L., Mas, A., Teixeira, I., Seguí, J., Perearnau, A., Thampi, B.N., Schapire, A.L., Rodrigo, N., La Verde, G., Manrique, S. and Coca, M., (2017) SUMOylation Inhibition Mediated by Disruption of SUMO E1-E2 Interactions

Confers Plant Susceptibility to Necrotrophic Fungal Pathogens. Molecular Plant, 10(5), pp.709-720.

Catala, R., Ouyang, J., Abreu, I.A., Hu, Y., Seo, H., Zhang, X., Chua, N. (2007) The Arabidopsis E3 SUMO ligase SIZ1 regulates plant growth and drought responses. Plant Cell, 19, 2952–2966.

Chisholm, S.T., Coaker, G., Day, B., Staskawicz ,B.J. (2006) Host–microbe interactions: shaping the evolution of the plant immune response. Cell, 124:803-814.

Chory, J., Ecker, J.R., Briggs, S. et al. (2000) National Science Foundationsponsored workshop report: ‘‘The 2010 Project’’ – Functional genomics and the virtual plant. A blueprint for understanding how plants are built and how to improve them. Plant Physiology, 123, 423–425.

Colby, T., Matthai, A., Boeckelmann, A., Stuible, H.P. (2006) SUMO-conjugating and SUMO deconjugating enzymes from Arabidopsis. Plant Physiology, 142, 318 –332.

Collier, S.M., Moffett, P. (2009) NB-LRRs work a “bait and switch” on pathogens. Trends in Plant Science, 14:521–529.

Conrath, U. (2006) Systemic Acquired Resistance. Plant Signaling & Behavior 1, 4, 179-184

Conti, L., Price, G., O’Donnell, E., Schwessinger, B., Dominy, P., Sadanandom, A. (2008) Small ubiquitin-like modifier proteases OVERLY TOLERANT TO SALT1 and -2 regulate salt stress responses in Arabidopsis. Plant Cell, 20, 2894–2908.

Delano, W. L., 2002. The PyMOL molecular graphics system. Delano Scientific, San Carlos, CA, USA

Dickinson, M. (2003). Molecular plant pathology. ADVANCED TEXT BIOS Scientific Publishers, 145–156.

Downes, B., Vierstra, R.D. (2005) Post-translational regulation in plants employing a diverse set of polypeptide tags. Biochemical Society Transactions, 33, 393–399.

Dreze, M., Carvunis, A-R., Charloteaux, B., Galli, M., Pevzner, S.J., and Tasan, M. (2011) Arabidopsis Interactome Mapping Consortium
Evidence for network evolution in an Arabidopsis interactome map.
Science, 333 pp. 601-607

Dunning, F.M., Sun, W., Jansen, K.L., Helft, L., Bent, A.F. (2007) Identification and Mutational Analysis of Arabidopsis FLS2 Leucine-Rich Repeat
Domain Residues That Contribute to Flagellin Perception. The Plant cell, 19(10):3297–3313.

Durrant, W.E., Dong, X. (2004) Systemic acquired resistance. Annual Reviews in Phytopathology; 42:185 – 209

Dye, B.T., Schulman, B.A. (2007) Structural mechanisms underlying posttranslational modification by ubiquitin-like proteins. Annual Review of Biophysics and Biomolecular Structure, 36, 131–150

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 32(5), 1792–1797.

Earley, K., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K., Pikaard, C.S. (2006) Gateway-compatible vectors for plant functional genomics and proteomics. The Plant Journal, 45:616-629

Edwards, J.M., Roberts, T.H., Atwell, B.J. (2012) Quantifying ATP turnover in anoxic coleoptiles of rice (Oryza sativa) demonstrates preferential allocation of energy to protein synthesis. The Journal Experimental Botany, 63, 4389–4402.

Feldmann, K.A. and Marks, M.D. (1987) Agrobacterium-mediated transformation of germinating seeds of Arabidopsis thaliana: a non-tissue culture approach. Molecular Genetics, 208, 1–9.

Feng, F., Zhou, JM. (2012) Plant-bacterial pathogen interactions mediated by type III effectors. Current Opinions Plant Biology, 15. 469-476

Finn, T. E., Wang, L., Smolilo, D., Smith, N. A., White, R., Chaudhury, A., Wang, M.B. (2011) Transgene Expression and Transgene-Induced Silencing in Diploid and Autotetraploid Arabidopsis. Genetics, 187(2), 409–423.

Gareau, J. R., & Lima, C. D. (2010) The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. Nature Reviews. Molecular Cell Biology, 11(12), 861–871.

Geer, L. Y., Marchler-Bauer, A., Geer, R. C., Han, L., He, J., He, S., Liu, C., Shi, W., and Bryant, S. H. (2010). The NCBI BioSystems database. Nucleic Acids Research, 38(Database issue):D492–6.

Gill, G. (2004) SUMO and ubiquitin in the nucleus: different functions, similar mechanisms? Genes and Development, 18, 2046–2059.

Gill. G. (2005) Something about SUMO inhibits transcription. Current Opinion in Genetic Development, 15, 536–541.

Gillies, J. and Hochstrasser, M. (2012) A new class of SUMO proteases. EMBO Reports, 13(4), 284–285.

Glazebrook, J. (2005) Contrasting Mechanisms of Defense Against Biotrophic and Necrotrophic Pathogens. Annual Review of Phytopathology, 43(1):205–227.

Gong L., Millas, S., Maul, G.G., and Yeh, E.T. (2000) Differential regulation of sentrinized proteins by a novel sentrin-specific protease. Journal of Biological. Chemistry, 275:3355-3359.

Gómez-Vásquez, R., Day, R., Buschmann, H., Randles, S., Beeching, J.R., Cooper, R.M., (2004) Phenylpropanoids, phenylalanine ammonia lyase and peroxidases in elicitor-challenged cassava (Maninhot esculenta) suspension cells and leaves. Annual Botony, 94 (1), 87–97.

Group, M., Ichimura, K., Shinozaki, K., Tena, G., Sheen, J., *et al*. (2002) Mitogen-activated protein kinase cascades in plants: a new nomenclature. Trends Plant Science, 7:301–8

Grzybek, T., Pietrzak, R., Wachowska, H. (2004) The Comparison of Oxygen and Sulfur Species Formed by Coal Oxidation with O2/Na2CO3 or Peroxyacetic Acid Solution. XPS Studies *Energy & Fuels* 2004 *18* (3), 804-809

Hanania, U., Furman-Matarasso, N., Ron, M., Avni, A. (1999) Isolation of a novel SUMO protein from tomato that suppresses EIX-induced cell death. The Plant Journal, 19, 533–541.

Haseloff, J. and Amos, B. (1995) GFP in plants. Trends Genetics, 11, 328–329.

Hay, R. (2005) SUMO: A history of modification. Molecular Cell, 18, 1–12.

Henry, E., Yadeta, K. A. and Coaker, G. (2013) Recognition of bacterial plant pathogens: local, systemic and transgenerational immunity. New Phytologist, 199: 908–915.

Hermkes, R., Fu, Y.F., Nürrenberg, K., Budhiraja, R., Schmelzer, E., Elrouby, N., Dohmen, R.J., Bachmair, A., Coupland, G. (2011) Distinct roles for Arabidopsis SUMO protease ESD4 and its closest homolog ELS1. Planta, 233, 63–73.

Hershko, A., Heller, H., Elias, S., Ciechanover, A. (1983) Components of ubiquitin-protein ligase system: Resolution, affinity purification and role in protein breakdown. The Journal of Biological Chemistry, 258, 8206–8214.

Hickey, C. M., Wilson, N. R., and Hochstrasser, M. (2012) Function and regulation of SUMO proteases. Nat. Rev. Mol. Cell Biol, 13, 755–766.

Hirano, S.S., Upper, D.C. (2000) Bacteria in the leaf ecosystem with emphasis on *Pseudomonas syringae*-a pathogen, ice nucleus, and epiphyte. Microbiol. Mol. Biol. Rev, 64:11624–653

Hoege, C., Pfander, B., Moldovan, G.L., Pyrowolakis, G., Jentsch, S. (2002) RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. Nature, 419, 135–141

Hotson, A., Chosed, R., Shu, H., Orth, K., Mudgett, M.B. (2003) Xanthomonas type III effector XopD targets SUMO-conjugated proteins in planta. Molecular Microbiology, 50, 377–389.

Isono, E., Nagel, M-K. (2014) Deubiquitylating enzymes and their emerging role in plant biology. Frontiers in Plant Science, 5, 56.

Johnson, L.N., Nobel, M.E.M., Owen, D.J. (1996) Active and Inactive Protein Kinases: Structural Basis for Regulation. Cell, 85: 149–158.LN

Johnson, E.S. (2004) Protein modification by SUMO. Annual Review of Biochemistry, 73, 355–382.

Jones, J.D. and Dangl, J.L. (2006) The plant immune system. Nature, 444:323-329.

Katagiri, F., Thilmony, R., He, S.Y. (2002) The Arabidopsis Thaliana Pseudomonas Syringae Interaction. The Arabidopsis Book, 20(1):e0039.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E. (2015) The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. Nature Protocols, 10, 845-858.

Khokhlatchev. A.V., Canagarajah, B., Wilsbacher, J., Robinson, M., Atkinson, M. (1998) Phosphorylation of the MAP kinase ERK2 promotes its homodimerization and nuclear translocation. Cell, 93:605–1

Kolli, N., Mikolajczyk, J., Drag, M., Mukhopadhyay, D., Moffatt, N., et al. (2010) Distribution and paralogue specificity of mammalian deSUMOylating enzymes. Biochemistry Journal, 430, 335–344.

Kraft, E., Stone, S.L., Ma, L., Su, N., Gao, Y., Lau, O.S., Deng, X.W., Callis, J. (2005) Genome analysis and functional characterization of the E2 and RING-type E3 ligase ubiquitination enzymes of Arabidopsis. Plant Physiology, 139, 1597–1611

Kerscher, O., Felberbaum, R., Hochstrasser, M. (2006) Modification of proteins by ubiquitin and ubiquitin-like proteins. Annual Review of Cell Developmental Biology, 22, 159–180.

Kombrink, A., Sánchez-Vallet, A., and Thomma, B. P. H. J. (2011) The role of chitin detection in plant–pathogen interactions. Microbes and Infection, 13(14-15):1168–1176.

Kurepa, J., Walker, J.M., Smalle, J., Gosink, M.M. Davis, S.J., Durham, T.L., Sung, D.Y., Vierstra, R.D. (2003) The small ubiquitin-like modifier (SUMO) protein modification system in Arabidopsis. Accumulation of SUMO1 and -2 conjugates is increased by stress. The Journal of Biological Chemistry, 278, 6862–6872.

Lamesch, P., Berardini, TZ., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, DL., Garcia-Hernandez, M., et al. (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids, 40. 1202-1210

Langridge, J. (1955) Biochemical mutations in the crucifer Arabidopsis thaliana (L) Heynh. Nature, 176, 260–261.

Laibach, F. (1907) Zur Frage nach der individualita¨t der Chromosomen im Pflanzenreich. Beit. Bot. Zentralbl, 22, 191–210.

Lee, E.M., Lee, S.S., Tripathi, B.N., Jung, H.S., Cao, G.P., Lee,Y., Chung, B.Y. (2015) Site-directed mutagenesis substituting cysteine for serine in 2-Cys peroxiredoxin (2-Cys Prx A) of *Arabidopsis thaliana* effectively improves its peroxidase and chaperone functions. Annals of Botany, 116(4), 713–725.

Lee, J., Nam, J., Park, H.C., et al. (2007) Salicylic acid-mediated innate immunity in Arabidopsis is regulated by SIZ1 SUMO E3 ligase. The Plant Journal, 49, 79–90.

Lois LM Lima CD Chua NH . 2003. Small ubiquitin-like modifier modulates abscisic acid signaling in Arabidopsis. The Plant Cell, 15, 1347–1359.

Leutwiler, L.S., Houghevans, B.R. and Meyerowitz, E.M. (1984) The DNA of Arabidopsis thaliana. Molecular General Genetics, 194, 15–23.

Li, H., Xu, H., Zhou, Y., Zhang, J., Long, C. et al. (2007) The phosphothreonine lyase activity of a bacterial type III effector family. Science, 315:1000–3

Li, S.J., Hochstrasser, M . (1999) A new protease required for cell-cycle progression in yeast. Nature, 398, 246–251

Li, S.J., Hochstrasser, M. (2000) The yeast ULP2 (SMT4) gene encodes a novel protease specific for the ubiquitin-like Smt3 protein. Molecular Cell Biology, 20, 2367–2377.

Li, S.J., Hochstrasser, M. (2003) The Ulp1 SUMO isopeptidase: distinct domains required for viability, nuclear envelope localization, and substrate specificity. J Cell Biology, 160: 1069–1081

Liu, T., Liu, Z., Song, C., Hu, Y., Han, Z., She, J., Fan, F., Wang, J., Jin, C., Chang, J., Zhou, J., and Chai, J. (2012b) Chitin-Induced Dimerization Activates a Plant Immune Receptor. Science, 336(6085):1160–1164.

Lois, L.M., Lima, C.D., Chua, N.H . (2003) Small ubiquitin-like modifier modulates abscisic acid signaling in Arabidopsis. The Plant Cell, 15, 1347–1359.

Lyst, M.J., Stancheva, I. (2007) A role for SUMO modification in transcriptional repression and activation. Biochemical Society Transactions, 35, 1389–1392.

Impens, F., Radoshevich, L., Cossart, P., Ribet, D. (2014) Mapping of SUMO sites and analysis of SUMOylation changes induced by external stimuli. The Proceedings of the National Academy of Science, U S A 111: 12432–12437. pmid:25114211

Macho, A. P. and Zipfel, C. (2014) Plant PRRs and the Activation of Innate Immune Signaling. Molecular Cell, 54(2):263–272.

Maldonado, A.M., Doerner, P., Dixon, R.A., Lamb, C.J., Cameron, R.K. (2002) A putative lipid transfer protein
involved in systemic resistance signalling in Arabidopsis. Nature, 419:399–403

Malinovsky, F.G., Fangel, J.U., and Willats, W.G.T. (2014) The role of the cell wall in plant immunity. Frontiers in plant science, 5:178.

Mann, M., Jenson, O.N. (2003) Proteomic analysis of post-translational modifications. Nature Biotechnology, 21, 255–261

Matzke, M. A., Aufsatz, W., Kanno, T., Mette M.F., Matzke, A.J. (2002) Homology-dependent gene silencing and host defense in plants. Advanced Genetics, 46 235–275.

McGinnis, K.M., Thomas, S.G., Soule, J.D., Strader, L.C., Zale, J.M., Sun, T-P., Steber, C.M. (2003) The Arabidopsis SLEEPY1 gene encodes a putative F-box subunit of an SCF E3 ubiquitin ligase. The Plant Cell, 15, 1120–1130.

Melchior, F. (2000) SUMO nonclassical ubiquitin. Annual Review Cellular & Developmental Biology, 16, 591–626.

Miller, M.J., Scalf, M., Rytz, T.C., Hubler, S.L., Smith, L.M., Vierstra, R.D. (2013) Quantitative proteomics reveals factors regulating RNA biology as dynamic targets of stress-induced SUMOylation in Arabidopsis. Molecular & Cellular Proteomics, 12, 449–463.

Miura, K., Hasegawa, P.M. (2010) Sumoylation and other ubiquitin-like post-translational modifications in plants. Trends in Cell Biology, 20, 223–232.

Miura, K., Jin, J.B., Hasegawa, P.M. (2007a). Sumoylation, a post-translational regulatory process in plants. Current Opinion in Plant Biology, 10, 495–502.

Miura, K., Jin, J.B., Lee, J., Yoo, C.Y., Stirm, V., Miura, T., Ashworth, E.N., Bressan, R.A., Hasegawa, P.M. (2007b) SIZ1-mediated sumoylation of ICE1 controls CBF3/DREB1A expression and freezing tolerance in Arabidopsis. The Plant Cell, 19, 1403–1414.

Miura, K., Rus, A., Sharkhuu, A., Yokoi, S, Karthikeyan, A.S., Raghothama, K.G., Baek, D., Koo, Y.D., Jin, J.B., Bressan, R.A. (2005) The Arabidopsis SUMO E3 ligase SIZ1 controls phosphate deficiency responses. The Proceedings of the National Academy of Science USA , 102, 7760–7765.

Mockaitis, K. & Howell, SH. (2000) Auxin induces mitogenic activated protein kinase (MAPK) activation in roots of Arabidopsis seedlings. Plant Journal, 24:785–96

Mukhopadhyay, D., Riezman, H. (2007) Proteasome-independent functions of ubiquitin in endocytosis and signaling. Science, 315, 201–205.

Muller, S., Hoege, C., Pyrowolakis, G., Jentsch, S. (2001) SUMO, ubiquitin's mysterious cousin. Nature Review Molecular Cell Biology, 2, 202–210.

Munns, R., Tester, M. (2008) Mechanisms of salinity tolerance. Annual Reviews in Plant Biology, 59, 651–681.

Murtas, G., Reeves, P.H., Fu, Y.F., Bancroft, I., Dean, C., Coupland, G. (2003) A nuclear protease required for flowering-time regulation in Arabidopsis reduces the abundance of SMALL UBIQUITIN- RELATED MODIFIER conjugates. Plant Cell, 15, 2308–2319.

Nacerddine, K., Lehembre, F., Bhaumik, M., Artus, J., Cohen, Tannoudji, M., Babinet, C., Pandolfi, P.P., Dejean, A. (2005) The SUMO pathway is essential for nuclear integrity and chromosome segregation in mice. Developmental Cell, 9, 769–779.

Naito, K., Taguchi, F., Suzuki, T., Inagaki, Y., Toyoda, K., Shiraishi, T., Ichinose, Y. (2008) Amino acid sequence of bacterial microbe-associated molecular pattern flg22 is required for virulence. Moleculer Plant Microbe Interactions, 21(9):1165-74.

Nayak, A., & Müller, S. (2014) SUMO-specific proteases/isopeptidases: SENPs and beyond. Genome Biology, 15(7), 422.

Nelis, S. (2014) An investigation into the role and mechanism of action of small ubiquitin-like modifier interacting motifs in Arabidopsis thaliana proteins. Unpublished PhD thesis, University of Warwick.

Nelis, S., Conti, L., Zhang, C., & Sadanandom, A. (2015) A functional Small Ubiquitin-like Modifier (SUMO) interacting motif (SIM) in the gibberellin hormone receptor GID1 is conserved in cereal crops and disrupting this motif does not abolish hormone dependency of the DELLA-GID1 interaction. Plant Signaling & Behavior, *10*(2), e987528.

Novy, R., Drott, D., Yaeger, K., and Mierendorf, R. (2001) Overcoming the codon bias of E. coli for enhanced protein expression. inNovations, 12, 1–3

Orosa, B., Yates, G., Verma, V., Sadanandom, A. (2017-Unpublished) The immune receptor FLS2 is SUMOylated under flg22 detection.

Park, S.W., Kaimoyo, E., Kumar, D., Mosher, S., Klessig, D.F. (2007) Methyl salicylate is a critical mobile signal for plant systemic acquired resistance. Science, 318:113–16

Pearce, S., Saville, R., Vaughan, S.P., et al. (2011) Molecular characterization of Rht-1 dwarfing genes in hexaploid wheat. Plant Physiology, 157, 1820–1831.

Peng, J.R., Richards, D.E., Hartley, N.M., et al. (1999) 'Green revolution' genes encode mutant gibberellin response modulators. Nature, 400, 256–261.

Raffaele, S., Kamoun, S. (2012) Genome evolution in filamentous plant pathogens: why bigger can be better. Nature Reviews,10, 417-430

Rafiqi, M., Bernoux, M., Ellis, JG., Dodds, PN. (2009) In the trenches of plant pathogen recognition: Role of NB-LRR proteins. Seminars in Cell and Developmental Biology, 20:1017–1024

Rasmussen, J.B., Hamerschmidt, R., Zook, M.N. (1991) Systemic induction of salicylic acid accumulation in cucumber after inoculation with Pseudomonas syringae pv syringae. Plant Physiology, 97:1342 – 1347

Reed, JM., Dervinis, C., Morse, AM., Davis, JM. (2010) The SUMO conjugation pathway in Populus: genomic analysis, tissue- specific and inducible SUMOylation and in vitro de-SUMOylation. Planta, 232, 51–59

Reeves, P.H., Murtas, G., Dash, S., Coupland, G. (2002) Early in short days 4, a mutation in Arabidopsis that causes early flowering and reduces the mRNA abundance of the floral repressor FLC. Development, 129, 5349–5361.

Rodriguez, M.C.S., Petersen, M., Mundy, J. (2010) Mitogen-Activated Protein Kinase Signaling in Plants. Annual Review of Plant Biology, 61(1):621-49 ·

Ross, S., Best, J.L., Zon, L.I., Gill, G. (2002) SUMO-1 modification represses Sp3 transcriptional activation and modulates its subnuclear localization. Molecular Cell, 10, 831–842.

Rajan, S., Plant, L.D., Rabin, M.L., Butler, M.H., Goldstein, S.A. (2005) SUMOylation silences the plasma membrane leak K+ channel K2P1. Cell, 121, 37–47.

Sadanandom, A., Adam, E., Orosa, B., Viczian, A., Klose, C., Zhang, C., Josse, EM., Kozma-Bognar, L., Nagy, F. (2015) SUMOylation of phytochrome-B negatively regulates light-induced signaling in Arabidopsis thaliana . Proceedings of National Academy of Science USA, 112, 11108–11113.

Sadanandom, A., Bailey, M., Ewan, R., Lee, J., Nelis, S. (2012) The ubiquitin-proteasome system: central modifier of plant signalling. New Phytologist 196, 13–28.

Scholl, R.L., May, S.T., Ware, D.H. (2000) Seed and molecular resources for Arabidopsis. Plant Physiology, 124(4):1477-80.

Seo, J., Lee, K. (2004) Post-translational modifications and their biological functions: proteomic analysis and systematic approaches. The Journal of Biochemical and Molecular Biology, 37, 35–44.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology, 7:539.

Schulz, S., Chachami, G., Kozaczkiewicz, L., Winter, U., Stankovic-Valentin, N., et al. (2012) Ubiquitin-specific protease-like 1 (USPL1) is a SUMO isopeptidase with essential, non-catalytic functions. EMBO Reports, 13, 930–938.

Somerville, C.R., Meyerowitz. E.M., (2002) eds. The Arabidopsis Book. Rockville, MD: The American Society of Plant Biologists.

Shen, L. N., Geoffroy, M. C., Jaffray, E. G. and Hay, R. T. (2009) Characterization of SENP7, a SUMO-2/3-specific isopeptidase. Biochemistry Journal, 421, 223–230.

Shi, H., Shen, Q., Qi, Y., Yan, H., Nie, H., Chen, Y., Zhao, T., Katagiri, F.,

and Tang, D. (2013) BR-SIGNALING KINASE1 Physically Associates with FLAGELLIN SENSING2 and Regulates Plant Innate Immunity in Arabidopsis. The Plant cell, 25(3):1143–1157.

Shin, E.J., Shin, H.M., Nam, E., Kim, W.S., Kim, J.H., Oh, B.H., Yun, Y. (2012) DeSUMOylating isopeptidase: a second class of SUMO protease. EMBO Reports, 13: 339-346.

Shiu, S.H., and Bleecker, A.B. (2002) Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in Arabidopsis. Plant Physiology, 132,530 -543.

Somerville, C.R., and Ogren, W.L. (1980) Photorespiration mutants of Arabidopsis thaliana deficient in serine-glyoxylate aminotransferase activity. Proceedings of the National Academy of Sciences USA, 77, 2684–2687.

Suh, H.Y., Kim, J.H., Woo, J.S., Ku, B., Shin, E.J., Yun, Y., Oh, B,H. (2012) Crystal structure of DeSI-1, a novel deSUMOylase belonging to a putative isopeptidase superfamily. Proteins, 2012, 80: 2099-2104.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution, 30(12), 2725–2729.

Thakur, M., Sohal, B.S., (2013) Role of elicitors in inducing resistance in plants against pathogen infection: a review. ISRN Biochemistry, 1–11.

Tomanov, K., Hardtke, C., Budhiraja, R., Hermkes, R., Coupland, G., Bachmair, A. (2013). SUMO Conjugating Enzyme with Active Site Mutation Acts as Dominant Negative Inhibitor of SUMO Conjugation in Arabidopsis. Journal of Integrated Plant Biology. 55, 75-82.

Tomanov, K., Zeschmann, A., Hermkes, R., Eifler, K., Ziba, I., Grieco, M., … Bachmair, A. (2014) *Arabidopsis* PIAL1 and 2 Promote SUMO Chain Formation as E4-Type SUMO Ligases and Are Involved in Stress Responses and Sulfur Metabolism. The Plant Cell, *26*(11), 4547–4560.

Trinchieri, G., Sher, A. (2007) Cooperation of Toll-like receptor signals in innate immune defence. Nature Reviews Immunology, 7 (3): 179-190.

Van den Burg, H.A., Kini, R.K., Schuurink, R.C., Takken, F.L.W. (2010) Arabidopsis small ubiquitin-like modifier paralogs have distinct functions in development and defense. Plant Cell, 22, 1998–2016.

Vernooij, B., Friedrich, L., Morse, A., Reist, R., Kolditz-Jawhar, R., Ward, E., Uknes, S., Kessmann, H., Ryals, J. (1994) Salicylic acid is not the translocated signal responsible for inducing systemic acquired resistance but is required in signal transduction. Plant Cell, 6:959 – 965

Vertegaal, A.C. (2011) Uncovering ubiquitin and ubiquitin-like signaling networks. Chemical Reviews, 111, 7923–7940.

Vierstra, R.D. (2009) The ubiquitin-26S proteasome system at the nexus of plant biology. Nature Review Molecular Cell Biology, 10, 385–397.

Vierstra, R.D. (2012) The expanding universe of ubiquitin and ubiquitin-like modifiers. Plant Physiology, 160, 2–14.

Vijn, I., Govers, F. (2003) Agrobacterium tumefaciens mediated transformation of the oomycete plant pathogen Phytophthora infestans. Molecular Plant Pathology, 4: 459–467.

Vukašinović, N., Cvrčková, F., Eliáš, M., Cole, R., Fowler, J. E., Žárský, V., & Synek, L. (2014) Dissecting a Hidden Gene Duplication: The Arabidopsis thaliana SEC10 Locus. PLoS ONE, 9(4), e94077.

Wang, H., Ngwenyama, N., Liu, Y., Walker, J.C., Zhang, S. (2007) Stomatal development and patterning are regulated by environmentally responsive mitogen-activated protein kinases in Arabidopsis. Plant Cell, 19:63–73

Wilkinson, K.A., Henley, J.M. (2010) Mechanisms, regulation and consequences of protein SUMOylation. Biochemical Journal, 428, 133–145.

Walsh, C.K. Sadanandom, A. (2014) Ubiquitin chain topology in plant cell signaling: a new facet to an evergreen story. Frontiers in Plant Science, 5, 122.

Wessling R, Epple P, Altmann S, He Y, Yang L, et al. (2014) Convergent targeting of a common host protein-network by pathogen effectors from three kingdoms of life. Cell Host Microbe, 16:364–75

Win, J., Chaparro-Garcia, A., Belhaj, K., Saunders, DG., Yoshida, K., Dong, S., Schornack, S., Zipfel, C., Robatzek, S., Hogenhout, SA., Kamoun, S. (2012) Effector biology of plant-associated organisms: concepts and perspectives Cold Spring Harbor Symposia on Quantitative Biology, 77, pp. 235-247

Xia, Y. (2004) Protease in pathogenesis and plant defence. Current Microbiology, 6, 905–913.

Xiang, T., Zong, N., Zou, Y., Wu, Y., Zhang, J., Xing, W., Li, Y., Tang, X., Zhu, L., Chai, J., Zhou, J.M. (2008) Pseudomonas syringae effector AvrPto blocks innate immunity by targeting receptor kinases. Current Biology, 2008 Jan 8; 18(1):74-80.

Xing, Y., Jia, W., Zhang, J. (2008) AtMKK1 mediates ABA-induced CAT1 expression and H2O2 production via AtMPK6-coupled signaling in Arabidopsis. Plant Journal, 54:440–51

Yates, G., Sadanandom, A. (2013) Ubiquitination in plant nutrient utilization. Frontiers in Plant Science, 4, 452.

Yates, G., Srivastava, A.K., Sadanandom, A. (2016) SUMO proteases: uncovering the roles of deSUMOylation in plants. Journal of Experimental Botany, 2016; 67 (9): 2541-2548.

Yoo, C.Y., Miura, K., Jin, J.B., Lee, J., Park, H.C., Salt, D.E., Yun, D-J., Bressan, R.A., Hasegawa, P.M. (2006) SIZ1 small ubiquitin-like modifier E3 ligase facilitates basal thermotolerance in Arabidopsis independent of salicylic acid. Plant Physiology, 142, 1548–1558.

Zhang, S., Klessig, DF. (1997) Salicylic acid activates a 48-kD MAP kinase in tobacco. Plant Cell, 9:809–24

Zhang, J., Li, W., Xiang, T., Liu, Z., Laluk, K., Ding, X., Zou, Y., Gao, M., Zhang, X., Chen, S., Mengiste, T., Zhang, Y., and Zhou, J.-M. (2010a) Receptor-like cytoplasmic kinases integrate signaling from multiple plant immune receptors and are targeted by a Pseudomonas syringae eector. Cell host & microbe, 7(4):290–301.

Zeng, W., & He, S.Y. (2010) A Prominent Role of the Flagellin Receptor FLAGELLIN-SENSING2 in Mediating Stomatal Response to *Pseudomonas syringae* pv *tomato* DC3000 in Arabidopsis. Plant Physiology, *153*(3),

Zipfel, C., Robatzek, S., Navarro, L., Oakeley, E.J., Jones, J., Felix, G., and Boller, T. (2004) Bacterial disease resistance in Arabidopsis through flagellin perception. Nature, 428, 764–767

Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J. D. G., Boller, T., and Felix, G. (2006) Perception of the Bacterial PAMP EF-Tu by the Receptor EFR Restricts Agrobacterium-Mediated Transformation. Cell, 125(4):749–760.

Zipfel, C. (2008) Pattern-recognition receptors in plant innate immunity. Current Opinion in Immunology, 20:10-16.